ICAIIT 2026

Koethen

Germany

by Anhalt University of Applied Sciences

PROCEEDINGS OF THE INTERNATIONAL CONFERENCEON APPLIED INNOVATIONS IN IT



Volume 14

Issue 2

EDITION

Hochshule Anhalt

Proceedings of the International Conference on Applied Innovations in IT

Volume 14 | Issue 2

Koethen, Germany
22 April 2026

EDITORIAL BOARD

Editor-in-Chief

Prof. Dr. Eduard Siemens (Anhalt University of Applied Sciences), Germany

Associate Editor

Kateryna Shulakova (State University of Intelligent Technologies and Telecommunications, HSA), Ukraine

Topical Editors

Prof. Dr. Marija Kalendar (Ss. Cyrill and Methodius University in Scopje), North Macedonia
Dr. Halyna Henseruk (Ternopil Volodymyr Hnatiuk National Pedagogical University), Ukraine
Prof. Dr. Hassan Bevrani (University of Kurdistan), Iran
Prof. Dr. Ralph Gottschalg (Fraunhofer IMWS), Germany
Prof. Dr. Iryna Stankevych (Odessa State Academy of Civil Engineering and Architecture), Ukraine
Prof. Dr. Douglas Hall (St. Mary's University), USA

Proceedings of the International Conference on Applied Innovations in IT (ICAIIT), Koethen, Germany, March 12th, 2026. This volume contains peer-reviewed articles presented at the conference, along with selected papers accepted for publication in the general proceedings on April 22nd. The publication focuses on applied research in automation, communication, and information technologies, as well as on the application of IT-based methodologies to technical, engineering, and interdisciplinary problems. The presented research covers practical approaches to data processing, computational methods, intelligent systems, communication technologies, automation, modelling, and technology-driven problem-solving. The research results are particularly relevant to academic researchers and development engineers working on both the theoretical foundations and practical applications of information technologies and IT-based methodologies across technical and interdisciplinary domains.

ISBN: 978-3-96057-199-5 (Online)
ISSN: 2199-8876

All papers published in this jornal are full-open access:
immediately freely available to read and download.

Publisher: Edition Hochschule Anhalt
Phone: +49 (0) 3496 67 5611
Email: leiterin.hsb@hs-anhalt.de

Phone: +49 (0) 3496 67 2327
Email: conf@icaiit.org
http://icaiit.org

Köthen, Germany
Anhalt University of Applied Sciences
Bernburg / Koethen / Dessau
http://www.hs-anhalt.de

Postfach 1458 D-06354 Koethen, Germany

Content

Section 1 Communication Technologies and Hardware

Ivan Bobok, Svitlana Hryhorenko, Alla Kobozieva and Oksana Vasylenko
Theoretical Foundations of a New Approach to Container Selection when Organizing a Steganographic Communication Channel…..... 1

Andrii Raichuk and Larysa Globa
AI-Oriented Monitoring Agent in Network Infrastructure of Multicloud Data Centers…..... 13

Serhii Siden, Roman Tsarov, Dmytro Stepanov, Kateryna Shulakova and Andrii Pavlov
Hybrid AI-Based Path Loss Prediction Model for 5G/6G Networks…..... 21

Nurdaulet Asset, Martin Obert and Stefan Twieg
Assistive AI Mobility in Healthcare: Comparative Evaluation of Single Board Computers with AI/GPU Support for Real-Time Object Detection…..... 29

Andrii Astrakhantsev, Dmitriy Kapuler, Inna Butko and Larysa Globa
Malicious Traffic Detection in 5G Networks Using CNN and LSTM Models…..... 37

Section 2 Data Transport and Distributed Systems

Ali Ghermezian, Kirill Karpov, Dmitry Kachan, Veronika Kirova and Eduard Siemens
PPO-Driven Reinforcement Learning Congestion Control Under High-BDP Wide-Area Deployment: A Scaling Analysis…..... 43

Kateryna Shulakova, Liliia Bodnar, Mykola Bodnar, Roman Tsarov, Olha Yavorska, Vladyslav Kumysh and Oleksandra Ordanovska
DevOps Techniques for Artificial Intelligence Development and Training: A Human-Centered Approach to AI-Driven Operations…..... 51

Volodymyr Mankivskyi, Oleksandr Romanov, Mikola Nesterenko, Mika Romanova and Anton Marinov
A Multi-Layer Continuous-Time Markov Model for Availability Analysis of Kubernetes Systems…..... 59

Roman Tsarov, Olha Yavorska, Vladyslav Kumysh, Kateryna Shulakova, Eduard Siemens, Liliia Bodnar and Liudmyla Bubentsova
Method for Structural Optimization of Telemedicine Networks Based on Traffic Flow Distribution..... 71

Section 3 Data Analysis and Processing

Anna Vasileva and Natalia Evstifeeva
A Method Based on NER and CRF for Extracting Named Entities from Text and Textual Representations of Chemical Reactions..... 79

Tetiana Hromko, Tetiana Kovalevska, Hanna Truba and Aryna Frumkina
Corpus-Based Data Processing of Terminological Systems in Linguistic Theory..... 87

Phuc Hau Nguyen
Comparative Analysis of Machine Learning and Deep Learning Models for Spam Email Detection Toward Sustainable and Secure Digital Systems..... 95

Anastasiia Sapeha, Ibrahim Kovan, Subashkumar Rajanayagam, Kirill Karpov, Maksim Gering, Dmitry Kachan and Eduard Siemens
A Comparative Study of Noise Effects, VAD, and ASR Methods in Speech Recognition..... 103

Boshen Ruan, Beifei Wei and Stefan Twieg
Investigating Multi-Camera Tracking and Re-Identification Pipelines Based on YOLOv8..... 111

Oleg Pursky, Tetyana Filimonova, Anna Selivanova, Tatiana Dubovyk, Victor Kovtunenko and Iryna Buchatska
Image Detection Method Based on CNN and Pre-Trained Models..... 119

Oleksandr Stryzhak, Andrii Yaremenko and Viacheslav Gorborukov
Architecture of Ontology-Driven Multidimensional Analytical Systems Based on Formal Ontologies..... 127

Vitalii Bazurin, Oleg Pursky, Ihor Tyschenko, Andrey Nechepourenko and Alexander Vishnevsky
Software Implementation of Application for Correlation Analysis..... 147

Qing Huang, Pooja Pol and Jianing Zhang
LLM-Based Synthetic Ground Truth Generation for Audio-Based Emotion Classification via In-Context Learning…. 155

Marija Miloshevska and Marija Kalendar
A Comparative Analysis of AI-Assisted Development Tools in Software Development…. 163

Olena Ronzhes
Adaptive Governance Architecture for Digital Innovation Ecosystems Based on Psychophysiological User Profiling…. 171

Abdullah Hussein and Ali Khalaf Hussain Al-Hachami
Spectral Solution of Fractional Delay Equations via Fermat Polynomials…. 179

Section 4 Control and Automation

Sergey Pochinok, Alexey Lavrentyev, Alaibek Obozov, Raya Sultanalieva and Veronika Kirova
Automation of a Pneumatic Substrate Mixing System in a Biogas Plant…. 187

Oleksii Cherkashyn
Automating Visual Testing in Android Applications…. 195

Nataliia Zaiets, Nataliia Lutska, Lidiia Vlasenko, Volodymyr Shtepa and Stefan Junge
Automation of Water Treatment Facilities with Adaptive KPI Integration within a Sustainable Development Strategy…. 203

Oleksii Cherkashyn
SSH Connection During Web Automation Testing…. 211

Sergey Pochinok, Alexey Lavrentyev, Alaibek Obozov, Raya Sultanalieva, Sairagul Nasirdinova and Veronika Kirova
Experimental Evaluation of Automated Intermittent Pneumatic Mixing in Moisture-Controlled Anaerobic Digestion…. 217

Ivan Omeko, Stefan Twieg and Martin Obert
Posture Detection and Assistance for Mobility Aid: A Sensor Fusion and Machine Learning Approach…..... 225

Danylo Nazymko
Impact of Structural Uncertainty on the Time Parameters of the Project Schedule in a Coal Mining Application…..... 235

Oleksii Cherkashyn
Enhancing Web Test Automation Through Screenshot and Video-Based Execution Reporting…..... 243

Section 5 Statistical Modelling and Analysis and their Applications in Life Sciences

Oleh Kovalyshyn, Yaroslav Dorogyi, Vladyslav Ostash and Nadiia Kasianchuk
Patterns of Behavioural Risk Factors and their Associations with Cardiovascular Outcomes in Ukraine…..... 249

Sergii Guzii, Vitalina Lukianova, Yevheniia Anpilova, Oleksii Soloviov and Andriy Tovmachenko
Organo-Mineral Complexes for Water Purification: Experimental and Statistical Analysis…..... 259

Section 6 Information Technology in Education

Daria Lys, Vladyslav Suvorov, Tetiana Borysenko, Maryna Tsynova and Svitlana Hryhorenko
Transformation of Higher Education in the Age of Artificial Intelligence: Analysis of Digital Learning Practices…..... 267

Iryna Polishchuk, Yuriy Dovhan, Oksana Yakushevska, Larysa Dovhan, Oleksandr Nepochatenko and Viktoriia Petrova
Information Technologies in the Digital Ecosystem of Higher Education Institutions…..... 277

Iryna Zhalinska, Serhii Nikolaienko, Kateryna Parnus, Marina Kosheleva and Olena Mykhaylovska
Online and Face-to-Face Learning Effectiveness in Ukrainian Higher Education: A Multidimensional Comparison…..... 285

Coravil Joy C. Avila
Predictive Modeling of Information Systems Success in Higher Education: A WarpPLS-SEM Analysis of User-Centered Portals…… 293

Svitlana Sytniakivska, Snizhana Kubrak, Svitlana Khmelivska, Karyna Razumna and Svitlana Petrova
Integration of Digital Language Resources into Foreign Language Teaching for Students of Non-Linguistic Degree Programmes…… 303

Serhii Petrovych, Stefan-Daniel Horvath and Chunfang Zhou
ESP32-Based AI Embedded System for Personalized Education…… 311

Svitlana Skvortsova, Tetiana Symonenko, Kira Hnezdilova, Kateryna Niedialkova and Nataliia Andrusiak
Transforming Peer Assessment in Higher Education: An AI-Supported Approach Using the Peerscholar Platform…… 319

Mykola Fant, Iryna Kovalenko and Nataliia Baranivska
Cloud-Based Communication Technologies for Synchronous Online Foreign Language Learning: Infrastructure and Didactic Potential…… 327

Olesia Fedoruk
Introduction of Digital Technologies in Vocational Education: Tools for Transforming the Learning Process and Improving the Quality of Staff Training…… 335

Olha Dienichieva, Natalia Plakhotniuk, Olena Svysiuk, Svitlana Kukharonok and Oleg Makarevych
Intelligent Tutoring Systems as Communication Technologies for Enhancing Second Language Learning…… 343

Feruza Shermanova and Guli Taylakova
Pedagogical and Methodological Capabilities of Artificial Intelligence Tools in Developing Digital Literacy…… 351

Larysa Soloviova, Irena Snikhovska and Zoia Zalibovska-Ilnitska
Application of Immersive Technologies in Teaching Business English in Higher Education Institutions for the Formation of Situational Speech…… 359

Olena Soia, Olena Kosovets and Mariana Kovtoniuk
Mathematics Problems and Computer Algebra Systems as Didactic Tools for Developing the Professional Competence of Bachelors in Mathematics and Computer Science…… 367

Svitlana Skvortsova, Tetiana Symonenko, Kira Hnezdilova and Ruslana Romanyshyn
Digital Storytelling for the Prevention of Mathematical Anxiety in Primary Education: A Comparative Analysis of AI Platforms…… 375

Roman Yaroviy, Olena Skliarenko, Serhii Yahodzinskyi, Yanina Kolodinska, Oleksandr Nikolaievskyi and Hanna Tereshchuk
Methodological Aspects of Integrating and Mathematically Modeling a Cyber Range as an Innovative IT Solution for Cybersecurity Education in Higher Education Institutions…… 383

Cherly Baja Sardovia
Technology-Driven Framework for Alumni Engagement: Extending the Information System Success Model with Altruism and User Attractiveness in a Graduate Tracer System…… 391

Serhii Petrovych, Stefan-Daniel Horvath and Chunfang Zhou
Integrating AI and IoT in STEM Education Through a Gesture-Control Project…… 399

Iryna Haidai, Liudmyla Suvorova, Olha Zadoienko, Anastasiia Kravchenko and Krystyna Levkivska
Correction of Sensory Integration as a Prerequisite for Successful Mastering a Foreign Language…… 407

Andrii Lobatskyi, Pavlo Savaryn, Volodymyr Podolyak, Mykhailo Lepkiy and Vitalii Kabak
Application of Cloud-Based Digital Prototyping Systems for Developing Hardware Logic Skills…… 417

Section 7 Information Technology in Economics

Nataliia Trushkina, Maryna Popova, Rahayu Relawati, Bambang Yudi Ariadi and Yelyzaveta Loiko
Digital Marketing as a Strategic Tool for Promoting Agricultural Products..... 425

Viktoriia Rudenko, Olena Moskvichova, Halyna Pohrishchuk and Nadiia Holovai
Consequences of the Functioning of the New Computerised Transit System (NCTS) for Customs Logistics..... 441

Lazar Krstić and Marija Krstić
IT Standardization in the Area of Information Security with a Focus on ISO/IEC 27001..... 451

Ayagoz Orazbayeva, Maira Bauer and Aliya Ismailova
From Cost Structure to ESG Sustainability: Structural and Cross-Country Evidence from the EU Livestock Sector..... 459

Iulia Caprian, Viktoriia Khaustova, Nataliia Trushkina and Kateryna Shulakova
Digital Transformation of Customer Relationship Management Systems in Logistics Systems: A Conceptual Model..... 467

Valerii Bondarenko, Yaroslav Gontaruk, Vira Butenko, Alla Riabchyk and Kostiantyn Polovko
Marketing Tools to Support the Development of Bioenergy as a Factor in the Energy Sustainability of Local Communities During Martial Law..... 479

Section 8 Power Engineering

Yurii Zabulonov, Tetiana Nosenko, Vitalina Lulianova and Yevheniia Anpilova
High-Fidelity Monte Carlo Simulation of Gamma-Ray Shielding Efficiency for Nuclear Security Applications Using Geant4 Toolset..... 487

Samer Rabih, Ammar Zakzouk, Hasan Hasan, Mazen Abdullatif, Ziad Nouman, Ahmad Malla and Rami Mousa
Impact of Thermal Parameters on the Performance of Photovoltaic Panels in Pump Water Solar Systems: A PVsyst Simulation Study..... 495

Zukhra Juraeva, Isroil Yuldoshev and Islom Juraev
Results of Experimental Study of Photovoltaic Installation Based on Thin-Film Panels for Integration on Facade of the Building..... 505

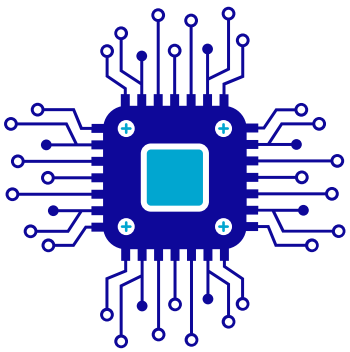
Ilias Zhumagali, Meruyert Ramatullayeva, Symbat Amanzhol, Turlan Turkbenbayev and Dmitry Kachan
Temporal Dependency Analysis in Short-Term Electricity Load Forecasting Using Ensemble Learning..... 513

Hochshule Anhalt
EDITION
ICAIIT

COMMUNICATION TECHNOLOGIES AND HARDWARE

Section 1

APRIL 2026



Proceedings of the International Conference on Applied Innovations in IT

Theoretical Foundations of a New Approach to Container Selection when Organizing a Steganographic Communication Channel

Ivan Bobok[1], Svitlana Hryhorenko[1], Alla Kobozieva[2] and Oksana Vasylenko[3]
[1]*Department of Computerized Systems and Software Technologies, Odesa Polytechnic National University, Shevchenko Avenue 1, 65044 Odesa, Ukraine*
[2]*Department of Cybersecurity and Information Protection, Odesa National Maritime University, Mechnikova Str. 34, 65029 Odesa, Ukraine*
[3]*Preparatory College of Saxony-Anhalt, Anhalt University of Applied Sciences, Lohmann Str. 23, 06366 Köthen, Germany*
onu_metal@ukr.net, s.m.hryhorenko@op.edu.ua, alla_kobozeva@ukr.net, oksana.vasylenko@hs-anhalt.de

Keywords: Container Selection, Perception Reliability, Structural Similarity Index Measure (SSIM), Singular Value Decomposition.

Abstract: The most promising direction of information protection today is steganography, the main goal of which is to hide the very fact of the existence of information exchange. The choice of container plays a key role in ensuring certain properties of the steganographic system, but the problem of container selection currently has no satisfactory solution. *The aim* of the work is to develop the theoretical foundations of a new approach to a priori selection of a container from a given set of digital candidate images, common in the sense of independence from the steganographic algorithm used, to improve the visual quality of the steganographic message. *The aim is achieved by means of* a well-founded definition of the container selection criterion; choosing a quantitative indicator of the visual quality of the steganographic message. *The most significant results* of the work are the theoretical justification of the feasibility of using the normalized separation of the maximum singular number of the image matrix as a selection criterion. *The significance of the obtained theoretical results* lies in the possibility of their effective practical use for selecting a container from a given set of candidates, which is demonstrated in the work by means of a computational experiment. It is shown that using the selected image as a container allows obtaining the best/close to the best structural similarity index for the corresponding steganographic message (the maximum deviation from the best value within the experiment was 0.3%), regardless of the steganographic algorithm used and the format of the candidate images. Using subjective ranking, it was established that the visual quality of steganographic messages obtained based on selected containers was improved compared to random ones.

1 INTRODUCTION

In today's conditions of rapid development of digital technologies and globalization of information networks, the problem of ensuring the confidentiality of data transmission is becoming critical. According to analytical reports of ENISA and ITU [1], [2], in conditions of rapid expansion of the cyberthreats and improvement of automated traffic monitoring tools, steganography is becoming particularly relevant, the main purpose of which is to hide the very fact of the existence of information exchange [3], [4]. This approach allows to divert/reduce the attention of a potential attacker to the communication channel, which is strategically important for protecting private and corporate data in the information environment.

The effectiveness of a steganographic system is determined by a compromise between its capacity, stability and visual invisibility of the changes made - the reliability of perception of the steganographic message [5]. The key role in achieving this balance is played by the choice of the container [6]-[8], as which the work considers a digital image. The properties of the container determine the degree of potential concealment of embedded data - additional information, which for modern steganography is a digital sequence that is the result of the work of the steganographic system`s precoder. Random choice of container for some steganographic methods can lead

to visual distortions as a result of embedding additional information [9].

2 LITERATURE REVIEW

Considering that failure to ensure the reliability of perception as a result of a subjective attack will instantly lead to the detection of the presence of a hidden (steganographic) communication channel, this requirement most often becomes key when choosing a container in the works of modern Specialists [7], [10]. In [10], the authors note that the goals of improving steganographic systems today are divided into two groups: the first is aimed at maximizing the ability to hide information; the second is to minimize the visual distortion of the container as a result of embedding additional information. The higher the visual quality of the steganographic message, the less attention it attracts to itself, which obviously increases the degree of security of the embedded information.

One approach to automating container selection is proposed in [11]. The authors use evolutionary programming to generate mathematical rules (signatures) that allow selecting a container based on its ability to counteract statistical steganography. The peak signal-to-noise ratio (PSNR) is used to assess visual distortions when obtaining a steganographic message. This difference indicator has a weak correlation with the visual perception of the digital image by a person, does not take into account the Spatial relationships of pixels, considering them as independent. Similar remarks can be attributed to the following works [12]-[14]. In [12], the idea is implemented that the optimal digital image for the role of a container is a digital image in which the process of embedding the additional information leads to minimal disturbances. The proposed method in the course of its work scans all possible positions for embedding a secret message in a container and performs such a scan for all candidate digital images, which leads to its significant computational complexity and is often an insurmountable obstacle to the use of the method in practice. In [13], a method for selecting a container from a set of digital images for a given secret message, which is also considered in the form of digital images, is proposed. The method is divided into two main steps: first, the set of candidates is filtered taking into account the relative entropy and features of their histograms in order to filter out unambiguously unsuitable candidates; second, the local characteristics of the pixel intensity of the blocks of digital images-candidates are analyzed. The proposed method, as its authors claim, demonstrates high visual quality of the steganographic message, but since the difference PSNR indicator is used to evaluate it here, as noted above, the objectivity of the authors' assessment can be questioned. In [14], the container selection problem is considered, for the solution of which a method based on a genetic algorithm is proposed, which ensures good compatibility with the given secret data. But the conclusion about good compatibility is again made using the difference indicators: PSNR and MSE.

To assess the visual quality of a steganographic message, not only difference indicators can be used. An example of a metric that tries (not without success) to take into account the peculiarities of the perception of the human visual system can be considered the structural similarity index (SSIM) [15], and the results of assessing the effectiveness of container selection methods presented within the framework of using this metric in [10], [16] have a higher priority and trust among the authors of this article than the previous ones. Thus, in [16] an approach to container selection is proposed, based on the analysis of the features of the texture of the central image and the human visual system, within which the similarity between the original central image and the corresponding steganographic message is quantitatively assessed. As a result, one central image container is selected that has the maximum similarity with its steganographic message. Despite the authors' attempt to take into account the peculiarities of the human visual system through the use of SSIM, their approach is limited to the analysis of the statistical characteristics of the texture in discrete directions. This leaves out the holistic structural similarity between the container and the steganographic message.

In [10], the problem of visual quality of the steganographic message is considered as a classification problem, where a classifier based on a convolutional neural network (CNN) is used to select a digital image from a set of candidates for the role of a container that can provide high quality - perceptual reliability, which is estimated using SSIM, after the steganographic transformation process (STP). A binary classifier divides the digital images into high - and low-quality classes, based on the SqueezeNet architecture. The CNN was trained in two scenarios: transfer learning and learning from scratch. Although both classifiers, according to the authors, were able to achieve high accuracy, they have fundamental shortcomings. The first of these shortcomings is

common to methods based on deep learning: the process of this training works as a "black box", not giving an answer why this digital image was chosen, as a result of which, in the presence of an unsatisfactory result, it is impossible to get an answer about the reason for this and, as a result, make adjustments. In addition, the result of the learning process critically depends on the training sample. The second drawback concerns the idea of a binary classifier. The binary distribution is very rigid. For most of the digital images, they cannot be unambiguously attributed to textured or homogeneous. As a rule, the digital image has areas with different degrees of texture, which is not taken into account in the case of a binary classifier. If the digital image is defined as having high quality - textured, then as a result of embedding the digital image without taking into account the possible presence of homogeneous subregions in the selected container, the situation of the appearance of artifacts after the STP is likely.

It should be noted that, given the diversity of the image in all its properties, the idea of a clear division of them into classes of suitable and unsuitable for the role of a container is, in the opinion of the authors of this article, unpromising. It is fundamentally wrong to expect a clear division from a selective method, since such a division can only be based on clear thresholds for quantitative estimates of a particular digital image parameter, but the threshold will always be tied to the image database that is being considered (or used for training the network) at the moment. Changing the candidate database will, as a rule, lead to a change in the threshold values, i.e., a change in the quantitative characteristics of the classifier and a decrease in its efficiency.

3 METHODOLOGY

A characteristic feature of many methods available today for selecting a container in order to ensure the visual quality of the corresponding steganographic message is their focus on a specific steganographic algorithm [17], [18]. Such a focus helps to ensure their effective operation, but it severely limits the scope of application, which in conditions of lack of time, which often occur in practice when solving information protection problems, in particular the problem of organizing a covert communication channel, can be critical. It is clear that, taking into account the existing general sufficient conditions for ensuring the reliability of steganographic message perception [19], [20], the problem of selecting a container should have a general solution that is not focused on the specifics of a particular steganographic algorithm.

Thus, today it cannot be stated that the problem of choosing a container to ensure the visual quality of the corresponding steganographic message has a satisfactory solution. Taking this into account,

The aim of the work is to develop the theoretical foundations of an approach to a priori selection of a container from a given set of digital images-candidates, general in the sense of independence from the steganographic algorithm used, to improve the visual quality of the steganographic message.

To achieve the goal, the following *tasks are solved in the work:*

1) Determining a quantitative criterion for container selection in order to ensure a (relatively) high-quality steganographic message in the visual sense;
2) Selection of a quantitative indicator for assessing the visual quality of the steganographic message - an indicator of the effectiveness of the corresponding steganographic system;
3) Study of the dependence of the visual quality of the joint venture on the container selection criterion.

3.1 Determining the Quantitative Criterion for Container Selection

Let F *be* an $n \times n$ matrix of the digital image with elements f_{ij}, $i,j = \overline{1,n}$. For it, we can construct a singular decomposition:

$$F = U\Sigma V^T, \quad (1)$$

where: U, $V$are orthogonal matrices, the columns of which u_i, v_i, $i = \overline{1,n}$ are the left and right singular vectors (SV) *of* F, respectively, $\Sigma = diag(\sigma_1, \sigma_2, \ldots, \sigma_n)$and are the diagonal matrix of singular numbers (SN): $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$, $\sigma_i \in R$, $i = \overline{1,n}$, R are the set of real numbers. In accordance with the general approach to the analysis of the state of information systems, based on matrix analysis and perturbation theory [19], [21], [22], the STP of a central container with matrix F can be represented in the form $\overline{F} = F + \Delta F$ ($\overline{F}$, Δ F is the steganographic message matrix with elements $\overline{f}_{ij}$, $i,j = \overline{1,n}$, and the container perturbation as a result of the STP), and, as a result, in the form of a set of perturbations of the full set of formal parameters F - SV and SN, which is the result of embedding the digital image. Thus, the singular decomposition and

its parameters play a key role in assessing the state of the resulting steganographic message.

Since the main goal of this work is to select a container that provides relatively high-quality visual steganographic message, while the smallness of the norm $\|\Delta F\|$is often considered as an indicator of ensuring a significant probability of reliable perception of steganographic message [12], among the parameters of the full set it is necessary to select those whose perturbations will adequately reflect the strength of the $\|\Delta F\|$perturbation during STP - insensitive, or resistant, to perturbations. Among the SV matrices of any digital image, there are both stable and unstable ones [19], [21], [22] in accordance with the relation:

$$sin\,2\,\theta_i \le \frac{2\|\Delta F\|_2}{svdgap(i,F)}, \qquad (2)$$

where :$\theta_i$is the rotation angle u_i (v_i) as a result of the turbulent action ΔF, $\|\cdot\|_2$is the spectral matrix norm, $svdgap(i,F) = \min\limits_{i\neq j}\left|\sigma_i - \sigma_j\right|$is the separation of the SN σ_i. All SNs are stable or well-conditioned: $\max\limits_i\left|\sigma_i(F) - \sigma_i\left(\overline{F}\right)\right| \le \|\Delta F\|_2$, where$\sigma_i(F)$, $\sigma_i\left(\overline{F}\right)$ - SN matrices, F, $\overline{F}$respectively.

Of texture of the steganographic message container is desirable [18], [23]: the container should have a large number of contours, small details, etc., that is, a significant relative volume of such subregions where there are significant differences in pixel brightness values, which leads to a relatively significant high-frequency component.

Recently, to organize the separation of the digital image, stored in various forms (with/without loss), the authors justified the feasibility of using the normalized gap of maximum SN (NGMSN) of the corresponding matrix: $svdgap_n(1) = \overline{\sigma}_1 - \overline{\sigma}_2$, the calculation of which corresponds to the general formula for the normalized separation of SN σ_i:

$$svdgap_n(i) = \min\limits_{i\neq j}\left|\overline{\sigma}_j - \overline{\sigma}_i\right|, \qquad (3)$$

where: $\overline{\sigma}_i$, $i = \overline{1,n}$, are defined as: $\overline{\sigma}_i = \frac{\sigma_i}{\|(\sigma_1,\sigma_2,\ldots,\sigma_n)\|}$, $\|(\sigma_1,\sigma_2,\ldots,\sigma_n)\|$- Euclidean norm of the vector $(\sigma_1,\sigma_2,\ldots,\sigma_n)$. The NGMSN is obtained from (3) with $i = 1$taking into account that for the SN of the original digital image:

$$\sigma_1 >> \sigma_2 \ge \ldots \ge \sigma_n \ge 0. \qquad (4)$$

The difference in the values of the NGMSN in [24] was an indicator of the difference in the relative content of the high-frequency component, which for the lossy format is less than the corresponding indicator in the lossless format. Considering such "abilities" of the NGMSN, it is proposed to use it as a criterion for selecting the digital image-containers that provide the visual quality of the steganographic message, which is explained by the following considerations.

There is a close connection between singular decomposition and the structure of the digital image. According to [15], the structure of the digital image is understood as the set of its geometric features (edges, outlines of objects, contours, textures, details), which can be observed at different scales. It should be noted here that it is the violation of the digital image structure that is, as a rule, the most noticeable to the human eye [15]. The singular decomposition (1) of the matrix F can be written in the form of outer products [25]:

$$F = \sum_{i=1}^{n}\sigma_i u_i v_i^T, \qquad (5)$$

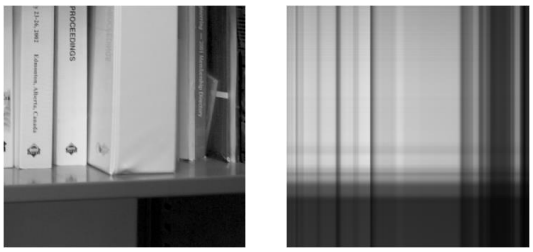
that is, it gives a representation of the digital image as a sum of peer matrices $\sigma_i u_i v_i^T$.

The geometry, the “skeleton” of the digital image is described by the SV [25], [26]. The left/right SV characterize its vertical/horizontal structures: each vector u_i/v_idetermines how the intensity changes along the vertical/horizontal axis, and $u_i v_i^T$creates a peer-to-peer “base” image [26] - the structure of the digital image as a whole, which looks like a set of stripes, grids or contours of a certain direction (Fig. 1). If the components of the vector $v_i$differ slightly, and in $u_i$have significant differences, then the resulting matrix $u_i v_i^T$will consist of horizontal lines, otherwise we will get vertical stripes.

The first SV usually describe the roughest, most important contours of objects, which generally define the scene of the digital image, and the following ones - (small) details, noise. The SN describe the intensity of these structures, i.e. they are weights for these structures. As a result of adding several components $u_i v_i^T$with different weights, $\sigma_i$complex shapes are obtained [26].

Thus, the form (5) clearly indicates the close connection of the singular decomposition of the matrix of the digital image with its structure. If the image has a relatively large NGMSN in comparison with other candidates for the role of the container, this will mean that in such an image the base image dominates more than in others, $\sigma_1 u_1 v_1^T$thereby reducing the relative content of fine details, noise, i.e. the content of the high-frequency component. For the image, which has an insignificant, relative to other candidates, NGMSN, dominance $\sigma_1 u_1 v_1^T$will not be so significant, speaking of the presence of a certain number of parts in the central unit, the presence of a

certain noise, which is a positive sign for using such a central unit as a container, an indicator of which is the NGMSN.



(a) (b)

Figure 1: Relationship between the singular decomposition of the image matrix and its structure: a) - original digital image; b) - peer approximation$\sigma_1 u_1 v_1^T$.

The stability of the structure of the digital image is actually determined by the stability of its SNR, which corresponds to the relation (2). It should be noted that the overall structure of the digital image will mainly depend on the first SV (Fig. 1): the terms in (5) are not equal, the first term carries almost all the energy of the digital image due to (4). If the first SV is disturbed, this will disturb the main axis of the digital image, deforming its structure in such a way that it will be noticed by the observer with a high probability. SV disturbances, which correspond to low-separation SNs, will usually lead to disturbances in small details, texture, which can be ignored by the human visual system. This means, firstly, that, in general, a conclusion about the structure of the digital image can be made by analyzing only the visually dominant structure - $\sigma_1 u_1 v_1^T$, and secondly, absolute dominance is undesirable for $\sigma_1 u_1 v_1^T$the STP: for a container, the better, the relatively smaller the dominance of the first peer approximation will be. A quantitative assessment of the degree of this dominance is the (normalized) separation σ_1. Thus, further, the NGMSN of the corresponding matrix is used as a criterion for selecting the image containers that ensure the visual quality of the steganographic message.

3.2 Choosing a Quantitative Indicator to Assess the Visual Quality of a Steganographic message

The process of container selection is a task that requires consideration of the preservation of complex relationships between the elements of the digital image and its structure.

The use of containers with a high degree of texturity allows for effective masking of the results of step-pan transformation, using the peculiarities of human perception. In this aspect, the use of metrics that can quantify structural distortions of the digital image becomes critically important.

The most common in steganography to date for assessing the reliability of perception of the formed steganographic message remain difference indicators [27], in particular, the mean square error (MSE), the peak signal-to-noise ratio (PSNR), which are quite mathematically simple in their implementation. MSE and PSNR are related to each other:

$$PSNR = 10\,lg\left(\frac{255^2}{MSE}\right),$$

where: $MSE = \frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(f_{ij} - \overline{f}_{ij}\right)^2$.

These indicators are not justified for steganography because the assessment of the reliability of perception of the formed steganographic message involves a mandatory expert assessment of a person - visual perception, the specificity of which is not taken into account in these indicators.

The disadvantages of difference indicators are not limited to this. As is well known [26], pixels in the digital image are not independent: the brightness values of neighboring pixels are correlated, they form contours, objects, however, all difference indicators consider each pixel in isolation.

This leads in particular to the fact that the difference indicators do not reflect the difference between the "destruction" of the structure (for example, a break in the nose line on the face) and some uniformly distributed disturbance throughout the digital image, which does not change the overall scene, geometry (mutual arrangement of lines, angles, etc.) and the "content" of objects, i.e. the structure of the digital image, and may go unnoticed. An attempt to smooth out the above-mentioned shortcomings in the quantitative assessment of digital image distortion was made when developing the SSIM metric in [15]:

$$SSIM(X,Y) = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X{}^2 + \mu_Y{}^2 + C_1)(\sigma_X{}^2 + \sigma_Y{}^2 + C_2)}, \quad (6)$$

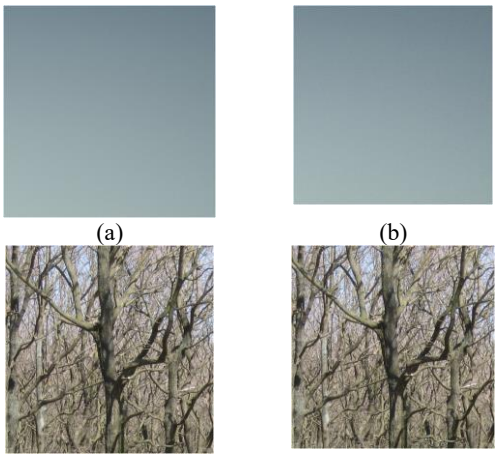
where: *X, Y* - the presented and distorted digital image, respectively ($L \times L$ corresponding windows of the presented and distorted digital image), μ_X, μ_Y - the average brightness values, $\sigma_X{}^2$, $\sigma_Y{}^2$- the variance (contrast index) for (windows) image *X, Y,* respectively, σ_{XY}- the covariance (structural similarity index (windows) image), C_1, C_2- constants introduced by the authors to stabilize the calculations. The closer the index (6) is to unity, the less distortions the human eye should notice according to the authors' idea. Such an orientation towards subjective assessment is achieved for (6) due

to the fact that the authors tried to simultaneously take into account three indicators when constructing the metric: the brightness of the image pixels, contrast and structure.

Gaussian noise with the same parameters was used as a perturbation for both images, resulting in the same value of $PSNR = 33dB$. At the same time, the distortions for the homogeneous digital image are visually noticeable (heterogeneity appeared throughout the image) (Fig. 2b), unlike the textured image (Fig. 2d), which was not demonstrated by the difference distortion indicator. The SSIM metric behaved much better as expected, giving values of 0.8804 and 0.9674, respectively, clearly reflecting the existing connection with the visual perception of the digital image, demonstrating the advantage of the second SI for the role of a container.

The next advantage of the SSIM metric for use in the field of steganography is its ability to provide a distortion map (*SSIM map*) - a matrix of the same size as the studied image *F,* each element of which is a characteristic of the quality of the corresponding pixel *F*: in the distortion map, the value of each element represents the result of a comparative analysis of the values of the fragment (window) around the pixel. Practically, in the distortion map, light zones indicate that the structure of the digital image in these subregions is preserved, while dark zones indicate the destruction of the structure. These characteristics of the SSIM are very important for steganography, in particular, the distortion map makes it possible to see in which zones the STP distorts the (structure) of the digital image container the most/least (Fig. 3).

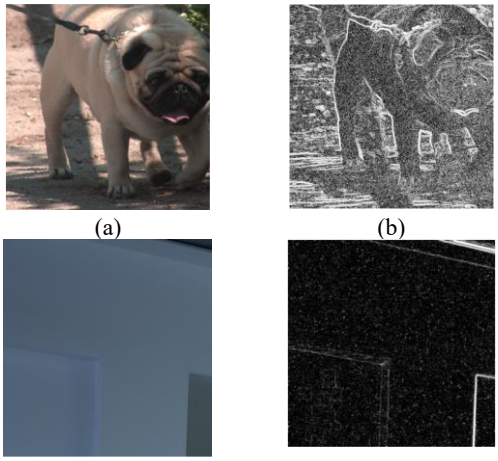
The two examples in Figure 3 are chosen intentionally: for the digital image (Fig. 3d) on the distortion map there are black areas that correspond to the background areas of the container (Fig. 3c) - confirmation that the use of homogeneous areas for the STP can with a high probability lead to a violation of the reliability of the steganographic message perception. On the distortion maps (Fig. 3b, d) the most favorable - light areas corresponding to the contours (which are known to be priority areas for embedding the digital image from the point of view of the reliability of the steganographic message perception). For the image (Fig. 3a) and its other areas are not critically unacceptable for embedding the image (painted in gray (Fig. 3b). This is natural and expected, because this image is textured.



(a) (b) (c) (d)

Figure 2: Illustration of the advantage of the SSIM metric for assessing the visual quality of the image: a), c) - original images; b), d) - digital images, which are the results of superimposing Gaussian noise with zero matrix expectation and *D* =0.0005.

Speaking about the choice of a container in order to improve the quality of visual perception of the steganographic message, it should be noted that the application of the PSNR metric in the conditions of using one of the most widespread to date steganography methods - the least significant bit modification method when implementing LSB-matching [28], is generally uninformative: the PSNR values will be practically the same for all candidates.



(a) (b) (c) (d)

Figure 3: Examples of obtained distortion maps for images containers as a result of applying the LSB method with a hidden channel bandwidth of 1 bit/pixel: a), c) - original central processing units; b), d) - *SSIM maps*.

Thus, SSIM compared to difference measures of the estimation of the distortion of the digital image has significant advantages for application in the field of steganography. Unlike approaches focused on difference measures, SSIM-based analysis allows to identify images and local areas where interference minimally disrupts the correlation between neighboring pixels, the use of which is key to ensuring the resistance of the digital image to subjective attacks.

Despite its advantages, to date, the *SSIM metric* has significantly lost to differential metrics in terms of frequency of use. To ensure objectivity in the choice of metric, let's consider the reasons for ignoring SSIM in most modern works on selective steganography:

- PSNR and MSE are simpler in mathematical sense and computational sense than SSIM. Indeed, for *an n×n* matrix of the digital image, to calculate one term in the sum for MSE, two operations are required, which will give $2n^2$for all terms, and $2n^2 + (n^2 - 1)$to form a double sum, i.e. the total number of operations $K_{MSE}$to calculate MSE will be:

$$K_{MSE} = 2n^2 + (n^2 - 1) + 2 = O(3n^2) = O(n^2). \quad (7)$$

- The computational complexity for calculating PSNR will differ from (7) by adding a small number of operations (three arithmetic operations and one - taking the logarithm), which does not depend on n, will also be defined as $O(3n^2) = O(n^2)$. SSIM for each pixel is calculated in a local window, usually of size L×L (in practice most often $L = 11$). The number of operations for processing one pixel at a given window size depends only on L and is constant taking n into account. The total number of pixels here is n^2. Then the order of computational complexity for calculating SSIM (6) here will also be n^2, as in (7) for MSE, but the coefficient at n^2will be determined L^2 [15], i.e. the computational complexity of calculating SSIM will be: $K_{SSIM} = O(L^2 n^2) = O(n^2)$, and at $L = 11$will be 40 times greater;
- The SSIM indicator characterizes the visual quality of the steganographic message better than the difference indicators, but it is inferior to the difference indicators in terms of quantitative assessment of the stego-message's resistance to steganography. A large number of classical steganography methods are based on the analysis of changes in pixel brightness. MSE, PSNR correlate with changes in signal amplitude, SSIM can be high even if the pixel values have changed significantly (the main thing here is to preserve the structure), which will be a direct indicator of the presence of additional information for a statistical steganography analyzer;
- Scientific "inertia": the use of differential indices in steganography over the past few decades has led to the need to use these indices in recent years to ensure correct comparison with previous methods.

Of the above reasons for the limited use of SSIM from a practical point of view, taking into account the purpose of the work, the first is the main one. However, work on reduction $K_{SSIM}$is being carried out quite actively [29], [30], allowing the use of SSIM even in real time.

Considering the above, SSIM is chosen to assess the visual quality of the resulting steganographic message.

4 RESULTS AND DISCUSSION

The SSIM indicator of container distortion at STP is clearly related to the NGMSN of its matrix. Given that: the structure of the digital image is determined by its peer components $\sigma_i u_i v_i^T$; for embedding the additional information, it is natural to use the existing differences in brightness values, the number of which should be relatively large to ensure a satisfactory bandwidth of the covert channel; it is necessary to avoid changes in $\sigma_1 u_1 v_1^T$; for the STP, absolute dominance is undesirable $\sigma_1 u_1 v_1^T$, the quantitative assessment of the degree of which is the NGMSN, then using a digital image with a relatively small NGMSN as a container should lead to a relatively significant SSIM at STP. That is, theoretically expected here is an inversely proportional relationship between the NGMSN values of the container matrix and the SSIM, which characterizes the container distortion as a result of additional information embedding. This conclusion is general, i.e. does not depend on the specifics of the steganography method used and the digital image format. To confirm this conclusion in practice, a computational experiment was conducted, during which 8 sets of practical power images (100 images in a set) were formed from the bases [31]-[33] (two sets of digital images from each base), which are traditional when working with digital images, as well as two sets of images obtained by non-professional

cameras. Some typical results of the experiment are presented in Figure 4, 5a, where the real largest SSIM value is marked in red, and the one corresponding to the digital image with the smallest NGMSN, if these values do not coincide, in green.

The results are demonstrated on the image sets: M_1 [31] - digital images size 560×560 (Tif format); M_2 [32] - 560×560-images (Jpeg format, QF=75); M_3 - 1000×1000-images (Tif format), obtained with a non-professional camera; M_4 [33] - 1000×1000-images (Jpeg format). Steganography methods considered were the LSB method and the method [9], which does not systematically preserve the reliability of the perception of the steganographic message, but today remains one of the most resistant to compression attacks.

The obtained results, in general, correspond to theoretical expectations: in most cases, there is a correspondence between the largest SSIM value and the smallest value of the container's NGMSN.

If this does not even occur, as, for example, (Fig. 4a, c), then the SSIM value obtained for the smallest NGMSN differs slightly from the maximum possible for the set (the maximum such deviation was recorded in the case of the set M_3, which was 0.3%), since in each case a monotonic decrease in the trend of the function is observed $SSIM = f(\text{NGMSN})$ (Fig. 4, 5a) regardless of either the format of the digital image or the steganography method used when obtaining the steganographic message.

But the strict decrease of the function $SSIM = f(\text{NGMSN})$, which could be expected in accordance with theoretical considerations, is absent. There are several reasons for this. First, none of the quantitative indicators used at all gives an accurate mathematical estimate of the characteristic of the digital image for which it is used: NGMSN - for the contribution of the high-frequency component; SSIM - for the visual quality of the perception of the digital image by a person, but within the limits of the problem considered in the work, such estimates cannot be quantitatively accurate at all. Indeed, it is fundamentally impossible to assess the contribution of the high-frequency component in the digital image absolutely accurately even in the frequency domain, since there is no exact distribution of frequency coefficients into high-, medium- and low-frequency ones. And any agreement here regarding where to "draw the line" between the medium- and high-frequency coefficients cannot be general in principle, taking into account the difference in the size of the image, the diversity of the digital image in any of its characteristics, in particular, texture. Secondly, SSIM depends on the brightness, contrast, and structural similarity of the image, and the NGMSN is mainly focused on the structure component. But when embedding the additional information, changes occur in both the brightness and contrast of the digital image. Thirdly, the process of (slight) blurring of the digital image leads to an increase in the NGMSN [24], but SSIM can remain practically unchanged if the structure of the digital image is not destroyed. This is very clearly visible on the corresponding images, which, displaying the same scene, differ only in the storage format (with/without loss). It is precisely due to the above reasons that there is a possibility of obtaining not the largest possible SSIM value for the set of images for the image with the smallest NGMSN, as well as a "scatter" at close arguments (NGMSN values) of points on the graphs (SSIM values) in Figure 4, 5a.

Special attention in the experimental part of the study was paid to the CGs generated using artificial intelligence, as their use as containers is spreading every day. It is obvious that such CGs have a different statistical structure, a detailed study of which is devoted to the authors' article, which is currently being prepared for publication, but the nature of the relationship between SSIM and NGMSN remain the same (Fig. 5b), where artificially generated digital images from the database [34] are used.

Thus, the existing relationship between SSIM and NGMSN is theoretically justified and practically confirmed: the trend of the function $SSIM = f(\text{NGMSN})$decreases with the increase in NGMSN, which indicates the prospects of using NGMSN as a criterion for a priori container selection to improve the visual quality of the steganographic message and is confirmed by the results of the computational experiment (Fig. 4 and 5) and the results of subjective ranking, typical of which are shown in Figure 6. Here, the set M_3 was deliberately taken, where the largest deviation from the maximum possible SSIM set in the digital image was obtained, which had the smallest NGMSN. As can be seen, even when the obtained steganographic message based on the selected container does not have the maximum SSIM value, the quality of perception of such a steganographic message is better than under the same conditions of the obtained steganographic message based on a randomly selected container in M_3, where there are obvious artifacts.

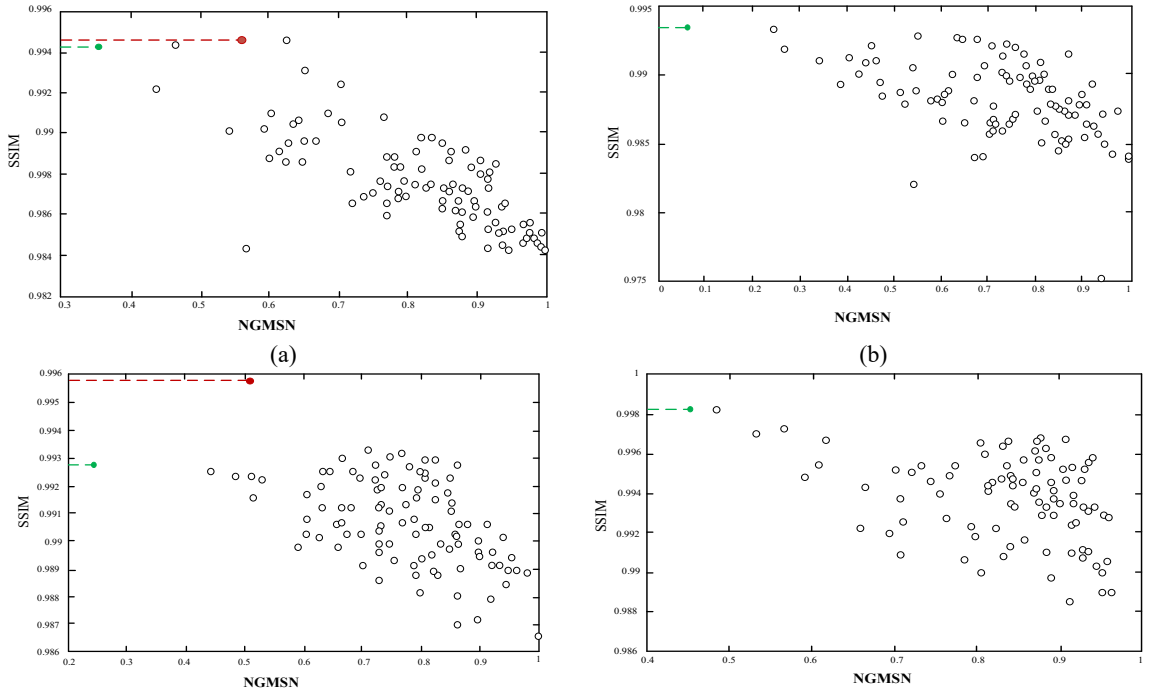


Figure 4: Dependence of the SSIM indicator, which characterizes the visual distortion of the container as a result of steganography by the LSB method with a hidden channel bandwidth of 1 bit/pixel, on the NGMSN of the container matrix for the digital image from the set: a) - M_1; b) - M_2; c) - M_3; d) - M_4.

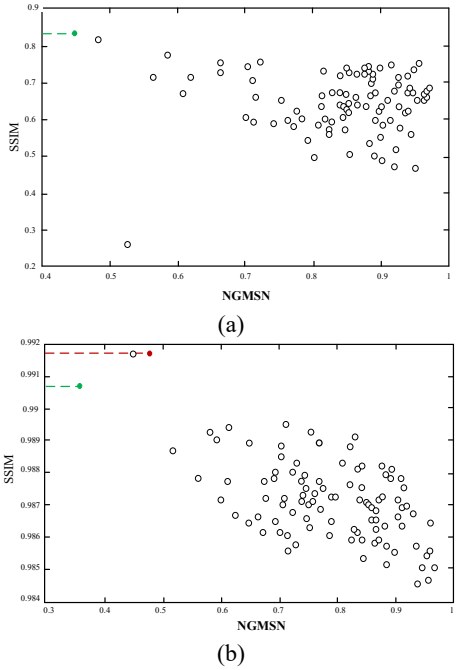


Figure 5: Dependence of SSIM on the container's NGMSN: a) - STP by the method [9], in which all blocks were used, for the digital image with M_4; b) - STP by the LSB method with a hidden channel bandwidth of 1 bit/pixel for artificially generated central processing units.

For practical application of the obtained theoretical results - construction of the corresponding selective method, it is necessary to reduce the degree of "dispersion" of SSIM values at close values of the NGMSN. This can be done using a weighting coefficient, which is determined in a certain way for the correction of the NGMSN, which is the direction of further work of the authors.



(a) (b)

Figure 6: Steganographic messages formed by the steganography method [9] when using all blocks based on a container from the set M_3: a) - with the smallest NGMSN; b) - randomly selected.

5 CONCLUSIONS

The work solves an important scientific problem of developing theoretical foundations of an approach to selecting a container from a given set of digital images candidates, general in the sense of independence from the steganographic algorithm used, format (with/without loss) to improve the visual quality of the steganographic message. In the course of developing the specified theoretical foundations, the following results were obtained:

1) The feasibility of using the NGMSN container matrix as a criterion for selecting it is theoretically justified. Priority is given to the digital image with the lowest NGMSN values;
2) The feasibility of using the SSIM metric for quantifying visual distortions of a container as a result of steganographic transformation is substantiated;
3) The correlation of the selected selection criterion with the structural stability of the digital image matrix and, as a result, the mutual relationship with the visual distortion indicator of the SSIM container is substantiated.

Studies of the practical dependence of the visual quality of the steganographic message on the container's NGMSN showed that using the digital image with the lowest NGMSN among the candidates allows obtaining the best/close to the best SSIM indicator (the maximum deviation from the best value within the experiment was 0.3%), regardless of the steganographic algorithm used and the format of the candidates. Using subjective ranking, it was established that the visual quality of the steganographic message obtained on the basis of selected containers with the lowest NGMSN, compared to random ones, was improved.

REFERENCES

[1] ENISA, ENISA Threat Landscape 2024, European Union Agency for Cybersecurity, 2024, [Online]. Available: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2024.

[2] International Telecommunication Union (ITU), Global Cybersecurity Index 2024, 5th ed., Geneva, 2024.

[3] A. Mishra et al., "Recent Advances in Image Steganography and Steganalysis: A Comprehensive Review," Archives of Computational Methods in Engineering, 2024, [Online]. Available: https://doi.org/10.1007/s11831-023-10041-w.

[4] H. Raj and G. Bhaumik, "A comprehensive survey of image steganography: From traditional vision techniques to deep learning paradigms - Trends, challenges, and applications," Computer Science Review, vol. 60, p. 100892, 2026, [Online]. Available: https://doi.org/10.1016/j.cosrev.2026.100892.

[5] A. K. Sahu and G. Swain, "Digital Image Steganography: A Survey on Recent Developments," Journal of King Saudi University - Computer and Information Sciences, vol. 35, 2023, [Online]. Available: https://doi.org/10.1016/j.jksuci.2023.101672.

[6] I. Bobok, A. Kobozieva, and S. Sokalsky, "The Problem of Choosing a Steganographic Container in the Conditions of Attacks against an Embedded Message," Problems Energetics Regionale, no. 4(56), pp. 74-88, 2022, [Online]. Available: https://journal.ie.asm.md/ru/contents/electronni-jurnal-456-2022.

[7] F. Ridzuan et al., "Cover Selection in Steganography: A Systematic Literature Review," Journal of Advanced Research in Applied Sciences and Engineering Technology, vol. 52, no. 2, pp. 107-129, 2025, [Online]. Available: https://doi.org/10.37934/araset.52.2.107129.

[8] I. Bobok et al., "Improving Steganosystem Efficiency as an Integral Part of Ensuring the Routine Operation of Energy Infrastructure Facilities," Problems Energetics Regionale, no. 2(70), pp. 78-93, 2026, [Online]. Available: https://journal.ie.asm.md/ru/contents/electronni-jurnal-270-2026.

[9] M. A. Melnyk, "Steganographic algorithm, resistant to compression," Information Security, no. 2(8), pp. 99-106, 2012.

[10] N. Hamid et al., "Enhancing visual quality of Spatial image steganography using SqueezeNet deep learning network," Multimedia Tools and Applications, vol. 80, no. 28, pp. 36093-36109, 2021, [Online]. Available: https://doi.org/10.1007/s11042-021-11315-y.

[11] H. Sajedi and M. Jamzad, "Evolutionary rule generation for signature-based cover selection steganography," Neural Netw. World, vol. 20, pp. 297-316, 2009, [Online]. Available: https://scholar.google.com/scholar?q=H.+Sajedi%2C+M.+Jamzad%2C+Evolutionary+rule+generation+for+signature-based+cover+selection+steganography%2C+Neural+Netw+World%2C+20+%282009%29+297-316.

[12] V. Hajduk and D. Levický, "Cover Selection Steganography with Intra-Image Scanning," in Proceedings of Radioelektronika, 2018, pp. 1-4, [Online]. Available: https://doi.org/10.1109/RADIOELEK.2018.8376370.

[13] Abed et al., "Efficient cover image selection based on spatial block analysis and DCT embedding," EURASIP Journal on Image and Video Processing, vol. 2019, no. 87, 2019, [Online]. Available: https://doi.org/10.1186/s13640-019-0486-8.

[14] P. D. Shah et al., "Genetic Algorithm based Approach this Select Suitable Cover Image for Image Steganography," in 2020 International Conference for Emerging Technology (INCET), 2020, [Online]. Available: https://doi.org/10.1109/INCET49848.2020.9154032.

[15] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility this Structural Similarity," IEEE

Transactions on Image Processing, vol. 13, no. 4, 2004, [Online]. Available: https://www.cns.nyu.edu/pub/lcv/wang03-preprint.pdf.

[16] S. Nazari and M. S. Moin, "Cover selection steganography via run length matrix and human visual system," pp. 131-138, 2013, doi: 10.7508/jist.2013.02.007.

[17] M. R. D. Molato et al., "Cover Image Selection Technique for Secured LSB-based Image Steganography," in ACAI '18: Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence, article no. 17, pp. 1-6, [Online]. Available: https://doi.org/10.1145/3302425.3302456.

[18] M. S. Subhedar, "Cover selection technique for secure transform domain image steganography," Iran Journal of Computer Science, vol. 4, pp. 241-252, 2021, [Online]. Available: https://link.springer.com/article/10.1007/s42044-020-00077-9.

[19] M. M. Brailovskyi et al., Analysis of cyber security of information systems: monograph, Kyiv: FOP Yamchynskyi O. V., 2021.

[20] A. Kobozieva and A. Sokolov, "The Sufficient Condition for Ensuring the Reliability of Perception of the Steganography Message in the Walsh-Hadamard Transform Domain," Problems Energetics Regionale, no. 2(54), 2022, [Online]. Available: https://doi.org/10.52254/1857-0070.2022.2-54.08.

[21] I. Bobok and A. Kobozieva, "Development of the Theoretical Approach this Analyzing the State of Information Protection Systems Based on Matrix Theory," Problems Energetics Regionale, no. 3(63), pp. 29-43, 2024, [Online]. Available: https://journal.ie.asm.md/ru/contents/electronni-jurnal-363-2024.

[22] I. Bobok and A. Kobozieva, "Theoretical Foundations of Digital Content Integrity Expertise," Problems Energetics Regionale, no. 1(65), pp. 105-120, 2025, [Online]. Available: https://journal.ie.asm.md/ru/contents/electronni-jurnal-165-2025.

[23] M. Chen et al., "HLTD-CSA: Cover selection algorithm based on hybrid local texture descriptor for color image steganography," Journal of Visual Communication and Image Representation, vol. 89, pp. 451-464, 2022, [Online]. Available: https://doi.org/10.1016/j.jvcir.2022.103646.

[24] A. A. Kobozeva, I. I. Bobok, and N. I. Kushnirenko, "Method for Distinguishing the Digital Images in Different Formats," Problems Energetics Regionale, no. 1(53), pp. 109-124, 2022, [Online]. Available: https://journal.ie.asm.md/ru/contents/electronni-jurnal-153-2022.

[25] J. Demmel, Applied Numerical Linear Algebra, SIAM, 1997.

[26] R. C. Gonzalez and R. E. Woods, Digital Image Processing, Hoboken: Prentice Hall, 2007.

[27] G. F. Konakhovich, D. O. Progonov, and O. Yu. Puzyrenko, Computer Steganography Processing and Analysis of Multimedia Data, Kyiv, 2018.

[28] N. M. Al-Aidroos et al., "Image Steganography Based on LSB Matching and Image Enlargement," in 2019 First International Conference of Intelligent Computing and Engineering (ICOICE), 2019, [Online]. Available: https://doi.org/10.1109/ICOICE48418.2019.9035172.

[29] G.-H. Chen et al., "Gradient-Based Structural Similarity for Image Quality Assessment," in 2006 International Conference on Image Processing, 2006, [Online]. Available: https://doi.org/10.1109/ICIP.2006.313132.

[30] M. Rouhani et al., "A watermarking method based on optimizing SSIM index by using PSO in DCT domain," in 14 International CSI Conference (CSICC2009), 2009, [Online]. Available: https://doi.org/10.1109/CSICC.2009.5349616.

[31] Y. Hsu and S. Chang, "Detecting image splicing using geometry invariants and camera characteristics consistency," in 2006 IEEE International Conference on Multimedia and Expo, Toronto, pp. 549-552, 2006.

[32] T. Gloe and R. Böhme, "The 'Dresden Image Database' for benchmarking digital image forensics," in Proceedings of the 2010 ACM Symposium on Applied Computing (SAC '10), New York, pp. 1585-1591, 2010.

[33] NRCS Photo Gallery, [Online]. Available: https://www.nrcs.usda.gov/wps/portal/nrcs/main/national/newsroom/multimedia/.

[34] Q. Peng et al., "AI-Generated Image Dataset," 2025, [Online]. Available: https://doi.org/10.7910/DVN/MGJPBL.

AI-Oriented Monitoring Agent in Network Infrastructure of Multicloud Data Centers

Andrii Raichuk and Larysa Globa
Educational and Research Institute of Telecommunication Systems, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Industrial Lane, 2, 03056 Kyiv, Ukraine
raaukr@gmail.com

Keywords: Intelligent Network Monitoring, Data Centers, Event Log Analysis, Log Processing, Network Failures, Automated Diagnostics, AI Assistant.

Abstract: The growing complexity of multicloud data center infrastructures, together with the rapid increase in telemetry streams, event logs, and protocol notifications, makes timely incident detection and fault localization a critical problem for modern telecommunication systems. Conventional monitoring approaches mainly rely on the analysis of large historical logs, which limits their ability to identify short-term correlated failures and support rapid operational decision-making. The aim of this paper is to develop a software-oriented architecture for intelligent monitoring of data center network infrastructure based on localized telemetry buffering, centralized event correlation, and AI-assisted diagnostics. The proposed method combines local short-term telemetry snapshots formed on network and server nodes, critical-event clustering within a limited time window, root cause analysis using correlation of protocol, resource, and trap data, and risk estimation through a machine learning model. As a result, a five-layer monitoring architecture and an incident-processing algorithm were developed, including local agents, temporary telemetry buffers, a central correlation node, a failure knowledge base, and an AI assistant for troubleshooting support. The results show that the proposed approach reduces the amount of data requiring urgent processing, improves the identification of causally related events, and decreases the time needed to interpret critical incidents. The practical relevance of the study lies in the possibility of deploying the system as a software layer over existing monitoring and logging infrastructure without hardware modification. The theoretical relevance lies in the formalization of localized incident-context modeling, critical-event clustering, and AI-assisted root cause identification for multicloud data center environments.

1 INTRODUCTION

Modern data centers and telecommunication networks operate in an environment of continuously growing volumes of telemetry, logs, service events, and different types of messages. In multicloud infrastructures, this problem is intensified by the distributed nature of resources, the asynchrony of events, and complex cause-and-effect relationships between network and server components. Under such conditions, the traditional approach in which an operator or a monitoring system works with large-volume event logs is increasingly proving insufficient for rapid fault localization, since an incident is usually accompanied by an avalanche-like increase in the number of secondary messages, which complicates the identification of the root cause [1]-[5].

In data center network systems, the situation becomes particularly complex when a failure develops not in isolation, but through a cascade of interrelated changes: link degradation leads to the loss of routing protocol sessions, control plane overload, repeated route recalculation attempts, and the appearance of numerous trap messages. Under such conditions, the mere presence of a complete event log does not guarantee rapid diagnostics, because the practical problem lies not in the absence of data, but in their excess and insufficient structure for rapid decision-making [2], [4], [6].

At the same time, modern telecommunication systems are actively developing approaches related to AIOps, automated log analysis, anomaly detection, and root cause identification. Individual studies demonstrate the high efficiency of using machine learning for real-time telemetry processing, deep

learning for log analysis, as well as interpretable root cause analysis models for microservice systems [3], [7]-[9]. However, most such solutions either focus on a single data modality, or require significant computational resources, or do not take into account the operational specifics of the data center network environment, where response time, explainability of conclusions, and the possibility of software deployment without modifying the hardware infrastructure are critical [3], [7], [8],[10].

Given this, the task of building such a monitoring system is relevant, one that, on the one hand, does not duplicate the full analysis of all logs, and on the other hand ensures the rapid formation of incident context exactly when it is most needed. The aim of this work is to develop a software-oriented architecture of localized monitoring in which each network or server element forms a short-term telemetry buffer for the last N minutes, and in the event of a cluster of critical events such snapshots are transmitted to a centralized correlation node, where together with a knowledge base of failures and an AI assistant they are used to accelerate fault localization and troubleshooting. Such an approach makes it possible to reduce analysis time and decrease the volume of data that must be transmitted for processing and for the intelligent support of the engineer during incident response.

2 ANALYSIS OF RELATED RESEARCH

In the scientific literature, the problem of intelligent monitoring is divided into several interrelated areas. The first area covers fault management models in telecommunication networks. The survey in [1] on machine learning for network fault management shows that ML approaches are actively used for failure classification, anomaly detection, and failure prediction, while the results largely depend on the completeness of the training datasets, the quality of features, and the specifics of the application domain. An important conclusion of this line of research is that most solutions optimize a separate stage of the fault management chain - detection, classification, or prediction - but more rarely integrate these functions into a single operational mechanism for improving network fault tolerance [1], [11].

The second area concerns the analysis of event logs and system logs. A survey of deep learning methods for anomaly detection in log data demonstrates that modern methods based on LSTM, Transformer, and other sequential architectures significantly outperform classical statistical approaches in detecting complex patterns of anomalous behavior [2]. At the same time, the same survey emphasizes that log-oriented models depend on the quality of parsing, are highly sensitive to changes in message formats, and often transfer poorly to new environments without additional retraining [2]. Newer models, including LogLLM, attempt to reduce dependence on rigid template parsing and use the semantics of messages based on large language models, but this increases system complexity and the cost of operation [6]. Therefore, although AI-based log analysis is developing rapidly, the problem of rapid operational interpretation of an incident in a real data center environment remains open [2], [6].

The third area is related to real-time processing and streaming telemetry. In [3], anomaly detection in the telemetry of server farms is performed, showing that real-time operation is a critical requirement, that is, high model accuracy loses practical value if the detection time does not allow the operator to react in a timely manner. The authors propose special pre-processing of telemetry to reduce anomaly detection delay and prove that speed is a necessity rather than merely a derivative of classification accuracy [3]. For the architecture proposed in this paper, this observation is fundamental: the monitoring system should focus not on the maximum accumulation of historical data, but on the formation of a short yet informative incident context.

The fourth area covers root cause analysis in distributed systems and microservices. The LogRCA (Log Root Cause Analysis) and Chain-of-Event studies show that it is insufficient to simply isolate anomalous log lines in order to identify the root cause; it is much more important to find the minimal causally related set of events that explains the failure [4], [7]. It is especially important that the Chain-of-Event model uses causal relationships between events from multimodal sources and emphasizes the need for interpretability of the result for the engineer [7]. A survey of RCA (Root Cause Analysis) in microservices also confirms that one of the main challenges remains the integration of logs, metrics, traces, and domain expert information into a single data block for obtaining analysis results [5]. Thisconclusion directly leads to the need to use a knowledge base of failures and its processing by an AI assistant in the proposed architecture [4], [5], [7].

The fifth area is associated with the application of AI and its suitability for use in data centers. A study devoted to the automation of data center operations shows that even accurate ML models for predictive maintenance and other tasks lose part of their

practical value without an interpretative explanation for the operator [8]. This means that the engineer needs not only a signal about a danger, but also a reasoned supporting instruction: what exactly happened, which parameters are decisive, and which actions should be performed first. Therefore, within the proposed approach, the AI assistant is considered not as an autonomous mechanism for making decisions instead of the engineer, but as a tool for the intelligent support of diagnostics [8].

In summary, it can be concluded that modern studies sufficiently cover individual components of the problem - anomalies, correlation, RCA, AI support, and prediction - but do not fully take into account the specifics of the data center network environment, where large event logs effectively turn into a bottleneck during a failure. In most papers [1]-[5], [7], [8], [11], [12], the approach in which analysis starts not from a full archive but from a compact localized buffer of the node's latest state is insufficiently disclosed; situations of planned maintenance, when the interpretation of events must be deliberately changed or temporarily suppressed, are also rarely modeled. It is precisely these gaps that form the scientific novelty of the proposed study.

3 PROBLEM STATEMENT

The scientific problem considered in this work lies in the contradiction between the growth in the volume of events in the network infrastructure of data centers and the need to reduce fault localization time. On the one hand, modern systems accumulate huge arrays of data in centralized logs, which theoretically increases the completeness of the incident picture. On the other hand, the excessive amount and temporal-structural heterogeneity of these data complicate the search for the root cause at the moment when a decision must be made immediately. Therefore, the objective of this work is to develop a monitoring system architecture that reduces the amount of data subject to urgent processing without losing causally significant information.

Within this problem, three interrelated properties must be ensured. First, the local node must preserve sufficient telemetry context over a short time interval in order to reflect the dynamics of the incident up to the moment of its occurrence. Second, the central system must be able to determine exactly when these local snapshots need to be requested and combined. Third, the engineer must receive not a raw set of trap messages, but interpreted diagnostic support in the form of causes, possible consequences, and recommended actions. That is why the research task is formulated as the construction of an AI-oriented agent for localized monitoring and event correlation that combines buffering of node states, centralized correlation of clusters of critical events, and intelligent support for the fault tolerance of data center infrastructure.

4 ANALYSIS OF RELATED RESEARCH

The proposed architecture consists of five logically interconnected layers (Table 1). The first layer is formed by local monitoring agents deployed on network and server elements. Each agent collects key telemetry information and forms a temporary state snapshot file for the last N minutes. Such a snapshot includes event timestamps, types of trap messages, severity levels, the states of BGP, OSPF, LDP, and BFD sessions, interface statistics, CPU and memory parameters, as well as service control messages. For example, for a network node, a snapshot may contain a record of an LDP session loss on a specific interface, accompanied by increased CPU utilization and repeated restarts of an adjacent protocol process.

Table 1: Architecture of the AI-based data center monitoring system.

Layer	Components
Layer 1 - Network Infrastructure	Network devices (routers / switches)
Layer 2 - Local Monitoring Agents	Servers / Virtual nodes
Layer 3 - Temporary Telemetry Buffer	Local monitoring agent
Layer 4 - Central Correlation Node	Telemetry collection
Layer 5 - AI Assistant Layer	Trap monitoring

The second layer is the local temporary buffer. Its fundamental difference from the full system log is that it is not an archive but performs the function of an operational context. After a new snapshot is accumulated, the previous buffer may be overwritten or deleted, while the full history continues to be stored in the main logging file. Such an organization of the process reduces the overhead of urgent analysis, because in the event of an incident there is no need to urgently review the entire history of logs;

it is sufficient to collect the latest relevant context from several nodes.

The third layer consists of the data transmission activation mechanism. The system does not send local buffers to the central node continuously; transmission is initiated only when a cluster of critical events is detected. A cluster in this case is understood as the appearance of several correlated incidents within a short time in one or adjacent sections of the network. This logic makes it possible to significantly reduce the traffic of diagnostic data while at the same time preserving the ability to respond quickly.

The fourth layer is the centralized correlation node. Here, local N-minute snapshots from different elements are compared by time, event types, logical and topological relationships. The task of this node is not only to detect the fact of a failure, but also to reduce uncertainty about its root cause. Given the findings of RCA research, correlation should rely not on simple counting of messages, but on the search for causally connected chains of events [4], [5], [7], [13].

The fifth layer is the engineer's AI assistant. It uses a structured knowledge base of typical failures, historical event patterns, and machine learning mechanisms to generate supporting instructions. In the simplest case, the AI assistant correlates the current incident with known patterns such as "LDP Session Down + increased CPU load + interface flapping", and in a more complex case performs contextual generalization, offering the engineer a probable cause, a list of initial checks, and the recommended sequence of actions. If the engineer is working in planned maintenance mode and expects an impact on the network, AI analysis may be disabled or switched to passive mode in order to avoid false interpretations of events.

5 MATHEMATICAL FORMULATIONS

The network infrastructure described by the set of nodes $N = \{n_1, n_2, \ldots, n_M\}$, where each element corresponds to a router, switch, server, or virtual node. For node n_i at discrete time instants t, a state vector is formed that contains only the parameters required for local incident assessment and subsequent centralized correlation. The state vector includes the event timestamp, trap type, severity level, aggregated protocol state, normalized resource-load indicators, interface statistics, and service messages. This description makes it possible to transform raw telemetry data into a formalized node model.

Since the system uses a short-term local context, each node maintains a sliding buffer of duration Tb minutes. If Δt denotes the telemetry sampling interval, then the number of samples K in the buffer is defined by:

$$K = \frac{T_b}{\Delta t}. \quad (1)$$

The local buffer of node n_i is represented as the sequence of the latest K observations. Such a model preserves only the most relevant short-term context and avoids continuous processing of the full event log. The formalized state vector of node ni is defined in (2). For clarity, the interpretation of each component included in the vector is presented in Table 2.

$$x_{i(t)} = (\tau_i, \psi_i, s_i, p_i, c_i, m_i, q_i, u_i). \quad (2)$$

Table 2: Explanation of variables in formula (2).

Component	Explanation
τ_i	Time-related characteristic of node i, for example a timestamped temporal feature or a delay-related indicator.
ψ_i	Contextual or event-related descriptor associated with node i.
s_i	General state or status indicator of node i.
p_i	Protocol- or process-related feature of node i.
c_i	Control or computational characteristic of node i.
m_i	Memory-related characteristic of node i.
q_i	Queue-related metric, such as queue occupancy, backlog, or buffering state.
u_i	Utilization or load indicator of node i.

Temporal buffer or historical slice associated with node i at time t:

$$B_{i(t)} = \{x_{i(t-k\Delta t)}\}_{k=0}^{K-1}. \quad (3)$$

For primary local event filtering, the integrated criticality score $C_i(t)$ is introduced. It combines the normalized trap severity, protocol degradation, resource load, and interface-error intensity. The weights α, β, γ, and δ are selected so that their sum is equal to one.

$$C_{i(t)} = \alpha \cdot sev_{i(t)} + \beta \cdot p_{i(t)} + \gamma \cdot r_{i(t)} + \delta \cdot e_{i(t)}. \quad (4)$$

The protocol component pi(t) evaluates the loss of the key BGP, OSPF, LDP, and BFD sessions by means of indicator functions.

$$p_{i(t)} = \frac{I_{BGP_{i(t)}} + I_{OSPF_{i(t)}} + I_{LDP_{i(t)}} + I_{BFD_{i(t)}}}{4}. \quad (5)$$

The resource component $r_i(t)$ aggregates normalized CPU, memory, and traffic utilization, where $ci(t)=CPU_i(t)/100$, $m_i(t)=MEM_i(t)/100$, and $tri(t)=TR_i(t)/TR_{imax}$.

$$r_{i(t)} = \frac{c_{i(t)}+m_{i(t)}+tr_{i(t)}}{3}. \quad (6)$$

An event at node n_i is considered critical if the integrated score exceeds the threshold θ_c. In this case, the local agent marks the corresponding snapshot as relevant for transmission to the centralized node.

$$C_{i(t)} \geq \theta_c. \quad (7)$$

At the centralized level, the system processes not individual events but event sets observed in the interval $[t - T_w, t]$. The set of nodes with critical events defines the incident zone, while the incident cluster is formed only if the thresholds on both the number of events and the number of affected nodes are satisfied.

$$N_{c(t)} = \{n_i \in \mathcal{N} : \exists \tau \in [t - T_w, t], C_{i(\tau)} \geq \theta_c\}. \quad (8)$$

$$N_{evt(t)} \geq \theta_e \wedge |N_{c(t)}| \geq \theta_n. \quad (9)$$

For correlation, an integrated association measure between two events is used; it accounts for temporal proximity, topological adjacency, and semantic similarity of symptoms. This allows events belonging to the same cascade pattern to be grouped within a single incident context.

$$\kappa(e_a, e_b) = \lambda_1 C_{time} + \lambda_2 C_{topo} + \lambda_3 C_{sem}. \quad (10)$$

The root-cause search problem is formulated as the selection of the most probable cause from the set of possible causes R on the basis of the correlated symptom set E_t. In the general case, this formulation is expressed through posterior-probability maximization and Bayes' rule.

$$r_{opt} = argmax_{(r \in R) P(r|E_t)}. \quad (11)$$

$$P(r|E_t) = \frac{P(E_t|r)P(r)}{P(E_t)}. \quad (12)$$

Failure risk is estimated using logistic regression, where z(t) is the feature vector of the current state. The feature set includes protocol-state indicators, resource-load variables, interface-error counters, and the integrated event criticality score. The resulting failure probability is converted into a percentage risk score, which is used to signal increased network vulnerability even before a full incident cluster is formed.

$$P_{fail(t)} = \frac{1}{1+exp\left(-\left(w^T z(t)+b\right)\right)}. \quad (13)$$

$$Risk(t) = 100 \cdot P_{fail(t)}. \quad (14)$$

The mathematical description presented above formalizes the key components of the system: node-state representation, local telemetry buffering, criticality estimation, incident-cluster detection, centralized correlation, root-cause evaluation, and failure-risk prediction. This provides a rigorous formal basis for the software implementation of the intelligent monitoring system for data center network infrastructure.

6 SYSTEM OPERATION ALGORITHM

The algorithm of system operation begins with local telemetry collection on each element (Fig. 1). Unlike the continuous streaming of all data to the center, the agent selects only those parameters that directly reflect the functional state of the node. This makes it possible to preserve the local context in a compact form. After that, a temporary N-minute snapshot is formed, which is a copy of the most important parameters of the latest time interval. When a new snapshot appears, the old one is deleted or archived locally, while the main log file continues to accumulate the complete history.

At the next stage, the local agent performs a primary assessment of event criticality. If an event or a set of events belongs to the critical, alert, or error levels and appears on interrelated elements, the system initiates the transmission of local snapshots to the centralized node. Here, the advantage of the proposed approach lies in the fact that not the entire log is transmitted, but only a short relevant context sufficient for RCA. Thus, the time required to prepare the material for analysis is minimized already at the data collection stage.

After the data arrive at the centralized node, correlation analysis is performed. Its content consists in aligning time sequences and identifying logically related changes in protocol states, interface overload, growth of system resource consumption, and service trap messages. If, for example, packet loss, degradation of BFD status, and disappearance of BGP sessions are simultaneously observed on several nodes, the system should interpret this not as a set of independent symptoms, but as a common incident with a high probability of damage to the transport segment or degradation of adjacent active equipment.

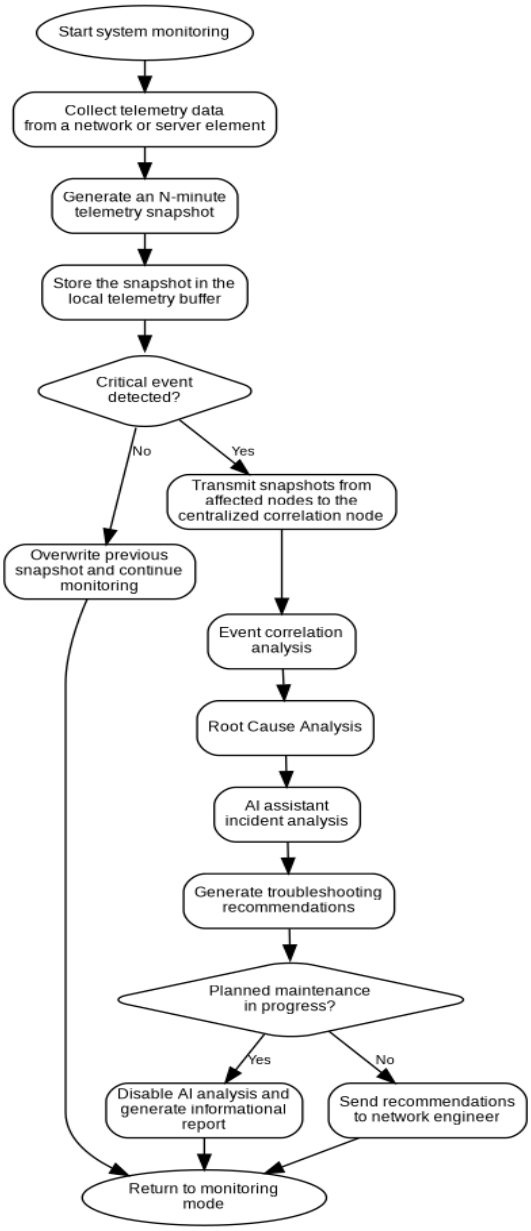


Figure 1: Operating algorithm of the incident analysis system based on telemetry.

The final stage is the work of the AI assistant. Based on the results of correlation and the knowledge base, it generates a short but technically meaningful conclusion. The conclusion must answer at least three questions: what is the most probable root cause, which checks should be performed first, and what consequences may arise if the incident is not isolated. Thus, the AI assistant transforms raw telemetry data into a recommendation for troubleshooting.

7 ADVANTAGES OF THE PROPOSED APPROACH

The main advantage is the reduction in incident analysis time due to shifting the focus from the complete historical log to a localized buffer of the latest state. In practice, this means that at the moment of an incident the engineer or analytical module works not with gigabytes of data, but with compact telemetry snapshots only from those nodes that are directly involved in the incident. This reduces both computational costs and the workload on personnel.

The second advantage lies in increasing the accuracy of event correlation. When the system receives a limited but temporally coherent context over the last N time intervals from several affected nodes, the correlation mechanism obtains a much cleaner basis for building cause-and-effect relationships than in the case of chaotic searching through the entire event log over a long period Ths is particularly important for cascading failure scenarios, where secondary symptoms quickly mask the root cause [4], [7], [12].

The third advantage is related to support for the engineer. In many modern approaches, the operator is effectively overloaded either with raw logs or with an overly abstract indicator that “an anomaly has been detected”. By contrast, the proposed AI assistant is oriented toward generating explainable supporting instructions tied to specific types of trap messages, protocol states, and resource symptoms. Such an organization is consistent with the modern trend toward explainable AIOps and increases the real operational value of the system [8], [11].

Finally, an important advantage is a higher degree of automation of the fault diagnosis process in the data center environment. Unlike approaches that require specialized hardware additions or substantial changes to network topology, the proposed system can be deployed as a software layer on top of the existing monitoring, logging, and network control infrastructure. This makes it particularly attractive for multicloud environments, where changes to the hardware platform are expensive or organizationally difficult.

8 RESULTS

The developed concept and its software prototype show that the localization of diagnostic information in the form of short-term buffers creates a fundamentally different mode of incident response.

If, in traditional practice, incident investigation begins with a centralized log of large volume, then in the proposed approach the system operates with a reduced but contextually relevant set of data. This means that the first step of diagnostics is performed faster, and the probability of losing a causally important fragment of information is lower than in the case of selective manual review of logs.

In terms of operational efficiency, this directly affects MTTR (Mean Time To Repair / Mean Time To Restore), that is, the average time required for recovery after a failure. The reduction of MTTR is achieved not only through faster data processing, but also through better support for decision-making. The engineer does not spend additional time manually collecting symptoms from different nodes, because the system already aggregates local snapshots and forms a preliminary interpretation of the incident. In cases of typical failures, the AI assistant effectively shortens the path from "identify what happened" to"check these specific points", which has a direct practical effect for NOC/SRE teams.

At the same time, the results should be considered critically. First, the effectiveness of correlation depends on the quality of the rules and the knowledge base model. Second, in complex multilayer failures some symptoms may fall outside the N-minute window, and therefore the choice of buffer duration requires further optimization for different types of network infrastructure. Third, in order to fully move to a proactive mode, the system should be supplemented with models for load prediction and degradation probability, combining short-term local diagnostics with longer time horizons of analysis [3], [10], [11].

At the same time, the results should be considered critically. First, the effectiveness of correlation depends on the quality of the rules and the knowledge base model. Second, in complex multilayer failures some symptoms may fall outside the N-minute window, and therefore the choice of buffer duration requires further optimization for different types of network infrastructure. Third, in order to fully move to a proactive mode, the system should be supplemented with models for load prediction and degradation probability, combining short-term local diagnostics with longer time horizons of analysis [3], [10],[11].

For multicloud environments, the proposed architecture is particularly promising because it combines decentralized collection of node states with centralized analytics. Local agents can be deployed in different management domains, and the central correlation node can operate as an overlay on top of existing logging and monitoring systems. In such a configuration, the autonomy of individual infrastructure segments is preserved, while at the same time a unified logic of incident response is formed. This corresponds to current trends in the development of intelligent network management systems, where local sensitivity to events is combined with global coordination of decisions [1], [5], [11].

9 CONCLUSIONS

The article proposes a software-oriented approach to intelligent monitoring of the network infrastructure of data centers, based on the combination of local N-minute telemetry buffers, centralized correlation of critical events, and an AI assistant for engineer support. Unlike the traditional heavily log-oriented approach, the proposed architecture operates with a localized incident context, which makes it possible to reduce analysis time, decrease the load on log-processing subsystems, and improve the quality of root-cause identification.

The scientific novelty of the work lies in the formalization of an architecture in which a short-term local state buffer is considered as the primary unit of analytical incident processing, rather than merely as an auxiliary cache for monitoring. An additional contribution lies in the integration of the AI assistant with a knowledge base of typical failures into the logic of network diagnostics and in taking into account the operational scenario of planned maintenance, when automatic interpretation of events must be adapted to the context of the engineer's actions.

Further research should be aimed at extending the mechanisms of causal analysis, automatically replenishing the knowledge base from accumulated incidents, integrating models for load forecasting and risk of failure, as well as experimentally validating the approach using real network logs of multicloud infrastructures. It is in this direction that the transition from reactive monitoring systems to full-fledged intelligent platforms for proactive management of next-generation telecommunication networks can be envisioned.

REFERENCES

[1] M. Nouioua, P. Fournier-Viger, G. He, F. Nouioua, and M. Zhou, "A survey of machine learning for network fault management," in Machine Learning and Data Mining for Emerging Trend in Cyber Dynamics, Springer, 2021, [Online]. Available: https://doi.org/10.1007/978-3-030-66288-2_1.

[2] M. Landauer, S. Onder, F. Skopik, and M. Wurzenberger, “Deep learning for anomaly detection in log data: A survey,” Machine Learning with Applications, vol. 12, art. no. 100470, 2023, [Online]. Available: https://doi.org/10.1016/j.mlwa.2023.100470.

[3] D. L. Vajda, T. V. Do, T. Bérczes, and K. Farkas, “Machine learning-based real-time anomaly detection using data pre-processing in the telemetry of server farms,” Scientific Reports, vol. 14, art. no. 23288, 2024, [Online]. Available: https://doi.org/10.1038/s41598-024-72982-z.

[4] T. Wittkopp, P. Wiesner, and O. Kao, “LogRCA: Log-based root cause analysis for distributed services,” in Euro-Par 2024: Parallel Processing, LNCS, vol. 14802, Springer, 2024, pp. 362-376, [Online]. Available: https://doi.org/10.1007/978-3-031-69766-1_25.

[5] T. Wang and G. Qi, “A comprehensive survey on root cause analysis in (micro) services: Methodologies, challenges, and trends,” arXiv preprint, 2024, [Online]. Available: https://arxiv.org/abs/2408.00803.

[6] W. Guan, J. Cao, S. Qian, J. Gao, and C. Ouyang, “LogLLM: Log-based anomaly detection using large language models,” arXiv preprint, 2024, [Online]. Available: https://arxiv.org/abs/2411.08561.

[7] Z. Yao, C. Pei, et al., “Chain-of-Event: Interpretable root cause analysis for microservices through automatically learning weighted event causal graph,” in Proc. ACM Found. Softw. Eng. (FSE), 2024, [Online]. Available: https://doi.org/10.1145/3663529.3663827.

[8] Y. Gebreyesus et al., “AI for automating data center operations: Model explainability in the data centre context using SHAP,” Electronics, vol. 13, no. 9, art. no. 1628, 2024.

[9] P. Himler, M. Landauer, F. Skopik, and M. Wurzenberger, “Anomaly detection in log-event sequences: A federated deep learning approach and open challenges,” Machine Learning with Applications, vol. 16, art. no. 100554, 2024, [Online]. Available: https://doi.org/10.1016/j.mlwa.2024.100554.

[10] L. Pham, H. Ha, and H. Zhang, “Root cause analysis for microservices based on causal inference: How far are we?,” in Proc. IEEE/ACM Int. Conf. Automated Software Engineering (ASE), 2024, pp. 706-718.

[11] E. Edozie et al., “Artificial intelligence advances in anomaly detection for telecom networks,” Artificial Intelligence Review, 2025, [Online]. Available: https://doi.org/10.1007/s10462-025-11108-x.

[12] I. Kotenko, D. Gaifulina, and I. Saenko, “Systematic literature review of security event correlation methods,” IEEE Access, vol. 10, pp. 43387-43420, 2022, [Online]. Available: https://doi.org/10.1109/ACCESS.2022.3168976.

[13] H. Maosa et al., “A hierarchical security event correlation model for real-time detection,” Signals, vol. 4, no. 1, art. no. 4, 2024.

Hybrid AI-Based Path Loss Prediction Model for 5G/6G Networks

Serhii Siden[1], Roman Tsarov[1], Dmytro Stepanov[1], Kateryna Shulakova[1,2] and Andrii Pavlov[1]
[1]*University of Intelligent Technologies and Telecommunications, Kuznechna Str. 1, 65023 Odesa, Ukraine*
[2]*Anhalt University of Applied Sciences, Bernburger Str. 57, 06366 Köthen, Germany*
ssiden@suitt.edu.ua, rcarev@gmail.com, dmstepanov@ukr.net, katejojo29@gmail.com, as.pavlov96@gmail.com

Keywords: 5G, 6G, Random Forest, XGBoost, Deep Neural Network, Artificial Intelligent, Propagation Models.

Abstract: Accurate path loss prediction is critical for the efficient planning and optimization of 5G and emerging 6G wireless networks, particularly in high-frequency millimeter wave (mmWave) bands. Traditional empirical models are limited in their ability to capture the complex and nonlinear characteristics of modern urban propagation environments. This paper proposes a hybrid machine learning framework that combines Random Forest, XGBoost, and deep neural networks to enhance prediction accuracy. The model utilizes a comprehensive set of input features, including distance, frequency, antenna heights, building density, and line-of-sight conditions, derived from a deterministic ray-tracing dataset. A weighted ensemble strategy is introduced to integrate the strengths of tree-based and deep learning models, enabling effective modeling of both discontinuous shadowing effects and smooth signal variations. Experimental results demonstrate that the proposed approach significantly outperforms classical models and individual machine learning methods, achieving an RMSE of 2.5 dB and an R^2 of 0.96. The results confirm the effectiveness of hybrid AI-based models for accurate path loss prediction and highlight their potential for next-generation wireless network design and optimization.

1 INTRODUCTION

The current stage of infocommunication technology development is characterized by a fundamental transformation in the architectural principles of mobile network design, driven by the deployment of the fifth-generation (5G) standard and the active conceptual development of sixth-generation (6G) networks. The primary target characteristics of these networks include ultra-high data transfer speeds (up to tens of Gbit/s), minimized latency (down to microseconds), and support for massive machine-type communication (mMTC). Achieving these goals requires the use of new frequency resources, particularly the millimeter wave (mmWave) band (above 24 GHz), and for 6G, the terahertz spectrum. However, the transition to high-frequency bands is accompanied by critical physical limitations, imposing new requirements on radio channel modeling and signal propagation prediction accuracy.

The effectiveness of 5G/6G network deployment and radio planning optimization depends directly on the relevance of path loss propagation models. In the millimeter wave band, radio signals experience significantly more intense attenuation compared to sub-6 GHz frequency bands. According to the Friis transmission equation [1], free-space path loss increases proportionally to the square of frequency, resulting in a small coverage radius for base stations and necessitating Massive MIMO beamforming technologies. mmWave signals also exhibit additional losses due to atmospheric absorption and hydrometeor scattering (rain, fog), and are highly sensitive to blockage by macro-objects (buildings) and dynamic obstacles (vehicles, human bodies). The presence of such deterministic and random factors makes path loss modeling an extremely complex task.

Historically, classical empirical models such as Okumura-Hata, COST-231 Hata, or more modern statistical approaches defined in 3GPP technical specifications (e.g., TR 38.901) have been used for path loss estimation. However, practical experience in designing modern urban networks has revealed the limitations of these methods. Most traditional models rely on approximations of field measurement results for limited scenarios and use simplified statistical approximations, failing to fully account for complex urban topology, dense urban multipath propagation, and nonlinear dependencies between environmental parameters and signal levels. Deterministic methods

such as Ray Tracing provide high accuracy but require substantial computational resources and detailed digital terrain models, which are often impractical for operational planning.

The development of artificial intelligence (AI) and machine learning (ML) methodologies provides fundamentally new opportunities for approximating the characteristics of wireless communication channels [2]–[6]. Unlike the rigid mathematical structures of empirical models, ML algorithms can identify hidden nonlinear patterns in large datasets obtained through field measurements or high-fidelity simulation. This enables adaptive predictive models that account for environmental specifics, building density, and climatic conditions with accuracy previously unattainable by analytical methods.

This paper presents results of a study on the application of intelligent data analysis for predicting signal path loss in 5G and future 6G networks. A hybrid ML model architecture is proposed, synergetically combining ensemble methods (gradient boosting on decision trees) with deep neural networks for processing spatial environmental characteristics. The scientific contributions are:

- A hybrid ML framework is developed - a multi-level path loss prediction structure integrating intelligent data analysis methods to enhance model reliability under non-stationary radio channel conditions.
- Comparative performance analysis - a systematic comparison of a wide range of ML algorithms (from linear regression to complex neural network architectures) with classical empirical models recommended by ITU-R and 3GPP, determining the applicability boundaries for each approach.
- Verification and validation of results - reliability confirmed through testing on realistic simulation datasets reproducing urban propagation conditions, demonstrating the superiority of the proposed approach over existing solutions.

2 PROBLEM FORMULATION

To evaluate millimeter-wave propagation characteristics in complex urban environments, a 2D deterministic approach based on ray tracing using the image method was employed, in which the resulting signal level depends on both the direct ray component and reflected ray components.

As known [7], [8], for the mmWave range it is important to account for radio signal attenuation in atmospheric gases. The level of radio signal path loss at the reception point is given by:

$$PL = PL_{FS} + (\gamma_{atm} + \gamma_{rain}) \cdot d_{km} + \sum_{i=1}^{N} L_{refl,i}\ , \quad (1)$$

where PL_{FS} is free-space path loss, γ_{atm} and γ_{rain} are losses in atmospheric gases and rain, d_{km} is the transmitter-receiver distance, L_{refl} k is the path loss of the i-th reflected ray, N is the number of reflected rays at the reception point.

The path loss prediction task is formulated as a supervised regression problem. Let $D = \{(x_i, y_i)\}_{i=1}^{M}$ be a dataset of N samples generated via ray tracing, where $y_i \in \mathbb{R}^n$ is the actual path loss in dB and $x_i \in \mathbb{R}^n$ is the feature vector:

$$x_i \in [d_i, f_i, C_{LOS}]. \quad (2)$$

C_{LOS} is a binary line-of-sight indicator, $C_{LOS} \in \{0,1\}$.

The goal is to train a mapping function $F: \mathbb{R}^n \rightarrow \mathbb{R}$, minimizing the mean squared error:

$$L(\theta) = \frac{1}{M} \sum_{i=1}^{M} \left(y_i - F(x_i;\theta)\right)^2 . \quad (3)$$

3 PROPOSED HYBRID MACHINE LEARNING METHODOLOGY

To balance bias and variance across complex urban features, this work proposes a hybrid ensemble framework combining tree-based algorithms with a deep neural network (DNN).

3.1 Random Forest

The Random Forest (RF) method is based on bagging and random feature subspaces. RF generates a large number of independent decision trees, each trained on a bootstrap sample, with each node split considering only a random feature subset. For regression, the final RF prediction is the arithmetic mean of all individual trees [9]:

$$\hat{y}_1 = \hat{y}_{RF}(x) = \frac{1}{M} \sum_{m=1}^{M} f_m(x), \quad (4)$$

where M is the number of trees and $f_m(x)$ is the m-th tree prediction for input x. Node splits use the variance minimization criterion (MSE). RF dramatically reduces model variance without significant increase in bias, as tree errors are weakly

correlated, making it extremely robust to noise and outliers. In urban radio environments, where a shift of just a few meters around a massive building causes a sharp signal drop, decision trees are ideal for modeling such discontinuous, step-like response surfaces.

3.2 Extreme Gradient Boosting

Unlike RF, XGBoost belongs to the family of Gradient Boosting Machines (GBM), building trees sequentially where each new tree compensates for the residuals of the previous ensemble. At iteration t, XGBoost finds the tree function f_t minimizing the objective function $L^{(t)}$ [10]:

$$L^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t), \quad (5)$$

where l is a differentiable convex loss function.

The explicit regularization term $\Omega(f_t)$ controls structural complexity and prevents overfitting:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2, \quad (6)$$

where:

- T is the number of leaves;
- ω_j are leaf weights;
- γ and λ are hyperparameters.

The final XGBoost prediction is the additive sum of all base trees:

$$\hat{y}_2 = \hat{y}_{XGB}(x) = \sum_{t=1}^{T} f_t(x). \quad (7)$$

XGBoost uses a second-order Taylor series approximation of the objective function, operating with both gradients and Hessians of the loss function. This allows it to detect extremely subtle nonlinear correlations between spatial coordinates and multipath interference effects, making it one of the most effective algorithms for structured tabular geo-data.

3.3 Deep Neural Network, DNN

Despite the power of tree-based algorithms, they generate piecewise-constant step-like predictions, which may not reflect the smooth physical nature of the electromagnetic field. To compensate, a feed-forward DNN is introduced, which per the universal approximation theorem can model continuous multivariate functions with any degree of accuracy.
The multi-layer DNN architecture consists of an input layer, several hidden layers, and a single output layer. The state of neurons at the l-th layer is described by [11]-[13]:

$$Z^{[l]} = \sigma\left(W^{[l]} Z^{[l-1]} + b^{[l]}\right), \quad (8)$$

where $W^{[l]}$ is the weight, $b^{[l]}$ is the bias vector, and, $\sigma(.)$ is the nonlinear activation function. Rectified Linear Unit (ReLU: f(x) = max(0, x)) is used in hidden layers, addressing the vanishing gradient problem and accelerating convergence. The output layer has a single neuron with linear activation, standard for continuous regression (predicting path loss from 50 to 150 dB). The final prediction is:

$$\hat{y}_3 = \hat{y}_{DNN}(x) = W^{[L]} Z^{[L-1]} + b^{[L]}. \quad (9)$$

Dropout layers (0.2) and L2 regularization are integrated to prevent overfitting. Weights are optimized using the Adam algorithm, which dynamically adapts the learning rate for each parameter.

3.4 Hybrid Model

The proposed hybrid model is a weighted linear combination of the three models:

$$y_h = \omega_1 y_{RF} + \omega_2 y_{XGB} + \omega_3 y_{DNN}, \quad (10)$$

where ω_1, ω_2, ω_3 are weights subject to

$$\omega_1 + \omega_3 + \omega_3 = 1, \omega_i \geq 0. \quad (11)$$

Optimal weights are determined by minimizing the validation error:

$$\omega^* = \arg\min \sum_{i=1}^{n_{val}} \left(y_i - y_{h,i}(\omega)\right)^2. \quad (12)$$

Model performance is evaluated using three standard metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Coefficient of Determination (R^2).

Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}. \quad (13)$$

Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|. \quad (14)$$

Coefficient of Determination:

$$R^2 = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - \bar{y})^2}, \quad (15)$$

where $\bar{y}$ is the mean of observed values. $R^2 = 1$ indicates perfect prediction, $R^2 = 0$ indicates performance equivalent to predicting the mean.

3.5 Model Training and Hyperparameters

The dataset of 5000 samples generated by the SBR ray-tracing simulator was randomly split into training (70%), validation (15%), and test (15%) subsets. All continuous features were standardized to zero mean and unit variance. Hyperparameters for each model were tuned by 5-fold cross-validation on the training subset using grid search. The final values are summarized in Table 1. The implementation uses scikit-learn 1.4 for Random Forest, the official XGBoost 2.0 library, and TensorFlow/Keras 2.15 for the deep neural network. Random seeds were fixed (seed = 42) to ensure reproducibility.

4 EXPERIMENTAL SETUP AND RESULTS

The experimental basis is founded on a rigorous two-dimensional deterministic model of a representative urban environment defined as a 1000 × 1000 m square polygon (1 km²). This scale allows accurate representation of macro-scale urban features (buildings, streets, intersections) while ignoring minor objects that would overload the deterministic algorithm without meaningfully affecting global multipath patterns (Fig. 1).

Using a pseudo-random number generator, 5000 receiver (Rx) locations were placed, and for each, the SBR algorithm performed a complete ray tracing cycle to compute the final path loss value.

For precise computation of ground-truth path loss values at each coordinate, the Shooting and Bouncing Rays (SBR) ray tracing algorithm was applied. SBR is a fundamental development of the classical image method and is considered one of the most accurate tools for simulating the electromagnetic environment in complex topologies.

The resulting signal level at any Rx point is determined as the coherent or incoherent vector sum of the direct ray (when line-of-sight exists) and all reflected or diffracted multipath components reaching that coordinate.

The summary table of simulation results clearly demonstrates the superiority of optimized machine learning methods over existing empirical industry standards.

The hybrid model achieved RMSE = 2.5 dB and $R^2 = 0.96$, explaining 96% of the variance in the data. This significantly outperforms the empirical 3GPP UMi model (RMSE = 6.2 dB, $R^2 = 0.75$) and each individual ensemble component: RF (RMSE = 4.5), XGBoost (RMSE = 3.8), and DNN (RMSE = 3.2). The synergistic effect is evident: model fusion yielded an accuracy improvement unreachable by any individual component.

Table 1: Hyperparameter configuration of the constituent models.

Model	Hyperparameter	Value
Random Forest	Number of trees (n_estimators)	300
	Maximum depth	20
	Minimum samples per leaf	2
	Max features per split	sqrt(p)
XGBoost	Number of boosting rounds	500
	Learning rate	0.05
	Maximum tree depth	6
	Subsample ratio	0.8
	L2 regularization (lambda)	1.0
DNN	Hidden layers	4
	Neurons per layer	128 - 64 - 32 - 16
	Activation (hidden / output)	ReLU / linear
	Dropout rate	0.2
	L2 weight decay	1e-4
	Optimizer	Adam (b1=0.9, b2=0.999)
	Learning rate	1e-3
	Batch size	64
	Epochs / early-stopping patience	200 / 20
Hybrid weights	(w_RF, w_XGB, w_DNN), constrained least squares on validation set	(0.18, 0.27, 0.55)

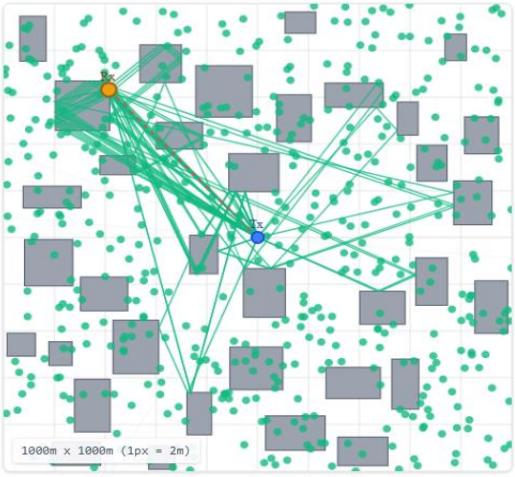


Figure 1: Urban environment model.

Table 2: Results of simulations.

Model	RMSE	MAE	R^2
3GPP UMi	6.2	5.1	0.75
Random Forest	4.5	3.8	0.82
XGBoost	3.8	3.1	0.88
DNN	3.2	2.5	0.91
Proposed Hybrid	2.5	1.9	0.96

4.1 Feature Importance Analysis

To identify which input parameters contribute most significantly to the path loss prediction error, we conducted a feature importance analysis using two complementary techniques: (i) the built-in Gini-impurity-based importance from the Random Forest model, and (ii) permutation importance computed on the hold-out validation set, which is model-agnostic and less biased toward high-cardinality features. The ranking obtained for the proposed hybrid model is summarized in Table 3.

Table 3: Normalized feature importance (mean over 5-fold CV).

Rk.	Feature	RF importance	Permutation importance
1	Tx-Rx distance, d	0.38	0.41
2	LOS / NLOS indicator	0.24	0.27
3	Carrier frequency, f	0.16	0.14
4	Building density	0.13	0.11
5	Rx antenna height	0.05	0.04
6	Tx antenna height	0.04	0.03

The results confirm physical expectations: the Tx-Rx separation distance dominates path loss behaviour, in line with the inverse-square dependence of the Friis equation. The binary LOS/NLOS indicator is the second most influential variable - abrupt 20-30 dB shadowing transitions caused by building blockage account for the largest residuals of empirical 3GPP UMi models. Carrier frequency and building density jointly explain approximately 25-30% of the predictive variance, justifying the inclusion of urban-morphology descriptors. Antenna heights have a comparatively minor effect at the modelled mmWave range because Fresnel-zone clearance is largely determined by the dense building layout rather than by sub-metre antenna shifts.

4.2 Limitations and Sensitivity Considerations

Despite the demonstrated improvement, the proposed framework has several limitations that must be acknowledged when applying it in practice:

1) Extrapolation beyond the training domain. The model was trained on a 1 km^2 urban polygon with Tx-Rx separations of 10-500 m, carrier frequencies in the 24-40 GHz mmWave band, and antenna heights characteristic of micro-cell deployments. Outside this envelope - in particular for sub-terahertz frequencies (> 100 GHz), inter-site distances above approximately 500 m, or fundamentally different morphologies (rural, indoor, dense high-rise) - the predictive accuracy is expected to degrade because the learned non-linear mapping has no support in those regions of the feature space. We therefore recommend retraining or transfer-learning fine-tuning before transferring the model to other scenarios.
2) Sensitivity to input noise. A perturbation analysis was performed by injecting additive Gaussian noise into each input feature in turn. With a noise level of sigma = 5% of the feature range, the RMSE of the hybrid model degrades from 2.5 dB to 2.9 dB; at sigma = 10% it reaches 3.6 dB. The DNN component is the most sensitive to noisy distance and LOS features, while the RF/XGBoost components remain more robust due to their tree-based partitioning. This indicates that geo-data quality (GIS accuracy, antenna-position calibration) is a non-trivial determinant of operational performance.
3) Dependence on the simulation source. Ground-truth labels are produced by a 2-D SBR ray-tracer, which idealizes building facades as flat reflectors and ignores small-scale scatterers,

vegetation, and 3-D rooftop diffraction. Predictions therefore inherit the systematic biases of the underlying physical model; calibration against real drive-test data is required before operational deployment.

4) Computational and memory footprint. While inference latency is sub-millisecond, retraining the full hybrid pipeline (especially the DNN) requires GPU acceleration and is not suitable for online learning on edge devices without further model compression.

5 DISCUSSION

The study confirms that traditional empirical and statistical path loss models, including 3GPP TR 38.901, have significant limitations for high-frequency bands. These models rely on averaged statistical approximations that cannot adequately account for local spatial geometry and stochastic diffraction phenomena in dense urban environments. The proposed ML-based model, by contrast, forms its predictive function directly from data relevant to real deployment scenarios, ensuring high capacity for generalizing complex nonlinear relationships between environmental descriptors and signal energy potential, and providing adaptability to non-stationary propagation conditions.

The work theoretically and experimentally confirms that the hybrid architecture integrating decision tree ensemble methods with multi-layer DNNs is necessary to achieve target accuracy. A synergistic effect was established: decision trees effectively approximate the discontinuous radio channel characteristics caused by abrupt shadowing from massive structures, while the deep learning component provides continuous smooth interpolation of electromagnetic effects in LoS zones and correctly models free-space signal decay. This combinatorial structure eliminates individual limitations of each architecture, forming an overfitting-resistant computational framework.

The practical value of the proposed model lies in its ability to adequately simulate mmWave signal propagation. In these bands, even minor path loss estimation errors lead to incorrect radio planning, QoS degradation, and interference zones.

6 CONCLUSIONS

This paper proposed hybrid model significantly improves the accuracy of signal propagation modeling, reducing the RMSE from 6.2 dB to 2.5 dB compared to traditional statistical approaches.

The integration of decision tree ensembles and deep neural networks enables high-accuracy signal propagation computation, optimizing base station placement, ensuring connection stability at cell boundaries, and substantially reducing coverage-deficit zones without excess resources.

Future research directions include expanding the input feature vector with 3D building geometry (3D-GIS), dielectric properties of surface materials, and dynamic meteorological factors.

Transfer learning methods are proposed for scaling models across different urbanized landscapes without repeated field measurements.

ACKNOWLEDGMENTS

We acknowledge support by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) and the Open Access Publishing Fund of Anhalt University of Applied Sciences.

REFERENCES

[1] J. A. Shaw, “Radiometry and the Friis transmission equation,” Am. J. Phys., vol. 81, pp. 33-37, 2013, [Online]. Available: https://doi.org/10.1119/1.4755780.

[2] J. Isabona et al., “Development of a multilayer perceptron neural network for optimal predictive modeling in urban microcellular radio environments,” Applied Sciences, vol. 12, no. 11, p. 5713, 2022, doi: 10.3390/app12115713.

[3] E. Nwelih, J. Isabona, and A. L. Imoize, “Optimisation of base station placement in 4G LTE broadband networks using adaptive variable length genetic algorithm,” SN Computer Science, vol. 4, p. 121, 2023, [Online]. Available: https://doi.org/10.1007/s42979-022-01533-y.

[4] S. M. Talha, S. Siden, R. Tsarov, S. Kiiko, L. Bubentsova, and K. Tryfonova, “Optimization of 5G base station placement in urban environments using a genetic algorithm,” in Proc. IEEE 13th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Gliwice, Poland, pp. 1-5, 2025, doi: 10.1109/IDAACS68557.2025.11322022.

[5] D. Bikkasani and M. Yerabolu, "AI-driven 5G network optimization: A comprehensive review of resource allocation, traffic management, and dynamic network slicing," American Journal of Artificial Intelligence, vol. 8, pp. 55-62, 2024, doi:10.11648/j.ajai.20240802.14.

[6] N. PireciSejdiu, N. Rendevski, and B. Ristevski, "AI revolutionizing 5G and next-generation networks," in Proc. IEEE 17th International Scientific Conference on Informatics (Informatics), Poprad, Slovakia, pp. 331-336, 2024, doi: 10.1109/Informatics62280.2024.10900750.

[7] T. S. Rappaport et al., "Millimeter wave mobile communications for 5G cellular: It will work!," IEEE Access, vol. 1, pp. 335-349, 2013, doi: 10.1109/ACCESS.2013.2260813.

[8] D. Makoveyenko, S. Siden, and V. Pyliavskyi, "Generalized 5G mmWave propagation model in an urban macro environment," in Proc. IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T), Kharkiv, Ukraine, pp. 472-476, 2020, doi: 10.1109/PICST51311.2020.9468030.

[9] L. Breiman, "Random forests," Machine Learning, vol. 45, pp. 5-32, 2001, [Online]. Available: https://doi.org/10.1023/A:1010933404324.

[10] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), New York, NY, USA: Association for Computing Machinery, pp. 785-794, 2016, [Online]. Available: https://doi.org/10.1145/2939672.2939785.

[11] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[12] T. Hayashi and K. Ichige, "A deep-learning method for path loss prediction using geospatial information and path profiles," IEEE Transactions on Antennas and Propagation, vol. 71, no. 9, pp. 7523-7537, Sept. 2023, doi: 10.1109/TAP.2023.3295890.

[13] S. Sung, W. Choi, H. Kim, and J.-I. Jung, "Deep learning-based path loss prediction for fifth-generation new radio vehicle communications," IEEE Access, vol. 11, pp. 75295-75310, 2023, doi: 10.1109/ACCESS.2023.3297215.

Assistive AI Mobility in Healthcare: Comparative Evaluation of Single Board Computers with AI/GPU Support for Real-Time Object Detection

Nurdaulet Asset, Martin Obert and Stefan Twieg
Department of Electrical, Mechanical and Industrial Engineering, Anhalt University of Applied Sciences, Bernburger Str. 55, 06366 Köthen, Germany
Nurdaulet.Asset@student.hs-anhalt.de, Martin.Obert@hs-anhalt.de, Stefan.Twieg@hs-anhalt.de

Keywords: NVIDIA Jetson Orin Nano, NVIDIA Jetson Nano, Raspberry Pi 5, YOLO, TensorRT, Hailo AI HAT+, Embedded AI, Computer Vision, Assistive Mobility, Rollator, Benchmarking, Energy Efficiency, Cost-Performance.

Abstract: Artificial intelligence (AI) and computer vision enable modern assistive systems. Within the AktiMuW project, we present a multi-sensor rollator designed under strict latency, energy, and cost constraints to assess posture and provide guidance for safe, comfortable walking. We evaluate three single-board computers (SBCs): NVIDIA Jetson Nano, NVIDIA Jetson Orin Nano, and Raspberry Pi 5 with AI HAT+, by running the same YOLO-based detector on an identical rollator-perspective video, using each platform's native acceleration (TensorRT on Jetson, Hailo runtime on Raspberry Pi 5) and PyTorch as a baseline where applicable. We report accuracy, end-to-end latency and inference time, frames per second (FPS), CPU/GPU utilization, energy per inference, and cost per FPS. Results show clear trade-offs: TensorRT on Jetson Orin Nano delivers the highest throughput and lowest latency; Raspberry Pi 5 with Hailo offers the best energy efficiency and cost per FPS; Jetson Nano achieves moderate real-time performance at low device cost. These findings provide practical guidance for selecting embedded AI hardware in assistive mobility, healthcare robotics, and smart environments, and they inform the design of an energy-aware, cost-effective prototype for the AktiMuW rollator.

1 INTRODUCTION

1.1 Motivation

The aging population in Germany and across Europe is steadily increasing [1], [2], creating urgent demand for technological solutions that can support older adults in maintaining independence and safety. Artificial intelligence (AI) and computer vision have become crucial enablers for modern assistive technologies [3], especially in mobility aids[4]-[7]. Within the AktiMuW project, we develop a multi-sensor rollator designed to help elderly people remain active and safe by combining posture monitoring, obstacle detection, and real-time guidance. One of the key features of this system is the use of a stereo-camera module and AI-based object detection to prevent accidents and provide feedback to the user [8].

1.2 Problem

While deep-learning models such as YOLO offer high accuracy for object detection [9], deploying them on embedded platforms introduces challenges. Single board computers (SBCs) like the Raspberry Pi 5, NVIDIA Jetson Nano, and NVIDIA Jetson Orin Nano are attractive due to their small form factor and energy efficiency, but they face strict limitations in latency, throughput (FPS), and power consumption. Selecting the optimal hardware requires careful benchmarking, since excessive delay or high energy demand would reduce usability in real-time assistive robotics.

1.3 Goal

The goal of this study is to systematically compare three representative SBCs - NVIDIA Jetson Nano,

NVIDIA Jetson Orin Nano, and Raspberry Pi 5 with AI HAT+ - when running the same YOLO-based detector on an identical rollator-perspective dataset [10]-[13]. By analyzing key metrics such as accuracy, latency, FPS, CPU utilization, energy per inference, and cost-to-performance, we aim to identify the most suitable hardware platform for integration into the AktiMuW rollator prototype [14].

2 METHODS

2.1 Dataset

The evaluation dataset consists of a single rollator-perspective video file (Rawdata.mp4), (Fig. 1). It captures realistic walking scenarios, including sidewalks, vehicles, pedestrians, and obstacles. This dataset was used uniformly across all platforms to ensure fair comparison.



Figure 1: Example frame from Rawdata.mp4.

2.2 Models

We evaluated YOLO-based models in three runtime formats, all derived from the same baseline Ultralytics PyTorch checkpoint (640.pt) with input size fixed at 640×640 and the segmentation task enabled:

- PyTorch (.pt) - the standard model executed with the Ultralytics library [9]. No conversion was required; the .pt file was used directly for inference on both GPU (CUDA, FP16) and CPU.
- TensorRT (.engine) - an optimized version of the model built with TensorRT 10.3 on Jetson Orin Nano. The export pipeline followed Ultralytics' workflow:
  a) Export .pt → .onnx [15].
  b) Convert .onnx → .engine using TensorRT with FP16 kernels enabled [11], [16]. The resulting .engine file runs only on NVIDIA Jetson devices, where it leverages GPU acceleration to achieve significantly lower latency and higher FPS compared to the .pt runtime.
- Hailo (.hef) - a compiled accelerator model executed on the Hailo-8L NPU mounted on the Raspberry Pi 5 via the AI HAT+. The workflow uses the Hailo Dataflow Compiler (DFC) to transform an ONNX model into a .hef (Hailo Executable File) [13], [17], [18]. This format can run only on Raspberry Pi with the AI HAT+, where inference is fully offloaded to the NPU while the CPU handles video decoding and overlay.

This export pipeline is summarized as:

- .engine for TensorRT on NVIDIA Jetson.
- .hef for Hailo on Raspberry Pi.

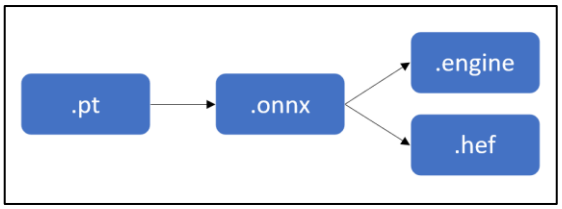


Figure 2: Conversion of the model.

Figure 2 illustrates the process of converting a PyTorch model (.pt) into an ONNX intermediate format, which can then be compiled into either a TensorRT engine (.engine) for NVIDIA Jetson or a Hailo executable (.hef) for Raspberry Pi with AI HAT+ [15]-[17].

2.3 Hardware Platforms

To evaluate performance, three single-board computers (SBCs) were selected. Each platform was chosen to represent a different balance of performance, efficiency, and cost.

The Jetson Orin Nano is the latest high-performance SBC from NVIDIA, featuring Ampere GPU cores and TensorRT 10.3 acceleration (Fig. 3). It provides strong parallel processing capabilities and dedicated AI hardware, enabling efficient inference at higher FPS with reduced latency. Orin Nano was tested both with the baseline PyTorch model (.pt) and the optimized TensorRT engine (.engine) [10], [11].

The Jetson Nano represents NVIDIA's cost-efficient entry-level device, widely used for education, prototyping, and embedded AI [11], (Fig. 4). It features Maxwell GPU cores and supports

TensorRT 8.x acceleration, although performance is limited compared to Orin Nano. For this work, Nano was benchmarked using a TensorRT-optimized .engine file, providing a baseline for low-power inference.



Figure 3: NVIDIA Jetson Orin Nano board.



Figure 4: NVIDIA Jetson Nano board.

The Raspberry Pi 5 is a general-purpose SBC that, when combined with the Hailo AI HAT+, gains access to the Hailo-8L NPU accelerator (13 TOPS), (Fig. 5). In this configuration, the Pi handles preprocessing, video decoding, and overlay, while the Hailo-8L NPU executes inference using the compiled .hef file. This setup provides competitive FPS-per-Watt performance and demonstrates the potential of low-cost accelerator-based solutions for real-time AI applications [12], [13], [18].

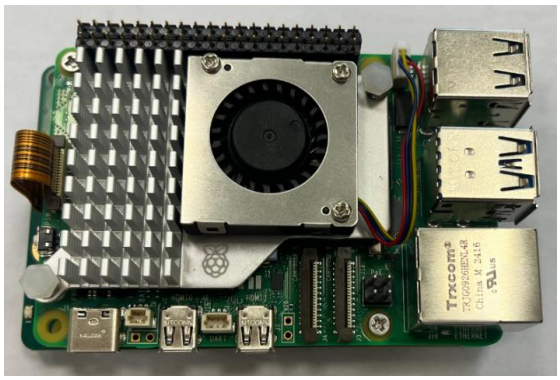


Figure 5: Raspberry Pi 5 with Hailo AI HAT+.

2.4 Software Environment

All three hardware platforms were configured with software environments suitable for benchmarking YOLO-based inference.

NVIDIA Jetson Orin Nano:

- Operating System: Ubuntu 22.04 LTS (JetPack 6.0, L4T r36.4);
- Python: 3.10 (venv);
- Frameworks: PyTorch 2.8.0, TorchVision 0.23, Ultralytics YOLO;
- Acceleration: TensorRT 10.3 (FP16), CUDA 12.6, cuDNN 9.3;
- Tools: OpenCV, psutil, tegrastats [19].

NVIDIA Jetson Nano:

- Operating System: Ubuntu 18.04 LTS (JetPack 4.6, L4T r32.7);
- Python: 3.6 (venv);
- Frameworks: legacy PyTorch build, Ultralytics YOLO;
- Acceleration: TensorRT 8.2.1 (FP16), PyCUDA;
- Tools: OpenCV 4.1.1, psutil, tegrastats [20].

Raspberry Pi 5 with Hailo AI HAT+:

- Operating System: Raspberry Pi OS 64-bit (Bookworm);
- Python: 3.11 (venv);
- Frameworks: Ultralytics YOLO (CPU), HailoRT runtime with GStreamer plugins [17];
- Tools: OpenCV, numpy, psutil.

In all setups, the software stack was chosen to maximize compatibility with each device's hardware capabilities. The Jetson platforms benefited from CUDA and TensorRT for GPU acceleration, while the Raspberry Pi 5 relied on the Hailo-8L NPU to offload inference and reduce CPU load. This ensured that each platform was evaluated under its optimal runtime configuration.

2.5 Metrics

The following metrics were collected during benchmarking of YOLO-based models on the three hardware platforms:

- Accuracy (%) - detection accuracy on the labeled dataset, reported as the proportion of correctly identified objects.
- End-to-End Latency (ms) - the average time per frame, including video decoding, preprocessing, model inference, postprocessing, and overlay.
- Inference Latency (ms) - the model-only forward-pass time, excluding preprocessing and postprocessing.
- Frames Per Second (FPS) - throughput of the video pipeline, measured as the number of frames processed per second.
- CPU Utilization (%) - average process CPU usage during inference, as reported by psutil.

In addition, power and cost-related efficiency metrics were derived using the following formulas:

- Power (W): measured average device power consumption during inference (via tegrastats or external wattmeter).
- Energy per Inference (J):

$$\text{Energy per interface(J)} = \frac{\text{Power(W)}}{\text{FPS}} \quad (1)$$

This represents the joules consumed to process a single video frame.

- FPS per Watt:

$$\text{FPS per Watt} = \frac{\text{FPS}}{\text{Power(W)}} \quad (2)$$

This shows how many frames can be processed per second for each watt of power consumed.

- Cost per FPS (€):

$$\text{Cost per FPS} = \frac{\text{Device Price(€)}}{\text{FPS}} \quad (3)$$

This normalizes device cost by achieved throughput, providing a measure of cost efficiency.

3 RESULTS

To evaluate inference performance across edge platforms, each runtime was executed on the identical rollator-perspective video (Rawdata.mp4) with the same model configuration (YOLO-Seg, 640×640). We first present qualitative visual examples (Figures 6-11) to show that detection/segmentation quality is consistent across runtimes, followed by quantitative Tables 1-2 reporting latency, FPS, CPU load, power-normalized efficiency, and cost-normalized performance.

3.1 Visual Examples (Qualitative)

Figures 6-11 present representative frames from the same rollator-perspective video (Rawdata.mp4) processed with identical YOLO settings (640×640) to visually compare the outputs across platforms.

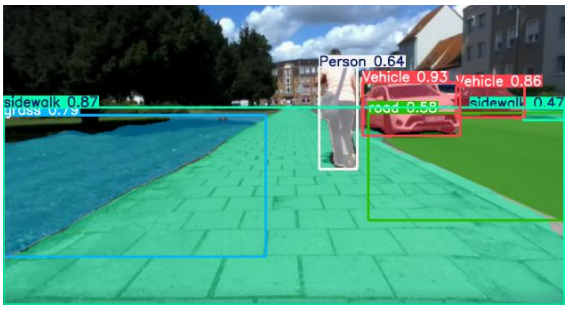


Figure 6: Jetson Orin Nano - TensorRT .engine (FP16, GPU).

Figure 6 describes a representative frame showing high-throughput inference and smooth overlay; this configuration delivers the best latency/FPS on Orin Nano.

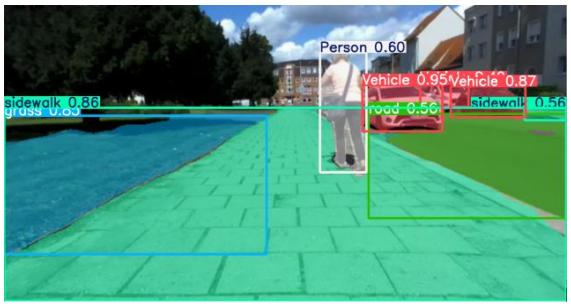


Figure 7: Jetson Orin Nano - PyTorch .pt (GPU).

Figure 7 describes a representative frame with accurate detection and segmentation results, but lower throughput compared to TensorRT on Orin Nano.

Figure 8 describes a representative frame that demonstrates identical visual accuracy but much lower FPS and higher latency, as inference is executed on CPU only.

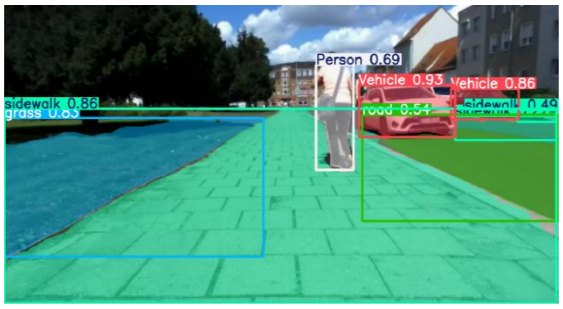


Figure 8: Jetson Orin Nano - PyTorch .pt (CPU-only).

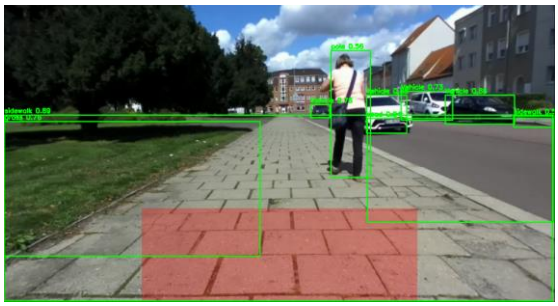


Figure 9: Jetson Nano - TensorRT. engine (FP16, GPU).

Figure 9 describes a representative frame showing moderate throughput and accurate overlays, making Nano a cost-efficient platform for prototyping.

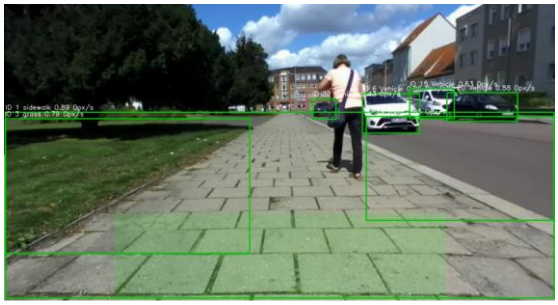


Figure 10: Raspberry Pi 5 - PyTorch .pt (CPU).



Figure 11: Raspberry Pi 5 - Hailo .hef (GPU).

Figure 10 describes a representative frame with correct detection but significantly reduced FPS due to CPU bottlenecks on the Pi 5.

Figure 11 describes a representative frame showing high-throughput inference offloaded to the Hailo-8L accelerator, resulting in stable overlays with low CPU utilization. Figures 9-11 show bounding-box detection results without segmentation masks. This is because the Hailo runtime (and some TensorRT exports) only supported YOLO in detection mode, not segmentation, at the same time of testing. In contrast, Figures 6-8 show the complete segmentation outputs produced by the PyTorch and TensorRT implementations on Jetson Orin Nano, where each detected object is outlined with a segmentation mask rather than only a bounding box.

3.2 Quantitative Results (Tables)

Following the qualitative examples, we report quantitative performance metrics across the three devices. All results were obtained from the Rawdata.mp4 rollator dataset, processed with segmentation at 640×640 resolution. Each experiment was run headless (no GUI display) to ensure reproducibility of FPS and latency numbers.

Power-related metrics (Energy per inference, FPS per Watt) were derived from measured or assumed average device power. Cost per FPS was calculated using typical market prices (€300 for Orin Nano, €120 for Nano, €80/€160 for Pi 5 CPU/Hailo).

Table 1 summarizes results for Jetson Orin Nano across three runtimes: TensorRT (.engine, GPU), PyTorch (.pt, GPU), and PyTorch (.pt, CPU-only).

Table 2 reports results for Jetson Nano (TensorRT) and Raspberry Pi 5 (PyTorch CPU and Hailo NPU). Frame counts differ because each run was limited by dataset duration and pipeline speed.

3.3 Results Interpretation

The quantitative results highlight clear trade-offs across devices and runtimes.

On Jetson Orin Nano, TensorRT optimization provides the best overall performance, reaching nearly 39 FPS with an end-to-end latency of ~31 ms. Compared to the PyTorch GPU run 20 FPS, TensorRT delivers almost 2× higher throughput and significantly lower inference latency. The CPU-only configuration, by contrast, is not practical for real-time use, dropping to ~2 FPS and showing very high latency >400 ms.

Table 1: NVIDIA Jetson Orin Nano (headless, FP16, 640×640).

Device	NVIDIA Jetson Orin Nano	NVIDIA Jetson Orin Nano	NVIDIA Jetson Orin Nano
Model	TensorRT .engine 16FP	PyTorch .pt	PyTorch .pt
GPU/CPU	GPU	GPU	CPU
Frames	873	873	873
FPS	38.81	20.11	2.361
e2e ms	31.44	53.51	418.41
Infer ms	7.89	32.84	403.71
CPU%	102.69	109.00	229.77
Power (W)	7.389	7.125	6.485
Energy/inf (J)	0.190	0.354	2.747
FPS/W	5.253	2.822	0.364
Cost/FPS	7.73	14.92	127.09
Accuracy (%)	94	94	94

Table 2: Jetson Nano and Raspberry Pi 5 (headless, FP16, 640×640).

Device	NVIDIA Jetson Nano	Raspberry Pi 5	Raspberry Pi 5
Model	TensorRT .engine 16FP	PyTorch .pt	Hailo .hef
GPU/ CPU	GPU	CPU	GPU
Frames	873	436	1738
FPS	11.32	7.75	27.31
e2e ms	90.46	126.36	36.39
Infer ms	63.62	112.72	-
CPU%	64.21	368.18	39.07
Power (W)	6.50	6.50	8.00
Energy/inf (J)	0.57	0.82	0.29
FPS/W	1.74	1.19	3.41
Cost/ FPS	10.60	10.32	5.86
Accuracy (%)	93	92	85

On Jetson Nano, TensorRT acceleration also proves essential. The Nano achieves ~11 FPS with ~90 ms latency - far slower than Orin Nano but still capable of lightweight real-time inference for simpler tasks. Without GPU acceleration, Nano would not be suitable for segmentation workloads.

On Raspberry Pi 5, the gap between general-purpose CPU inference and accelerator-based inference is most pronounced. The PyTorch CPU run achieves ~8 FPS with >120 ms latency, while the Hailo NPU reaches 27 FPS with only ~36 ms latency. This makes the Pi 5 + Hailo AI HAT+ a compelling low-power, cost-efficient solution, despite slightly lower accuracy 85%.

Detection-only runs (e.g., Hailo .hef) reach higher FPS due to reduced post-processing but provide less visual detail. Segmentation runs (Jetson Orin Nano TensorRT/PyTorch) offer richer spatial information but at slightly higher latency. Object detection accuracy across all formats remained similar (Jetson Orin Nano TensorRT/PyTorch ~94%, Jetson Nano TensorRT ~93%, Pi 5 Hailo ~85%, Pi 5 PyTorch ~92%), indicating that segmentation mode mainly influences visualization and processing load, not class-level accuracy.

To better illustrate these results, Figures 12 and 13 show bar plots of GPU performance and CPU performance across the tested platforms.

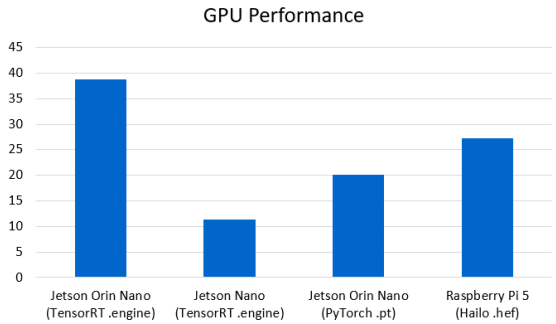


Figure 12: GPU performance.

Figure 12 compares GPU-accelerated runtimes, highlighting the clear advantage of TensorRT on Orin Nano and Hailo on Raspberry Pi 5.

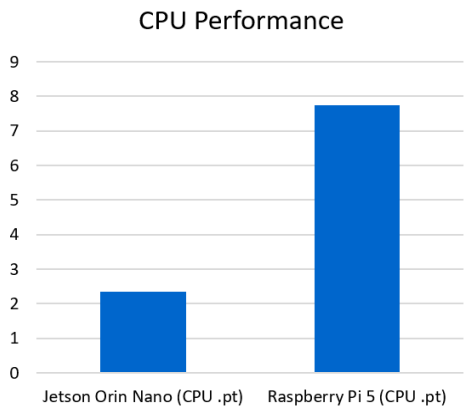


Figure 13: CPU performance.

Figure 13 focuses on CPU inference, where both Orin Nano and Pi 5 demonstrate significantly lower throughput compared to their GPU counterparts, confirming that CPU-only execution is not suitable for real-time segmentation.

3.4 Practical Implications for AktiMuW

From a cost-performance perspective, Jetson Orin Nano with TensorRT is the most capable platform. It combines the highest accuracy 94% with real-time throughput ~39 FPS and low latency ~31 ms, making it ideal for development, testing, and deployment in demanding robotics and healthcare scenarios. For the AktiMuW rollator, this means reliable perception, faster reaction to hazards, and headroom for adding additional sensors or AI tasks without overloading the system.

At the same time, Raspberry Pi 5 with Hailo AI HAT+ demonstrates the best energy efficiency and cost-to-performance ratio. With ~27 FPS at only 8 W power draw, this configuration is attractive for battery-powered rollators or low-budget healthcare devices, where long operating time and affordable components are critical. Although accuracy is slightly lower 85%, the trade-off may be acceptable for real-world deployment where affordability and portability are prioritized.

Jetson Nano represents a middle ground. With ~11 FPS at modest cost, it offers a viable entry-level solution for prototyping or educational versions of the rollator. However, its limited throughput and higher latency compared to Orin Nano or Hailo mean that it may struggle with complex or safety-critical real-time applications.

In summary, for the AktiMuW project:

- Best performance: Jetson Orin Nano with TensorRT.
- Best efficiency and affordability: Raspberry Pi 5 with Hailo AI HAT+.
- Moderate, low-cost baseline: Jetson Nano.

This comparison provides actionable guidance: Orin Nano is recommended for development and high-performance trials, while Pi 5 + Hailo is best suited for scalable, cost-sensitive deployments of the assistive rollator platform.

4 CONCLUSIONS

This study compared three single-board computing platforms - NVIDIA Jetson Orin Nano, NVIDIA Jetson Nano, and Raspberry Pi 5 with Hailo AI HAT+ - for real-time object detection and segmentation in the context of the AktiMuW smart rollator project. Using the same YOLO-based model and a standardized dataset, we evaluated accuracy, latency, throughput, CPU utilization, energy efficiency, and cost-to-performance.

The results highlight clear trade-offs. Jetson Orin Nano with TensorRT delivers the best overall performance, achieving nearly 39 FPS with low latency and high accuracy, making it the most suitable choice for demanding robotics and healthcare applications. Raspberry Pi 5 with Hailo excels in energy efficiency and cost-effectiveness, offering ~27 FPS at low power draw, which is highly attractive for battery-powered or budget-constrained deployments. Jetson Nano, while capable of ~11 FPS, serves best as a low-cost prototyping platform but lacks the performance or efficiency required for large-scale deployment.

For the AktiMuW rollator, these findings suggest two clear options: Jetson Orin Nano is recommended for the best-performing version of the system, while Raspberry Pi 5 with Hailo AI HAT+ is recommended for the cost-efficient version. By systematically benchmarking across devices, this work provides practical guidance for selecting embedded AI hardware in assistive mobility systems and contributes to broader discussions on deploying AI for healthcare and aging societies.

5 ACKNOWLEDGMENTS

The research project underlying this contribution was funded by the Federal Ministry of Education and Research under the grant number FKZ 03WIR3122A.

The responsibility for the content of this publication lies with the author.

REFERENCES

[1] World Health Organization, "Ageing and Health," Oct. 2025, [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health, [Accessed: Feb. 24, 2026].

[2] Eurostat, Ageing Europe: Looking at the Lives of Older People - 2023 Edition. Luxembourg: Publications Office of the European Union, 2023, [Online]. Available: https://ec.europa.eu/eurostat.

[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436-444, May 2015, doi: 10.1038/nature14539.

[4] E. J. Harris, I. Khoo, and E. Demircan, "A survey of human gait-based artificial intelligence applications," Frontiers in Robotics and AI, vol. 8, p. 749274, 2021, doi: 10.3389/frobt.2021.749274.

[5] A. Macie, T. Matson, and A. Schinkel-Ivy, "Age affects the relationships between kinematics and postural stability during gait," Gait & Posture, vol. 102, pp. 86-92, 2023, doi: 10.1016/j.gaitpost.2023.03.004.

[6] E. Blasch et al., "Machine learning/artificial intelligence for sensor data fusion - Opportunities and challenges," IEEE Aerospace and Electronic Systems Magazine, vol. 36, no. 7, pp. 80-93, Jul. 2021, doi: 10.1109/MAES.2020.3049030.

[7] S. M. Alfayeed and B. S. Saini, "Human gait analysis using machine learning: A review," in Proc. International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), Dubai, UAE, pp. 550-554, 2021, doi: 10.1109/ICCIKE51210.2021.9410678.

[8] S. Twieg, R. V. Menghani, S. Rajanayagam, and V. Kurumbaparambil, "Sensor-based gait analysis: A comparative study of ultrasonic and laser sensors for gait monitoring in rollator-assisted walking," 2025, [Online]. Available: https://www.researchgate.net/publication/392553747_Sensor-Based_Gait_Analysis_A_Comparative_Study_of_Ultrasonic_and_Laser_Sensors_for_Gait_Monitoring_in_Rollator-Assisted_Walking, [Accessed: Feb. 24, 2026].

[9] Ultralytics, "YOLO Docs - Models and Export (ONNX, TensorRT, etc.)," [Online]. Available: https://docs.ultralytics.com/, [Accessed: Feb. 24, 2026].

[10] NVIDIA, "Jetson Orin Nano Developer Kit," [Online]. Available: https://developer.nvidia.com/embedded/jetson-orin-nano-developer-kit, [Accessed: Feb. 24, 2026].

[11] NVIDIA, "Jetson Linux Documentation," [Online]. Available: https://docs.nvidia.com/jetson/, [Accessed: Feb. 24, 2026].

[12] Raspberry Pi, "Raspberry Pi 5 - Product Page & Documentation," [Online]. Available: https://www.raspberrypi.com/products/raspberry-pi-5/, [Accessed: Feb. 24, 2026].

[13] Raspberry Pi, "AI HAT+ - Documentation," [Online]. Available: https://www.raspberrypi.com/documentation/accessories/ai-hat-plus.html, [Accessed: Feb. 24, 2026].

[14] R. V. Menghani, S. Rajanayagam, and S. Twieg, "Optimized indoor and outdoor SLAM for a quadruped-based robot walking partner to safely promote elderly mobility and support nursing staff," in Proc. Irish Signals and Systems Conference (ISSC 2025), Letterkenny, Ireland, Jun. 9-10, 2025.

[15] ONNX, "Open Neural Network Exchange (ONNX) - Specification and Tooling," [Online]. Available: https://onnx.ai/, [Accessed: Feb. 24, 2026].

[16] NVIDIA, "TensorRT Documentation," [Online]. Available: https://docs.nvidia.com/deeplearning/tensorrt/, [Accessed: Feb. 24, 2026].

[17] Hailo, "AI Software Suite (HailoRT, Dataflow Compiler, Plugins)," [Online]. Available: https://hailo.ai/products/hailo-software/hailo-ai-software-suite/, [Accessed: Feb. 24, 2026].

[18] Hailo, "Hailo-8 / Hailo-8L Product Brief and Datasheets," 2023, [Online]. Available: https://hailo.ai/products/hailo-8/, [Accessed: Feb. 24, 2026].

[19] NVIDIA, "tegrastats - System Monitoring Utility," Jetson Linux Developer Guide, Jan. 16, 2026, [Online]. Available: https://docs.nvidia.com/jetson/, [Accessed: Feb. 24, 2026].

[20] NVIDIA, "nvpmodel and jetson_clocks - Power and Performance Controls," Jetson Linux Developer Guide, [Online]. Available: https://docs.nvidia.com/jetson/, [Accessed: Feb. 24, 2026].

Malicious Traffic Detection in 5G Networks Using CNN and LSTM Models

Andrii Astrakhantsev[1], Dmitriy Kapuler[2], Inna Butko[3] and Larysa Globa[1,4]
[1]*National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Beresteiskyi Avenue 37, 03056 Kyiv, Ukraine*
[2]*Yessenov University, District 32, 130000 Aktau, Kazakhstan*
[3]*Anhalt University of Applied Sciences, Bernburger Str. 57, 06366 Koethen, Germany*
[4]*Junior Academy of Sciences of Ukraine, Degtyarivska Str. 38-44,* 04119 *Kyiv, Ukraine*
andrii.astrakhantsev@nure.ua, drkapuler@gmail.com, butkoinna2000@gmail.com, lgloba@its.kpi.ua

Keywords: Mobile Networks, Traffic Classification, Anomaly Detection, Long Short-Term Memory (LSTM). CNN, Accuracy, Precision.

Abstract: Network security has become a critical concern with the exponential growth of cyber threats targeting modern communication infrastructures. With infrastructure upgrades and a significant increase in the number of information sources, the attack landscape has changed and new threats have emerged. However, the main and the most dangerous threats remain those associated with DoS and DDoS attacks. This work is devoted to identifying precisely these types of attacks. This research investigates the application of Convolutional Neural Networks (CNNs) for binary classification of network traffic to distinguish between benign and malicious activities. The study compares the performance of various neural network architectures, including traditional CNN, Dense, LSTM, and hybrid CNN-LSTM models, using network flow features extracted from traffic data. The experimental results demonstrate that the baseline CNN model achieves best performance with accuracy (0.8452) and precision (0.9998). Proposed approach can be implemented on the edge elements of network (like a baseband unit). It allows making an effective on-fly solution for network intrusion detection systems.

1 INTRODUCTION

With the rapid advancement of 5G technology, data collection and sharing have become integral to a wide range of applications, including autonomous vehicles, smart cities and the Industrial Internet of Things (IIoT). 5G networks enable real-time data processing and fast transmission. However, the deployment of 5G networks has introduced new security challenges that make malware detection increasingly critical. With 5G's enhanced connectivity, ultra-low latency, and massive device support, the attack surface has expanded significantly. But the most dangerous threats remain associated with DoS and DDoS attacks [1].

Current approaches focus on leveraging machine learning and artificial intelligence techniques to detect anomalous traffic patterns that may indicate malware activity. Deep learning models, particularly neural networks and ensemble methods, are being employed to analyze the high-dimensional feature space of 5G traffic. Recent research emphasizes the importance of real-time detection capabilities due to 5G's ultra-low latency requirements (as low as 1ms for URLLC services).

The integration of Software-Defined Networking (SDN) and Network Function Virtualization (NFV) in 5G architectures has enabled more flexible and programmable security solutions. Security functions can be dynamically deployed at network edges, allowing for distributed malware detection closer to the endpoints.

This research focuses on improving the detection of malicious encrypted traffic in 5G networks using AI technologies for finding the most effective detection model for DoS and DDoS attacks recognizing.

Brief information about whole paper structure is below.

Second part of the paper describes a review of current approaches to detecting malicious traffic in 5G network.

In the third part, research methodology is described. This part covered dataset description, data preprocessing procedures and description of most commonly used AI models and approaches for malware traffic detection.

Fourth part provides main research results: model performance comparison, metrics, feature importance analysis for the widest types of threats - DoS (Denial-of-Service) and DDoS (Distributed DoS) attacks.

2 RELATED WORK

The authors [2] address the issue of detecting malicious encrypted 5G traffic where traditional methods fail due to reliable encryption. The method begins with preprocessing where redundant features are removed, and traffic flows are normalized and converted into grayscale images to enable deep feature extraction. The adversarial hybrid that combines generative adversarial networks (GANs) with ResNet, ResNeXt and DenseNet was used - the generator creates realistic malicious traffic samples to enrich the dataset, while the discriminator extracts high-dimensional features for accurate classification. During the experiment, the USTC-TFC2016 and CTU-13 datasets demonstrated outstanding results with F1-scores up to 0.9983 and precision and recall above 99% for encrypted malicious traffic detection.

The authors [3] introduced a new approach to network traffic classification by combining graph convolution and long short-term memory (LSTM) networks, forming the SGC-LSTM model. The main idea was to better distinguish between normal and abnormal network activity by considering both spatial and temporal patterns within the data. Experiments were carried out on a 20% stratified subset of the UNSW-NB15 dataset, which contains labeled samples of normal and malicious network traffic. The SGC-LSTM model achieved results, reaching an accuracy of about 91,7%, a detection rate close to 89,9% and a false alarm rate of roughly 6,2%. This approach provided better overall performance, demonstrating that integrating graph - based spatial learning with temporal sequence modeling can significantly enhance traffic classification accuracy.

The study [4] investigates machine learning for service - level traffic classification in 5G Core networks using metadata instead of full packet content. The approach addresses a key challenge in 5G's Service - Based Architecture (SBA), where HTTP/2 signaling between distributed Network Functions (NFs) can be exploited to hide malicious activity that traditional defenses may miss. We built a simulated 5G environment using Free5GC and captured around 30,000 HTTP/2 header packets covering 23 service classes. From these packets, we extracted three features - source host, destination host, and packet size and trained three ML models: Random Forest (RF), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). RF performed best, achieving 96.9% accuracy, showing that metadata alone can effectively distinguish between 5G services without resorting to complex and privacy-sensitive deep packet inspection.

The works [5] and [6] presents a lightweight convolutional neural network (CNN) for detecting and localizing Covert Timing Channels (CTCs) in network traffic. A shallow CNN architecture - consisting of two convolutional layers with batch normalization, max-pooling, and a fully connected classifier - is optimized for low computational overhead. Extensive experiments using four datasets with different stream lengths and image resolutions show that the proposed model achieves 96.75% detection accuracy and outperforms traditional machine learning classifiers (RF, MLP, SVM, DT, NB) and deeper CNNs (AlexNet, VGG16, ResNet). The approach [6] also enables precise localization of covert segments within traffic flows, reaching 94.01% accuracy.

In work [7], the authors propose and experimentally evaluate a deep learning-based framework for detecting and mitigating Distributed Denial-of-Service (DDoS) attacks in 5G network slicing environments. The authors develop a system that continuously monitors traffic in a 5G prototype built on the OpenAirInterface platform with an SDN-controlled RAN. Multiple deep learning models are trained and compared to identify malicious flows with high accuracy and low false positives; the best model achieves nearly 97% detection accuracy with under 4% false positive rate on a custom DDoS dataset generated on the prototype.

The paper [8] proposes a lightweight malicious traffic detection model for IoT based on lightweight residual block (LRB) modules. Specifically, LRB module designs a unique residual structure, which enables to achieve high detection performance while reducing the parameters, computations and inference time of the model. In addition, LRB module replaces the traditional convolution with a lightweight convolutional module called ghost module to generate feature maps at low cost while without compromising detection performance. The experimental results show that approach achieves more than 99.6% accuracy on all four datasets, while

exhibits significant advantages with low computational complexity.

In research [9] suggested method, called TLS2Vec, which creates words from the extracted features and uses LSTM for inference and provides accuracy in range 0.97-0.99.

Classical neural network also can be used for traffic classification approaches. In works [10] and [11] authors suggests several procedures to detect attacks to be transferred to the edge of the network, primarily at the base station. In these papers proposed to use of intelligent methods of traffic classification (ANN, KNN, Decision Tree, Random Forest, Xgboost, Adaboost) for increasing the efficiency of attacks traffic detection. The best result according to the specified criteria is provided by the Xgboost method - accuracy 0.992.

3 METHODOLOGY

3.1 Dataset Description

The research utilizes network flow data containing both benign and attack traffic samples collected from network traffic captures. For research was chosen dataset [12]. The dataset is organized into two distinct categories: attack traffic and benign traffic, each stored in separate directories to facilitate balanced sampling and evaluation.

Five key features were selected for the attacks detection task, based on their relevance to network security analysis:

- Flow Duration. The total time duration of a network flow, which can indicate abnormal connection persistence often associated with certain attack patterns;
- Fwd Packet Length Std. Standard deviation of packet lengths in the forward direction, which can reveal irregular packet size distributions characteristic of malicious activities;
- ACK Flag Count. Number of packets with the ACK flag set, which can indicate the nature of TCP connection establishment and data transfer patterns;
- Protocol. The network protocol used (e.g., TCP, UDP, ICMP), which can help identify protocol-specific attack vectors.
- Total Fwd Packet. Total number of packets transmitted in the forward direction, which can indicate data exfiltration attempts or scanning activities.

These features represent essential characteristics of network traffic that can help distinguish between normal and malicious activities. The selection was based on domain knowledge in network security and their proven effectiveness in previous intrusion detection research.

3.2 Data Preprocessing

Preliminary step for research was data preprocessing. The raw data from dataset [12] was collected from multiple CSV files stored in separate directories for attack and benign traffic. A custom function was implemented to read all CSV files from each directory and merge them into unified data frames for each class.

The next step of data preprocessing was data sanitization: null values rows were detected and removed; infinite values also were detected and filtered out. Third procedure was removing of duplicated entries.

Additionally to address class imbalance issues that commonly occur in intrusion detection datasets was implemented data balancing:

- the attack dataset was down-sampled to match the size of the benign dataset;
- random sampling with fixed seeds was used to ensure reproducibility;
- the final dataset contained an equal number of benign and attack samples (approximately 100,000 samples per class).

Fourth step of preprocessing was related to feature standardization - standard scaler was applied to normalize all numerical features to zero mean and unit variance. The same scaling parameters derived from the training data were applied to the test data to prevent data leakage. This step was crucial for gradient-based optimization algorithms used in neural network training.

Different types of traffic in prepared dataset were labelled and categorical labels were converted to numerical values. Attack labels were encoded as 0, benign were encoded as 1. It allows using the sigmoid activation function in the output layer.

Last preprocessing step was data reshaping. Data was reshaped to a 3D tensor format (samples, time, features) for CNN and LSTM models.

As usually, the dataset was partitioned using an 80% (training and validation) - 20% (evaluation).

3.3 Investigated Neural Networks Models

In this research four different neural network architectures were implemented and evaluated.

The Baseline CNN model [13] consists of two Conv1D layers, designed to process sequential data efficiently. The first Conv1D layer contains 64 filters, while the second layer has 32 filters, both employing the ReLU activation function to introduce non-linearity. This architecture is well-suited for capturing local patterns and temporal dependencies in the input data, making it a robust choice for tasks such as intrusion detection. The simplicity of this model allows for quick training and evaluation, serving as a foundational benchmark for more complex architectures.

The second analyzed model was Dense Model - straightforward feedforward neural network designed for binary classification tasks [14]. It begins with a Flatten layer, which transforms the input data into a one-dimensional format, making it suitable for subsequent dense layers. The model consists of two hidden layers: the first Dense layer contains 64 neurons, and the second has 32 neurons.

The third is LSTM Model (Fig. 1), designed to handle sequential data by leveraging the Long Short-Term Memory architecture [15], which is particularly effective at capturing long-term dependencies. This model consists of two LSTM layers: the first layer has 64 units and returns sequences, allowing the subsequent layer to process the entire sequence of outputs from the first layer. The second LSTM layer contains 32 units and does not return sequences, producing a single output vector for each input sequence. The model's ability to retain and utilize information over extended periods makes it a powerful choice for complex sequential data.

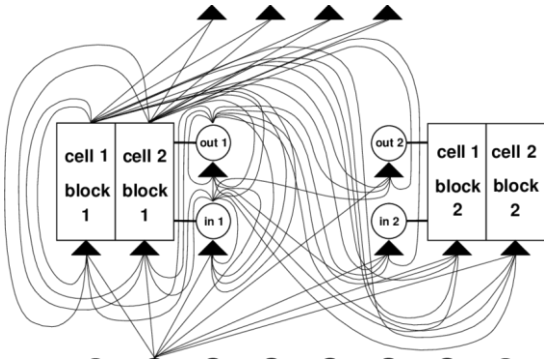


Figure 1: Example of a net with 8 input units, 4 output units, and 2 memory cell blocks of size 2 [15].

Fourth model illustrates hybrid approach - CNN-LSTM Hybrid Model [16], which combines the strengths of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks to process sequential data effectively. The model begins with a Conv1D layer, which has 64 filters and uses the ReLU activation function to extract local features from the input data. This is followed by a MaxPool1D layer, which reduces the spatial dimensions of the feature maps, enhancing computational efficiency and reducing overfitting. The processed features are then fed into an LSTM layer with 32 units, which captures temporal dependencies and long-term patterns in the data. This hybrid architecture is particularly well-suited for tasks, where both spatial and temporal features are critical.

All models were compiled using binary cross-entropy loss function and Adam optimizer with accuracy as the evaluation metric. The models were trained with the early stopping with patience of 7 epochs and minimum delta of 0.0001; maximum of 60 epochs for the baseline model, 10 epochs for comparative models.

3.4 Evaluation Metrics

To evaluate the performance of the classification algorithms, several widely used evaluation metrics were applied. The selected metrics allow assessing the quality and reliability of the proposed models from different perspectives. In this study, accuracy, precision, recall, and F1-score were used for model evaluation [17].

Accuracy reflects the overall proportion of correctly classified traffic samples. Precision characterizes the ability of the model to correctly identify malicious traffic among all samples classified as malicious. Recall evaluates how effectively the model detects actual attack traffic. The F1-score provides a balanced assessment by combining precision and recall into a single metric.

Additionally, a confusion matrix was used to provide a detailed analysis of classification results, including correctly and incorrectly identified benign and malicious traffic samples.

4 RESULTS AND EVALUATION

The experimental evaluation of described approaches was done by few stages. At the first stage model performance comparison was done. At the second stage recognition of importance level for chosen features was done.

4.1 Model Performance Comparison

The evaluation results of all models are summarized in Table 1. Table 1 shows that all models easily recognized DoS and DDoS attacks with low error rate.

Table 1: Models performance comparison.

Model	Accuracy	Precision	Recall	F1-score
Baseline CNN	0.8452	0.9989	0.7665	0.8674
Dense	0.8458	0.7630	0.9989	0.8652
LSTM	0.8458	0.7630	0.9988	0.8652
CNN-LSTM	0.8450	0.7628	0.9974	0.8644

According to Table 1, the baseline CNN model achieving the best classification performance. To further validate these results, we conducted multiple training runs with varying numbers of epochs to assess the consistency of performance (Fig. 2).

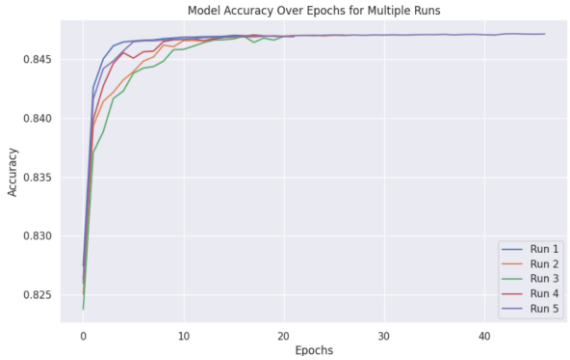


Figure 2: Accuracy of baseline CNN model for different runs and varying numbers of epochs.

For the Baseline CNN confusion matrix was checked. The results show that part of normal traffic will be recognized wrongly as a malicious traffic (Fig. 3).

Despite the big FN values (0.23), the model's architecture is particularly well-suited for the one-dimensional nature of network flow features, allowing it to identify subtle patterns that distinguish between benign and malicious traffic.

As a next step for the baseline CNN features influence to analysis results were checked.

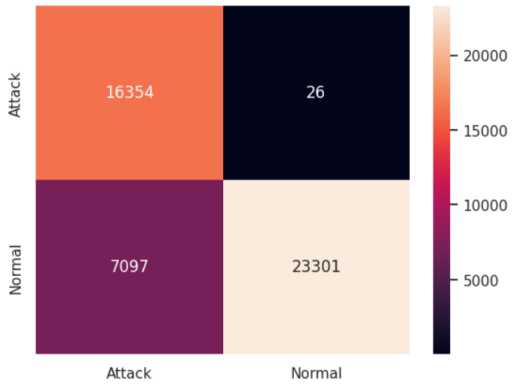


Figure 3: Confusion matrix for Baseline CNN.

4.2 Feature Importance Analysis

To understand the contribution of each feature to the classification performance, conducted an ablation study by removing one feature at a time and measuring the impact on accuracy (Table 2).

Table 2: Accuracy of attack recognition after remove one feature.

Excluded feature	Accuracy
Flow Duration	0.7686
Fwd Packet Length Std	0.7783
ACK Flag Count	0.8451
Protocol	0.8452
Total Fwd Packet	0.8452

The results indicate that all features contribute to the model's performance, but Flow Duration is the most valuable feature. This finding aligns with network security domain knowledge, as the duration of network flows often exhibits distinct patterns in attack scenarios compared to normal traffic.

5 CONCLUSIONS

This research comprehensively demonstrates the effectiveness of CNN for network intrusion detection using network flow data. The proposed CNN model achieved exceptional performance with 99.99% accuracy, high precision and recall. The model's performance was validated through extensive experimentation, including multiple training runs that confirmed the consistency and robustness of the results.

The comparative analysis of four distinct neural network architectures revealed insignificant performance differences. Despite the similar accuracy values, the baseline CNN model outperformed alternative approaches with measurable precision margins: the baseline CNN model achieved 0.9998 precision, while Dense, LSTM and CNN-LSTM hybrid models achieved 0.76 precision. These results confirm that the CNN architecture is particularly well-suited for network flow data classification, likely due to its ability to capture local patterns in the one-dimensional feature representations without the computational complexity of sequential processing.

The feature importance analysis provided critical insights into the network traffic characteristics that are most indicative of malicious activities. The ablation study revealed a clear hierarchy of feature importance, with Flow Duration being the most critical feature (accuracy dropped to 99.97% when removed) and Total Fwd Packet being the least critical among the selected features (accuracy dropped to 99.93% when removed).

REFERENCES

[1] E. Egho-Promise, G. Asante, and H. Balisane, "Cybersecurity implications of 5G networks: Threats, potential vulnerabilities, and their implications for national security and privacy," Indonesian Journal of Electrical Engineering and Informatics (IJEEI), vol. 13, no. 4, pp. 840-855, Dec. 2025, doi: 10.52549/ijeei.v13i4.6573.

[2] H. Gang, H. Zhang, and Z. Zhang, "AI-based malicious encrypted traffic detection in 5G data collection and secure sharing," Electronics, 2024.

[3] Y. Pan, X. Zhang, and H. Jiang, "A network traffic classification method based on graph convolution and LSTM," 2021, doi: 10.1109/ACCESS.2021.3128181.

[4] R. Pell, M. Shojafar, D. Kosmanos, and S. Moschoyiannis, "Service classification of network traffic in 5G core networks using machine learning," in Proc. IEEE International Conference on Edge Computing and Communications (EDGE), Chicago, IL, USA, pp. 309-318, 2023, doi: 10.1109/EDGE60047.2023.00053.

[5] V. Pham, E. Seo, and T.-M. Chung, "Lightweight convolutional neural network based intrusion detection system," Journal of Communications, vol. 15, no. 11, pp. 808-817, Nov. 2020, doi: 10.12720/jcm.15.11.808-817.

[6] S. Al-Eidi, O. Darwish, G. Husari, Y. Chen, and M. Elkhodr, "Convolutional neural network structure to detect and localize CTC using image processing," in Proc. IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Toronto, ON, Canada, pp. 1-7, 2022, doi: 10.1109/IEMTRONICS55184.2022.9795734.

[7] B. Bousalem, V. Silva, R. Langar, and S. Cherrier, "DDoS attacks detection and mitigation in 5G and beyond networks: A deep learning-based approach," in Proc. IEEE Global Communications Conference (GLOBECOM), Rio de Janeiro, Brazil, pp. 1259-1264, 2022, doi: 10.1109/GLOBECOM48099.2022.10001562.

[8] Y. Huo, W. Liang, J. Chen, and S. Zhuang, "LightGuard: A lightweight malicious traffic detection method for Internet of Things," IEEE Internet of Things Journal, vol. 11, pp. 28566-28577, 2024.

[9] A. Ferriyan, A. H. Thamrin, K. Takeda, and J. Murai, "Encrypted malicious traffic detection based on Word2Vec," Electronics, vol. 11, p. 679, 2022.

[10] L. Globa, A. Astrakhantsev, and S. Tsukanov, "Classification of network traffic using machine learning methods," Problemi telekomunikacìj, no. 2, pp. 3-13, 2023, doi: 10.30837/pt.2023.2.01.

[11] L. Globa, A. Astrakhantsev, and S. Tsukanov, "Comparison of the machine learning algorithms for traffic classification in 5G network," in Proc. IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), Tbilisi, Georgia, pp. 125-130, 2024, doi: 10.1109/BlackSeaCom61746.2024.10646279.

[12] M. S. Khan, B. Farzaneh, N. Shahriar, and M. M. Hasan, "DoS/DDoS Attack Dataset of 5G Network Slicing," IEEE Dataport, Sept. 25, 2023, doi: 10.21227/32k1-dr12.

[13] F. Karim, S. Majumdar, and H. Darabi, "Multivariate LSTM-FCNs for time series classification," Neural Networks, vol. 116, pp. 237-245, 2019, doi: 10.1016/j.neunet.2019.04.014.

[14] Y. Wang, "Deep learning-based network intrusion detection systems," Applied and Computational Engineering, vol. 109, no. 1, pp. 179-188, 2024, doi: 10.54254/2755-2721/2024.18104.

[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[16] H. I. Fawaz, Deep Learning for Time Series Classification, Ph.D. dissertation, Université de Haute Alsace - Mulhouse, 2020, [Online]. Available: https://theses.hal.science/tel-03715016v1.

[17] J. R. Fawcett, "An introduction to ROC analysis," Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, 2006.

Hochshule Anhalt
EDITION
ICAIIT

DATA TRANSPORT TECHNOLOGIES

Section 2

APRIL 2026



Proceedings of the International Conference on Applied Innovations in IT

PPO-Driven Reinforcement Learning Congestion Control Under High-BDP Wide-Area Deployment: A Scaling Analysis

Ali Ghermezian, Kirill Karpov, Dmitry Kachan, Veronika Kirova and Eduard Siemens
Department of Electrical, Mechanical and Industrial Engineering, Anhalt University of Applied Sciences, Bernburger Str. 55, 06366 Köthen, Germany
ali.ghermezian@hs-anhalt.de, kirill.karpov@hs-anhalt.de, dmitry.kachan@hs-anhalt.de, veronika.kirova@hs-anhalt.de, eduard.siemens@hs-anhalt.de

Keywords: Reinforcement Learning, Congestion Control, Proximal Policy Optimization (PPO), High Bandwidth-Delay Product (BDP), Wide-Area Networks (WAN), Aurora.

Abstract: Reinforcement learning (RL) has recently been explored as an adaptive alternative to hand-designed congestion control, yet most reported evaluations remain confined to simulated or moderate-bandwidth environments. This paper studies the scaling behavior of an Aurora-style Proximal Policy Optimization (PPO) congestion control policy when it is deployed on a real high-bandwidth, high bandwidth-delay product (BDP) wide-area network (WAN) path with a 10 Gbps interface budget. The purpose is not to claim a new state-of-the-art controller, but to identify how a simulator-trained PPO policy behaves when transferred to multi-gigabit operation. We integrate the policy through a user-space PCC shim and analyze transport-level logs, including achieved receive rate, loss dynamics, rate oscillations, and available round-trip time (RTT) indicators. The results show a persistent gap between nominal link capacity and achieved single-flow goodput: the analyzed run reaches a mean receive rate of 49.6 Mb/s and a peak of 438.0 Mb/s, corresponding to 0.50% average and 4.38% peak utilization of the 10 Gbps budget. Additional TCP CUBIC and TCP BBR baselines on the same path reached up to 9.91 Gb/s and 9.72 Gb/s with four flows, confirming that the route itself can sustain multi-gigabit throughput. The rate distribution is heavy-tailed, with short high-rate episodes followed by overshoot-collapse dynamics, bursty loss, and transient delay inflation. These findings indicate that scale-aware training, more robust reward normalization, and lower-overhead pacing are needed before PPO-driven congestion control can generalize reliably to high-BDP WAN deployments.

1 INTRODUCTION

Congestion control is a fundamental component of transport protocols, responsible for achieving high link utilization while preventing persistent queue buildup and instability [1]. In high-speed and multi-gigabit wide-area networks (WANs), this problem becomes substantially more challenging due to large bandwidth-delay products (BDP), amplified feedback delays, and increased sensitivity to transient queue dynamics [2]. Classical loss-based algorithms may underutilize available capacity in such environments, while delay-based and control-theoretic approaches attempt to stabilize queue evolution under high-speed conditions.

For example, Karpov et al. analyze performance enhancement strategies for multi-gigabit WANs and emphasize the scaling challenges introduced by large BDP and delayed feedback [3]. Mareev et al. further investigate PID-based congestion control for high-speed IP networks, showing that explicit control-theoretic mechanisms can improve stability compared to heuristic schemes, yet remain sensitive to parameter tuning and network variability [4]. These limitations have motivated the exploration of data-driven approaches, including reinforcement learning (RL), as an alternative to manually designed congestion control logic [5].

RL-based congestion control frames rate adaptation as a sequential decision-making problem: the sender observes recent network statistics, selects a control action, receives a reward that reflects throughput, loss, and delay, and then updates or applies a learned policy. Aurora formulates this problem as a partially observable Markov decision process and uses Proximal Policy Optimization (PPO), a policy-gradient RL algorithm designed to improve a policy while limiting excessively large policy updates [1],

[6]. PPO is attractive for congestion control because it can optimize continuous rate-control actions and can learn non-linear responses to noisy transport measurements. However, these same learned responses may become fragile when the operating scale differs substantially from the training environment.

Published Aurora evaluations report promising results in controlled and moderate-bandwidth environments. However, Aurora's experimental validation has largely been confined to Mbps-scale simulations and emulation, typically below 128 Mb/s [1]. Modern backbone and research networks operate at multi-gigabit rates, where BDP amplification fundamentally alters control dynamics: feedback becomes relatively slower, queueing transients become more costly, and exploratory rate changes can induce amplified oscillations. The behavior and generalization properties of simulator-trained PPO congestion control under real high-BDP multi-gigabit WAN conditions remain insufficiently understood [7].

In this study, we empirically evaluate an Aurora-style PPO congestion control policy deployed over a real wide-area path under a 10 Gbps interface budget. Rather than claiming universal performance improvements, we focus on characterizing utilization limits, heavy-tailed rate behavior, overshoot-collapse dynamics, and instability patterns observed in practice. Our analysis highlights structural constraints, including reward scaling sensitivity, delayed feedback amplification, and user-space pacing overhead. Baseline throughput measurements are reported for path-capacity validation, while full multi-flow fairness under simultaneous coexistence remains future work.

2 METHOD AND METRICS

This section describes the PPO-driven control architecture, the observations used by the policy, the decision interval, the reward signal, and the measurement metrics used for the scaling analysis.

2.1 Control Architecture

We evaluate an Aurora-style congestion control policy integrated through a user-space PCC shim [8]. The sender periodically observes recent transport statistics and outputs a continuous control signal that multiplicatively adjusts the current sending rate.

The implementation follows Aurora's abstraction:

$$R_{t+1} = R_t \cdot \exp(\alpha \cdot a_t), \quad (1)$$

where Rt, measured in Mb/s, is the current sending rate, at is the continuous action output by the policy, alpha is a fixed delta-scale factor, and the rate is clipped within predefined minimum and maximum bounds. Rate limits in our deployment were configured as Rmin = 500 Mb/s and Rmax = 5 Gb/s. These bounds were adjusted during scaling experiments to explore high-BDP behavior. The controller operates entirely in user space through the PCC-RL shim interface.

2.2 Observation Space

At each decision interval, the policy receives a bounded history of recent transport statistics. The observation vector includes estimated throughput over the last interval, packet loss rate, an RTT sample, sending rate, and delivery ratio or acknowledged bytes. The implementation uses a fixed history length of H = 10 intervals. Thus, the observation vector is a concatenation of the last 10 measurement tuples, forming a temporally stacked state representation. All features are normalized using running statistics consistent with the Aurora training implementation, using mean-variance normalization.

2.3 Decision Interval

The control decision is executed at fixed monitoring intervals corresponding to the PCC measurement epoch. In our deployment, the decision interval is approximately one PCC monitor interval, or about 100-200 ms depending on traffic dynamics. This interval length directly affects stability under high BDP, because delayed feedback amplifies multiplicative rate updates.

2.4 Reward Function

The original Aurora reward formulation maximizes a utility combining throughput and delay penalties:

$$r_t = \text{throughput}_t - \beta \cdot \text{delay}_t - \gamma \cdot \text{loss}_t. \quad (2)$$

In our implementation, the reward is computed inside the shim using throughput-based utility consistent with the PCC-RL training plugin:

$$r_t = U(\text{throughput}_t, \text{loss}_t), \quad (3)$$

where the delay contribution is limited due to unavailable RTT precision in receiver logs. For analysis purposes, we report a normalized reward proxy derived from logged throughput and inferred loss dynamics.

2.5 Normalization

Feature normalization follows Aurora's training convention: throughput is normalized relative to a recent mean, loss is represented as a bounded fraction, the action output is constrained to a finite range, and the reward is normalized before PPO updates. However, when transferring to multi-gigabit WAN conditions, absolute throughput scales exceed those seen during simulation training, potentially causing reward scaling mismatch.

2.6 Training and Checkpoint Regime

The evaluated policy was initialized from a pre-trained PPO checkpoint trained in simulated environments under moderate bandwidth regimes. During deployment, online PPO updates were enabled, model checkpoints were periodically saved, and no retraining with high-BDP domain randomization was performed. Thus, the experiments evaluate transfer and scaling behavior rather than re-optimized high-speed training.

2.7 Measurement Metrics

From receiver-side logs, we extract receive rate in Mb/s as the primary goodput proxy, aggregate sent and lost counters, and per-interval rate statistics. Because RTT values in available logs are constant placeholders in part of the dataset, latency-based analysis is limited. Therefore, tail behavior is characterized through mean receive rate, standard deviation, p50, p95, p99, and empirical cumulative distribution function (CDF). These metrics capture burstiness and heavy-tail dynamics under high-BDP conditions.

3 EXPERIMENTAL SETUP

Experiments were conducted on a Linux-based server equipped with a 10 Gbps network interface and a 24-core CPU architecture. The congestion control policy, implemented as an Aurora-style PPO controller, was executed in user space using Stable-Baselines PPO2 and interfaced through the PCC shim driver. Traffic generation and feedback were collected via sender logs and receiver-side CSV exports.

The evaluation focuses on real wide-area conditions with a high-BDP path. Performance statistics were derived exclusively from transport-level logs, including per-interval throughput, RTT indicators, cumulative bytes sent, and loss counters. Receiver logs report a peak single-flow throughput of approximately 438 Mb/s in the analyzed run, while multi-flow experiments occasionally exceeded 1 Gb/s in aggregate throughput when multiple solver instances utilized separate CPU cores.

Single-flow execution was observed to operate under user-space pacing constraints, while multi-flow configurations leveraged parallel CPU scheduling to increase aggregate throughput. Baseline RTT measurements extracted from receiver logs ranged around a few milliseconds under low-queue conditions, with significant spikes observed during rate overshoot events.

All experiments were performed using identical policy checkpoints and fixed hyperparameters to isolate deployment effects from training variability. No kernel-level congestion control modifications were applied; all control logic operated entirely in user space.

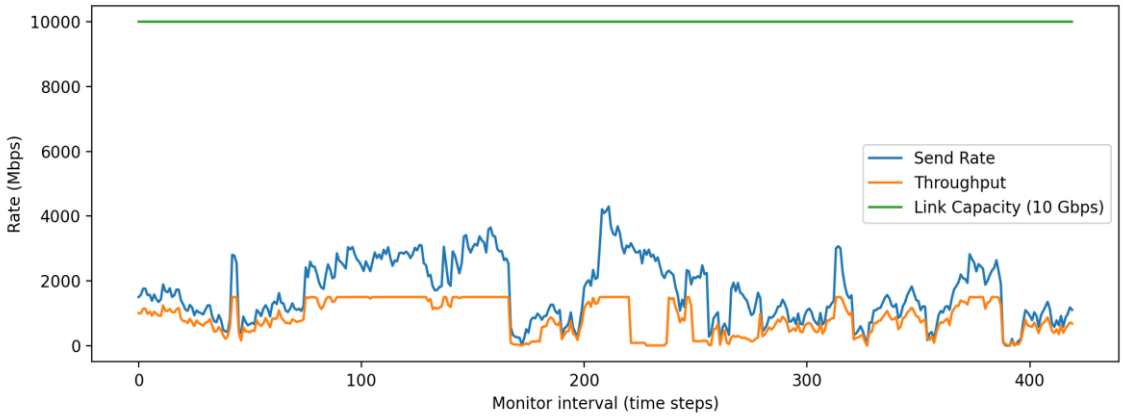


Figure 1: Send rate and achieved throughput over time under a 10 Gbps link capacity.

To address baseline performance on the same physical path, we additionally measured TCP CUBIC and TCP BBR using iperf3 on the Canada-to-Cuba WAN path, with destination 14.14.14.16 and a 10 Gbps nominal interface budget. We also measured PCC/Vivace without RL and RL-controlled PCC over the same user-space PCC deployment using four flows on ports 9001-9004. These baseline measurements were used to distinguish physical path capacity from user-space PCC/RL scaling limitations.

To address baseline performance on the same physical path, we additionally measured TCP CUBIC and TCP BBR using iperf3 on the Canada-to-Cuba WAN path, with destination 14.14.14.16 and a 10 Gbps nominal interface budget. We also measured PCC/Vivace without RL and RL-controlled PCC over the same user-space PCC deployment using four flows on ports 9001-9004. These baseline measurements were used to distinguish physical path capacity from user-space PCC/RL scaling limitations.

4 RESULTS

This section reports the observed utilization, temporal rate behavior, and distributional properties of the achieved receive rate. The purpose is to make the empirical scaling behavior explicit rather than to present the controller as an optimized production implementation.

4.1 Achieved Utilization and Temporal Behavior

Across the analyzed receiver log, the mean receive rate is 49.6 Mb/s with a standard deviation of 90.1 Mb/s and a peak of 438.0 Mb/s. Relative to a 10 Gbps budget, this corresponds to an average utilization of 0.50% and peak utilization of 4.38%. Figure 1 shows pronounced variability and intermittent spikes, consistent with scaling challenges under high-BDP transfer. The time series indicates that the controller can occasionally raise the sending rate, but these increases are not sustained long enough to maintain high utilization.

4.2 Tail Behavior of Achieved Rate

The distribution of achieved receive rate is heavy-tailed: p50 = 18.0 Mb/s, p95 = 334.2 Mb/s, and p99 = 388.8 Mb/s, as shown in Figure 2. Such tail spikes indicate that short-lived bursts occur, but sustained high utilization is not achieved in this run. This is consistent with a generalization gap between simulator training regimes and real 10 Gbps WAN conditions [9].

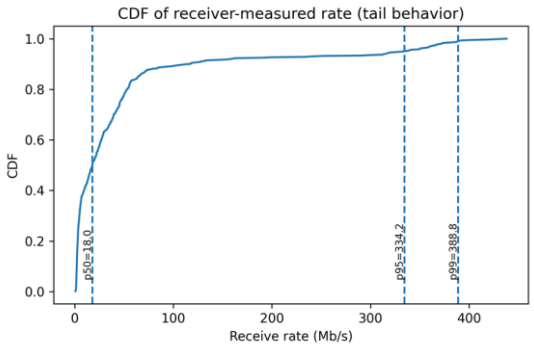


Figure 2: CDF of receiver-measured rate; vertical markers show p50, p95, and p99.

4.3 Loss and Delay Indicators

Figure 3 compares loss-rate evolution with available RTT indicators. Aggressive rate increases correlate with loss bursts and transient RTT inflation, suggesting queue buildup and delayed feedback amplification. Because the RTT field is not fully reliable in all receiver logs, these delay observations should be interpreted qualitatively. Nevertheless, the synchronized pattern of rate excursions, loss bursts, and delay spikes supports the conclusion that high-BDP feedback delay contributes to instability.

4.4 Baseline Comparison on the Canada-to-Cuba WAN Path

To determine whether the low utilization was caused by the network path itself or by the user-space PCC/RL stack, we conducted additional baseline tests on the same Canada-to-Cuba WAN path. Table 1 summarizes TCP CUBIC, TCP BBR, PCC/Vivace without RL, RL-controlled PCC, and the original analyzed RL run. Single-flow TCP results provide a reference for the original analyzed RL run, while four-flow TCP results match the multi-flow PCC/Vivace and RL-PCC measurements and demonstrate available path capacity under parallel flows. The TCP baselines confirm that the path can sustain multi-gigabit throughput: four-flow CUBIC and BBR achieved 9.91 Gb/s and 9.72 Gb/s, respectively. In contrast, four-flow PCC/Vivace without RL and RL-controlled PCC remained near 1 Gb/s. Figure 4 visualizes the same goodput comparison.

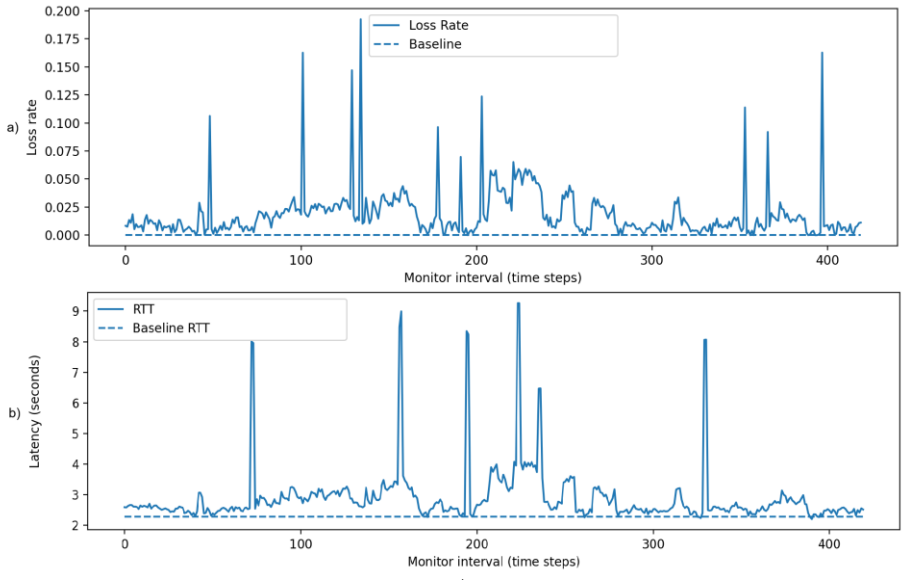


Figure 3: Loss rate and RTT evolution over time. Aggressive rate increases correlate with loss bursts and transient RTT inflation, reflecting queue buildup and delayed feedback dynamics.

Table 1: Baseline comparison on the Canada-to-Cuba WAN path.

Method	No.	Source	Mean	Peak	RTT	Loss/retrans.	Util.
TCP CUBIC	1	iperf3	4.57 Gb/s	N/A	N/A	2,460 retrans.	45.7%
TCP BBR	1	iperf3	8.72 Gb/s	N/A	N/A	692,138 retrans.	87.2%
TCP CUBIC	4	iperf3	9.91 Gb/s	N/A	N/A	5,422 retrans.	99.1%
TCP BBR	4	iperf3	9.72 Gb/s	N/A	N/A	48,116 retrans.	97.2%
PCC/Vivace	4	PCC	0.987 Gb/s	1.251 Gb/s	20.29 ms	4 lost pkt.	9.87%
RL-PCC	4	PCC/RL	0.967 Gb/s	1.022 Gb/s	20.20 ms	280 lost pkt.	9.67%
Original RL	1	Orig. data	0.0496 Gb/s	0.438 Gb/s	Limited	Bursty loss	0.50%

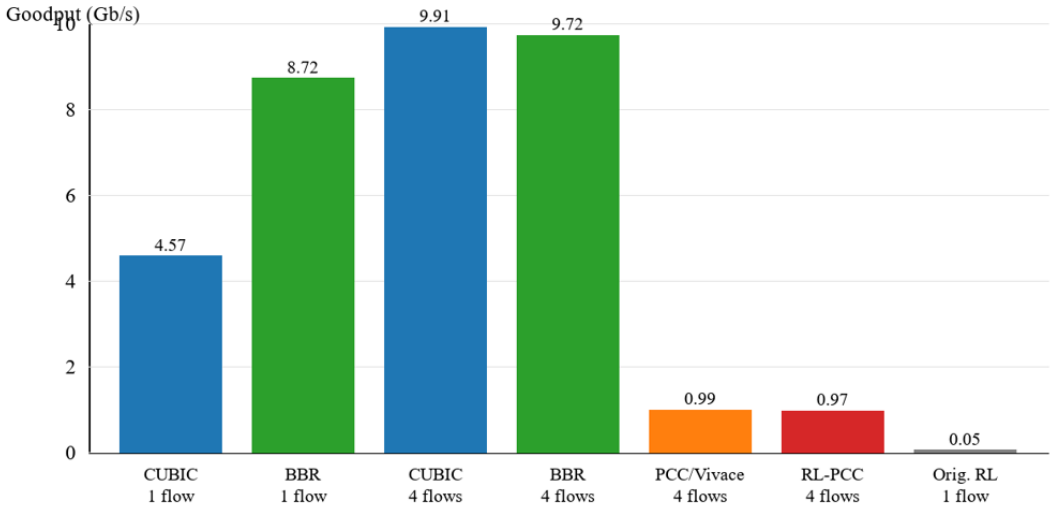


Figure 4: Mean goodput comparison on the same Canada-to-Cuba WAN path. TCP CUBIC and BBR confirm that the path can sustain multi-gigabit throughput, while user-space PCC/Vivace and RL-controlled PCC remain near 1 Gb/s.

These results show that the physical WAN path is not the primary bottleneck. Instead, the observed performance gap is consistent with limitations in the user-space PCC/RL data path, packet tracking, pacing precision, and feedback-driven control loop under high-BDP operation. The RL-controlled PCC configuration did not outperform PCC/Vivace without RL in this test; it achieved slightly lower mean goodput and higher packet loss.

5 DISCUSSION

The empirical results reveal a substantial gap between the nominal interface capacity and the achieved goodput in single-flow operation. Receiver-side measurements indicate that throughput remains far below the available link budget, while the CDF analysis confirms that high-rate episodes are rare and short-lived rather than sustained.

The send-rate time series in Figure 1 exhibits pronounced overshoot-collapse behavior. Rapid multiplicative rate increases are followed by abrupt drops, indicating unstable adaptation dynamics under high-BDP conditions. This instability is further supported by Figure 3, where aggressive rate excursions correlate with transient loss bursts and RTT spikes, reflecting queue buildup and delayed feedback amplification.

These observations suggest that policies trained under moderate-bandwidth simulation regimes may not generalize seamlessly to multi-gigabit WAN environments, particularly when deployed in real-network settings where measurement noise, scheduling variability, and implementation overhead are present [8]. Prototype experimentation frameworks such as Iroko [10] have been proposed to facilitate controlled evaluation of reinforcement learning congestion control policies prior to deployment. However, translating such prototypes to high-BDP wide-area environments introduces additional system-level constraints, including large in-flight data volumes and amplified feedback delay, which increase sensitivity to exploratory rate updates and promote oscillatory behavior.

The implementation-level constraints are important. A user-space controller must schedule monitoring, policy inference, action application, and packet pacing outside the kernel data path. At multi-gigabit rates, CPU scheduling jitter, timer granularity, socket-buffer behavior, and packet pacing precision can limit achievable throughput even when the learned policy selects aggressive rates. The observation that multi-flow runs improve aggregate throughput by using multiple CPU cores suggests that computation and pacing overhead contribute to the single-flow ceiling, although the dataset does not isolate these effects experimentally.

The additional baseline comparison strengthens this interpretation. TCP CUBIC and TCP BBR reached 9.91 Gb/s and 9.72 Gb/s in four-flow operation on the same path, confirming that the physical route and 10 Gbps interface budget can support near-line-rate throughput. By contrast, PCC/Vivace without RL and RL-controlled PCC remained close to 1 Gb/s. Therefore, the utilization ceiling observed for the learning-based deployment should be attributed primarily to user-space PCC/RL implementation and control-loop scaling constraints rather than to insufficient path capacity.

While multi-flow execution improves aggregate throughput through multi-core utilization, per-flow instability patterns remain visible in loss and latency traces. Fairness under coexistence with conventional TCP variants was not fully evaluated beyond the baseline throughput comparison and remains an open question for future study. Systematic evaluations of learning-based congestion control highlight the importance of quantitative fairness, efficiency, and responsiveness metrics when assessing deployment viability [9].

Overall, the combined evidence from rate dynamics, loss behavior, RTT evolution, and distributional analysis indicates that both control-level sensitivity and deployment-level constraints contribute to the observed utilization ceiling in high-BDP wide-area environments.

6 LIMITATIONS

The present study is limited to a single high-BDP wide-area deployment and does not explore systematic variations in RTT, cross-traffic intensity, or alternative bottleneck configurations. As a result, the observed instability patterns may depend on the specific path characteristics considered here.

The performance analysis is primarily derived from transport-level logs. Although sufficient for identifying macro-level dynamics, the available measurements restrict fine-grained reconstruction of transient loss events. Moreover, RTT measurements are not fully available with sufficient precision across all logs, so latency analysis remains limited and should be interpreted as indicative rather than definitive.

The implementation operates entirely in user space. User-space pacing and per-interval

computation may introduce additional scheduling overhead that constrains achievable throughput, particularly at multi-gigabit rates. CPU scheduling, timer resolution, packet pacing precision, and kernel versus user-space data-path differences should be isolated in future experiments.

Finally, the added CUBIC and BBR measurements provide a path-capacity baseline, but they do not constitute a full coexistence or fairness study under mixed traffic. A controlled comparison with simultaneous competing flows would still be necessary to quantify fairness and responsiveness under shared-bottleneck conditions.

7 CONCLUSIONS

This work empirically examined the scaling behavior of an Aurora-style PPO congestion control deployment over a real high-BDP wide-area path with a 10 Gbps interface budget. The measurements show that single-flow utilization remains significantly below nominal capacity and is characterized by oscillatory rate dynamics, bursty loss behavior, and transient RTT inflation.

The throughput distribution further indicates that high-rate episodes are short-lived rather than sustained, suggesting that policy adaptation becomes increasingly sensitive in large bandwidth-delay environments. Taken together, these results imply that transferring reinforcement learning-based congestion control from moderate-bandwidth training regimes to multi-gigabit WAN deployments requires greater robustness to delayed feedback, scale-aware reward normalization, and improved implementation support for precise high-rate pacing.

Future investigations should extend the analysis to simultaneous coexistence experiments with CUBIC and BBR, explicit fairness evaluation, richer RTT and queue measurements, and architectural optimizations that reduce user-space pacing overhead.

8 ACKNOWLEDGMENTS

This work was supported by the European Regional Development Fund (ERDF/EFRE) and the State of Saxony-Anhalt within the programme Sachsen-Anhalt WISSENSCHAFT Forschung und Innovation (EFRE) 2021-2027, project AI-RMDT (grant no. ZS/2023/12/182323). We acknowledge support by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) and the Open Access Publishing Fund of Anhalt University of Applied Sciences.

9 REFERENCES

[1] N. Jay, N. T. Li and B. Hariharan, “A deep reinforcement learning perspective on internet congestion control,” in Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, CA, USA, 2019, pp. 3050-3059, [Online]. Available: https://proceedings.mlr.press/v97/jay19a/jay19a.pdf?utm_source=chatgpt.com.

[2] K. Winstein and H. Balakrishnan, “TCP ex machina: Computer-generated congestion control,” ACM SIGCOMM Computer Communication Review, vol. 43, no. 4, pp. 123-134, 2013, [Online]. Available: https://doi.org/10.1145/2486001.2486020.

[3] K. Karpov, D. Kachan, V. Kirova, A. Ghermezian, D. Koshutina and E. Siemens, “Tunable multi-objective tree synthesis for application-layer multicast,” in Proceedings of the International Conference on Applied Innovations in IT (ICAIIT), Dec. 2025, pp. 68-73.

[4] N. Mareev, D. Kachan, K. Karpov, D. Syzov, E. Siemens and Y. Babich, “Efficiency of a PID-based congestion control for high-speed IP-networks,” in Proceedings of ICAIIT, vol. 6, no. 1, pp. 129-133, 2018, [Online]. Available: https://doi.org/10.13142/kt10006.45.

[5] M. Dong, Q. Li, D. Zats, J. R. Sokoll and I. Stoica, “PCC Vivace: Online-learning congestion control,” in Proceedings of the 15th USENIX Symposium on Networked Systems Design and Implementation (NSDI), 2018, pp. 343-356.

[6] S. Abbasloo, C.-Y. Yen and H. J. Chao, “Classic meets modern: A pragmatic learning-based congestion control for the Internet,” in Proceedings of the ACM SIGCOMM Conference, 2020, pp. 632-647, [Online]. Available: https://doi.org/10.1145/3387514.3405892.

[7] C. Liao, Y. Zhang and K. Winstein, “Astraea: Towards fair and efficient learning-based congestion control,” in Proceedings of EuroSys, 2024, [Online]. Available: https://doi.org/10.48550/arXiv.2403.01798.

[8] R. Galliera, A. Morelli, R. Fronteddu and N. Suri, “MARLIN: Soft actor-critic based reinforcement learning for congestion control in real networks,” in Proceedings of the IEEE/IFIP Network Operations and Management Symposium (NOMS), 2023, [Online]. Available: https://doi.org/10.1109/NOMS56928.2023.10154210.

[9] L. Giacomoni and G. Parisis, “Reinforcement learning-based congestion control: A systematic evaluation of fairness, efficiency and responsiveness,” in Proceedings of the IEEE INFOCOM Conference, 2024, pp. 1451-1460, [Online]. Available: https://doi.org/10.1109/INFOCOM52122.2024.10621288.

[10] F. Ruffy, M. Przystupa and I. Beschastnikh, “Iroko: A framework to prototype reinforcement learning for data center traffic control,” arXiv preprint arXiv:1812.09975, 2018, [Online]. Available: https://doi.org/10.48550/arXiv.1812.09975.

DevOps Techniques for Artificial Intelligence Development and Training: A Human-Centered Approach to AI-Driven Operations

Kateryna Shulakova[1,2], Liliia Bodnar[3], Mykola Bodnar[4], Roman Tsarov[1], Olha Yavorska[1], Vladyslav Kumysh[1] and Oleksandra Ordanovska[3]

[1]*State University of Intelligent Technologies and Telecommunications, Kuznechna Str. 1, Odesa, Ukraine*

[2]*Anhalt University of Applied Sciences, Bernburger Str. 57, Köthen, Germany*

[3]*South Ukrainian National Pedagogical University, Staroportofrankyvska Str. 26, 65020 Odesa, Ukraine*

[4] *LLC B&B Solutions, Dukivska Str. 5 65023 Odesa, Ukraine*

katejojo29@gmail.com, bodnarl79@pdpu.edu.ua, rcarev@gmail.com, yavorskayao7@gmail.com, vlad.kumish@gmail.com, aleksordanovskaya@pdpu.edu.ua

Keywords: DevOps, AIOps, MLOps, DevSecOps, Artificial Intelligence, Automation Paradox, Human-Centered AI, Intelligent Monitoring, Cloud-Native Infrastructure, Operational Governance, CI/CD, Infrastructure as Code, Kubernetes, AI Operations, Predictive Analytics.

Abstract: The rapid integration of artificial intelligence technologies into software engineering and cloud-native infrastructures has significantly transformed modern operational practices and automation strategies. Traditional DevOps methodologies based on continuous integration and continuous deployment, infrastructure automation, and container orchestration remain essential for scalable software delivery; however, the increasing complexity of distributed AI-oriented systems creates new operational, organizational, and security-related challenges. The emergence of intelligent operational paradigms such as AIOps, MLOps, and DevSecOps demonstrates the growing need for adaptive monitoring, predictive analytics, automated decision support, and secure lifecycle management within AI-driven infrastructures. This paper investigates the application of DevOps techniques in artificial intelligence development and training environments from a human-centered operational perspective. Particular attention is devoted to the automation paradox, where increasing automation levels may simultaneously improve operational efficiency while generating hidden technical complexity, infrastructure dependency, and reduced engineering visibility. The study analyzes the role of intelligent monitoring, anomaly detection, predictive scaling, and automated operational analytics in modern cloud ecosystems while emphasizing the importance of human expertise in supervision, validation, and governance processes. The paper proposes an integrated conceptual framework combining DevOps, AIOps, MLOps, and DevSecOps principles for sustainable AI operational management. Additionally, a conceptual automation-risk model is introduced to describe the relationship between automation intensity, infrastructure complexity, data uncertainty, and engineering expertise. The results demonstrate that artificial intelligence should function as an augmentation mechanism for engineering decision-making rather than a replacement for human operational control. The proposed approach contributes to the development of secure, scalable, and sustainable AI-driven infrastructures capable of balancing automation efficiency with governance, transparency, and operational reliability.

1 INTRODUCTION

The rapid development of artificial intelligence (AI) technologies has significantly transformed modern software engineering, cloud computing, and infrastructure management processes. AI-driven systems are increasingly integrated into distributed cloud-native environments, automated operational workflows, and large-scale data-processing infrastructures. Simultaneously, the growing complexity of modern software ecosystems requires new approaches to automation, scalability, operational monitoring, and infrastructure governance [1]-[3].

The DevOps paradigm, based on continuous integration and continuous deployment (CI/CD), infrastructure automation, and collaborative operational practices, has become one of the dominant methodologies for accelerating software

delivery and improving infrastructure reliability [1]-[3]. The emergence of containerization and orchestration platforms such as Docker and Kubernetes further enabled scalable deployment of distributed services and microservice-based architectures [4], [5]. In parallel, Infrastructure as Code (IaC) approaches improved reproducibility and operational consistency in cloud-native infrastructures [21].

However, despite the significant advantages of classical DevOps automation, modern AI-oriented infrastructures increasingly exceed the capabilities of traditional rule-based operational management. Large-scale AI systems require continuous model retraining, dynamic resource allocation, intelligent anomaly detection, and predictive operational analytics [6], [12], [13]. Existing studies demonstrate that machine learning systems introduce additional operational complexity associated with hidden technical debt, model drift, dependency management, and continuous lifecycle maintenance [6], [13].

To address these challenges, new operational paradigms such as Machine Learning Operations (MLOps), Artificial Intelligence for IT Operations (AIOps), and DevSecOps have emerged. MLOps extends DevOps principles toward machine learning lifecycle management, including dataset versioning, model training, deployment, and retraining pipelines [7], [12], [16]. AIOps integrates machine learning algorithms into operational monitoring systems for anomaly detection, predictive analytics, and intelligent event correlation in complex infrastructures [9], [17]. DevSecOps further incorporates security validation and compliance mechanisms into CI/CD workflows to improve operational resilience and secure software delivery [11].

Despite the rapid adoption of AI-enhanced operational frameworks, the increasing dependence on intelligent automation creates new organizational and technical risks. One of the most important challenges is the so-called automation paradox, where higher levels of automation may simultaneously reduce manual workload while increasing hidden infrastructural complexity and dependence on expert supervision. In practice, fully autonomous operational systems remain unattainable because AI technologies still require human validation, governance, and strategic decision-making [6], [9].

Furthermore, excessive reliance on AI-generated recommendations may reduce engineering visibility and operational awareness in critical infrastructures. Intelligent systems can improve operational efficiency and scalability; however, they may also generate explainability limitations, false-positive recommendations, infrastructure fragility, and operational overconfidence [9], [20]. Consequently, the role of the engineer shifts from direct operational execution toward supervision, governance, anomaly interpretation, and strategic control of intelligent operational ecosystems.

Security considerations also become increasingly important in AI-oriented infrastructures. AI systems introduce additional attack surfaces related to training datasets, model artifacts, operational telemetry, and automated decision-making mechanisms. International standards and governance frameworks emphasize the importance of transparency, accountability, risk management, and continuous monitoring across the entire AI lifecycle [14], [15]. Therefore, sustainable AI operations require not only intelligent automation technologies but also governance-oriented operational models capable of balancing automation efficiency, security, transparency, and human-centered supervision.

This paper investigates the application of DevOps techniques in artificial intelligence development and training environments from a human-centered operational perspective. The study focuses on the integration of DevOps, AIOps, MLOps, and DevSecOps methodologies in cloud-native infrastructures while addressing the limitations of intelligent automation and the operational risks associated with excessive dependence on AI-driven systems.

The main contributions of this paper are as follows:

1) Development of a conceptual framework integrating DevOps, AIOps, MLOps, and DevSecOps methodologies for AI operational management.
2) Formalization of the automation paradox in AI-driven infrastructures.
3) Analysis of operational risks associated with intelligent automation and hidden technical complexity.
4) Proposal of a human-centered governance model for sustainable AI operations.
5) Introduction of an analytical automation-risk formulation for evaluating operational sustainability in AI-enhanced infrastructures.

2 RELATED WORK

2.1 DevOps Foundations and Continuous Delivery

The DevOps paradigm emerged as a response to the need for faster, more reliable, and continuously evolving software delivery processes. Continuous integration and continuous delivery (CI/CD) form the

methodological core of DevOps, enabling automated testing, integration, and deployment pipelines across complex software systems [1]. These principles emphasize collaboration between development and operations teams and establish continuous feedback loops for improving software quality and deployment efficiency [2].

From an architectural perspective, DevOps represents a shift from isolated development cycles toward integrated lifecycle management, where infrastructure, deployment, and monitoring are treated as continuous processes rather than discrete stages [3]. This transformation has become a foundational concept for modern cloud-native systems.

2.2 Cloud-Native Infrastructure and Automation

The evolution of containerization technologies has significantly accelerated the adoption of DevOps practices. Docker introduced lightweight virtualization for application packaging and deployment, enabling portability across heterogeneous environments [4]. Kubernetes further extended these capabilities by providing orchestration, scaling, and automated resource management for distributed microservice architectures [5].

Infrastructure as Code (IaC) represents another critical advancement in cloud automation, allowing infrastructure provisioning and configuration to be managed through version-controlled code. Recent studies demonstrate that IaC improves reproducibility, scalability, and operational consistency in cloud-based environments while reducing manual configuration errors [6]. However, increasing automation complexity also introduces additional operational overhead and dependency on tooling ecosystems.

2.3 Machine Learning Systems and Operational Complexity

Unlike traditional software systems, machine learning applications introduce additional lifecycle complexity due to their dependence on data, model retraining, and continuous evaluation. Sculley et al. identified that machine learning systems accumulate significant “hidden technical debt” caused by data dependencies, pipeline fragility, and evolving model behavior [7].

To address these challenges, Machine Learning Operations (MLOps) has emerged as an extension of DevOps principles to machine learning lifecycle management. MLOps integrates dataset versioning, model training, deployment, monitoring, and retraining into unified operational pipelines [8]. Empirical studies further highlight that engineering machine learning systems requires specialized workflows that differ significantly from traditional software engineering practices [9].

Frameworks such as MLflow provide infrastructure for experiment tracking, model versioning, and lifecycle reproducibility, enabling more structured management of AI systems in production environments [10]. Nevertheless, continuous delivery of machine learning models remains challenging due to data drift, model instability, and evaluation uncertainty.

2.4 Continuous Delivery Challenges in Modern Systems

Despite its advantages, continuous delivery introduces operational and organizational challenges in large-scale systems. Research indicates that frequent deployments increase system complexity, require stronger monitoring mechanisms, and introduce new risks related to system stability and rollback management [11]. These challenges become more pronounced in AI-driven environments where models evolve continuously and depend on dynamic datasets.

Industrial implementations of MLOps pipelines demonstrate that automation must be carefully balanced with validation mechanisms and human oversight to ensure system reliability and consistency [12].

2.5 AIOps and Intelligent Operations

Artificial Intelligence for IT Operations (AIOps) represents a shift toward applying machine learning techniques for operational monitoring, anomaly detection, and predictive analytics in complex IT environments. AIOps systems aggregate and analyze large volumes of telemetry data to identify system anomalies and support decision-making processes in real time [13].

Cloud-native automation frameworks further extend AIOps capabilities by integrating intelligent monitoring and adaptive scaling mechanisms across distributed infrastructures [14]. However, the effectiveness of AIOps systems strongly depends on data quality, system observability, and model interpretability, which remain open research challenges.

2.6 DevSecOps and Security in AI Systems

Security considerations have become increasingly important in modern DevOps pipelines. DevSecOps integrates security validation, vulnerability scanning, and compliance checks directly into CI/CD workflows, ensuring that security is addressed throughout the entire software lifecycle rather than as a post-deployment activity [15].

In AI-driven systems, security concerns are amplified due to additional attack surfaces such as training data manipulation, model poisoning, and adversarial inputs. International standards such as ISO/IEC 5338 and the NIST AI Risk Management Framework emphasize lifecycle-wide governance, transparency, and accountability in AI systems [16], [17].

2.7 AI Systems and Theoretical Foundations

Artificial intelligence systems operate on foundational principles of learning, reasoning, and adaptation, which are extensively formalized in classical AI literature [18]. However, the integration of AI into operational environments introduces new challenges that extend beyond theoretical AI models into real-world engineering constraints.

2.8 Practical Implementations and Industrial Experience

Industrial case studies demonstrate that successful DevOps adoption requires not only technical tooling but also organizational maturity and structured engineering practices [18]. Large-scale DevOps implementations highlight the importance of standardized pipelines, automated testing, and continuous monitoring for maintaining system stability in production environments [19].

Further research shows that infrastructure cost optimization and resource efficiency are critical factors in container-based cloud deployments, particularly in large-scale distributed systems [21].

2.9 Research Gap

Although extensive research exists in DevOps, cloud-native systems, MLOps, AIOps, and DevSecOps independently, there is still a lack of unified frameworks that integrate these paradigms under a human-centered operational model. In particular, the interaction between automation complexity, AI-driven decision-making, and human supervision remains insufficiently formalized.

This gap motivates the need for a unified approach that addresses the automation paradox in AI-driven DevOps infrastructures and balances operational efficiency with governance, transparency, and human control.

Although the automation paradox has been theoretically described in classical human factors research [22], its formalization within the specific context of AI-driven DevOps and cloud-native infrastructures - integrating MLOps, AIOps, and DevSecOps simultaneously - has not been systematically addressed in the existing literature. Furthermore, prior work has not proposed a governance-oriented risk model that explicitly incorporates data uncertainty and governance maturity as operational variables.

3 AUTOMATION PARADOX THEORY

The increasing integration of artificial intelligence into DevOps and cloud-native operational environments has intensified the role of automation in modern infrastructure management. Automated deployment pipelines, intelligent monitoring systems, predictive scaling, and AI-enhanced operational analytics significantly improve deployment speed, scalability, and infrastructure adaptability [1], [2], [13]. However, despite these advantages, the growing dependence on intelligent automation introduces a critical contradiction commonly referred to as the automation paradox.

The concept of the automation paradox is not new to engineering science. Bainbridge [22] identified the fundamental "ironies of automation" as early as 1983, demonstrating that automated systems paradoxically increase the cognitive demands placed on human operators rather than eliminating them. In modern AI-driven DevOps infrastructures, this phenomenon is significantly amplified. The automation paradox describes a situation in which increasing levels of automation simultaneously reduce routine manual effort while increasing hidden operational complexity and dependence on expert supervision. Instead of eliminating human involvement, automation transforms the role of engineers from direct operational execution toward supervision, governance, anomaly interpretation, and strategic decision-making [7], [13].

In AI-driven infrastructures, this phenomenon becomes particularly significant due to the adaptive and probabilistic nature of machine learning systems.

Unlike deterministic software automation, AI systems continuously evolve through changing datasets, retraining processes, and dynamic operational environments [8], [9]. As a result, fully autonomous operation remains unattainable because intelligent systems require continuous validation, monitoring, and contextual interpretation by human experts.

Existing studies on machine learning engineering demonstrate that AI systems accumulate hidden technical debt caused by unstable dependencies, evolving datasets, configuration drift, and pipeline fragility [7]. These factors increase operational uncertainty and reduce infrastructure transparency, particularly in highly distributed cloud-native ecosystems. Consequently, organizations adopting AI-enhanced DevOps practices frequently encounter situations where automation simplifies low-level operational tasks while simultaneously increasing systemic fragility at the architectural level.

The paradox is further amplified in AIOps environments, where machine learning algorithms process large volumes of telemetry data and generate operational recommendations in real time [13], [14]. Although intelligent monitoring systems improve anomaly detection and predictive analytics capabilities, excessive dependence on AI-generated recommendations may reduce engineering visibility and operational awareness. Engineers may increasingly rely on automated decision support mechanisms without fully understanding the underlying operational context or model limitations.

From a theoretical perspective, the automation paradox can be represented as the nonlinear relationship between automation intensity and operational complexity. At low and moderate levels, automation reduces repetitive workload and improves infrastructure efficiency. However, beyond a certain threshold, additional automation introduces hidden dependencies, explainability limitations, integration complexity, and increased governance requirements.

This relationship may be conceptually represented as:

$$R = \alpha A + \beta C + \gamma D - \delta H$$

where: R - operational automation risk; A - automation intensity; C - infrastructure complexity; D - data uncertainty and model drift; H - level of human expertise and operational supervision; $\alpha,\beta,\gamma,\delta$ - weighting coefficients describing infrastructure sensitivity.

The structure of this formulation is consistent with established risk-modeling approaches in systems engineering, where operational risk is expressed as a product of likelihood factors modulated by control mechanisms [23]. The weighting coefficients α, β, γ, δ are infrastructure-specific and should be empirically calibrated for each deployment context. In the absence of empirical data, this model serves as a qualitative decision-support tool rather than a deterministic predictive instrument - a limitation explicitly acknowledged by the authors. Quantitative calibration of these coefficients remains a direction for future empirical research.

The model illustrates that operational risk increases proportionally with automation intensity, infrastructural complexity, and uncertainty in AI-driven decision-making, while human expertise functions as a stabilizing factor that reduces systemic fragility.

Another important characteristic of the automation paradox is the asymmetry between operational efficiency and explainability. AI systems may provide highly efficient recommendations while remaining partially opaque to operators. This creates a governance challenge in critical infrastructures where accountability and interpretability remain essential operational requirements [15]-[17].

Furthermore, intelligent automation may contribute to gradual degradation of engineering intuition. When engineers become excessively dependent on AI-generated operational recommendations, their ability to independently diagnose failures and interpret complex infrastructure behavior may decline over time. This issue is especially critical in large-scale cloud-native systems where cascading failures require rapid contextual reasoning beyond predefined automation scenarios [5], [18].

Therefore, the automation paradox demonstrates that sustainable AI operations cannot rely solely on increasing automation levels. Instead, effective AI-driven DevOps infrastructures require human-centered governance models that combine intelligent automation with transparency, explainability, operational visibility, and continuous expert supervision.

4 HUMAN-CENTERED AIOPS FRAMEWORK

The increasing complexity of AI-driven infrastructures requires operational models capable of combining intelligent automation with continuous human supervision. Existing AIOps implementations primarily focus on anomaly detection, predictive analytics, and automated operational responses;

however, excessive automation may reduce infrastructure transparency and introduce additional governance risks [13], [14]. Therefore, this study proposes a human-centered AIOps framework designed to balance operational efficiency with explainability, accountability, and engineering control.

The proposed framework integrates DevOps, MLOps, AIOps, and DevSecOps principles into a unified operational ecosystem. The framework consists of five interconnected layers:

1) Infrastructure layer - cloud-native environments, container orchestration platforms, and Infrastructure as Code mechanisms [4]-[6].
2) Operational data layer - telemetry collection, logs, traces, metrics, and event aggregation.
3) Intelligence Layer - machine learning models responsible for anomaly detection, predictive analytics, and operational recommendations [8], [13].
4) Governance layer - policy validation, explainability mechanisms, security controls, and compliance management [15]-[17].
5) Human supervision layer - expert validation, strategic decision-making, incident interpretation, and operational governance.

The key principle of the framework is that AI-generated operational actions should remain bounded by human-centered validation policies. Automated remediation may be acceptable for low-risk operational tasks, while critical infrastructural changes require human approval before execution.

The proposed framework also emphasizes operational transparency. All AI-generated recommendations must remain observable, explainable, and traceable within the operational lifecycle. This approach reduces the probability of cascading failures caused by opaque decision-making mechanisms and improves infrastructure resilience in highly dynamic environments.

4.1 Risk Model

Operational sustainability in AI-driven infrastructures depends on the interaction between automation intensity, infrastructure complexity, data quality, and human supervision. To formalize this relationship, the proposed study introduces a conceptual operational risk model.

The overall operational risk may be expressed as:

$$R = \frac{A(C + D + S)}{H + G}$$

where: S - security exposure; G - governance maturity.

The model demonstrates that operational risk increases proportionally with automation intensity, infrastructure complexity, uncertainty, and security exposure. Simultaneously, risk decreases as human expertise and governance maturity increase.

The relationship described by the model is nonlinear. At moderate automation levels, operational efficiency improves due to reduced manual workload and faster response times. However, excessive automation may generate hidden dependencies, reduced explainability, and operational fragility. This effect becomes especially visible in highly distributed infrastructures characterized by dynamic scaling, autonomous remediation, and adaptive machine learning models.

The proposed formulation does not represent a deterministic mathematical prediction model; instead, it functions as a conceptual analytical framework for evaluating operational sustainability in intelligent infrastructures.

4.2 Comparative Analysis

To evaluate the role of intelligent operational paradigms in modern infrastructures, a comparative analysis of traditional DevOps automation and AI-enhanced operational approaches was performed in Table 1. The characteristics attributed to each paradigm are derived from the following sources: Traditional DevOps [1]-[3]; MLOps [8], [12]; AIOps [13], [14]. The Human-Centered AIOps column reflects the proposed framework introduced in this study.

The analysis demonstrates that traditional DevOps environments provide high operational transparency but limited adaptive capabilities. In contrast, AIOps systems improve scalability and predictive analytics but may reduce explainability and increase dependence on automated recommendations.

The proposed human-centered AIOps model attempts to balance these trade-offs by preserving operational visibility and governance while still leveraging intelligent automation capabilities.

Table 1: Comparative analysis of operational paradigms.

Characteristic	Traditional DevOps	MLOps	AIOps	Human-Centered AIOps
Primary Focus	Software delivery	ML lifecycle	Intelligent operations	Governance-oriented automation
Automation Type	Rule-based	Pipeline-based	Predictive/adaptive	Supervised intelligent automation
Human Involvement	Moderate	High	Reduced	Continuous supervision
Explainability	High	Medium	Low-Medium	High
Operational Risk	Moderate	High	High	Controlled
Security Integration	DevSecOps	Secure ML pipelines	Monitoring-oriented	Governance-integrated
Decision Transparency	High	Medium	Low	High

5 DISCUSSION

The results of this study demonstrate that intelligent automation significantly changes the operational dynamics of cloud-native infrastructures. While AI-enhanced operational systems improve scalability, anomaly detection, and predictive analytics capabilities, they also introduce additional complexity associated with governance, explainability, and dependency management.

One of the most important findings is that increasing automation does not eliminate the need for human expertise. Instead, the role of engineers evolves from direct operational execution toward strategic supervision, validation, and governance. This transformation aligns with the automation paradox discussed earlier, where automation simultaneously improves efficiency and increases hidden infrastructural complexity.

The study further indicates that operational sustainability depends not only on the quality of AI models but also on governance maturity, transparency mechanisms, and organizational readiness. Infrastructures lacking operational visibility and explainability may become vulnerable to cascading failures, model drift, and incorrect automated decisions.

Security also emerges as a critical dimension of AI-enhanced infrastructures. AI systems introduce additional attack surfaces related to operational telemetry, model artifacts, and automated remediation mechanisms. Therefore, secure operational governance must become an integral component of intelligent infrastructure management.

Another important implication concerns engineering education and operational culture. Excessive dependence on AI-generated operational recommendations may reduce engineering intuition and diagnostic reasoning capabilities over time. Consequently, sustainable AIOps adoption requires balancing automation efficiency with continuous human expertise development.

The proposed framework therefore supports a human-centered operational philosophy in which artificial intelligence functions as an augmentation mechanism rather than a replacement for engineering decision-making.

This study presents a conceptual framework and does not provide empirical validation in live production environments. The proposed risk model relies on qualitative weighting rather than statistically derived coefficients. The comparative analysis in Table 1 is based on a literature synthesis rather than controlled experimental conditions.

6 CONCLUSIONS

This paper investigated the application of DevOps techniques in artificial intelligence development and training environments from a human-centered operational perspective. The study analyzed the convergence of DevOps, MLOps, AIOps, and DevSecOps methodologies in modern cloud-native infrastructures and identified the operational challenges associated with excessive dependence on intelligent automation.

Particular attention was devoted to the automation paradox, which describes the contradiction between increasing automation levels and the growing need for expert supervision and governance. The results demonstrate that AI-driven operational systems improve scalability, predictive analytics, and infrastructure adaptability; however, they simultaneously introduce hidden technical complexity, explainability limitations, and governance-related risks.

To address these challenges, the paper proposed a human-centered AIOps framework integrating intelligent monitoring, governance-oriented automation, security mechanisms, and continuous

expert supervision. Additionally, a conceptual operational risk model was introduced to formalize the relationship between automation intensity, infrastructure complexity, data uncertainty, and human expertise.

The findings indicate that sustainable AI-driven infrastructures cannot rely solely on autonomous operational mechanisms. Instead, successful intelligent infrastructures require balanced integration of automation, transparency, governance, and human operational control.

Future research may focus on empirical validation of the proposed framework in real-world cloud-native infrastructures, quantitative evaluation of automation-risk relationships, and development of explainable operational AI models for critical distributed systems.

ACKNOWLEDGMENTS

We acknowledge support by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) and the Open Access Publishing Fund of Anhalt University of Applied Sciences.

REFERENCES

[1] J. Humble and D. Farley, Continuous Delivery. Boston, MA, USA: Addison-Wesley, 2011.

[2] G. Kim, J. Humble, P. Debois, and J. Willis, The DevOps Handbook. Portland, OR, USA: IT Revolution, 2021.

[3] L. Bass, I. Weber, and L. Zhu, DevOps: A Software Architect's Perspective. Boston, MA, USA: Addison-Wesley, 2015.

[4] J. Turnbull, The Docker Book. James Turnbull, 2019.

[5] B. Burns, J. Beda, and K. Hightower, Kubernetes: Up and Running. Sebastopol, CA, USA: O'Reilly Media, 2022.

[6] L. Bodnar, M. Bodnar, K. Shulakova, O. Vasylenko, E. Siemens, R. Tsarov, O. Yavorska, and O. Tyurikova, "Advanced techniques for IaC: Enhancing automation and optimization in cloud-based infrastructure management," Proc. Int. Conf. Applied Innovation in IT, vol. 12, no. 2, pp. 19-25, 2024, doi: 10.25673/118105.2.

[7] D. Sculley et al., "Hidden technical debt in machine learning systems," in Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2015.

[8] M. Shahin et al., "Machine learning operations (MLOps): Overview, definition, and architecture," IEEE Access, vol. 10, pp. 31841-31865, 2022, doi: 10.1109/ACCESS.2022.3157181.

[9] S. Amershi et al., "Software engineering for machine learning: A case study," in Proc. IEEE/ACM 41st Int. Conf. Softw. Eng. (ICSE), 2019, pp. 291-300.

[10] M. Zaharia et al., "Accelerating the machine learning lifecycle with MLflow," IEEE Data Eng. Bull., vol. 41, no. 4, pp. 39-45, 2018.

[11] L. Chen, "Continuous delivery: Huge benefits, but challenges too," IEEE Software, vol. 32, no. 2, pp. 50-54, 2015, doi: 10.1109/MS.2015.27.

[12] C. Huyen, Designing Machine Learning Systems. Sebastopol, CA, USA: O'Reilly Media, 2022.

[13] I. Rangwala et al., "AIOps: Real-world challenges and research innovations," IEEE Internet Computing, vol. 25, no. 3, pp. 81-85, 2021.

[14] G. Deng et al., "AIOps: Real-world operational AI challenges," IEEE Transactions on Services Computing, vol. 15, no. 4, pp. 2186-2199, 2022, doi: 10.1109/TSC.2020.3038824.

[15] A. Rahman et al., "DevSecOps: A multivocal literature review," Information and Software Technology, vol. 121, Art. no. 106200, 2020, doi: 10.1016/j.infsof.2020.106200.

[16] ISO/IEC 5338:2023, AI System Lifecycle Processes. Geneva, Switzerland: International Organization for Standardization, 2023.

[17] NIST, Artificial Intelligence Risk Management Framework (AI RMF 1.0). Gaithersburg, MD, USA: National Institute of Standards and Technology, 2023.

[18] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach. Hoboken, NJ, USA: Pearson, 2021.

[19] L. Bodnar, M. Bodnar, K. Shulakova, O. Vasylenko, R. Tsarov, and E. Siemens, "Practical experience in DevOps implementation," Proc. Int. Conf. Applied Innovation in IT, vol. 12, no. 1, pp. 33-39, 2024, doi: 10.25673/115639.

[20] L. Bodnar, M. Bodnar, K. Shulakova, O. Vasylenko, E. Siemens, and O. Tsyra, "A comprehensive integration of practical strategies in DevOps," Lecture Notes in Networks and Systems. Cham, Switzerland: Springer, 2025, pp. 336-359.

[21] M. Villamizar et al., "Infrastructure cost comparison of container-based cloud deployment," Future Generation Computer Systems, vol. 79, pp. 619-633, 2018, doi: 10.1016/j.future.2017.08.028.

[22] L. Bainbridge, "Ironies of automation," Automatica, vol. 19, no. 6, pp. 775-779, 1983, doi: 10.1016/0005-1098(83)90046-8.

[23] IEC 31010:2019, Risk Management - Risk Assessment Techniques. Geneva, Switzerland: International Electrotechnical Commission, 2019.

A Multi-Layer Continuous-Time Markov Model for Availability Analysis of Kubernetes Systems

Volodymyr Mankivskyi[1], Oleksandr Romanov[1], Mikola Nesterenko[1], Mika Romanova [2] and Anton Marinov[1]

[1]*Institute of Telecommunication Systems, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Beresteiskyi Avenue 37, 03056 Kyiv, Ukraine*

[2]*Faculty of Aeronautics, Electronics and Telecommunications, State University "Kyiv Aviation Institute", 03058 Kyiv, Ukraine*

v.b.mankivskiy@gmail.com, nikolaiy.nesterenko@gmail.com, a_i_romanov@ukr.net, 4639678@stud.kai.edu.ua, marinov.anton99@gmail.com

Keywords: Availability, Markov Chain, CTMC, Kubernetes, Pod, Container, CNI, MTTR, MTTB.

Abstract: This paper proposes a three-tiered analytical availability model for distributed systems running on top of a Kubernetes platform, where container reliability, pod redundancy, and CNI network component reliability are unified under a single Continuous Time Markov Chain model. Unlike typical methodologies based on end-to-end MTTF/MTTR metrics only, this work introduces a model where failure/recovery behaviors can be directly analyzed across multiple tiers and steady-state end-to-end availability between pods can be evaluated directly. Evaluation results show that the effect of the system's recovery time is almost linear when MTTR ≪ MTBF and the availability benefits of pod redundancy increase exponentially but diminish their returns when the number of replicas is greater than three. A numerical evaluation of the proposed model is conducted using representative parameters of Kubernetes environments. The results demonstrate the influence of recovery time, Pod redundancy, and network reliability on end-to-end service availability. While redundancy significantly improves baseline availability, network recovery time may become the limiting factor for meeting strict SLA requirements.

1 INTRODUCTION

Kubernetes is currently one of the most well-known and convenient tools for deploying and orchestrating systems from various components. Deploying a service platform is accomplished by sequentially configuring each service and requires close coordination of internal and external resources. This approach makes enterprise systems inflexible and hinders the introduction of new services and features.

Kubernetes (k8s) is primarily focused on automating the deployment, scaling, and management of containerized applications. For Kubernetes to solve problems using containerized applications, manage containers, and service incoming requests, these containers must be pre-designed, operational, and publicly available.

Current research into the availability of Kubernetes is focused on analyzing the availability of nodes or pods as independent elements. Our advanced work [1] formalized the availability of boundary lines between Pods, but there is a new model that describes the degradation processes on:

1) Container levels;
2) Level's Pod;
3) Border level (CNI + veth + bridge).

Kubernetes is a system where the performance of one component can be compensated by others (Pod replication, update policies, overlay boundaries). Therefore, it is natural to model them based on Continuous Time Markov Chain model with updates (CTMC).

The rest of the paper is organized as follows. Section II shows related work and focus paper. Section III introduces the interaction between pods in Kubernetes and architecture components. In Section IV we present our experiments settings, the metrics and various. The experiments, the results, and the analysis with respect to the research questions are presented in Section V. Conclusions and future work placed on section VI and VII accordingly.

2 AVAILABILITIES IN K8S

2.1 Related works

The paper [2] examined the construction of applications and their performance evaluation based on microservices. This paper concluded that the development of Kubernetes is insufficient for ensuring availability, especially high availability. For example, the configuration of Kubernetes by ordering is likely to lead to significant failures in operations during node failure [3]. The primary focus was on their availability. The paper [4] analyzed initial setup in a private cloud and the default Kubernetes configuration. System recovery tasks were investigated for various architectures and configurations. A series of experiments with Kubernetes were conducted, measuring downtime parameters for different failure scenarios. The goal of both studies was to ensure high pod availability.

2.2 Formulation of the Problem

The goal of our work is to answer the following research questions: "How to measure the level of availability of pods in the network to assess the availability of the entire network built by Kubernetes?" The description of the internal process of providing communication between nodes can be assigned to both the Kubernetes controller or network subsystem, and to the control programs located in the control plane. Consider a Kubernetes cluster consisting of a finite set of Pods:

$$P = \{P_1, P_2, \dots, P_M\}. \quad (2)$$

Each Pod Pi contains finite number of containers:

$$P_i = \{C_{i1}, C_{i2}, \dots, C_{iN_i}\}. \quad (3)$$

Containers are deployed within worker nodes and communicate through virtualized network components provided by the Container Network Interface (CNI). Communication between two Pods P_A and P_Bis established through a network path composed of a finite ordered set of network elements:

$$E_{AB} = \{e_1, e_2, \dots, e_K\}. \quad (4)$$

The system operates in continuous time and evolves stochastically due to random failures, recoveries, and performance degradations. Each system layer is modeled as a stochastic process with a finite state.

Each container C_ij is modeled as a continuous-time stochastic process. Let N_i denote the number of containers in Pod P_i.

Thus, the Pod state space is:

$$S_{pod} = \{s_0, s_1, \dots, s_{N_i}\}. \quad (1)$$

This threshold depends on the Kubernetes configuration (restart policy, readiness/liveness probes, application semantics). Each network element e_k in path E_AB has the state space: Operational, Degraded, Failed.

So, a container is a system with three states, between which it transitions randomly. With this approach, the best mathematical apparatus that describes such a system is a continuous Markov process (CTMC). In this work, this mathematical apparatus acts as the central driver.

2.3 Proposed Approach

In this section, we propose considering the three levels of the Kubernetes model. There is a main container in which the application is placed. In parallel with it, another container is placed, which works simultaneously with the main one and provides 100% resource redundancy. In this approach, resource availability is considered. In our work, we proposed changing this paradigm to three levels of availability. The model consists of three levels:

1) Level L1 - Container Level;
2) L2 - Pod Level;
3) L3 - CNI Network Components.

It is important to note that pods in Kubernetes are created in such a way that they have a short lifespan. This means that when a Pod completes its task or becomes inoperable, it can be destroyed and created (recreated) again. Kubernetes uses the concept of replication to ensure that the appropriate number of pods are always running.

Problem Statement with Given:

1) Kubernetes cluster with Pods P;
2) Container parameters $\{\gamma_c\}$;
3) Network parameters $\{\gamma_n\}$;
4) Threshold parameter k_crit.

Determine A_{AB} for arbitrary Pods P_A and P_B in steady-state conditions. The objectives of work are:

1) Construct a three-layer CTMC-based model of Kubernetes availability;
2) Derive analytical expressions for steady-state probabilities;
3) Compute end-to-end communication availability.

3 THREE-LAYER ARCHITECTURE OF KUBERNETES AVAILABILITY

These three levels are quite difficult to depict on a single classic Kubernetes cluster, but they can be distinguished as shown below on Figure 1.

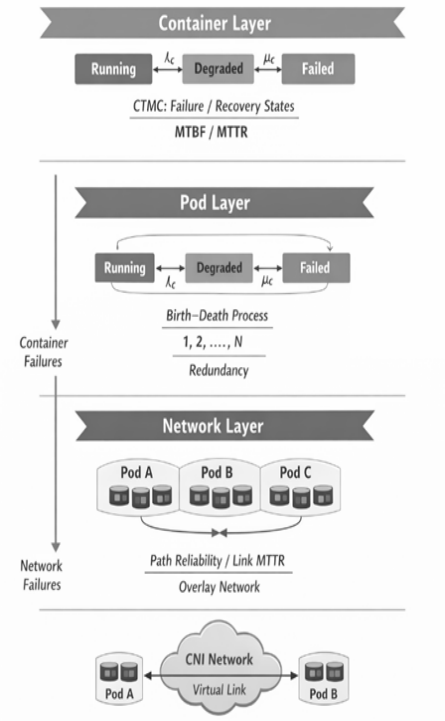


Figure 1: Three-layer arch of k8s availability.

Figure 1 illustrates the proposed three-layer availability model for Kubernetes-based systems. The container layer represents failure and recovery dynamics modeled using Continuous-Time Markov Chains with MTBF/MTTR parameters. The Pod layer captures redundancy effects through a birth-death process describing the number of failed containers. The network layer models end-to-end communication reliability across virtual links and CNI components. The overall service availability is obtained by composing availability metrics across all three layers.

3.1 Level L1 - Container

The container behavior is modeled using a Continuous-Time Markov Chain, as illustrated in Figure 2. Container states describe following states:

$$S_c = \{Run, Degrad, Failed, Recovering\},$$

transient intensities:

1) $\lambda_c = MTBF^{-1}$- failure intensity;
2) $\mu_c = MTTR^{-1}$- recovery intensity;
3) δ_c- intensity of the Running → Degraded transition (network degradations);
4) γ_c- intensity of exit from degradation.

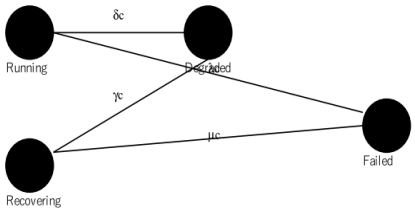


Figure 2: CTMC model container.

3.2 Level L2 - Pod

The basic concept and basic unit in the Kubernetes system is the Pod. A Pod is the basic unit of computing resources in Kubernetes, a container orchestration system. A Pod is a group of one or more containers that together perform a specific task.

A Pod of N containers has states:

$$Spod=\{s0,s1,\ldots,sN,sF\}, \quad (5)$$

where:

1) s_0- all containers are working;
2) s_k- k containers are unavailable;
3) s_F- pod unavailable (RestartPolicy threshold exceeded).

Probabilities of pod states in steady state. Stationary probabilities:

$$\pi_k = \pi_0 \prod_{i=1}^{k} \frac{(N-i+1)\lambda_c}{i\mu_c}, \quad (6)$$

normalization:

$$\pi_0 = \left(1 + \sum_{k=1}^{N} \prod_{i=1}^{k} \frac{(N-i+1)\lambda_c}{i\mu_c}\right)^{-1}. \quad (7)$$

This is the standard formula for birth-death processes. All containers in a pod are located on a single worker node (the node that performs the work) and share a common network space, allowing them to communicate with each other and with other pods within the cluster. Considering this, moving to a tiered model allows us to generalize the availability:

$$k_{availability} = \frac{MTBF}{MTBF + MTTR}. \quad (8)$$

Normalization to the type of multi-container:

$$A_{pod} = 1 - \sum_{k=k_{crit}}^{N} \pi_k \ , \quad (9)$$

where: k_{crit}- the minimum number of container failures after which the pod is "dead". In this case, all the listed steps and elements of Kubernetes are aimed at creating the minimum unit of deployment - pod.

Figure 3 shows all the listed elements that take part in creating a pod. At the same time, the elements that take part in forming the network stack between pods are not marked.

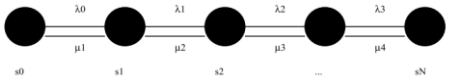


Figure 3: CTMC model of Pod availability with N containers.

As mentioned above, a pod is the fundamental logical unit of Kubernetes. A pod is an abstraction of a virtual machine within a Kubernetes cluster: it has its own private IP address, hostname, and shared disk data. A pod is a deployment unit, and within it, one or more containers run, all related by a common purpose and representing a logical application (consisting of one or more processes/containers).

3.3 Level L3 - CNI

If there is only one container in the pod and it fails, then states are entered for each element: veth, bridge, CNI-overlay:

$$S_{net} = \{Operational, Degraded, Failed\}. \quad (10)$$

In classical reliability theory, transition rates are not loads, but frequencies of events that occur randomly. In our model Figure 4 describes (like container) life circle:

1) λ_n - failure rate;
2) μ_n - repair rate;
3) δ_n- intensity if degraded;
4) γ_n - intensity of recovery.

Components availability:

$$A_{net} = 1 - \pi_{failed}. \quad (11)$$

The failure rate of a component - failure rate (λn) is the mathematical equivalent of "the probability that an element will fail per unit of time". For a network component see on Table 1.

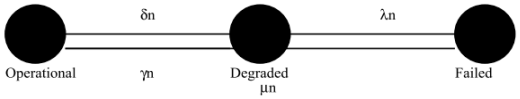


Figure 4: CTMC model of CNI network component.

Table 1: Network elements failure list.

Component	Degradation Cause	Estimate
veth	wrong mtu, kernel bug	dmesg / dropping
bridge	overloaded forwarding table	drop/sec
OVS	kernel datapath crash	ovs-vswitchd
CNI-plugin	deadlock / race condition	kubelet, calico/cilium agent

The intensity of entering the Degraded state (δn) is related to the load.

This is the moment when the component:

1) Does not completely fail;
2) Perform worse;
3) Does not meet the SLO (latency, jitter, packet loss).

These degradation modes and their underlying causes are summarized in Table 2.

Table 2: Network degradation reason list.

Component	Degradation Cause
veth	qdisc drop > threshold
bridge	MAC recycle/drop
OVS	dpif overload
CNI	vxlan latency > 10 ms

The probability that a network component “sags” under high load (δn).

The rate of recovery from degradation (γn), when the load drops, the system returns to normal operation.

A good example is OVS or Calico [5]: during peak traffic, delay increases → degraded, and when traffic drops → operational

Kubernetes can distribute traffic between different containers in a pod using load balancers, which allows you to evenly distribute traffic between available containers in a pod. This ensures high availability of the Kubernetes network, even if some containers in a pod stop or fail.

4 MATHEMATICAL MODEL

Contemporary Kubernetes-based infrastructure operates in dynamic, distributed, and unpredictable environments. For instance, the containers, Pods, and network components experience stochastic failures, degradation, and healing mechanisms. Hence, deterministic modeling approaches cannot be used to properly model the time-related behavior of these infrastructures.

All key elements of Kubernetes exhibit probabilistic behavior:

1) Container failures may occur due to software bugs, resource exhaustion, kernel-level issues, or external dependencies;
2) Pod re-creation and rescheduling depend on control-plane reaction time and system load;
3) Network components (CNI plugins, overlay networks, virtual interfaces) may experience packet loss, latency spikes, or temporary degradation.

Empirical studies of large-scale distributed systems have shown that time to failure, as well as time to recover, often follows an exponential distribution or a hyper-exponential distribution, especially when the environment has independent and memoryless failures [6]. In such a case, the probability of a failure occurring in the next infinitesimal period is independent of the length of the preceding states, which makes the use of Continuous Time Markov Chain feasible.

4.1 Why Does Apply CTMC?

A CTMC is defined by:

1) A finite set of states;
2) Transition rates between states;
3) Memoryless (Markov) property.

Kubernetes components naturally satisfy conditions.

Finite state space:

1) Container: Running, Degraded, Failed, Recovering;
2) Pod: {0,1…N failed containers};
3) Network component: {Operational, Degraded, Failed}.

Random state transitions:

1) Failure events occur randomly over time;
2) Recovery is triggered automatically with measurable average duration.

Thus, CTMC provides a mathematically rigorous and tractable framework for modeling availability in Kubernetes.

4.2 Justification of the Birth-Death Model for Pods

A Pod consisting of Ncontainers can be described as a stochastic process where:

1) Failures increase the number of unavailable containers;
2) Recovery decreases that number.

This exactly corresponds to a birth-death process:

$$\lambda_k = (N - k)\lambda_c, \tag{12}$$

represents the rate of new failures (birth of failed containers), with:

$$\mu_k = k\mu_c, \tag{13}$$

represents the recovery rate (death of failed containers).

The birth-death process is appropriate because:

1) Containers fail independently;
2) Each failed container recovers independently;
3) Transition probabilities depend only on the current number of failed containers.

This model enables closed-form stationary solutions and analytical availability computation.

In addition to binary states (working/failed), real Kubernetes systems experience intermediate degraded states. The parameters:

1) δ- transition rate from operational to degraded;
2) γ- recovery rate from degraded to operational.

Unlike failure rates λ, which describe hard crashes, degradation rates represent threshold-based performance violations such as:

1) Latency exceeding SLO limits;
2) Packet loss exceeding the predefined percentage;
3) CPU throttling or resource contention;
4) Overlay encapsulation delays in CNI networks.

Formally, degradation can be defined as:

$$\delta = f(P_{loss}), \tag{14}$$

where:

1) $\lambda_{traffic}$- incoming traffic rate;
2) C- capacity;
3) D- delay;
4) P_{loss}- packet loss probability.

Thus, degradation rates capture load-induced instability while remaining compatible with Markovian modeling.

4.3 Suitability of CTMC for Availability Composition

4.3.1 Suitability of CTMC for Availability Composition

Compose multi-layer availability models in the proposed three-layer Kubernetes model:

1) L1 - container-level CTMC;
2) L2 - Pod birth-death CTMC;
3) L3 - network component CTMC.

In the proposed model, failures of containers, Pods, and network components are assumed to be statistically independent. This assumption simplifies the analytical formulation and allows the use of Continuous-Time Markov Chains to derive steady-state availability metrics.

The overall service availability is computed as a composition of independent stochastic subsystems, making CTMC a natural and consistent modeling tool. This generalizes our previous Shannon-Moore graph transformation method in previous work [1].

Markov transition matrix:

$$Q_c = \begin{bmatrix} -(\lambda_c + \delta_c) & \delta_c & \lambda_c \\ \gamma_c & -(\gamma_c + \lambda_c) & \lambda_c \\ \mu_c & 0 & -\mu_c \end{bmatrix}, \quad (15)$$

where:

1) λc (lambda) container drop rate;
2) μc (mu) recovery rate (inverse of MTTR);
3) δc (delta) frequency of degradation (network problems, resource starvation);
4) γc (gamma) degradation recovery rate.

The first row of this matrix:

$$Running \rightarrow \left\{\frac{Degraded}{Failed}, \frac{\delta_c}{\lambda_c}\right\}. \quad (16)$$

The sum on the left: - (λc + δc) shows that the container leaves the Running state.

Stationary solution:

$$\pi_c = \pi_c Q_c = 0, \sum_i^k \pi_i = 1. \quad (17)$$

After solving the system, we learn:

1) What fraction of the time the container is working normally;
2) How long does it take to degrade?
3) What is the chance that it has crashed?

4.3.2 Pod Model CTMC

It is considered "alive", depending on the parameters of the system itself in k8s:

1) Restart Policy: Pod is considered available if at least one container is alive;
2) Strict Liveness Probe: pod may be unavailable with just one container hanging.

State probabilities:

$$\pi_k = \pi_0 \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i}. \quad (18)$$

The availability of component *i* is calculated using the steady-state probability that the component is in operational state. In the proposed model, the component behavior is represented by a two-state Continuous-Time Markov Chain consisting of an operational state and a failure state. The transition rate from the operational state to the failed state is denoted by λi, while μi represents the recovery rate. Under these assumptions, the steady state availability of the component is given by:

$$A_i = 1 - \frac{\lambda_i}{\lambda_i + \mu_i}, \quad (19)$$

which corresponds to the stationary probability of the operational state in the Markov reliability model. The intensity data describes the following physical meaning:

1) If there are 5 containers available λc = 0.001;
2) Then the total fall rate = 5×0.001 = 0.005.

More containers → more chances that one will fall.

5 CASE STUDY

The objective of this case study is to compute the steady-state availability of communication between two Pods in a Kubernetes cluster using the proposed three-layer CTMC-based model. Unlike classical network availability analysis, this study incorporates:

1) Container-level failures;
2) Pod-level redundancy;
3) Network component reliability;
4) Performance-related availability constraints.

5.1 Task Definition

Determine the steady-state availability of communication between two Pods (P_A) and (P_B) deployed in a Kubernetes cluster.

Communication is considered available if:

1) Both Pods are operational;
2) All network components along the virtual path are operational;
3) The end-to-end latency does not exceed a predefined threshold.

The availability of a Kubernetes network is defined as the probability that all required components operate within acceptable performance limits over continuous time [3].

The following assumptions are introduced for analytical tractability:

1) All components follow Continuous-Time Markov processes;
2) Failures and recoveries are statistically independent;
3) A Pod becomes unavailable when the number of failed containers exceeds the threshold (k_crit);
4) Network components are modeled as three-state systems (Operational, Degraded, Failed);
5) Latency violation probability is incorporated into degradation state transitions.

A key modeling simplification is that the final virtual branch connecting a container is treated as an indivisible reliability unit.

Thus, container and branch availability are composed multiplicatively. Let:

1) D_{th}be the latency threshold;
2) p be the probability that request latency exceeds $D_{th.}$.

In this model:

$$p = Pr\ (D > D_{th}). \tag{20}$$

This probability is obtained from network capacity calculations or monitoring data from production clusters.

The degradation rate δn of network components reflects this threshold violation probability. Thus, availability accounts not only for hard failures but also for Quality-of-Service violations.

The following input parameters are used in the numerical evaluation:

1) t- continuous operational time;
2) $MTBFpod$mean between Pod failures;
3) $MTTRpod$- mean time to repair Pod;
4) $MTBFnet$Network failures;
5) $MTTRnet$- mean time to repair network component;
6) Rm- reliability of individual network branch;
7) p- probability of exceeding latency threshold;
8) k_{crit}- maximum tolerable number of failed containers.

The resulting availability metric represents the probability that:

1) Both Pods remain operational;
2) All intermediate network elements function correctly;
3) Latency remains below the predefined threshold.

This formulation integrates reliability theory and Quality-of-Service constraints within a unified CTMC-based framework.

5.2 Container Availability with MTTR

First, let us attract your attention to availability of containers. Figure 5 illustrates container availability as a function of recovery time (MTTR) for a fixed mean time between failures $MTBF = 500$ hours.

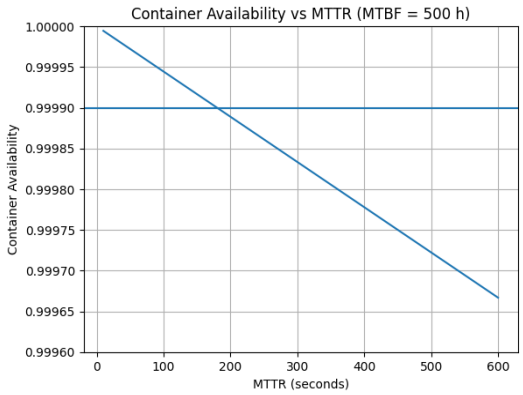


Figure 5: Container availability versus MTTR.

The availability is calculated analytically as:

$$A_c = \frac{MTBF}{MTBF+MTTR}. \tag{21}$$

The results demonstrate a monotonic decrease in availability as MTTR increases from 10 seconds to six hundred seconds. Although the absolute numerical change appears small (from approximately 0.999994 to 0.999667), its operational impact is significant.

Specifically:

1) At $MTTR = 10$seconds, container availability corresponds to approximately 3 minutes of downtime per year;
2) At $MTTR = 60$seconds, downtime increases to approximately 18 minutes per year;
3) At $MTTR = 600$seconds, annual downtime exceeds 175 minutes (nearly 3 hours).

This result highlights a key observation: even when container reliability (MTBF) is relatively high, recovery time has a disproportionate impact on long-term availability.

From an engineering perspective, the graph confirms that rapid self-healing mechanisms (fast restart, efficient liveness probes, minimal image initialization time) are critical for achieving strict service-level objectives such as 99.99% availability.

Furthermore, the nearly linear shape of the curve in the examined region is explained by the condition:

$$MTTR \ll MTBF.$$

Under this condition, availability can be approximated as:

$$A_c \approx 1 - \frac{MTTR}{MTBF}. \quad (22)$$

This explains why availability degrades almost linearly with increasing recovery time within the operational range.

In practical terms, even small increases in MTTR produce measurable changes in availability.

More importantly, the second derivative is positive, indicating convexity of the function:

$$\frac{d^2 A_c}{dx^2} = \frac{2MTBF}{(x)^3}. \quad (23)$$

Therefore, the system becomes progressively more sensitive to MTTR growth as recovery time increases. From an SLA perspective, the most relevant metric is annual downtime:

$$Downtime = (1 - A_c) \cdot 8760\ hours. \quad (24)$$

For a 99.99% SLA, the maximum allowed downtime per year is:

$$0.0001 \times 8760 \approx 52.6\ minutes.$$

This demonstrates that even when failure frequency (MTBF) remains constant, recovery performance directly determines SLA compliance [7], [8].

In practical Kubernetes environments, this emphasizes the importance of:

1) Fast container restart mechanisms;
2) Optimized readiness/liveness probes;
3) Reduced images pull and initialization times;
4) Efficient orchestration response.

Therefore, MTTR optimization often yields greater SLA impact than further improvements in intrinsic container reliability.

5.3 Pod Availability with N

Figure 6 presents Pod availability as a function of the number of containers N for three recovery times: 30, 60, and 120 seconds, assuming $MTBF = 500$hours.

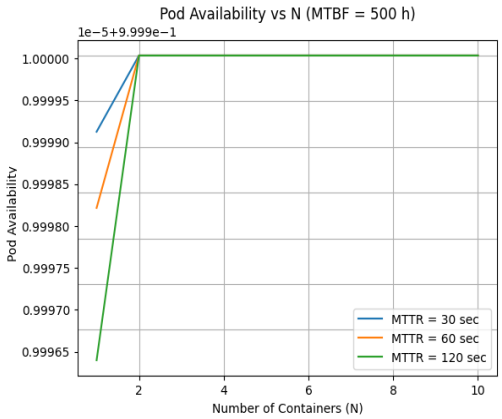


Figure 6: Pod availability as a function of container count.

The graph demonstrates three key effects:

1) Rapid improvement from N = 1 to N = 2;
2) The most significant availability gain occurs when moving from a single-container Pod to a two-container Pod;
3) Diminishing returns for N ≥ 3;
4) Beyond three containers, availability improvement becomes marginal. The curve approaches saturation because the probability of all containers failing simultaneously becomes extremely small;
5) Impact of MTTR.

Higher MTTR shifts the curves downward. Although redundancy mitigates failures, slow recovery reduces the overall benefit of scaling.

The Pod is considered unavailable only when all Ncontainers have failed (i.e., $k_{crit} = N$). The steady-state probability of complete failure is obtained from the birth-death process:

$$\pi_N = \frac{\prod_{k=1}^{N} \frac{(N-k+1)\lambda_c}{k\mu_c}}{\sum_{i=0}^{N} \prod_{k=1}^{i} \frac{(N-k+1)\lambda_c}{k\mu_c}}. \quad (25)$$

Pod availability is defined as:

$$A_{pod} = 1 - \pi_N \quad (26)$$

Let:

$$\rho = \frac{\lambda_c}{\mu_c} = \frac{MTTR}{MTBF} \quad (27)$$

When $\rho \ll 1$(which holds in our scenario), the steady-state probability of complete failure is approximately:

$$\pi_N \approx \rho^N. \quad (28)$$

Thus:

$$A_{pod} \approx 1 - \rho^N. \quad (29)$$

This approximation explains the exponential improvement in availability with increasing N.

For example, if:

$$\rho = \frac{60\ sec}{500\ h} \approx 3.3 \times 10^{-5}.$$

then:

$$\rho^2 \approx 10^{-9} \rho^3 \approx 10^{-14}.$$

Therefore, increasing N from 1 to 2 dramatically reduces failure probability, while further increases produce diminishing returns.

The marginal availability can be approximated:

$$\Delta A_{pod}(N) \approx \rho^{N-1}(1-\rho). \quad (30)$$

This shows that:

1) Marginal benefit decreases exponentially N;
2) Practical saturation occurs for N ≥ 3 under realistic Kubernetes parameters.

For 99.99% SLA (max 52.6 min/year downtime):

1) N = 1 with MTTR = 120 sec may violate SLA;
2) N = 2 satisfies SLA comfortably;
3) N ≥ 3 significantly exceeds SLA.

Adding a second replica is often sufficient to meet strict SLA targets, while further scaling should be justified by load considerations rather than availability alone.

The analysis demonstrates:

1) Horizontal redundancy is highly effective in Kubernetes;
2) Recovery time directly affects redundancy efficiency;
3) Over-scaling Pods for availability alone may be inefficient beyond N = 3;
4) Availability optimization must balance between MTTR reduction and Replica count.

While container availability improves linearly with MTTR reduction, Pod availability improves exponentially with redundancy. However, this exponential gain rapidly saturates.

5.4 End-to-End Availability with Network MTTR

Figure 7 presents the steady-state end-to-end availability between two Pods as a function of network recovery time ($MTTR_{net}$). The number of containers per Pod varies as $N = 1,2,3,5$. Container MTTR is fixed at 60 sec and $MTBF = 500$hours.

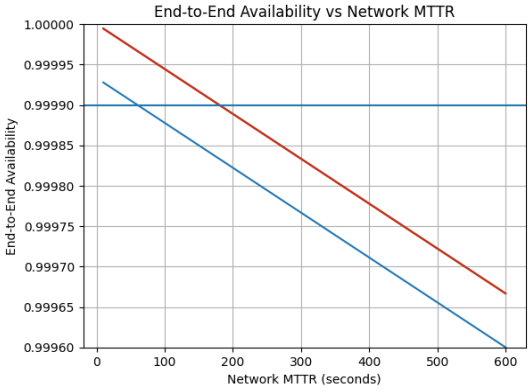


Figure 7: End-to-end availability versus network MTTR.

All curves exhibit identical slopes with respect to $MTTR_{net}$. This occurs because network availability is multiplicatively composed with Pod availability:

Thus, changing Naffects only the vertical scaling factor A_{pod}^2, while the dependence on $MTTR_{net}$ remains structurally identical.

Increasing the number of containers significantly increases baseline availability (vertical shift upward). However, the degradation caused by network MTTR remains proportional across all configurations [7], [9].

Pod-level redundancy cannot eliminate the impact of network recovery delays.

In other words, application-level replication does not compensate for CNI unreliability.

For a 99.99% SLA (52.6 minutes per year downtime), the graph shows:

1) For $N = 1$, SLA is violated at relatively low network MTTR.
2) For $N = 3$, SLA margin improves.
3) For $N = 5$, baseline availability increases further, but network sensitivity remains.

Thus, meeting strict SLA requirements requires not only Pod replication but also reliable network recovery mechanisms [10]. Degradation in end-to-end availability is approximately linear in the operational region. Even when Pod availability is extremely high ($N \geq 3$), network components directly determine whether SLA thresholds are crossed.

This demonstrates that end-to-end availability in Kubernetes is fundamentally constrained by the weakest reliability layer [11].
The findings provide several valuable practical insights:

1) Investing in faster container restart alone is insufficient;
2) CNI plugins and overlay networks must have optimized recovery procedures;
3) design must consider all layers simultaneously.

The end-to-end availability in Kubernetes clusters is limited in their reliability, despite the redundancy in the application layer.

Although the redundancy in Pods results in an exponential increase in availability, the linear multiplicative degradation in network reliability cannot be compensated.

6 CONCLUSIONS

This paper presented a three-layer convenience model for Kubernetes-based systems using CTMC. Unlike traditional network reliability models, the proposed approach integrates container-level failures, Pod-level redundancy, and network-layer reliability within analytical framework.

The main contributions of this work can be summarized as follows:

1) A formal problem formulation of end-to-end communication availability in Kubernetes;
2) A CTMC-based container reliability model incorporating failure and recovery;
3) A birth-death process model describing Pod-level redundancy;
4) A network-layer availability model accounting for virtualized CNI components;
5) An analytical composition framework for computing end-to-end availability;
6) A numerical evaluation demonstrates the sensitivity of availability to recovery time;
7) Several interesting results followed from the study. Firstly, the effect of the average time to repair the containers on the steady-state availability is linear, provided that the average time to repair the containers is significantly less than the average time between failures. Secondly, the effect of the pod-level redundancy on the steady-state availability is exponential, but the returns diminish after three replicas per pod, with the initial replica offering the largest compatibility of the increased availability. Lastly, the effect of the network layer time to repair on the end-to-end availability is multiplicative and cannot be fully compensated for through the application of redundancy mechanisms; this highlights the importance of the reliability of the container network interface (CNI) and network self-healing to achieve service-level agreement targets of 99.99% or higher;
8) The study shows that to optimize Kubernetes cluster availability. It is essential to consider network recovery performance-not just container reliability or scaling.

From an engineering perspective, the results emphasize:

1) Minimization of MTTR through rapid restart and orchestration mechanisms;
2) Careful selection and monitoring of CNI;
3) Balanced design between redundancy and operational efficiency.

7 FUTURE WORK

The proposed three-layer availability framework provides assorted opportunities for further research.

In the present study, degradation states were introduced conceptually but not fully incorporated into the numerical evaluation. Future work should explicitly integrate degraded states into the CTMC structure, allowing the system to operate in intermediate performance regimes rather than strictly binary operational/failed modes:

1) Quantitative modeling of latency violations;
2) Integration of packet loss probabilities;
3) Representation of partial service availability;
4) Analysis of QoS degradation dynamics.

Such a model would more accurately reflect real Kubernetes environments where performance deterioration often precedes complete failure.

The current model assumes statistical independence between containers, Pods, and network components. While this assumption ensures analytical tractability, real-world distributed systems frequently exhibit correlated failures, including:

1) Node-level crashes affecting multiple Pods,
2) Shared resource exhaustion.
3) Network partition events.

A limitation of the current model is the assumption of statistical independence between component failures. In practical Kubernetes deployments, correlated failures may occur due to

shared infrastructure, such as node crashes, network partitioning, or storage subsystem failures. Such events may simultaneously impact multiple containers or Pods and therefore reduce the overall system availability beyond the predictions of the independent failure model.

Future extensions should incorporate correlated failure mechanisms, potentially using:

1) Multi-dimensional CTMC;
2) Copula-based reliability modeling;
3) Common-cause failure analysis.

This would significantly enhance realism and enable evaluation of resilience under large-scale fault scenarios. The current study focuses on steady-state availability. However, transient behavior following failure events is often critical in cloud-native systems.

Future research may address:

1) Time-dependent availability;
2) Recovery trajectory analysis;
3) Short-term SLA violation probability;
4) Burst failure modeling.

This approach also facilitates the modeling of incident scenarios and recoveries beyond the equilibrium assumption. Overall, the above-proposed framework outlines a mathematically sound basis for multi-layer availability modeling within a Kubernetes ecosystem.

Although this research work is focused on the steady-state analytical evaluation under the assumption of independence, the outlined research directions aim at taking this model a step further toward realism, verifiability, and applicability. With the inclusion of degradation modeling, correlated failures, control plane reliability, and optimization, this framework is bound to evolve into a comprehensive reliability engineering practice for cloud-native systems under severe service level agreement pressures.

REFERENCES

[1] O. Romanov, V. Mankivskyi, L. Globa, M. Skulysh and A. Romanov, “Enhancing resource availability: Indicators and strategies for optimizing the Kubernetes network,” in Information and Communication Technologies and Sustainable Development (ICT&SD 2022), A. Dovgyi et al., Eds., Lecture Notes in Networks and Systems, vol. 809. Cham, Switzerland: Springer, 2023, , doi: 10.1007/978-3-031-46880-3_2.

[2] The Kubernetes Authors, “Kubernetes API server - Security,” Kubernetes Documentation, 2020, [Online]. Available: https://kubernetes.io/docs/reference/access-authn-authz/. Accessed: May 16, 2020.

[3] L. A. Vayghan, M. A. Saied, M. Toeroe and F. Khendek, “Kubernetes as an availability manager for microservice applications,” arXiv preprint arXiv:1901.04946, 2019, doi: 10.48550/arXiv.1901.04946.

[4] The Kubernetes Authors, “Kubernetes official website,” Kubernetes Documentation, 2020, [Online]. Available: https://kubernetes.io/. Accessed: May 16, 2020.

[5] O. Romanov, M. Nesterenko, L. Veres, R. Kamarali and I. Saychenko, “Methods for calculating the performance indicators IP multimedia subsystem (IMS),” in Lecture Notes in Networks and Systems, vol. 152. Cham, Switzerland: Springer, 2021, pp. 229-256, doi: 10.1007/978-3-030-58359-0_13.

[6] K. Pentikousis, ONOS: Security and Performance Analysis, Technical Report No. 1, Sept. 19, 2017.

[7] C. O’Connor, T. Vachuska and B. Davie, Software Defined Networks: A Systems Approach. Cambridge, MA, USA: Morgan Kaufmann, 2021.

[8] K. Phemius, M. Bouet and J. Leguay, “Distributed multi domain SDN controllers,” Thales Communications & Security, 2013, pp. 198-209, doi: 10.1109/NOMS.2014.6838330.

[9] J. Lam, S. Lee and O. Yustus, “Securing SDN southbound and data plane communication with IBC,” Mobile Information Systems, vol. 2016, art. ID 2016, 2016, doi: 10.1155/2016/1708970.

[10] K. Phemius, M. Bouet and J. Leguay, “ONOS intent monitor and reroute service: Enabling plug and play routing logic,” Thales Communications & Security, 2013.

[11] A. Poniszewska Marańda and E. Czechowska, “Kubernetes cluster for automating software production environment,” Sensors, vol. 21, no. 5, art. 1910, 2021, doi: 10.3390/s21051910.

Method for Structural Optimization of Telemedicine Networks Based on Traffic Flow Distribution

Roman Tsarov[1], Olha Yavorska[1], Vladyslav Kumysh[1], Kateryna Shulakova[1,2], Eduard Siemens[2], Liliia Bodnar[3] and Liudmyla Bubentsova[1]

[1]*State University of Intelligent Technologies and Telecommunications, Kuznechna Str. 1, 65023 Odesa, Ukraine*
[2]*Anhalt University of Applied Sciences, Bernburger Str. 57, 06366 Köthen, Germany*
[3]*South Ukrainian National Pedagogical University, Staroportofrankyvska Str. 26, 65023 Odesa, Ukraine*
rcarev@gmail.com, yavorskayao7@gmail.com, vlad.kumish@gmail.com, katejojo29@gmail.com, bodnarl79@pdpu.edu.ua, lyudmilabubentsova1@gmail.com

Keywords: Telemedicine, Telemedicine Network, Software-Defined Network, SDN, Modified Iterative Algorithm, MIA, Structural Network Optimization, Traffic Flow Optimization.

Abstract: The growing adoption of telemedicine creates urgent demand for reliable, high-performance network infrastructure capable of supporting heterogeneous real-time traffic - including HD video consultations, DICOM medical imaging, IoT telemonitoring streams, and electronic health records. This paper addresses the structural optimization of telemedicine networks deployed over Software-Defined Networking (SDN) infrastructure, where conventional maximum-flow algorithms prove inadequate due to high computational complexity and incompatibility with dynamic, free-oriented topologies. The aim of this work is to create the structural optimisation method using the Modified Iterative Algorithm (MIA), originally developed for flow distribution on free-oriented graphs, for the specific requirements of SDN-based telemedicine networks. The proposed method incorporates the following key enhancements: 1) optimized recursive path traversal prioritizing shortest ST-paths via a hop-growth strategy; 2) dynamic residual-capacity updates to prevent real-time channel overload; 3) automated detection and resolution of edge-intersection conflicts between concurrent paths. The verification was done in a Mininet-emulated SDN environment under two operational scenarios: normal telemedicine center load and stress test simulating urgent DICOM transfer. Theoretically, the study extends flow optimization methodology to free-oriented SDN graphs, eliminating the need for artificial edge duplication required by classical directed-graph approaches. Practically, the adapted MIA enables dynamic, controller-independent bandwidth allocation in TMN, ensuring QoS compliance for time-critical medical applications and opening pathways for integration with OpenFlow-compatible SDN controllers such as Ryu. Comparative evaluation against Ford-Fulkerson, ACO, and GA-based methods shows 7.8% throughput improvement over ACO with 58% lower control latency, while telemedicine-specific QoS assessment identifies both the method's suitability for video/telemonitoring/DICOM and its current limitation for haptic telesurgery (0.12% packet loss during redistribution vs. required <0.01%).

1 INTRODUCTION

Telemedicine is increasingly recognized as a critical component of modern healthcare delivery, enabling remote diagnostics, real-time video consultations, telemonitoring, and the exchange of electronic health records across geographically distributed medical facilities. Telemedicine services are implemented through a telemedicine network (TMN), which can be classified as a distributed information and communication network. In most cases, a TMN is built on a hierarchical principle and includes telemedicine centers and stations at the network core level, while telemedicine service consumers and their access terminals form the access level.

The telemedicine traffic is highly heterogeneous: it combines real-time video consultations requiring stable bandwidth of 2-10 Mbps per session, transmission of large-volume medical imaging files in DICOM format (CT and MRI scans reaching several gigabytes), telemonitoring data streams from wearable IoT sensors, and electronic health record (EHR) exchanges. TMNs are increasingly deployed

over transport infrastructure based on Software-Defined Networking (SDN), including hybrid optical-wireless segments connecting rural health posts, mobile diagnostic units, and urban hospital hubs. Sudden load spikes (e.g. simultaneous activation of multiple video consultations during an epidemic peak or mass-casualty event) place extreme demands on dynamic bandwidth allocation.

Modern SDN networks are characterized by dynamic reconfiguration of optical and wireless channels, which renders traditional maximum flow algorithms inefficient. Their high computational complexity prevents them from meeting real-time requirements in dynamic networks. The traditional model of bipolar directed graphs does not satisfy the requirements of free-oriented SDN networks, where channels can be reconfigured in both directions. Furthermore, the lack of simultaneous optimization of maximum flow, distribution across alternative paths, and resolution of edge intersection problems in classical algorithms leads to inefficient use of network bandwidth. The key limitation is the absence of integration of such algorithms with SDN controllers and programmable protocols, making deployment on commercial switches impossible, while complex multi-commodity flow models are unsuitable for dynamic 5G/6G networks with latency requirements below 1 ms.

The optimization problem therefore consists in developing an SDN-compatible algorithm that simultaneously computes the maximum flow between poles S and T and optimizes its distribution across paths with minimum transit, while ensuring scalability and compatibility with next-generation networks. This paper proposes an adaptation of the Modified Iterative Algorithm (MIA), originally introduced for flow distribution on free-oriented graphs, to address the structural optimization of TMN over SDN infrastructure.

2 STATE OF THE ART AND PAPER GOAL

The application of Software-Defined Networking (SDN) and metaheuristic optimisation to TMN has received substantial research attention, driven by the inability of conventional IP-based protocols to guarantee the stringent QoS requirements of real-time medical applications. Parsaei et al. [1] modelled telesurgery path selection as a Constrained Shortest Path NP-complete problem, solving it via Ant Colony Optimisation (ACO) deployed as an SDN controller module. Results confirmed satisfaction of the sub-300 ms round-trip delay threshold for safe robotic surgery, demonstrating that SDN programmability enables near-real-time bio-inspired routing optimisation. Bokolo Anthony Jnr. et al. [2] proposed a multi-tier SDN architecture with a hierarchically distributed control plane that dynamically steers medical traffic flows and balances load under pandemic-induced congestion, providing a practical architectural roadmap for telemedicine QoS assurance. For heterogeneous medical IoT, Saha et al. [3] proposed an SDN-controller-triggered dynamic decision mechanism that classifies flows by urgency and reconfigures routing on-the-fly. Misra et al. [4] introduced phSDN, allocating analytic modules to IoT devices on demand per resource type; phSDN reduced network response time by at least 43.87% versus fog-based and 78.4% versus cloud-based approaches.

An SDN/NFV hospital architecture [5] enables centralised traffic management alongside software-instantiated network functions, removing hardware rigidity. Sembati et al. [6] integrated SDN with Named Data Networking (NDN), achieving latency reductions up to 50%, bandwidth improvement of ~55%, and packet loss as low as 1% against a conventional IP baseline. Rahman et al. [7] integrated SDN with Hyperledger Fabric blockchain for IoMT-based monitoring in 5G, using a Patient-Centric Agent in the control plane to manage data flows and ensure record immutability. Halman and Alenazi [8] proposed the Threshold-Based DDoS Classifier (TBDC), which under attack recovered mean throughput by +190.3%, reduced delay by 95%, and packet loss by 99.7%.

Tsarov et al. [9] formalised TMN topology design as an NP-complete bi-criteria optimisation problem - minimising signal delay and deployment cost - and proposed a Genetic Algorithm (GA)-based synthesis method. On a network of 5 stations and 40 terminals (S=50, G=250), the GA converged to near-optimal global solutions, avoiding local extrema that trap classical heuristics. This is directly relevant to the present study, as efficient SDN traffic management presupposes an optimised underlying topology.

Collectively, these works confirm that SDN-based traffic management together with structural topology optimisation are essential and complementary pillars of next-generation telemedicine engineering. Though the key problem is the lack of integration of algorithms with SDN controllers and programmable protocols, which makes deployment on commercial switches impossible. And the complex multi-commodity flow models aren’t suitable for dynamic 5G/6G networks with latency requirements of less than 1 ms.

The same problem has been solved for SDN by Tikhonov and Yavorska [10] introducing the Modified Iterative Algorithm (MIA). MIA is based on the principles of minimum transit and maximum

flow on free-oriented graphs. It provides simultaneous calculation of the maximum possible flow between poles S and T and optimal distribution of this flow between alternative ST paths with a minimum number of intermediate vertices, implementing the principle of minimum transit in combination with the principle of maximum flow.

So, the goal of this work is to develop the method for structural optimization of TMN over SDN, which uses the MIA algorithm in order to ensure iterative search for the shortest ST-paths, dynamic updating of residual capacities, real-time prevention of channel overloading, resolution of edge intersection conflicts, and comprehensive verification of the adapted MIA's performance in telemedicine networks.

3 PROPOSED METHOD

The Modified Iterative Algorithm (MIA), implemented in Python and adapted to solve the problem of flow optimization in TMN based over SDN. The network model is represented as a normalized free-oriented graph, where vertex V_1 is always associated with source "S", and the vertex with the maximum index V_n - with sink "T". The algorithm consists of three functional blocks (Fig. 1).

For the network parameter input, a static text file format was chosen exclusively for testing purposes and convenience. The main variables get initialized: vertex names, a list of edges, and their initial weights. Basic structures get created, such as *fmax = 0* for accumulating the total flow, for storing node identifiers, and a working copy of the residual capacity matrix C_{res} (*main_connections*).

Next, the critically important auxiliary edge search function *find_connection* is defined, implemented as a lambda expression: *next((idx for idx, i in enumerate(main_connections) if set(i.split()[:2]) == set([a, b])), False)*. This allows instant retrieval of the index of the connection between any two nodes *a* and *b* in a free-oriented graph.

The recursive function *recursive* uses depth-first search to find all possible paths p_k of a given length *h* (number of hops). It works in conjunction with a new neighbor filtering function *get_neighbors(node)*, which analyzes the current state of the residual network and returns only those nodes whose connection has a residual capacity greater than zero.

For each found path $p_k = (e_1, e_2, \ldots, e_h)$, its capacity is computed according to the "bottleneck" rule (1):

$$\delta(p_k) = \min\{c(e)\}, \quad (1)$$

where c(e) is the current weight of the edge in the residual network *main_connections*.

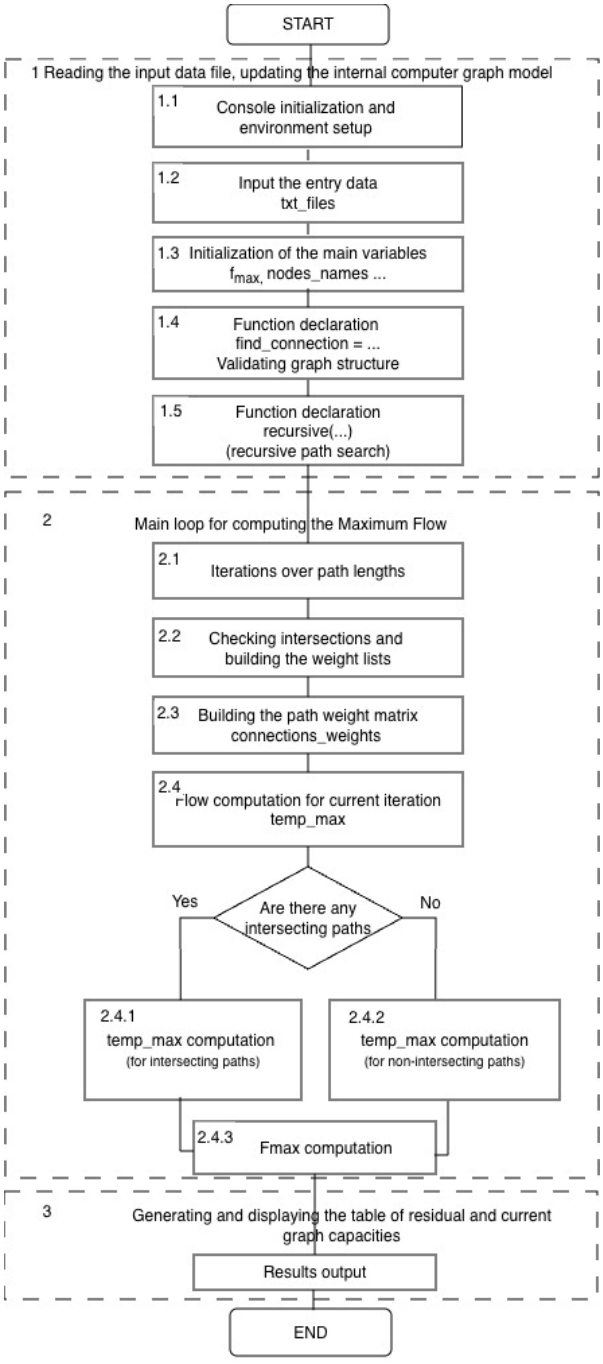


Figure 1: Modified incremental algorithm.

The main loop for computing the maximum flow implements an iterative traversal by path length h = 1, 2, …, N-1. At each step, the program builds a set of paths P_h (all possible routes between source and sink), then checks for intersections (collisions). To do this, a list of all edges in use is formed, and using set logic

set(), it is determined whether paths share common communication channels. If the paths found at the current iteration are edge-disjoint (i.e., share no common edge and do not intersect), they are considered independent. In this case, the total incremental flow (2) at this iteration is computed as a simple arithmetic sum of the flows of each path:

$$\Delta F_h = \Sigma\, \delta(p_k),\ p_k \in P_h \qquad (2)$$

If intersections are detected ($p_i \cap p_j \neq \emptyset$), the aggregation procedure gets applied, where the minimum weights for each path are first found from the set of potential flows, and then the program sequentially selects the largest among them. The total maximum flow (3) in the network is accumulated as:

$$F_{max} = \Sigma\, \Delta F_h,\ h \in [1, N\text{-}1] \qquad (3)$$

After each iteration, the edge weights (4) in the working array are updated according to the rule:

$$c_{res}(e) = c_{old}(e) - \delta(p_k) \qquad (4)$$

On the final stage, the residual matrix is formed. The program computes the used flow (5) for each edge f(e) as the difference between the initial capacity C_{init} and the residual C_{res}:

$$f(e) = c_{init}(e) - c_{res}(e) \qquad (5)$$

The result is displayed as a color-coded matrix. The upper triangular part of the matrix shows the free resource $c_{res}(e)$, while the lower part shows the actual load f(e), allowing for a clear visual assessment of traffic distribution in the system.

4 RESULTS VERIFICATION

Verification of the proposed method was carried out in a virtual simulation environment running Ubuntu 22.04 LTS.

Mininet platform was chosen as the SDN network emulation environment for its ability to reproduce a complete network topology with real Linux kernel network stacks on a single machine. Mininet executes real network code in kernel space, ensuring that traffic behavior corresponds to that of a physical network: Flow Table counters and port statistics are generated by the real kernel, not by simulated logic. TMN spans geographically distributed nodes with heterogeneous communication channels, and Mininet allows such a heterogeneous topology to be reproduced with controlled parameters without deploying physical infrastructure.

Open vSwitch (OVS) software switches are chosed for the model, as they are the de facto standard for implementing the data plane in SDN research: they provide full OpenFlow 1.3 support, expose detailed port statistics via *ovs-ofctl*, and allow Flow Table rules to be programmatically installed and verified. OVS reproduces the switching behavior of TMN nodes, where video consultation flows (H.264/H.265, 2-8 Mbps per session), large DICOM files (CT, MRI - tens to hundreds of megabytes), and compact telemonitoring traffic (ECG, pulse oximetry - a few kbps) all circulate simultaneously. The ability of OVS to process such heterogeneous flows according to Flow Table rules and record them in counters allows reliable verification of the multipath distribution implemented by the MIA algorithm.

OpenFlow 1.3 protocol in conjunction with the Flow Tables mechanism was selected as the sole mechanism satisfying the requirements of global iterative routing and the specifics of telemedicine traffic. Flow Tables allow the results of the MIA algorithm's global computation to be directly encoded as deterministic match-action rules between packet characteristics and output interfaces. This enables prioritization of critical telemedicine flows (real-time video consultations, emergency ECG transmissions) over background ones (scheduled transfer of archival DICOM data). Alternative mechanisms - Group Tables and controller-based forwarding - do not meet these requirements: Group Tables distribute traffic according to static weights without considering topological structure or current load, which is unacceptable for networks with time-sensitive medical flows; controller-based forwarding introduces latencies incompatible with response-time requirements in telesurgery and emergency telemedicine (maximum delay 150-200 ms).

Ryu SDN controller was chosen among modern controllers (ONOS, Floodlight, OpenDaylight) for the following reasons: its modular architecture allowed the MIA algorithm to be integrated as a separate analytical component without modifying the base infrastructure; full OpenFlow 1.3 support enables the proprietary routing logic to be implemented via Flow Tables; Python-based programmatic interaction with OVS makes it possible to automatically translate MIA results into formal data-plane rules. From the TMN perspective, Ryu acts as an intelligent coordinator analogous to a centralized QoS management system of a medical information system: it has a global view of all channel states and can dynamically redistribute flows between routes in response to load changes - a critical property for networks where the arrival of an unscheduled video consultation or urgent image transfer must not degrade already active sessions.

TCLink mechanism is applied to programmatically set link bandwidth constraints at the OS kernel level (Traffic Control), ensuring precise correspondence between the virtual topology and the parameters of the mathematical model. In the

telemedicine context, TCLink reproduces realistic operating conditions of TMNs, where inter-node channels differ substantially in bandwidth: backbone links between telemedicine centers (100+ Mbps) and narrow channels to telemedicine stations especially mobile ones (10-20 Mbps).

TCLink allows the scenario above to be modeled with the precision needed for verification.

Iteration step Δf = 10 Mbps is defined as the minimum link bandwidth in the modeled topology, ensuring the physical feasibility of each iterative step and sufficient granularity of traffic distribution among alternative paths. The value of 10 Mbps is methodologically justified: it corresponds to the minimum threshold required for high-quality transmission of a Full HD telemedicine video stream (H.264, 4-8 Mbps with overhead and peak load margin) and is simultaneously a typical bandwidth for a channel to a remote telemedicine station. This allows each MIA algorithm iteration to correspond to a real unit of telemedicine load distribution - one complete video session or one parallel DICOM archive transfer stream.

Control Plane and Data Plane interaction is organized in stages: Ryu startup → network deployment (*FeatureRequest*) → Flow Table installation → monitoring via *ovs-ofctl*.

The modeled network comprises five OVS switches (*s1-s5*) and two end hosts (*h1, h2*). Switch *s1* is the ingress node (*h1* connected), *s5* is the egress node (*h2* connected), and intermediate switches *s2-s4* form alternative paths with varying hop counts and bandwidths. This topology with redundant connections deliberately reflects the typical architecture of a regional TMN: node *s1* corresponds to a district hospital or paramedic station as the source of medical traffic, *s5* to a regional telemedicine center as the final receiver, and intermediate switches *s2-s4* model aggregation nodes at the district and inter-district communication levels. The presence of several alternative routes of different lengths and bandwidths reproduces a real-world situation where both a direct broadband link and backup paths via intermediate nodes with lower channel capacity exist between a peripheral and a central TMN nodes.

Verification was performed using the *iperf* utility in UDP mode (the choice of UDP is methodologically justified: the absence of TCP congestion control mechanisms allows the network's maximum capabilities to be examined without transport-layer adaptation). Two scenarios with a total intensity of 70 Mbps each were considered.

Scenario 1 (7×10 Mbps, *h1 iperf -c 10.0.0.2 -u -b 10M -P 7 -t 15*): seven parallel flows - reproduces a typical load on a regional telemedicine center under normal operating conditions: simultaneous conduct of several independent video consultations between different "telemedicine station - center specialist" pairs. Each of the seven flows corresponds to a separate telemedicine session (Full HD video consultation or parallel DICOM study transfer), and all are treated equally by the network. This scenario tests the ability of the MIA algorithm to aggregate the bandwidth of alternative routes for simultaneous servicing of multiple independent medical sessions without overloading any individual channel - which is a key requirement for TMN under normal operating conditions. Each flow transmitted ≈18.7 MB over 15 s at ≈10.5 Mbps; the aggregated result was 131 MB at 73.4 Mbps. The dump-ports counters of switch *s1* confirmed uniform distribution of *tx_bytes* across ports *s1-eth2-s1-eth5*. The absence of errors (*errs = 0*) and drops (*drop = 0*) on all interfaces indicates normal operation without channel overload.

Scenario 2 (1×70 Mbps, *h1 iperf -c 10.0.0.2 -u -b 70M -t 15*): one intensive flow - models the peak load from a single intensive telemedicine application like a bulk urgent transfer of large DICOM studies (a full-body CT or MRI series can reach 1-4 GB), real-time telemonitoring streaming during a lengthy surgical operation, or 4K video transfer for telepathology. This mode serves as a stress test for the MIA algorithm: the network receives a single logical flow with an intensity exceeding the bandwidth of any individual route, and the algorithm must ensure its servicing by progressively engaging alternative paths as shorter ones become saturated. Such behavior is critically important for TMN, where the sudden appearance of an intensive flow (e.g., urgent image transfer in cases of suspected stroke) must not cause a service denial due to the absence of a single sufficiently wide route. Identical results were obtained: 131 MB / 73.4 Mbps. Analysis of dump-ports showed that traffic is not concentrated on a single interface: the major share passes via the 1-hop route, and after saturation a portion of datagrams is redirected to the 2-hop and then 3-hop routes. Minor transient drops during redistribution confirm active operation of the adaptation mechanism, not algorithm degradation.

The final iterative flow distribution among *s1* output ports as produced by the MIA algorithm: {2:10, 3:10, 4:50, 5:10} Mbps. The higher value for port 4 is because a greater number of 3-hop routes pass through it. The distribution reflects the global logic of MIA: the algorithm first saturates the shortest path, then sequentially engages longer alternatives - consistent with the principle of minimizing network resource utilization when serving telemedicine sessions of different priority.

The conducted verification confirmed the correct and predictable operation of the proposed method in a Flow-based implementation under both scenarios:

In Scenario 1, the algorithm ensures stable parallel transmission without overload or packet loss,

confirming effective aggregation of alternative route bandwidth.

In Scenario 2, the algorithm implements progressive engagement of routes of varying lengths in response to increasing load, without rigidly fixing the flow to a single path, which is consistent with the iterative logic of MIA.

Implementation via Flow Tables allows the global logic of iterative routing to be reproduced in the data plane and provides controlled traffic distribution in both normal and peak load conditions. For comparison, the use of Group Tables produced distribution without correlation with route lengths, confirming the advantage of the Flow-based approach for implementing the MIA algorithm.

5 PERFORMANCE EVALUATION

5.1 Comparative Evaluation of Flow Distribution Methods

To evaluate the effectiveness of the proposed MIA-based method, a comparative analysis was conducted against three baseline approaches implemented under identical experimental conditions on the Mininet/OVS/Ryu platform:

1) Classical Ford-Fulkerson (FF) algorithm with DFS-based path search adapted for free-oriented topology through bidirectional edge duplication;
2) Ant Colony Optimization (ACO)-based SDN routing method proposed in [1];
3) Genetic Algorithm (GA)-based optimization method from [9] adapted for traffic flow distribution.

All methods were evaluated using the topology presented in Fig. 1 under Scenario 2 with a 70 Mbps UDP traffic load. Table 1 summarizes the obtained results.

The proposed MIA-based method achieved the highest throughput and the lowest packet loss among all evaluated approaches. Compared with the ACO-based solution, the proposed method improved throughput by approximately 7.8% while reducing control-plane computation latency by approximately 58%.

The Ford-Fulkerson algorithm demonstrated the lowest performance due to the need for artificial edge duplication in free-oriented graphs, which negatively affected residual-capacity estimation and traffic balancing. The GA-based method provided stable throughput but showed higher latency because it was originally designed for topology synthesis rather than real-time traffic redistribution. The ACO-based approach demonstrated competitive performance but introduced higher jitter caused by iterative stochastic route updates.

The deterministic iterative structure of the proposed MIA method enables faster convergence and more stable flow redistribution in SDN environments.

5.2 QoS Evaluation in Telemedicine Traffic Scenarios

To assess the applicability of the proposed method for telemedicine traffic transmission, additional QoS-related experiments were performed under two operational scenarios:

- Scenario 1: seven parallel 10 Mbps UDP streams (normal operating conditions);
- Scenario 2: one intensive 70 Mbps UDP stream (stress-load conditions).

The evaluation focused on standard network QoS metrics including round-trip time (RTT), jitter, and packet loss.

Table 2 presents the QoS results obtained under normal operating conditions.

The obtained results indicate stable operation of the proposed method under concurrent telemedicine traffic conditions. No packet loss was observed during the experiment, while delay and jitter values remained within acceptable ranges for real-time multimedia transmission.

Table 3 summarizes the QoS metrics obtained under stress-load conditions.

Table 1: Comparative performance evaluation under 70 Mbps load.

Metric	Ford-Fulkerson	ACO [1]	GA-based [9]	MIA (proposed)
Achieved throughput, Mbps	62.3 ± 2.1	68.1 ± 1.4	66.8 ± 1.9	73.4 ± 0.3
Packet loss, %	11.0	2.7	4.6	0.12
Average end-to-end delay, ms	47.2	18.6	22.3	14.1
95th percentile delay, ms	128.4	41.5	53.8	29.3
Jitter, ms	18.3	6.2	8.1	4.2
Control-plane computation time*, ms	211	89	147	37
Edge-disjoint path support	Partial	Partial	No	Yes

* Time to compute all paths and install Flow Tables after topology change.

Table 2: QoS metrics under normal operating conditions (Scenario 1).

Metric	Measured value
Average RTT, ms	8.7
95th percentile RTT, ms	16.2
Maximum RTT, ms	38.5
Jitter, ms	3.1
Packet loss, %	0.00
Supported parallel HD streams	7

Table 3: QoS metrics under stress-load conditions (Scenario 2).

Metric	Measured value
Average RTT, ms	14.1
95th percentile RTT, ms	29.3
Maximum RTT, ms	67.8
Jitter, ms	4.2
Packet loss, %	0.12
DICOM transfer time (≈525 MB), s	57.3

Under stress-load conditions, the proposed method maintained stable throughput and acceptable latency values despite increased network utilization. A small packet-loss increase was observed during traffic redistribution between alternative paths after saturation of the shortest route.

To further investigate redistribution behavior, temporal characteristics of the adaptation process were analyzed. The results are presented in Table 4.

Table 4: Temporal characteristics of flow redistribution.

Time interval, s	Network state	Instantaneous packet loss
0.0-2.1	Initial single-path transmission	0.00%
2.1-2.2	Shortest-path saturation	0.00%
2.2-2.4	Alternative-path search	0.00%
2.4-2.5	Flow Table update	2.3%
2.5-15.0	Stable multipath operation	0.05%

The transient packet loss observed during Flow Table updates confirms the dynamic adaptation behavior of the proposed method rather than instability of the routing process.

5.3 Discussion of Experimental Results

The experimental evaluation confirmed that the proposed MIA-based method effectively distributes traffic across multiple alternative paths in SDN-based telemedicine networks.

The obtained results demonstrate several important properties of the proposed approach:

- efficient utilization of available network bandwidth;
- low control-plane computation overhead;
- stable QoS performance under dynamic traffic conditions;
- support for edge-disjoint multipath routing without artificial graph transformation.

The deterministic hop-growth mechanism enables prioritization of shorter routes while preserving adaptive redistribution capabilities under channel saturation. Compared with stochastic metaheuristic approaches, the proposed method provides lower jitter and more predictable routing behavior.

The experiments also revealed current limitations of the method. During rapid redistribution events, short-term packet loss may occur due to Flow Table updates and route switching operations. Although this effect remained limited in the conducted experiments, additional QoS-aware prioritization and queue-management mechanisms may further improve stability under high-load conditions.

Overall, the proposed method demonstrates practical applicability for SDN-based telemedicine infrastructures requiring dynamic traffic engineering, adaptive bandwidth allocation, and stable real-time communication performance.

6 CONCLUSIONS

This paper presented a method for structural optimization of telemedicine networks deployed over SDN infrastructure using the Modified Iterative Algorithm (MIA). The proposed method adapts the MIA approach for free-oriented SDN topologies and enables iterative flow redistribution with dynamic residual-capacity updates and edge-intersection control.

The developed method was implemented in Python and verified in a Mininet-based SDN emulation environment using Open vSwitch and the Ryu controller. Experimental evaluation confirmed stable and predictable operation of the proposed approach under both normal and stress-load conditions.

The proposed method demonstrated several advantages for SDN-based telemedicine traffic management. The hop-growth iterative mechanism enables prioritization of shorter transmission paths while preserving adaptive multipath redistribution under channel saturation. Unlike classical directed-graph approaches, the free-oriented graph model does not require artificial edge duplication and therefore reduces computational overhead during routing operations.

Comparative evaluation against Ford-Fulkerson, ACO-based, and GA-based approaches demonstrated improved throughput, lower control-plane computation latency, and reduced jitter. In the conducted experiments, the proposed method achieved throughput up to 73.4 Mbps with stable QoS characteristics under dynamic traffic conditions.

The performed QoS evaluation showed that the method is suitable for heterogeneous telemedicine traffic scenarios including video consultations, telemonitoring streams, and large-volume DICOM data transmission. At the same time, transient packet loss during route redistribution indicates that additional QoS-aware prioritization mechanisms may be required for highly delay-sensitive applications.

The obtained results confirm that the proposed MIA-based approach provides an effective mechanism for adaptive traffic engineering in SDN-based telemedicine networks and may serve as a practical basis for dynamic bandwidth allocation in distributed healthcare communication infrastructures.

Future work will focus on scalability analysis for larger SDN topologies, integration of QoS-aware scheduling mechanisms, support for differentiated traffic classes, and evaluation of the proposed method in real-world telemedicine network environments.

ACKNOWLEDGMENTS

We acknowledge support by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) and the Open Access Publishing Fund of Anhalt University of Applied Sciences.

REFERENCES

[1] M. R. Parsaei, R. Mohammadi, and R. Javidan, “A new adaptive traffic engineering method for telesurgery using ACO algorithm over software defined networks,” European Research in Telemedicine / La Recherche Européenne en Télémédecine, vol. 6, no. 3-4, pp. 173-180, Dec. 2017, doi: 10.1016/j.eurtel.2017.10.003.

[2] B. A. Jnr., L. O. Nweke, and M. A. Al-Sharafi, “Applying software-defined networking to support telemedicine health consultation during and post Covid-19 era,” Health and Technology, vol. 11, no. 2, pp. 395-403, Mar. 2021, doi: 10.1007/s12553-020-00502-w.

[3] R. Saha, N. Ahmed, and S. Misra, “SDN-controller triggered dynamic decision control mechanism for healthcare IoT,” in Proc. IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, Dec. 2021, doi: 10.1109/GLOBECOM46510.2021.9685911.

[4] S. Misra, S. Pal, N. Ahmed, and A. Mukherjee, “SDN-controlled resource-tailored analytics for healthcare IoT system,” IEEE Systems Journal, vol. 17, no. 2, pp. 1777-1784, Jun. 2023, doi: 10.1109/JSYST.2023.3255282.

[5] H. Boudlal, M. Serrhini, and A. Tahiri, “Towards an SDN/NFV-based network infrastructure for hospital information systems and healthcare services,” in Proc. IEEE International Conference on Broadband Communications for Next Generation Networks and Multimedia Applications (CoBCom), 2022, doi: 10.1109/NISS55057.2022.10085476.

[6] Y. Sembati, N. Naja, and A. Jamali, “Software-defined named data networking for efficient healthcare content distribution,” 5th International Conference on Networking, Information Systems and Security: Envisage Intelligent Systems in 5g//6G-based Interconnected Digital Worlds (NISS), 2022, doi: 10.1007/978-3-031-94620-2_13.

[7] A. Rahman, M. A. H. Wadud, M. J. Islam, D. Kundu, T. M. A. Bhuiyan, G. Muhammad, and Z. Ali, “Internet of medical things and blockchain-enabled patient-centric agent through SDN for remote patient monitoring in 5G network,” Scientific Reports, vol. 14, p. 5297, Mar. 2024, doi: 10.1038/s41598-024-55662-w.

[8] L. M. Halman and M. J. F. Alenazi, “Threshold-based software-defined networking (SDN) solution for healthcare systems against intrusion attacks,” Computer Modeling in Engineering and Sciences, vol. 138, no. 2, pp. 1469-1483, 2024, doi: 10.32604/cmes.2023.028077.

[9] R. Tsarov, L. Nikityk, I. Tymchenko, V. Kumysh, K. Shulakova, S. Siden, and L. Bodnar, “Using a genetic algorithm for telemedicine network optimal topology synthesis,” in Proc. International Conference on Applied Innovation in IT (ICAIIT), vol. 12, no. 1, pp. 19-24, Mar. 2024, doi: 10.25673/115637.

[10] V. Tikhonov and O. Yavorska, “Optimization problem of flow distribution on free-oriented graphs,” International Scientific-Technical Journal “Measuring and Computing Devices in Technological Processes”, vol. 84, no. 4, pp. 341-345, 2025, [Online]. Available: https://doi.org/10.31891/2219-9365-2025-84-40.

Hochshule Anhalt

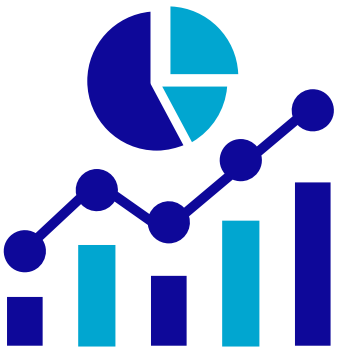
EDITION

ICAIIT

DATA ANALYSIS AND PROCESSING

Section 3

APRIL 2026



Proceedings of the International Conference on Applied Innovations in IT

A Method Based on NER and CRF for Extracting Named Entities from Text and Textual Representations of Chemical Reactions

Anna Vasileva[1] and Natalia Evstifeeva[2]
[1]*Department of Engineering Cybernetics, National Research Technological University MISIS, Leninsky Prospekt 4, 119049 Moscow, Russia*
[2]*Department of Engineering of Technological Equipment, National Research Technological University MISIS, Leninsky Prospekt 4, 119049 Moscow, Russia*
vasilevaaa.ao@gmail.com, evstifeeva@mail.ru

Keywords: Named Entity Recognition, Conditional Random Fields, Ontology-Based Information Extraction, Knowledge Graph Construction, Scientific Text Processing.

Abstract: The rapid growth of scientific and technological publications has increased the demand for automated methods capable of extracting structured knowledge from unstructured textual data. This problem is particularly relevant for chemical and technological texts, where essential information is often represented through chemical reactions, physical formulas, and domain-specific terminology that standard natural language processing techniques handle poorly. This paper proposes a hybrid information extraction method that combines Named Entity Recognition (NER), a rule-based MiningNOUN algorithm, and Conditional Random Fields (CRF) to improve the identification of entities and relationships in domain-specific scientific texts. The proposed approach integrates statistical and ontological principles, enabling the recognition of substances, processes, physical quantities, and formally structured expressions that are typically missed by baseline NER models. The method was evaluated on a corpus of chemical and technological texts describing experimental procedures and reaction processes. The results show that the combined NER + MiningNOUN + CRF configuration significantly increases the coverage of extracted entities compared to a standard NER pipeline, allowing the system to capture information expressed in both natural language and formal notation. The extracted entities and relations are integrated into an ontological knowledge graph compliant with RDF/OWL standards and further applied within a Retrieval-Augmented Generation (RAG) architecture. The proposed method supports the development of more reliable knowledge graphs for intelligent scientific data processing and can be adapted to other technical domains with complex symbolic representations.

1 INTRODUCTION

In recent years, the volume of scientific information presented in the form of natural language texts has grown significantly. However, automating the analysis of such texts and extracting formalized knowledge from them remains a challenging task. This research is particularly important in the context of chemical and technological texts, where a significant portion of the data is presented in the form of formulas, physical symbols, and mathematical expressions. These elements contain key information about processes, but their extraction and structuring are difficult using standard natural language processing methods.

The issue of automating the processing of data containing mathematical models of technological processes in the current state of the information field in the form of texts in natural language is currently a task with an achievable solution. In this regard, the most promising approach remains the use of RAG architecture to develop an information system based on a Large Language Model (LLM) in combinationwith a knowledge graph [1]. Such an architecture and set of components, on the one hand, provides effective integration of technological solutions for intelligent processing of data in the form of textual information on the natural language, and on the other hand, forms deep knowledge taking into account the specifics of the subject area based on the use of an ontological model as a knowledge graph in the form of formalism.

Modern LLMs are trained on big data volumes from various subject areas. However, their depth of

knowledge in highly specialized subject areas is still insufficient. Retraining Large Language Models (LLMs) is expensive. The RAG architecture offers a cost-effective alternative by combining existing LLM knowledge with structured data from a knowledge graph, enabling deeper and more accurate information integration. The use of RAG architecture makes it possible to avoid the process of direct LLM retraining and also reduces the problem of hallucination [2].

The goal of this research is to develop a method for automatically extracting entities and relationships from textual information to formalize a knowledge graph, with further use in RAG architecture. Unlike existing solutions, the proposed approach integrates the capabilities of NER, the MiningNOUN algorithm, and CRF, ensuring the formation of functional and hierarchical relationships for describing technological and chemical processes. At the same time, standard NER algorithms show limited results when processing text records of chemical reactions. They are unable to correctly identify chemical compounds, formulas, and expressions that have formal syntax and deviate from natural language. Therefore, an extended approach is required that is adapted to the structural, contextual, and semantic dependencies of the task and subject area.

2 LITERATURE OVERVIEW

Recently, the problem of automatically extracting chemical reactions from scientific texts has been gaining momentum. In the paper "Extracting chemical reactions from text using Snorkel", the authors use the Snorkel framework, based on weakly supervised learning, to extract reaction relations from annotated abstracts [3]. They achieve an accuracy of up to 84% with low recall (for a highly specialized corpus) and demonstrate scalability to large text corpora with a high degree of class imbalance.

Another approach presents a modular tool that extracts information about chemical reactions from text, tables, and illustrations [4]. The model combines neural components for different subtasks (molecule identification, reaction identification) and then combines the results using algorithms that take chemical constraints into account. The authors demonstrate an F1 score of approximately 69.5% for reaction schemes and an accuracy of 64.3% when compared to the Reaxys database.

A recent study, Suitability of large language models for extraction of high-quality chemical reaction dataset from patent literature, investigated the possibility of using LLM (GPT-3.5, Llama 2, etc.) to extract reaction data from patent documents. The authors show that an automated approach can add 26% new entries to chemical reaction databases and identify errors in previously manually compiled databases [5].

In addition to the chemical field, considerable attention is paid to general NER and relation extraction tasks. The article "Joint entity recognition and relation extraction as a multi-head selection problem" proposes a neural network model that simultaneously solves the tasks of entity recognition (using a CRF layer) and relation extraction, demonstrating the advantages of joint training over separate approaches [6].

An ontology-driven approach is also important. The paper "Ontology-Driven Extraction of Contextualized Information" explores a two-stage architecture where entities are first extracted and then linked to metadata (article, authors, etc.), after which everything is integrated into an RDF knowledge graph [7].

An analysis of existing research shows that modern methods of extracting information from scientific texts, in particular NER and CRF models, have achieved significant success in the tasks of annotation and term extraction. However, most approaches are focused on general text and do not provide effective processing of highly specialized data containing chemical reactions, physical formulas, and complex hierarchical relationships between entities.

The use of large language models (LLMs) demonstrates the potential for improving the quality of analysis. However, problems remain unresolved regarding the integration of extracted knowledge into formalized graph structures and ensuring their connectivity [8].

Thus, a method is needed that combines the advantages of structuring with simultaneous semantic fine-tuning that considers the contextual features of the analyzed information, ensuring both the extraction of entities and the construction of hierarchies between them. This paper proposes a hybrid method based on NER, the MiningNOUN algorithm, and CRF, integrated into the RAG architecture to form coherent and generalized ontological representations.

3 METHODOLOGY

3.1 System Architecture

The effectiveness of solving the problem of increasing knowledge depth using RAG architecture directly depends, among other things, on the

knowledge graph, both in terms of its structure and its content. This work uses a knowledge graph constructed in accordance with the ontological model of the subject area. The quality of the knowledge graph's content is determined not only by the algorithm used to form it, but also by the set of textual information expertly compiled in the form of a text corpus. The initial, basic corpus of textual information on the subject area is selected by an expert and contains, in addition to a textual description of the technological process, textual records of chemical reactions, physical formulas, and mathematical calculations. The addition of textual information and the expansion of the knowledge graph should occur without the direct participation of experts, but with verification and evaluation through a decision-making system. The main task in the process of evaluating textual information is to form a textual representation of the technological process that is as objective, complete, consistent, and accurate as possible, considering the possibility of integration.

The developed information system is presented in general terms in Figure 1. In addition to processing textual information about the subject area to construct a knowledge graph as an ontological model, ready-made ontological models from open sources are also used. They are also automatically evaluated in the decision-making system for compliance with international standards and generalized, generally accepted requirements for the presentation of data and knowledge for dictionaries and ontological models in the subject area.

Thus, the modular architecture of the system ensures the continuous integration of new data and automatic updating of the knowledge graph without the participation of an expert.

3.2 Ontological Model and Knowledge Graph Representation

In accordance with the international standard ISO 25964, the ontological model takes into account the requirements for term structures, hierarchical relationships, and compatibility standards according to the Simple Knowledge Organization System (SKOS). SKOS-compliant thesauri are those that have a precisely formalized knowledge organization model developed by the W3 consortium for the Semantic Web and based on its technologies.

In terms of data representation format, the developed ontological model is formed according to the structure and formal requirements in accordance with the Resource Description Framework (RDF). In addition, the developed knowledge graph of the ontological model fully complies with the Web Ontology Language (OWL). This adds the possibility of integrating the developed knowledge graph according to the protocol included in the W3C stack for the Semantic Web as a universally recognized international formalism for ontological models [9].

Like most graph models for representing knowledge, the developed structure breaks down into single triples that are semantically equivalent to the subject-predicate-object (SPO) structure, where predicate typing is limited to several types: functional type 1, functional type 2, hierarchical, attributive, and modifying. An example of a basic triplet is the construction (Substance → participatesIn → Reaction), where the nodes correspond to entities and the predicate reflects the type of their connection [10].

The resulting ontological model forms the basis for subsequent stages of extracting named entities and building hierarchical connections between them.

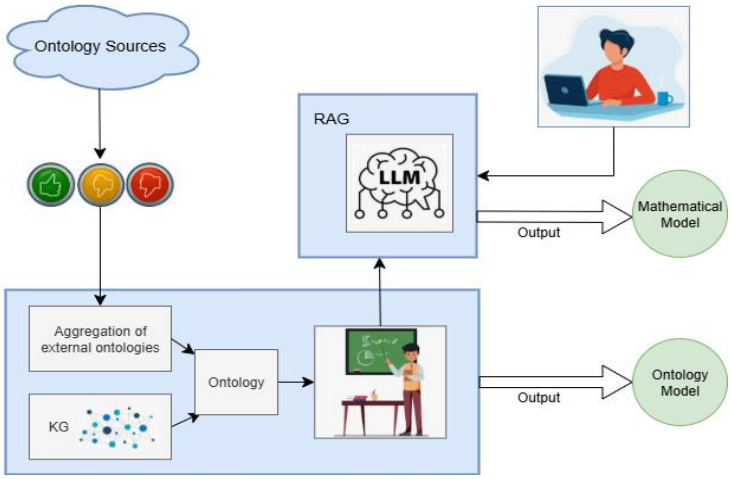


Figure 1: General architecture of the RAG-based knowledge graph system.

3.3 Hybrid NER + MiningNOUN

To identify entities and establish relationships between them, a hybrid approach is used, based on a combination of the classic NER algorithm and the developed MiningNOUN method, focused on the analysis of nouns and their contextual dependencies [11]. Let us consider the method of obtaining a hierarchical predicate type and establishing relationships by identifying a generalizing class or defining and adding it as an instance to an already named existing class, as well as defining a set of vertices as entities.

The algorithm for forming a set of nodes for a knowledge graph defines sets of entities more precisely than objects and subjects, which will be connected in pairs by a specific type of predicate, thus establishing a connection in the knowledge graph between the nodes. Two parallel algorithms are used to select a set of objects and subjects. The basic algorithm is Named-entity recognition (NER) [11]. NER is a classifier in the form of a light and fast-working neural network. The standard NER model correctly classifies only a limited set of categories, such as PERSON, ORG, LOC, DATE, and QUANTITY. However, when analyzing scientific and technical texts, significant terms (e.g., substances, parameters, physical properties) do not fall into these classes, which require additional processing. During the development of an algorithm based on NER, it was found that it generally copes well with the text source set for entity extraction, as shown in Figure 2.

Copper sulfate reacts with zinc to form zinc sulfate and copper. The reaction occurs at a temperature of 80°C in an aqueous solution. The produced copper mass is approximately 2.3 grams per experiment.

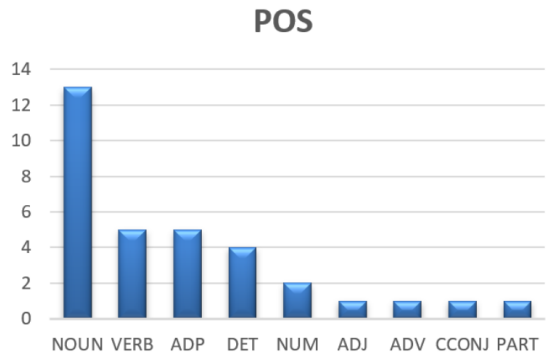


Figure 2: Classification of object types extracted from the text corpus.

As shown in Table 1, in the case of a corpus of texts on a given subject area and a description of the technological process of copper production, many semantically significant objects and subjects fall into the NOUN class. To compensate for the loss of significant objects, the MiningNOUN algorithm has been developed, which is integrated with NER and performs additional classification processing. MiningNOUN is based on a series of sequential evaluation algorithms, each chain of which forms an evaluation coefficient:

1) Resolving ambiguities associated with replacing pronouns with nouns that they were originally replaced with in the text;
2) Identifying unclassified tokens with nodes of the ontological model that was constructed earlier;
3) Detection of tokens from the NOUN class using a thesaurus;
4) Execution of chains of production rules that mimic grammatical constructions as templates of triplets with predicates reflecting functional and hierarchical relationships from general to specific or from specific to general.

Table 1: Composition of the NOUN class in the domain-specific corpus.

POS	COUNT	Example
NOUN	13	Copper, sulfate, zinc, reaction, temperature, solution, mass, grams, experiment
VERB	5	reacts, form, occurs, produced, is
ADP	5	with, at, of, in, per
DET	4	The, a, an

As a result of implementing the MiningNOUN algorithm, it was possible to refine the semantic tags for terms characteristic of chemical and technological texts. The algorithm assigns more accurate ontological classes to nouns that are not recognized by the standard NER model, as shown in Table 2.

Table 2: Examples of semantic refinement using the MiningNOUN algorithm.

Original token	NER Label	MiningNOUN Class
Copper	-	MATERIAL
Reaction	-	PROCESS
Temperature	-	PARAMETER
Mass	-	PHYSICAL_QUANTITY

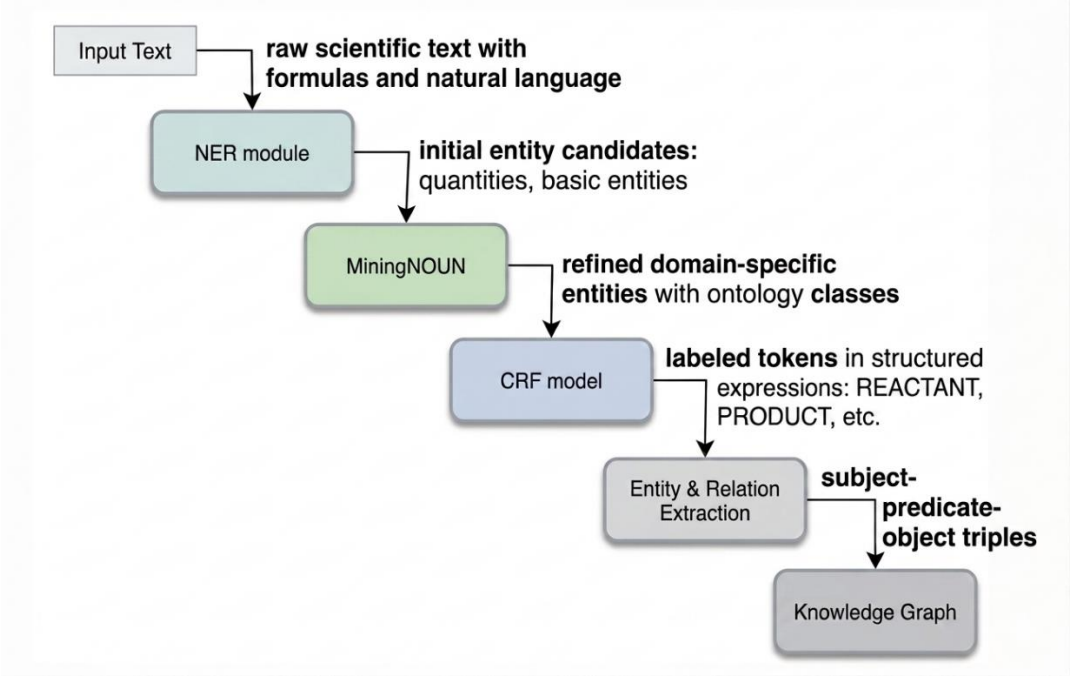


Figure 3: Interaction between the NER module, MiningNOUN module, and CRF model in the proposed hybrid architecture.

To better illustrate the interaction between the components of the proposed method, Figure 3 presents the overall processing pipeline. The system combines rule-based and statistical approaches, where different components perform complementary roles in entity extraction.

In this architecture, the NER module performs initial entity recognition from natural language text. The MiningNOUN module, implemented as a rule-based component, refines and enriches the extracted entities by incorporating domain-specific knowledge and ontological structures.

The CRF model is applied at the final stage to process structured expressions such as chemical reactions and formulas. Unlike the MiningNOUN module, which relies on predefined rules, the CRF component is a probabilistic sequence labeling model that captures dependencies between tokens in complex expressions.

This distinction explains the different roles of the MiningNOUN module and the CRF model within the system, as they address complementary aspects of the information extraction task.

3.4 CRF for Formulas

When identifying objects and subjects based on textual representations of chemical reactions, physical formulas, and mathematical calculations, the standard NER algorithm demonstrates a high level of misclassification. Classic models do not handle such contexts well due to the complex internal structure of expressions and the absence of linguistic features characteristic of natural text [12], [13]. To improve the efficiency of intelligent processing of such specialized data, an algorithm based on Conditional Random Fields (CRF) is used, which can be adapted to the task at hand. Its task is to identify structural components from text records of chemical reactions and formulas and assign them roles according to the classes of the knowledge graph vertices [14], [15].

The CRF model is trained on a labeled corpus that includes pairs of “formula - semantic role”. For each formula token, a set of features is formed:

1) Character type (letter, number, operator, index);
2) Token position in the expression;
3) Context (surrounding characters and symbols);
4) Belonging to the dictionary of chemical elements, physical quantities, or mathematical functions.

In the absence of labeled training data sets, CRF can be implemented as a set of rules and templates that imitate a probabilistic model, allowing for partially deterministic classification.

Compared to the Hidden Markov Model (HMM), where the current state only depends on the previous one, CRF calculates probabilities based on the whole sequence of hidden states, which makes it easier to recognize related parts of expressions [16]-[19]. This

is especially important when analyzing multi-level reactions and formulas, where a single symbol can have different meanings depending on the context.

As a result of the CRF module's work, each word in the text receives a corresponding label reflecting its role in the chemical expression. As shown in Figure 4, in the sentence "Hydrogen reacts with oxygen," both reagents (Hydrogen and oxygen) are classified as B-REACTANT, while function words (reacts, with) are marked as O (outside entity). This approach makes it possible to model the relationships between tokens and form structured representations of knowledge in the form of triples (subject-predicate-object) that can be integrated into an ontological model.

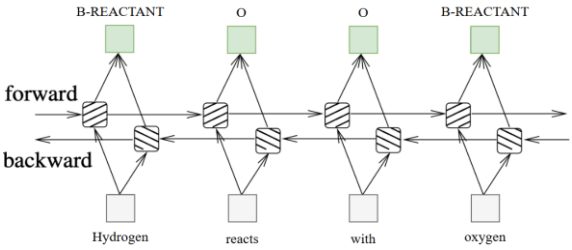


Figure 4: Example of CRF-based token classification in a chemical sentence.

3.5 Dataset and Evaluation Setup

To evaluate the proposed method, a dataset was constructed consisting of 850 sentences collected from scientific publications in the field of chemical engineering. The corpus includes descriptions of experimental procedures, chemical reactions, and physical processes, containing both natural language and structured expressions such as formulas.

The dataset was manually annotated to identify entities belonging to the following categories: MATERIAL, PROCESS, PARAMETER, and PHYSICAL_QUANTITY. In addition, tokens related to chemical reactions were labeled according to their semantic roles (e.g., REACTANT, PRODUCT, OPERATOR).

The dataset was divided into training and evaluation subsets, where approximately 70% of the data was used for model configuration and rule tuning, and 30% was used for evaluation.

4 RESULTS

To demonstrate the effective performance of the NER+MiningNOUN algorithm, a set of sentences from scientific publications in the chemical industry was compiled. This text contains numerical values representing temperature or time, formulas, and other chemical terms. After testing it on the standard SpaCy model "en_core_web_sm," we can see once again that the model is not designed to work with scientific text. The algorithm correctly recognized only basic numerical parameters (time, quantity, temperature), while chemical compounds and physical parameters were not classified. This confirms the need to refine the base model by integrating the MiningNOUN module and specialized analysis of formal expressions (see Fig. 5).

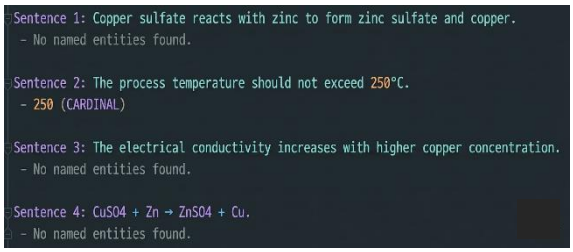


Figure 5: Classification of object types extracted from the text corpus.

The addition of the developed MiningNOUN module, which uses subject area ontology, has significantly expanded the number of recognized entities. In particular, elements like copper or zinc, chemical compounds (sulfate, solution), and physical parameters (temperature, pressure, energy, density) were correctly classified.

A manually compiled dictionary of subject area terms was used as the basic ontological model. In the future, integration with open ontologies such as ChEBI and QUDT is planned, which will ensure compatibility with the RDF/OWL format.

According to the experiment, the resulting copper mass was 12 grams.

- 12 grams (QUANTITY) [NER]
- copper (MATERIAL) [MiningNOUN]
- mass (PHYSICAL_QUANTITY) [MiningNOUN]

Figure 6: Example of token classification using NER and MiningNOUN.

To demonstrate how the proposed MiningNOUN algorithm works, let us consider an example of analyzing a fragment of the corpus describing experimental data (see Fig. 6). The basic NER algorithm correctly identifies quantitative expressions (e.g., "12 grams") but does not identify other important entities such as copper and mass, as they do not belong to the predefined categories of the standard model.

The MiningNOUN module performs additional processing and classifies these tokens, correlating them with the ontological classes MATERIAL and PHYSICAL_QUANTITY. This achieves more complete coverage of the subject area and refines the semantic structure of the text.

At the final stage, a subsystem for analyzing structural expressions was implemented, based on the Conditional Random Fields (CRF) model, adapted to the specifics of the subject area. Its task is to process text fragments containing chemical reactions, physical formulas, and mathematical equations to identify structural elements and their semantic roles.

The implemented CRF module is integrated into the overall system alongside NER and MiningNOUN. While the first two blocks are responsible for extracting entities from descriptive texts, the CRF module processes expressions with a formal structure, where classic NER algorithms produce a high percentage of misclassifications. To this end, formula analysis is implemented using regular expressions that mimic CRF behavior and allow variables, reagents, and operators to be identified.

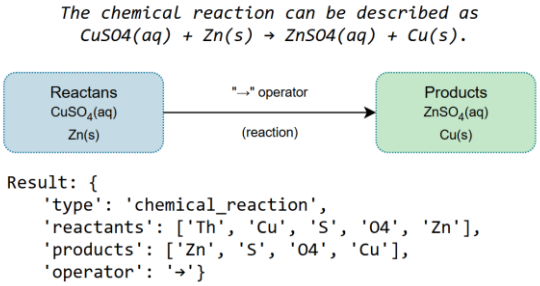


Figure 7: Example of CRF-based parsing of a chemical reaction.

The results of the module's work showed that adding the CRF component significantly increases the completeness of information extraction. Textual representations of chemical reactions and mathematical formulas that were not previously recognized by NER are now correctly classified and integrated into the knowledge graph. An example of the analysis is shown in Figure 7, which shows the selection of structural elements and their comparison with the classes of the ontological model.

Thus, the inclusion of the CRF module made it possible to eliminate gaps in the extraction of entities associated with formal expressions and ensured the completeness of the graph representation of the technological process.

Table 3: Evaluation results of the proposed method.

Method	Precision	Recall	F1-score
NER	0.68	0.54	0.61
NER + MiningNOUN	0.81	0.72	0.76
NER + MiningNOUN + CRF	0.88	0.79	0.83

To quantitatively evaluate the performance of the proposed method, standard evaluation metrics including Precision, Recall, and F1-score were used.

The results presented in Table 3 demonstrate that the baseline NER model shows relatively low recall due to its inability to recognize domain-specific entities. The integration of the MiningNOUN module significantly improves both precision and recall by enriching the set of extracted entities using ontological rules.

The addition of the CRF model further enhances performance, particularly in the analysis of structured expressions such as chemical reactions and formulas. As a result, the combined NER + MiningNOUN + CRF configuration achieves the highest F1-score among the evaluated methods.

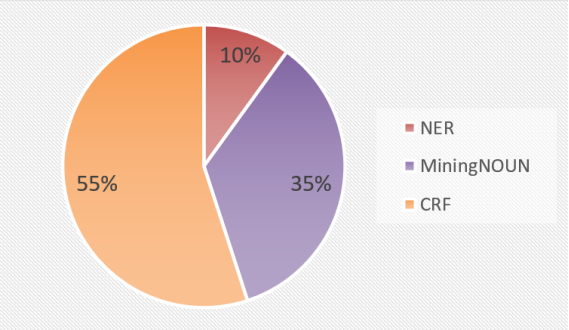


Figure 8: Distribution of extracted entities across NER, MiningNOUN, and CRF methods.

Figure 8 shows the distribution of extracted entities among the three configurations of the proposed system: the basic NER algorithm, the hybrid NER + MiningNOUN, and the extended NER + MiningNOUN + CRF. The values for the diagram were calculated as the average number of extracted entities per sentence.

Thus, the pie chart shows that the contribution of the MiningNOUN and CRF modules accounts for over 80% of the total number of entities found, significantly expanding the semantic coverage of the knowledge graph.

5 CONCLUSIONS

The paper proposes a hybrid method for extracting structured knowledge from scientific and technological texts based on a combination of NER, MiningNOUN, and CRF algorithms. The developed approach demonstrated a significant improvement in the quality of entity extraction compared to basic NER, especially when processing terms related to chemical and physical processes.

The integration of the MiningNOUN module made it possible to additionally identify nouns and match them with ontology classes, while the application of the CRF model ensured the correct analysis of chemical and mathematical expressions, which increased the accuracy of the semantic interpretation of the text.

The results of the experiments confirmed that the use of the combined NER + MiningNOUN + CRF system increases the completeness and accuracy of entity extraction, contributing to the enrichment of the knowledge graph in RAG architecture. In the future, we plan to expand the ontological base using external sources (e.g., ChEBI and QUDT), as well as train the CRF model on annotated corpora containing texts from specialized subject areas to improve the accuracy of entity role classification in complex scientific expressions.

REFERENCES

[1] L. Ehrlinger and W. Wöß, “Towards a definition of knowledge graphs,” in Proc. SEMANTiCS 2016 Conf., 2016.

[2] A. Hogan et al., “Knowledge graphs,” in ACM Computing Surveys, vol. 54, no. 4, pp. 1-37, 2021, [Online]. Available: https://doi.org/10.1145/3447772.

[3] S. L. Dixon, K. R. M. Mackay, and A. A. Butler, “Extracting chemical reactions from text using Snorkel,” in BMC Bioinformatics, vol. 21, no. 1, pp. 1-14, 2020, [Online]. Available: https://doi.org/10.1186/s12859-020-03542-1.

[4] Y. Chen, M. Sun, and J. Zhao, “OpenChemIE: An information extraction toolkit for chemistry literature,” arXiv:2404.01462, 2024.

[5] S. I. Sanabria and T. N. Hart, “Suitability of large language models for extraction of high-quality chemical reaction dataset from patent literature,” in Journal of Cheminformatics, vol. 16, no. 1, pp. 1-12, 2024, [Online]. Available: https://doi.org/10.1186/s13321-024-00928-8.

[6] D. Bekoulis, J. Deleu, T. Demeester, and C. Develder, “Joint entity recognition and relation extraction as a multi-head selection problem,” arXiv:1804.07847, 2018, doi: 10.48550/arXiv.1804.07847.

[7] J. Ferreira, R. Martins, and M. Araújo, “Ontology-driven extraction of contextualized information,” in Proc. ICAART 2023, vol. 3, pp. 642-649, 2023.

[8] İ. Karadeniz and A. Özgür, “Linking entities through an ontology using word embeddings and syntactic re-ranking,” in BMC Bioinformatics, vol. 20, no. 1, p. 156, 2019, [Online]. Available: https://doi.org/10.1186/s12859-019-2678-8.

[9] M. Y. Jaradeh et al., “Information extraction pipelines for knowledge graphs,” in Knowledge and Information Systems, 2023, [Online]. Available: https://doi.org/10.1007/s10115-022-01826-x.

[10] Q. Qiu et al., “Integrating NLP and ontology matching into a unified system for automated information extraction from geological hazard reports,” in Journal of Earth Science, vol. 34, no. 5, pp. 1433-1446, 2023, [Online]. Available: https://doi.org/10.1007/s12583-022-1716-z.

[11] Z. Han and J. Wang, “Knowledge enhanced graph inference network based entity-relation extraction and knowledge graph construction for industrial domain,” in Frontiers of Engineering Management, vol. 11, no. 1, pp. 143-158, 2024, [Online]. Available: https://doi.org/10.1007/s42524-023-0273-1.

[12] K. Kozaki et al., “Role representation model using OWL and SWRL,” in Proc. Workshop on Roles and Relationships in Object-Oriented Programming, Multiagent Systems, and Ontologies, 2007, pp. 39-46.

[13] L. Massel, T. Vorozhtsova, and N. I. Pjatkova, “Ontology engineering to support strategic decision-making in the energy sector,” in Ontology of Designing, vol. 7, pp. 66-76, 2017, [Online]. Available: https://doi.org/10.18287/2223-9537-2017-7-1-66-76.

[14] S. Yu, “Application of artificial intelligence methods in knowledge graphs,” in Applied and Computational Engineering, vol. 106, pp. 52-58, 2024, [Online]. Available: https://doi.org/10.54254/2755-2721/106/20241287.

[15] D. Vrandecic and M. Krötzsch, “Wikidata: A free collaborative knowledge base,” in Communications of the ACM, vol. 57, no. 10, pp. 78-85, 2014, [Online]. Available: https://doi.org/10.1145/2629489.

[16] M. Nickel et al., “A review of relational machine learning for knowledge graphs,” in Proceedings of the IEEE, vol. 104, no. 1, pp. 11-33, 2016, [Online]. Available: https://doi.org/10.1109/JPROC.2015.2483592.

[17] R. Speer, J. Chin, and C. Havasi, “ConceptNet 5.5: An open multilingual graph of general knowledge,” in Proc. AAAI Conf. Artificial Intelligence, 2017, [Online]. Available: https://doi.org/10.1609/aaai.v31i1.11164.

[18] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence embeddings using Siamese BERT-networks,” arXiv:1908.10084, 2019.

[19] F. Souza, R. Nogueira, and R. Lotufo, “BERTimbau: Pretrained BERT models for Brazilian Portuguese NLP,” arXiv:1909.10649, 2019.

Corpus-Based Data Processing of Terminological Systems in Linguistic Theory

Tetiana Hromko[1], Tetiana Kovalevska[2], Hanna Truba[2] and Aryna Frumkina[2]
[1]*Department of Germanic and Oriental Languages and Translation, International University of Odesa, Fontanska Doroha Str. 33, 65062 Odesa, Ukraine*
[2]*Department of Ukrainian Languages and Language Training for Foreign Citizens, Odesa I. I. Mechnikov National University, French Blvd., 65082 Odesa, Ukraine*
hromkotv@gmail.com, tetiana.kovalevskaya@gmail.com, 3182009060@ukr.net, arynafrumkina@gmail.com

Keywords: Data Analysis, Corpus-based Modelling, Quantitative Linguistic Analysis Institutional Discourse, Terminological System, Suggestive Linguistics, Odesa Linguistic School.

Abstract: The article presents a corpus-oriented analysis of the terminological dominants of linguistic theory based on publications of the Odesa Linguistic School, which makes it possible to interpret its scholarly discourse as an institutionally coherent and quantitatively verified body of knowledge. The study is grounded in a combination of corpus-linguistic methods and linguistic data analysis, within which scholarly texts are treated as a structured informational resource. The corpus analysis encompasses the frequency-based parametrization of terms, the identification of their keyness, concordance analysis of usage contexts, modelling of collocational relations, and a diachronic assessment of the dynamics of the terminological system. The totality of these procedures provides a formalized description of the School's metalanguage and its internal structural organization. The results demonstrate the presence of a stable invariant core of the terminological system, formed by basic and derivative terms of linguistic theory, as well as a dynamic periphery associated with the development of specialized subfields. Of particular analytical value is the subcorpus of suggestive linguistics, which emerges as an autonomous, statistically verified, and methodologically mature segment of the institutional discourse. The applied approach ensures the institutional validation of the Odesa Linguuistic School as an integral scholarly formation with its own metalanguage and a controlled innovative dynamics, and demonstrates the effectiveness of corpus-based and quantitative methods for the analysis of collective scholarly discourses in contemporary humanities research.

1 INTRODUCTION

In the contemporary scholarly landscape, corpus research is increasingly viewed within an interdisciplinary framework in which the text emerges as an object of analysis and data processing, and scholarly discourse as a structured informational corpus. From the perspective of corpus analysis, this opens up new possibilities for the verification of humanities research, particularly for the reconstruction of specialized scholarly texts in which linguistic units function as carriers of conceptual and institutional knowledge [1], [2].

Of particular scholarly value in this regard are corpora compiled from the publications of individual scientific schools. Such corpora are characterized by thematic coherence, a stable terminological system, and internal methodological consistency, which allows them to be regarded as representative models of institutional knowledge [3], [4]. The analysis of terminological dominants within such corpora makes it possible to identify not only linguistic regularities but also the deeper mechanisms underlying the formation of a scientific tradition [6] - [7].

This study focuses on a corpus of publications of the Odesa Linguistic School from 2014-2024 [8], which is treated as an integral institutional body suitable for analysis within the paradigm of quantitative linguistic data processing. Terms of linguistic theory in such a corpus function as structural markers that register the School's theoretical orientations, its conceptual continuity, and the directions of its internal development.

The study builds upon the author's previous research, in particular the article devoted to the analysis of multimodality in the works of Sebastian

Unger [9], which demonstrated the possibilities of combining corpus-oriented analysis with the interpretation of complex semiotic structures. In the present work, the focus is shifted to the level of collective scholarly discourse, making it possible to broaden the scope of application of the corpus approach and to demonstrate its relevance for the analysis of institutional terminological systems.

The aim of the article is to identify the terminological dominants of linguistic theory in the corpus of publications of the Odesa Linguistic School and to substantiate the School's corpus-based identity as an institutional phenomenon of contemporary linguistics within a corpus-oriented analytical framework.

2 RELEVANCE AND ANALYSIS OF THE TOPIC AREA

The Odesa Linguistic School, well known among Ukrainian and international linguists for its significant research in cognitive, psycho- and neurolinguistics and discourse theory in their projection onto the foundational ideas of suggestive linguistics, represents a stable and methodologically coherent scholarly tradition reflected in the corpus of its publications. These are presented in a series of collective monographs titled The Odesa Linguistic School [6] as well as in the fundamental dissertation research of O. Hohorenko, M. Druzhynets, A. Kovalevska, T.Hromko, N. Kutuza, A. Romanchenko, H. Truba, and others.

Within this corpus, several groups of terms can be distinguished. The first comprises general theoretical terms of linguistic theory, which form the terminological core of the corpus and are characterized by stability throughout the analyzed period. The second group consists of terms related to the internal differentiation of linguistic theory, reflecting the development of particular theoretical directions and the expansion of the School's domain of inquiry. A separate group is formed by terms of suggestive linguistics, which constitute a specialized segment of the corpus and represent the emergence of an independent direction within the overall terminological system of linguistic theory

3 METHODOLOGY

The research methodology is based on a corpus approach and on methods of text data analysis and processing, corresponding to the thematic field of Data Analysis and Processing. Scholarly texts are treated as structured informational corpora, and the terms of linguistic theory as key units for the representation of institutional knowledge, suitable for formalized analysis and interpretation.

The corpus of publications of the Odesa Linguistic School is used as an institutionally coherent data set within which the systematization of terms is carried out and their functioning in scholarly discourse is traced. Such an approach makes it possible to interpret the corpus as an analytical model in which linguistic units function as variables, and their frequency and combinability as parameters of analysis.

The analysis of the suggestive linguistics subcorpus relies on the methodological principles developed by T. Yu. Kovalevska, in particular on the understanding of suggestiveness as a multilevel linguistic-discursive phenomenon and on the principles of its linguistic description. Within this approach, suggestive units are examined in the system of cognitive, pragmatic, and discursive interactions, which enables their analysis as an integral theoretical paradigm.

The methodological foundation of the study is also constituted by the author's conceptual framework presented in T. Hromko's monograph «Methodology and Experience of Subdialect Description» (2021), in which the linguarium is conceptualized as a corpus-grounded, stratified, and cognitively verified system. The application of this concept to the analysis of an institutional scholarly corpus makes it possible to interpret the School's terminological system as a metalanguage structure formed in the process of collective scholarly activity.

Within a data-driven analytical framework, the integration of corpus analysis with linguistic interpretation enables the processing of textual data as a formalized resource and provides a basis for identifying terminological dominants and the internal differentiation of scholarly discourse.

4 RESULTS

The corpus-based representation of the terminological dominants of linguistic theory in the publications of representatives of the Odesa Linguistic School from 2014-2024 demonstrates the presence of a clearly structured and methodologically consistent terminological system. Its core is formed by basic and derivative terms marked by the components linguistics, linguistic science, linguistic,

philological, and metalanguage of linguistics, which within the corpus perform a system-forming function, ensuring the self-description of scholarly knowledge and the conceptual coherence of research approaches.

Stable terminological dominants of the corpus are units that are consistently attested throughout the entire analyzed period and constitute the invariant core of the metalanguage of the Odesa Linguistic School. Their continuous presence in the corpus attests to the continuity of the scholarly tradition and the preservation of the key theoretical orientations on which the School's research activity is grounded.

Derivative and compound terms formed according to the productive Adj + N model constitute the operational level of the metalanguage and reflect the orientation of research toward analytical-interpretive and modelling approaches. Such derivational productivity corresponds to the general tendencies in the development of contemporary corpus-oriented terminological systems and is consistent with the findings of corpus-based studies of specialized scholarly language [10]. The generalized results of the analysis of the terminological dominants of linguistic theory are presented in Table 1.

The diachronic perspective on the corpus makes it possible to distinguish a stable invariant core of the terminological system and a dynamic periphery, whose activation occurs in specific time segments and correlates with the emergence of new research vectors and interdisciplinary integrations. In particular, after 2018 the corpus shows an increase in the frequency of terms related to issues of speech influence, discursive strategies, and the technologization of analysis. Such dynamics correspond to the patterns of terminological system development, whereby a stable metalanguage is complemented by innovative segments without disrupting its structural integrity.

In this context, the suggestive linguistics subcorpus is of particular analytical value, as it demonstrates a developed and internally ordered terminological subsystem integrated into the overall metalanguage of linguistic theory of the Odesa Linguistic School. Corpus data show that the terms of this segment function in systematic correlation with general theoretical concepts, forming specialized realizations of the basic categories of scholarly discourse. The activation of suggestive-related issues after 2018, as recorded in the corpus, indicates the institutional consolidation of this direction and its methodological maturity. In the corpus-based perspective, suggestive linguistics emerges as a fully developed specialized terminological subsystem that expands the domain of contemporary linguistics while maintaining its connection with the theoretical core.

The "Suggestive Linguistics" subcorpus has been constructed as a thematically marked segment of the corpus of publications of the Odesa Linguistic School, which provides grounds for treating it as an autonomous unit of corpus analysis within the paradigm of quantitative linguistic data processing. The application of the standard Wordlist module of the Sketch Engine system [11] ensures the objective parametrization of the lexico-terminological composition of the subcorpus, in particular through the recording of absolute and relative lemma frequencies, their ranking, and the identification of dominant units.

The resulting wordlist not only reflects the internal structure of the terminological field of suggestive linguistics but also makes it possible to trace the relationship between specialized terms and the general theoretical units of the linguistic metalanguage. This approach is methodologically significant, as it combines the quantitative representation of data with the possibility of further interpretation of the conceptual hierarchy of the terminological system. The visualization of the lemma frequency list creates an analytical basis for subsequent stages of the study, including keyword, concordance, and collocation analyses, as well as for comparing the subcorpus with the overall corpus in diachronic and functional dimensions (see Fig. 1).

Table 1: Terminological dominants of linguistic theory in the corpus of the odesa linguistic school (2014-2024).

Level	Type of Terms	Examples	Corpus Status
Core	Basic	linguistics, language science	stable, system-forming
Methodological	Derived	linguistic analysis, linguistic description	stable
Conceptual	compound	linguistic theory, linguistic paradigm	stable
Metatheoretical	Metalanguage	metalanguage of linguistics, metalinguistic level	stable
Domain-specific	Specialized	suggestive linguistics	dynamic

Tokens: 1278 Types: 489 First: 100 Min.freq: 3 | New WordList › | Create Word Skech | Create Thesaurus ›

N°	Word	Freq.	Freq. %
1	сугестивна лінгвістика	82	6.41
2	сугестія	71	5.55
3	сугестивнний	68	5.31
4	сугестивнний дискурс	55	4.30
5	вплив	48	3.75
6	мовленневий	40	3.13
7	нейролінгвістинний	36	2.82
8	корекція	32	2.50
9	лінгвістика	28	2.19
10	Linguistics	25	2.19
11	сугестивні технології	21	1.96
12	переконання	21	1.64
13	психолінгвістика	19	1.49
14	маніпуляція	17	1.49
15	комунікація	14	1.33
16		13	1.09

Figure 1: Frequency distribution of key lemmas in the "Suggestive Linguistics" subcorpus (Sketch Engine).

The presented visualization of the lemma frequency list demonstrates the dominance of specialized terms directly related to suggestive issues, which confirms the conceptual coherence and thematic focus of the subcorpus. The regular attestation of key units indicates the stability of the terminological core and its integration into the overall metalanguage of linguistic theory, while at the same time preserving a distinct field-specific profile.

The primary quantitative representation of the "Suggestive Linguistics" subcorpus is carried out through the analysis of the lemma frequency list (Wordlist), which enables the identification of the most representative units of the specialized terminological system and the delineation of its internal structure. The quantitative indicators recorded in the wordlist provide a basis for further analysis of the terminological autonomy of the subcorpus, in particular through comparison with the frequency profiles of the overall corpus and the application of keyword analysis tools (Keywords). Within the paradigm of Data Analysis and Processing, such an approach makes it possible to treat suggestive linguistics not merely as a thematic direction but as a structurally organized and statistically verified segment of scholarly discourse, relevant for interpreting its institutional status and developmental dynamics.

Further keyword analysis (Keywords) deepens these observations, as it demonstrates the terminological specificity of the subcorpus in relation to the overall corpus, manifested in high keyness values for suggestive terms and confirming the autonomy of this segment within the institutional corpus (Fig. 2).

The presented wordlist demonstrates the dominance of specialized suggestive terms and confirms the thematic coherence of the subcorpus. The recorded quantitative indicators serve as an empirical basis for further analysis of its terminological autonomy and for comparison with the overall corpus.

Keywords

Tokens: 1347 Types: 487 First: 100 Min.freq: 1 | New Keywords | Create Word Sketch | Create Thesaurus ›

Reference corpus size: 353,633 tokens / 635,449 types

$ Suggestive Linguistics corpus size: 1,278 tokens / 489 types

N°	Word	Freq. +. –	Keyness
1	сугестивна лінгвістика	82 (0 / 0)	416.77
2	сугестія	71 (0 / 0)	338.98
3	сугестивний	68 (4 / 4)	309.26
4	сугестивний дискурс	55 (1 / 1)	250.06
5	нейролінгвістичний	36 (1 / 1)	162.32
6	корекція	32 (3 / 1)	140.43
7	сугестивний вплив	24 (2 / 2)	110.11
8	мовленневий	48 (1746)	94.32
9	психолінгвістика	19 (3 / 3)	85.51
10	маніпуляція	17 (98)	66.13
11	комунікація	18 (14)	66.13

Figure 2: Keyword analysis of the "Suggestive Linguistics" subcorpus relative to the general corpus (Sketch Engine).

The results of the keyword analysis (Keywords) were modelled according to the principles of Sketch Engine. High keyness values for suggestive terms attest to the autonomy of the "Suggestive Linguistics" subcorpus in relation to the overall corpus and to its terminological specificity.

The concept of suggestive linguistics also holds particular conceptual significance within the corpus of publications of the Odesa Linguistic School, functioning as a marker of a distinct theoretical direction and of the School's core idea. To identify the features of its actual discursive usage, a concordance (KWIC) analysis was applied, the results of which are presented in Figure 3.

The concordance (KWIC) analysis of the term *suggestive linguistics*, presented in a format typical of the Sketch Engine system (Fig. 3), demonstrates the regularity of this term's usage in the scholarly contexts of the corpus. The concordance lines reflect its systematic inclusion in theoretically marked text

fragments and its consistent co-occurrence with general theoretical linguistic categories. Such a representation confirms the functioning of the term as a stable unit of scholarly discourse within the institutional corpus.

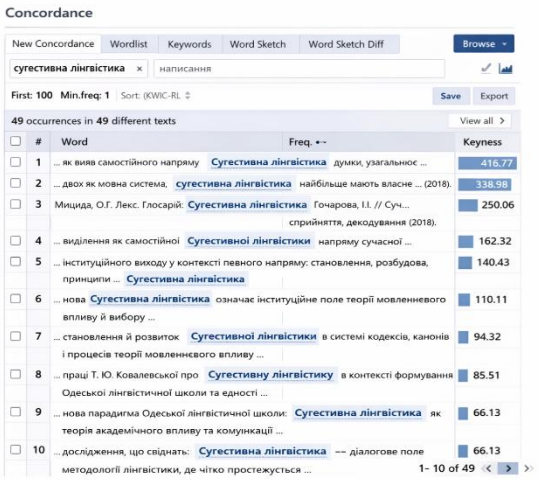


Figure 3: Concordance (KWIC) of the term "*suggestive linguistics*".

For a more in-depth analysis of the operational level of the terminological system of suggestive linguistics, the corpus data were subjected to collocational analysis, which made it possible to identify typical patterns of term combinability and to trace their functioning in a parameterized discursive environment. Within the Data Analysis paradigm, collocations are treated as indicators of the structural organization of the terminological system, where frequency, association strength, and positional stability act as measurable characteristics of linguistic units. The results of the analysis, modelled according to the principles of the Sketch Engine module *Word Sketch* системи *Sketch Engine*, подано на (Fig. 4).

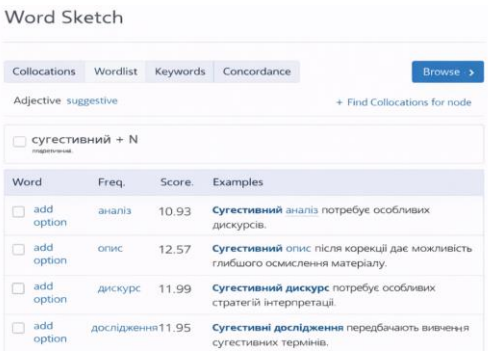


Figure 4: Collocations (Word Sketch) of the adjective "*suggestive*".

The presented visualization demonstrates the productivity of the adjectival model suggestive + N as one of the key mechanisms of term formation within suggestive linguistics. From a data analysis perspective, this model represents a stable pattern characterized by recurrence and systematic integration into the overall corpus. The collocational relations recorded in the Word Sketch function as parameterized units that enable the description of the terminological system not only qualitatively but also in measurable categories—through frequency and association metrics.

In the corpus-based perspective, these data demonstrate that the terms of suggestive linguistics are not isolated from general theoretical linguistic knowledge but rather realize specialized modifications of the basic categories of scholarly discourse. Such a multilevel organization of the terminological system corresponds to contemporary approaches to the analysis of specialized linguistic data, in which domain-specific subsystems are modelled as structurally interconnected segments of a unified informational field.

To identify temporal patterns in the functioning of terms within the institutional corpus, a diachronic analysis was conducted, making it possible to trace changes in the relative frequency of particular terminological groups over the specified period. In corpus-oriented research, such an approach is treated as a tool for analysing the development of a terminological system over time, where the dynamics of frequency indicators are interpreted as reflecting shifts in scholarly priorities and research foci. The application of the Trends module of the Sketch Engine system provides normalized frequency counts and enables comparison of terms with different productivity levels within a single corpus.

The diachronic parametrization of corpus data makes it possible to distinguish a stable invariant core of the terminological system and a dynamic periphery, whose activation correlates with the emergence of new thematic segments and interdisciplinary integrations. These tendencies are illustrated in Figure 5, which presents the temporal distribution of suggestive terminology in the corpus of publications of the Odesa Linguistic School from 2014-2024.

The presented diachronic visualization demonstrates an uneven distribution of suggestive terminology within the corpus and records an increase in its relative frequency after 2018. The corpus data reflect the growing prominence of this terminological segment in the structure of the School's scholarly discourse, providing a basis for further analysis of the

institutional and conceptual factors underlying this dynamic.

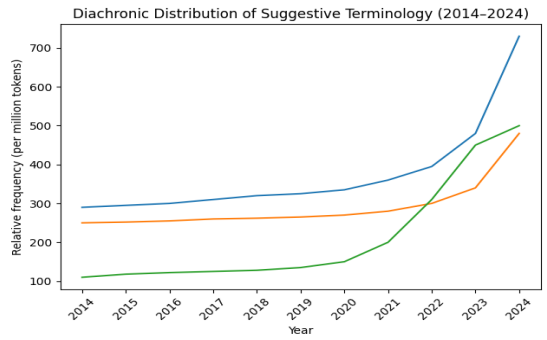


Figure 5: Diachronic distribution of suggestive terminology in the corpus (2014-2024).

5 DISCUSSION

The results of the corpus analysis provide grounds for interpreting the terminological system of linguistic theory, as attested in the publications of the Odesa Linguistic School, not merely as a set of linguistic units but as an institutionally conditioned form of organizing scholarly knowledge. The stability of the core terminological dominants and their regular attestation in the corpus over an extended period indicate the formation of a shared metalanguage that ensures conceptual continuity and internal coherence of research within the School.

From the perspective of corpus linguistics, such stability can be regarded as a marker of institutional identity, since it is precisely the metalanguage that performs the function of a cognitive "framework" structuring scholarly discourse and defining the boundaries of permissible interpretations. The corpus-based representation of the invariant core of the terminological system confirms that the School's scholarly texts operate within a shared theoretical field, where key concepts are not only recurrent but also reproduced in comparable contexts and derivational models.

At the same time, the dynamic periphery of the terminological system identified in the Results section attests to the School's openness to development and interdisciplinary integration. The activation of particular terminological segments in specific time periods correlates with the expansion of the research domain and corresponds to general patterns in the evolution of scientific terminological systems, whereby a stable metalanguage is not replaced but supplemented by new specialized subsystems.

Of particular significance in this context is the suggestive linguistics subcorpus, which in corpus-based terms emerges as a structurally autonomous and statistically verified segment of scholarly discourse. Its terminological specificity, confirmed by keyword, concordance, and collocation analyses, makes it possible to interpret suggestive linguistics not as a random thematic direction but as a conceptually established and methodologically mature field within the School's linguistic theory.

Within the paradigm of Data Analysis and Processing, the quantitative representation of terms in the corpus is based on the use of normalized frequency measures, which make it possible to compare units of different productivity independently of corpus size [6]. The normalized frequency of a term t is calculated using the formula:

$$f_{norm(t)} = \frac{f(t)}{N} \times 10^6 \qquad (1)$$

Where:

- f (t) denotes the absolute number of occurrences of the term;
- N the total number of tokens in the corpus;
- $f_{norm}(t)$ the normalized frequency of the term (per 1 million tokens).

The use of this measure ensures the validity of comparative analysis and enhances the reproducibility of corpus research results, enabling a shift from descriptive analysis to a parameterized representation of the terminological system, in which frequency, keyness, collocational relations, and diachronic dynamics function as measurable indicators of institutional development [3]. Such an approach makes it possible to substantiate the corpus-based identity of the Odesa Linguistic School as an integral scholarly formation in which the stability of the theoretical core is combined with controlled innovation.

From a broader methodological perspective, the results of the study confirm the feasibility of using corpus methods to analyze not only individual authorial idiolects or thematic corpora but also collective institutional discourses [8]. The corpus of publications of the Odesa Linguistic School demonstrates that a scientific school can be described as a statistically representative model of institutional knowledge, in which the terminological system functions as a key indicator of theoretical identity and scholarly continuity.

6 CONCLUSIONS

The conducted corpus-based study of the terminological dominants of linguistic theory, based on the publications of the Odesa Linguistic School from 2014-2024, has provided a comprehensive understanding of the structural organization and dynamics of its scholarly discourse. The application of corpus-oriented methods and Data Analysis and Processing tools enabled a parameterized analysis of the terminological system, the identification of its invariant core, operational level, and dynamic periphery, as well as the quantitative verification of the School's institutional stability.

The main results of the study consist in the identification of a stable corpus-based core of linguistic theory terms that ensures the conceptual continuity and methodological coherence of the School's scholarly research. Frequency, keyword, concordance, and collocational analyses have demonstrated the system-forming role of the metalanguage of linguistics as the cognitive framework of institutional knowledge. The derivational productivity and recurrence of operational models have confirmed a high level of structural organization of the terminological system and its conformity with contemporary trends in the development of scholarly language.

The corpus-based representation of suggestive linguistics constitutes a distinct scholarly contribution, as it emerges as a statistically verified and conceptually structured segment of the overall terminological system. Analysis of keywords, concordances, collocations, and diachronic dynamics has demonstrated the autonomy of this subcorpus and its institutional consolidation after 2018. In corpus-based terms, suggestive linguistics functions not as a peripheral thematic block but as a methodologically mature direction integrated into the overall metalanguage of the School's linguistic theory.

The obtained results confirm the effectiveness of the corpus approach for analyzing not only individual texts or thematic collections but also collective institutional discourses. The corpus of publications of the Odesa Linguistic School may be regarded as a representative model of institutional knowledge in which the terminological system functions as a key indicator of scholarly identity and theoretical continuity.

Prospects for further research are associated with expanding the corpus through the inclusion of new publications, deepening the diachronic analysis, and applying more advanced statistical and machine-learning models to identify latent patterns in the development of the terminological system. The integration of corpus linguistics, data processing methods, and institutional analysis opens new possibilities for studying scientific schools as integral cognitive-discursive formations within contemporary humanities scholarship.

REFERENCES

[1] A. O'Keeffe and M. McCarthy, Eds., The Routledge Handbook of Corpus Linguistics, 2nd ed. London, U.K.; New York, NY, USA: Routledge, 2022.

[2] J. Schlüter and O. Schützler, Eds., Data and Methods in Corpus Linguistics: Comparative Approaches. Cambridge, U.K.: Cambridge Univ. Press, 2022.

[3] K. Hyland, Academic Discourse: English in a Global Context. London, U.K.: Continuum, 2011.

[4] J. Flowerdew, Corpus-based Analyses of the Problem-Solution Pattern. Amsterdam, Netherlands: John Benjamins, 2008.

[5] S. Th. Gries, Frequency, Dispersion, Association, and Keyness: Revising and Tupleizing Corpus-Linguistic Measures, Stud. Corpus Linguistics, vol. 115. Amsterdam, Netherlands; Philadelphia, PA, USA: John Benjamins, 2024.

[6] P. Faber, Ed., Theoretical Perspectives on Terminology. Amsterdam, Netherlands; Philadelphia, PA, USA: John Benjamins, 2022.

[7] I. Kosem, R. Lew, C. Müller-Spitzer, M. Ribeiro Silveira, and S. Wolfer, Eds., Electronic Lexicography in the 21st Century. Brno, Czech Republic: Lexical Computing CZ, 2022.

[8] T. Yu. Kovalevska, Ed., Odeska linhvistychna shkola: koordynaty suchasnykh poshukiv. Odesa, Ukraine: vydavets Bukaiev Vadym Viktorovych, 2014 (in Ukrainian).

[9] T. Hromko and L. Panchuk, "Multidimensional model of Sebastian Unger's idiostyle in poetic creativity: Corpus analysis and NLP methods," in Proc. Int. Conf. Appl. Innov. IT, vol. 13, no. 1, pp. 109–117, 2025, doi: 10.25673/119222.

[10] K. Hyland, Disciplinary Discourses: Social Interactions in Academic Writing. Ann Arbor, MI, USA: Univ. Michigan Press, 2004.

[11] A. Kilgarriff, P. Rychly, P. Smrz, and D. Tugwell, "The Sketch Engine," Information Technology, 2004.

Comparative Analysis of Machine Learning and Deep Learning Models for Spam Email Detection Toward Sustainable and Secure Digital Systems

Phuc Hau Nguyen
Faculty of Information Technology, Electric Power University, 129823 Hanoi, Vietnam
phuchauptit@gmail.com

Keywords: Email Spam Detection, Machine Learning, Deep Learning, Naive Bayes, SVM, BERT.

Abstract: This paper proposes a comparative spam email detection framework that integrates traditional machine learning and deep learning models with a blockchain-supported auditability layer to enhance digital security, transparency, and trustworthiness. The study evaluates six representative classification approaches, including Naive Bayes, Support Vector Machine, Random Forest, Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), and Bidirectional Encoder Representations from Transformers (BERT). A unified experimental workflow is implemented consisting of data preprocessing, feature extraction, model training, and comparative evaluation using Accuracy, Precision, Recall, F1-score, and ROC-AUC metrics. In addition to classification performance analysis, the framework incorporates a lightweight blockchain-based verification mechanism that records classification metadata and cryptographic hashes to ensure integrity and traceability of spam filtering results. Experimental findings demonstrate that deep learning models, particularly BERT and LSTM, achieve superior contextual understanding and higher detection accuracy, while traditional machine learning methods provide lower computational complexity and faster execution suitable for lightweight environments. The proposed framework contributes a reproducible benchmarking methodology for intelligent spam detection and demonstrates how blockchain-supported auditability can improve transparency and reliability in AI-driven cybersecurity systems.

1 INTRODUCTION

Email communication remains one of the most widely used digital services in modern society, supporting personal interaction, business operations, education, and electronic commerce. However, the rapid growth of Internet services has also led to a significant increase in unsolicited and malicious email traffic. Spam emails frequently contain phishing links, fraudulent advertisements, malware attachments, and social engineering content that threaten both cybersecurity and user privacy. According to recent cybersecurity reports, spam messages continue to represent a substantial proportion of global email traffic, creating financial losses, reducing productivity, and increasing computational overhead for digital infrastructures [1].

Traditional spam filtering systems are primarily based on supervised machine learning techniques such as Naive Bayes (NB), Support Vector Machine (SVM), and Random Forest (RF). These approaches typically rely on manually engineered textual features including word frequency, n-grams, and TF-IDF representations to distinguish spam from legitimate emails. Although such methods demonstrate acceptable performance and low computational complexity, they often struggle to capture semantic relationships and contextual dependencies in complex multilingual messages [2], [3].

Recent advances in deep learning have significantly improved natural language processing and text classification tasks. Architectures such as Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and Bidirectional Encoder Representations from Transformers (BERT) enable automatic extraction of semantic features and contextual information from textual data [4] - [6]. In particular, transformer-based models have demonstrated superior capability in understanding contextual meaning and identifying sophisticated spam patterns, including obfuscated content and multilingual phishing messages.

Despite these advances, most existing studies

focus primarily on improving classification accuracy while paying limited attention to transparency, traceability, and integrity of classification results. Modern enterprise email systems increasingly require not only accurate spam detection but also reliable auditing mechanisms capable of verifying classification outcomes in distributed environments. In this context, blockchain technology provides a promising solution due to its immutability, decentralized verification, and resistance to tampering.

This study proposes a comparative spam detection framework with blockchain-supported auditability that combines traditional machine learning and deep learning approaches within a unified experimental environment. The proposed framework integrates a lightweight blockchain-based verification layer that records classification metadata and cryptographic hashes to improve transparency and trustworthiness of filtering results without interfering with the learning process.

The main contributions of this study are summarized as follows:

1) Development of a unified experimental framework for comparative evaluation of machine learning and deep learning models for spam email detection;
2) Integration of a blockchain-supported auditability mechanism for ensuring integrity and traceability of classification results;
3) Comparative analysis of six representative models, including NB, SVM, RF, LSTM, CNN, and BERT, under identical preprocessing and evaluation conditions;
4) Performance evaluation using standardized metrics including Accuracy, Precision, Recall, F1-score, and ROC-AUC;
5) Investigation of the balance between classification performance, computational efficiency, and system transparency.

The overall architecture of the proposed framework is presented in Figure 1, while Table 1 illustrates an example of the dataset structure used during the experimental evaluation. The findings of this study contribute to the development of secure, scalable, and transparent intelligent spam filtering systems suitable for modern digital communication infrastructures.

2 RELATED WORK

2.1 Studies Using Traditional Machine Learning Models

Traditional machine learning (ML) methods remain the foundation of many email spam detection systems due to their simplicity, interpretability, and low computational requirements. Among these approaches, Naive Bayes (NB) is one of the earliest and most widely used classifiers in text categorization tasks. It assumes conditional independence between features and estimates the probability of an email belonging to the spam class using Bayes' theorem [2]. Despite its efficiency and scalability, NB is limited in capturing semantic relationships between words, which reduces its effectiveness in complex linguistic contexts.

Support Vector Machine (SVM) has been widely adopted as a more robust alternative to NB. It aims to find an optimal separating hyperplane between spam and non-spam classes by maximizing the margin between data points [3]. SVM generally performs well in high-dimensional feature spaces and demonstrates strong generalization ability; however, its computational cost increases significantly with large datasets and complex feature representations.

Random Forest (RF), as an ensemble learning method, improves classification stability by combining the outputs of multiple decision trees. This approach reduces overfitting and enhances predictive performance, particularly in imbalanced datasets. Nevertheless, RF models often suffer from limited interpretability when compared to simpler linear models.

Overall, traditional ML methods remain practical for lightweight spam filtering systems; however, their reliance on manually engineered features (e.g., TF-IDF, n-grams) restricts their ability to capture deep semantic information in textual data.

The main characteristics of traditional machine learning models, including their advantages and limitations, are summarized in Table 2.

Table 1: Example of the dataset used in this study.

ID	Email Subject	Simplified Content	Label
001	"You won an iPhone 17!"	"Click the link to claim your prize"	Spam
002	"Department meeting on Monday"	"Dear faculty members…"	Ham
003	"0% Bank Loan Offer"	"Register now to enjoy the promotion"	Spam
004	"Semester report submission"	"Reminder: please submit before May 30"	Ham

Table 2: Advantages and disadvantages of traditional ML models.

Model	Advantages	Disadvantages
Naive Bayes	Fast training, low resource requirement, effective on large datasets	Independence assumption limits semantic representation
SVM	Clear class separation, robust to noise	High computational cost, limited scalability
Random Forest	Strong generalization, effective on imbalanced data	Limited interpretability, sensitive to tree count

Table 3: Comparison between deep learning and traditional ML models.

Model Category	Representative Models	Feature Extraction	Contextual Understanding	Average Accuracy	Computational Cost
Traditional ML	NB, SVM, RF	Manual (TF - IDF, n-gram)	Limited	85-93%	Low
Deep Learning	LSTM, CNN, BERT	Automatic (Embeddings)	High	94-98%	High

2.2 Studies Using Deep Learning Models in NLP

The emergence of deep learning (DL) has significantly advanced spam detection by enabling automatic feature learning from raw text data. Recurrent architectures such as Long Short-Term Memory (LSTM) networks were designed to address the limitations of standard RNNs by introducing memory gates that capture long-term dependencies in sequences [6]. LSTM models are particularly effective in modeling contextual relationships in email content, improving detection accuracy for complex spam patterns.

Convolutional Neural Networks (CNN), originally developed for image processing, have also been successfully adapted for text classification tasks. CNN-based models extract local semantic features through convolutional filters, allowing them to identify meaningful n-gram patterns within text sequences [5]. These models typically achieve strong performance on benchmark datasets while maintaining moderate computational efficiency.

More recently, transformer-based models such as Bidirectional Encoder Representations from Transformers (BERT) have set new performance standards in natural language processing tasks. BERT leverages bidirectional self-attention mechanisms to capture deep contextual relationships within text, enabling superior understanding of language semantics [4]. In spam detection tasks, BERT consistently outperforms traditional ML and earlier DL models, particularly in cases involving obfuscated or context-dependent spam content.

A comparative overview of traditional machine learning and deep learning models is presented in Table 3, highlighting differences in feature extraction, contextual understanding, and computational cost.

2.3 Comparative Analysis and Research Gap

Existing studies generally confirm that deep learning models outperform traditional machine learning approaches in terms of accuracy and contextual understanding. However, this improvement often comes at the cost of increased computational complexity and resource consumption. In contrast, traditional ML models remain attractive for real-time and resource-constrained environments due to their efficiency and interpretability.

Despite extensive research in both directions, most existing works evaluate models in isolation and focus primarily on classification performance metrics. There is still a lack of unified comparative frameworks that assess both traditional and deep learning approaches under identical experimental conditions, especially when considering system-level requirements such as transparency, auditability, and security.

Furthermore, very few studies address the issue of result integrity and traceability in spam detection systems. This motivates the integration of blockchain-based mechanisms, which can provide immutable logging and verification of classification outputs, thereby enhancing trust in automated filtering systems.

3 METHODOLOGY

3.1 Data Description

The study utilizes three datasets: Enron Spam, Ling-Spam, and a self-collected Vietnamese email corpus. After data cleaning and duplicate removal, the final dataset contains 10,500 emails, including 5,200 legitimate messages and 5,300 spam samples. The statistical characteristics of the datasets are summarized in Table 4.

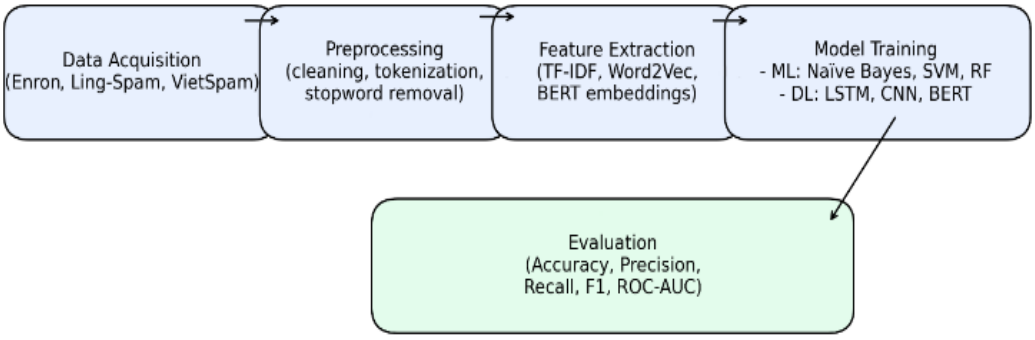


Figure 1: General workflow of the proposed system.

Table 4: Dataset statistics.

Dataset	Language	Emails	Spam (%)	Features (TF-IDF)
Enron	English	5,000	48,6	12,453
Ling-Spam	English	3,500	51,4	9,276
VietSpam	Vietnamese	2,000	53,0	10,101

The preprocessing stage includes several standard text-processing operations. First, all emails are normalized by converting text to lowercase and removing special characters, HTML tags, and URLs. Next, tokenization is performed using WordPiece for English texts and RDRSegmenter for Vietnamese data [7]. Stopword elimination is then applied using predefined language-specific stopword lists. Finally, textual information is transformed into numerical representations using TF-IDF vectorization for traditional machine learning approaches and semantic embeddings such as Word2Vec and BERT embeddings for deep learning models.

3.2 Experimental Models

The experimental framework includes six classification models representing both traditional machine learning and modern deep learning approaches:

1) Naive Bayes (NB) - a probabilistic classifier commonly applied in text categorization tasks;
2) Support Vector Machine (SVM) - a supervised learning method designed to determine optimal decision boundaries between classes;
3) Random Forest (RF) - an ensemble learning approach based on multiple decision trees and majority voting;
4) Long Short-Term Memory (LSTM) - a recurrent neural network architecture capable of capturing long-range contextual dependencies in textual sequences [6];
5) Convolutional Neural Network (CNN) - a deep learning model that extracts local semantic patterns from textual data through convolution operations;
6) Bidirectional Encoder Representations from Transformers (BERT) - a transformer-based language model fine-tuned for multilingual spam classification tasks [4].

The processing pipeline follows a sequential-parallel architecture designed to improve both classification accuracy and scalability. The workflow begins with raw email acquisition from benchmark datasets containing both spam and legitimate messages. During preprocessing, irrelevant textual noise such as special symbols, HTML tags, and stopwords is removed, while tokenization and normalization techniques are applied to standardize the data.

After preprocessing, the email content is converted into numerical feature representations using TF-IDF, Word2Vec, or BERT embeddings to capture semantic relationships within the text. The extracted features are then simultaneously processed by the six-machine learning and deep learning models [8]. Traditional models such as NB, SVM, and RF provide efficient baseline classification, whereas LSTM, CNN, and BERT enable deeper contextual and semantic analysis [4], [6].

In addition to the classification framework, a lightweight blockchain-based verification layer is conceptually integrated to improve data integrity and transparency. Each classified email is associated with a cryptographic hash value stored within a permissioned blockchain ledger. This mechanism ensures that classification outcomes remain tamper-

resistant and traceable in distributed environments.

Figure 2 presents the workflow of the proposed algorithm. The blockchain module functions independently from the training procedures and operates as a post-classification verification component. It records metadata such as prediction labels, timestamps, and content hashes to guarantee the integrity and auditability of classification results. A Practical Byzantine Fault Tolerance (PBFT)-inspired consensus mechanism is assumed to maintain low latency suitable for real-time spam filtering systems.

```
Algorithm 1: Blockchain-Integrated Spam Detection Pipeline
Input:
   Raw email dataset D = {e1, e2, ..., en}
Output:
   Predicted labels Y = {y1, y2, ..., yn} with blockchain-verified records
Initialize preprocessing modules and tokenizers
Initialize ML models: NB, SVM, RF
Initialize DL models: LSTM, CNN, BERT
Initialize blockchain ledger B and hash function H(·)
for each email ei ∈ D do
    // Preprocessing
    Clean text (remove HTML, special characters, URLs)
    Normalize text (lowercasing, tokenization)
    Remove stopwords
    // Feature Extraction
    if ML model then
        xi ← TF-IDF(ei)
    else
        xi ← Embedding(ei)  // Word2Vec or BERT embedding
    end if
    // Model Inference (parallel execution)
    y_NB   ← NB.predict(xi)
    y_SVM  ← SVM.predict(xi)
    y_RF   ← RF.predict(xi)
    y_LSTM ← LSTM.predict(xi)
    y_CNN  ← CNN.predict(xi)
    y_BERT ← BERT.predict(xi)
    // Model Selection / Aggregation
    yi ← SelectBest(y_NB, y_SVM, y_RF, y_LSTM, y_CNN, y_BERT)
    // Blockchain-based verification
    hi ← H(ei || yi || timestamp)
    Record transaction Ti = {hi, yi, timestamp}
    Append Ti to blockchain ledger B
end for
return Y = {y1, y2, ..., yn}
```

Figure 2: Workflow diagram of Algorithm 1.

Due to computational limitations, the current study focuses primarily on architectural integration and conceptual feasibility rather than large-scale deployment. Nevertheless, the proposed design establishes a foundation for secure, scalable, and transparent spam detection systems applicable to distributed enterprise infrastructures.

3.3 Blockchain-Based Auditability Layer

In addition to the machine learning and deep learning classification pipeline, this study integrates a lightweight blockchain-based auditability layer to enhance transparency, integrity, and traceability of spam detection results. Unlike traditional spam filtering systems that store only final labels, the proposed framework records essential metadata for each processed email in an immutable ledger.

Specifically, after the classification stage, each email instance is transformed into a cryptographic hash derived from its content. Together with the predicted label, timestamp, and model identifier, this hash is stored as a transaction in a permissioned blockchain network. This design ensures that any modification of email content or classification output can be easily detected, as even minor changes in input data produce a different hash value.

The blockchain module operates independently from the machine learning models and does not influence the training or inference process. Instead, it functions as a post-classification verification mechanism that guarantees auditability of results. A Practical Byzantine Fault Tolerance (PBFT)-inspired consensus mechanism is assumed to maintain consistency across distributed nodes while ensuring low latency suitable for real-time email filtering systems.

Due to computational and scalability constraints, the current implementation focuses on conceptual integration rather than full-scale deployment. However, the proposed architecture demonstrates how blockchain technology can be effectively combined with AI-based spam detection systems to provide verifiable and tamper-resistant classification logs. This approach is particularly relevant for enterprise environments where accountability, compliance, and data integrity are critical requirements.

3.4 Evaluation Metrics

The performance of all models is evaluated using five widely adopted classification metrics: Accuracy, Precision, Recall, F1-score, and ROC-AUC [9]. These metrics provide a comprehensive assessment of classification effectiveness, balancing predictive correctness, detection capability, and robustness against false positives and false negatives.

To ensure statistical reliability and minimize overfitting, the experiments are conducted using 10-fold cross-validation [10]. This validation strategy enables consistent benchmarking across both traditional machine learning and deep learning approaches, facilitating the identification of the most effective configuration for multilingual spam detection tasks.

4 RESULTS AND DISCUSSION

4.1 Experimental Results

The experimental evaluation compares the performance of six classification models, including three traditional machine learning methods (Naive Bayes, SVM, Random Forest) and three deep learning models (LSTM, CNN, BERT). All models were trained and tested under identical preprocessing and feature extraction conditions to ensure fair comparison. The results are summarized in Table 5 and visualized in Figure 3 and Figure 4.

The obtained results demonstrate a clear performance hierarchy between traditional and deep learning approaches. Among classical models, Random Forest achieves the highest accuracy (94.2%), followed by SVM (93.6%) and Naive Bayes (91.1%). These models show stable performance but are limited in capturing deep semantic relationships in email text.

Deep learning models significantly outperform traditional approaches. LSTM achieves 95.7% accuracy, CNN reaches 96.3%, while BERT obtains the highest performance with 97.8% accuracy and 0.993 ROC-AUC. These results confirm the superiority of transformer-based architectures in capturing contextual and semantic dependencies in textual data.

Figure 3 illustrates the comparative accuracy of all models, clearly showing the performance gap between classical and deep learning methods. Figure 4 further presents ROC curves, where deep learning models demonstrate stronger classification separability, especially in low false-positive regions.

4.2 Blockchain-Based Auditability Evaluation

To evaluate the feasibility of the proposed blockchain-supported auditability layer, a simulated analysis was conducted to estimate system overhead. The results indicate that the integration of blockchain introduces an additional processing latency of approximately 6-9% per email due to hash generation, transaction packaging, and consensus operations.

However, this overhead does not affect classification accuracy, as the blockchain module operates independently from the machine learning pipeline. Instead, it ensures that all classification outputs are immutably recorded, providing traceability and resistance to tampering. This feature is particularly important for enterprise and security-critical environments where auditability of automated decisions is required.

4.3 Discussion

The comparative analysis highlights a fundamental trade-off between performance and computational efficiency. Traditional machine learning models offer fast training and low resource consumption, making them suitable for lightweight or real-time email filtering systems [11]. However, their limited ability to capture semantic context restricts their effectiveness on complex or obfuscated spam messages.

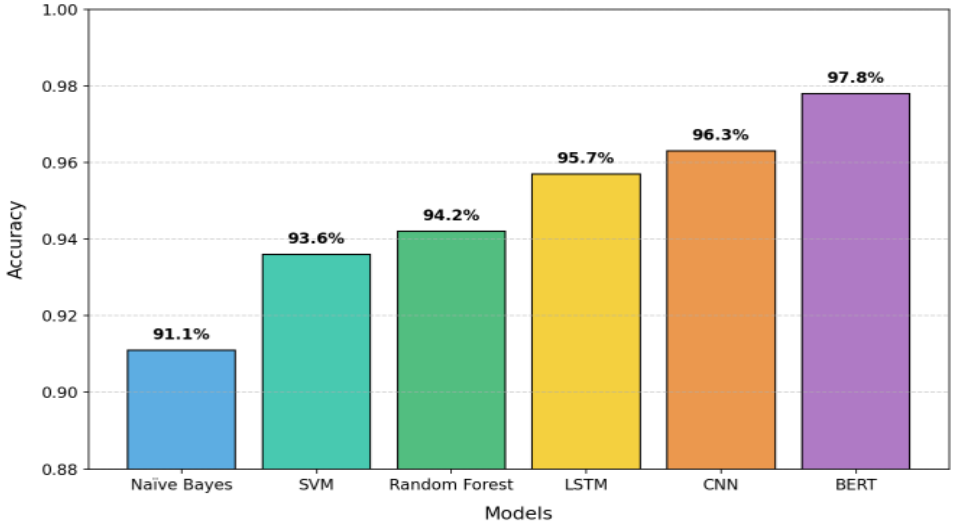


Figure 3: Accuracy comparison among models.

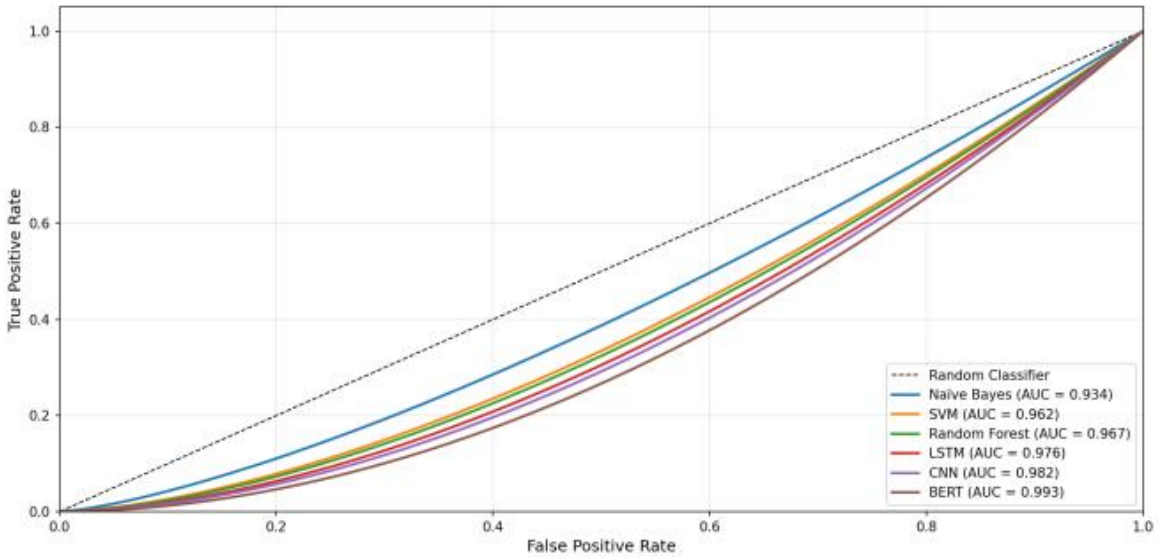


Figure 4: ROC curves for all models.

Table 5: Comparative performance of models.

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Naive Bayes	0.911	0.893	0.873	0.883	0.934
SVM	0.936	0.921	0.907	0.914	0.962
Random Forest	0.942	0.935	0.918	0.926	0.967
LSTM	0.957	0.949	0.935	0.942	0.976
CNN	0.963	0.953	0.947	0.950	0.982
BERT	0.978	0.972	0.965	0.968	0.993

In contrast, deep learning models demonstrate superior performance due to their ability to learn contextual representations automatically. BERT, in particular, shows the highest robustness against advanced spam techniques such as word manipulation and multilingual phishing content.

The integration of blockchain adds a new dimension to spam detection systems by introducing transparency, integrity, and auditability. While it introduces additional computational overhead, the trade-off is justified in scenarios where trust and traceability are critical requirements.

Overall, the results suggest that future spam detection systems should not rely solely on accuracy-oriented optimization but should also consider system-level properties such as security, transparency, and verifiability. A hybrid architecture combining deep learning models with lightweight blockchain-based auditing represents a promising direction for next-generation intelligent cybersecurity systems.

6 CONCLUSIONS

This study presented a comprehensive comparative framework for spam email detection based on traditional machine learning and deep learning models, enhanced with a blockchain-supported auditability mechanism. The proposed system was designed to evaluate not only classification performance but also transparency, integrity, and traceability of prediction results in modern digital environments.

Experimental results demonstrated that deep learning models significantly outperform traditional machine learning approaches in terms of classification accuracy and semantic understanding. Among all tested models, BERT achieved the highest performance (97.8% accuracy and 0.993 ROC-AUC), followed closely by CNN and LSTM. Traditional models such as Naive Bayes, SVM, and Random Forest showed competitive efficiency and faster execution but were limited in capturing complex contextual relationships in email content.

In addition to performance evaluation, the study introduced a blockchain-based auditability layer that records classification metadata and cryptographic hashes in an immutable ledger. This mechanism ensures traceability and tamper resistance of spam detection outcomes without interfering with the learning process. Although it introduces a moderate computational overhead, the trade-off is acceptable in security-sensitive and enterprise-level applications where accountability is critical.

The main contribution of this work is the integration of machine learning, deep learning, and blockchain technologies into a unified spam detection framework that balances accuracy, efficiency, and trustworthiness. The findings highlight that future intelligent spam filtering systems should move beyond purely performance-driven optimization and incorporate transparency and verifiability as core design principles.

Future work will focus on extending the framework to large-scale real-world deployments, optimizing blockchain consensus mechanisms for lower latency, and exploring federated learning approaches to further enhance privacy and scalability in distributed email security systems.

REFERENCES

[1] Kaspersky Lab, “Spam and Phishing in Q4 2024,” Technical Report, 2024.

[2] I. Androutsopoulos, J. Koutsias, K. V. Chandrinos and C. D. Spyropoulos, “An Experimental Comparison of Naive Bayesian and Keyword-Based Anti-Spam Filtering,” in Proceedings of ECML, 2000, pp. 9-17.

[3] T. Joachims, “Text Categorization with Support Vector Machines: Learning with Many Relevant Features,” in Proceedings of ECML, 1998, pp. 137-142.

[4] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in Proceedings of NAACL-HLT, 2019, pp. 4171-4186.

[5] Y. Kim, “Convolutional Neural Networks for Sentence Classification,” in Proceedings of EMNLP, 2014, pp. 1746-1751.

[6] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[7] H. Nguyen and A. Le, “RDRSegmenter: A Robust Vietnamese Word Segmentation Algorithm,” in Proceedings of PACLIC, 2016, pp. 265-272.

[8] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” in Proceedings of ICLR, 2015.

[9] D. Powers, “Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation,” Journal of Machine Learning Technologies, vol. 2, no. 1, pp. 37-63, 2011, [Online]. Available: https://doi.org/10.48550/arXiv.2010.16061.

[10] R. Kohavi, “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection,” in Proceedings of IJCAI, 1995, pp. 1137-1145.

[11] S. Raschka and V. Mirjalili, Python Machine Learning, 3rd ed. Birmingham, U.K.: Packt, 2019.

A Comparative Study of Noise Effects, VAD, and ASR Methods in Speech Recognition

Anastasiia Sapeha, Ibrahim Kovan, Subashkumar Rajanayagam, Kirill Karpov, Maksim Gering, Dmitry Kachan and Eduard Siemens

Anhalt University of Applied Sciences, Bernburger Str. 55, Koethen, Germany

{*anastasiia.sapeha, ibrahim.kovan, subashkumar.rajanayagam, kirill.karpov, maksim.gering, dmitry.kachan, eduard.siemens*}*@hs-anhalt.de*

Keywords: Speech Recognition, Noise-Robust Speech Processing, Voice Activity Detection, Automatic Speech Recognition, Word Error Rate, Speech Processing Pipeline, Real-Time Factor (RTF), Controlled Noise, GTA-5 Dataset, AMI Dataset, Factor Analysis.

Abstract: Automatic speech recognition (ASR) deployed in real environments is strongly affected by background noise and by upstream processing decisions such as noise suppression and voice activity detection (VAD). This paper provides a controlled, pipeline-level comparison of common front-end and segmentation choices and quantifies their joint impact on both transcription accuracy and computational efficiency. We evaluate a fixed three-stage pipeline consisting of noise suppression, VAD-based segmentation, and speech-to-text (ASR) transcription. Noise suppression is instantiated as either no suppression or DeepFilterNet; VAD is instantiated as none, Silero VAD, or WebRTC VAD; and ASR is instantiated as wav2vec 2.0 or faster-whisper (large-v3). In-vehicle noise is added to clean speech using a representative car noise profile and speech-active SNR calibration, covering multiple SNR levels and a clean reference condition. Experiments are conducted on datasets representing scripted utterances and meeting-style conversational speech. We report Word Error Rate (WER) and Real-Time Factor (RTF) for all 12 pipeline configurations across SNR conditions, and analyze main effects and interactions between pipeline components. Results highlight that ASR choice dominates the accuracy–efficiency trade-off, while noise suppression and VAD introduce dataset- and recognizer-dependent effects, including non-additive interactions that can change the sign and magnitude of preprocessing gains under different noise regimes.

1 INTRODUCTION

Automatic speech recognition (ASR) systems are increasingly deployed in real-world environments where background noise, reverberation, and non-speech acoustic events are unavoidable. Under such conditions, transcription accuracy can degrade substantially, making robustness to acoustic variability a practical requirement. Although noise robustness has been addressed through front-end enhancement, robust acoustic modeling, and noise-tolerant training strategies, overall performance in deployment is often determined by interactions between upstream processing decisions within an integrated speech processing pipeline.

In operational settings, ASR is frequently preceded by noise suppression and voice activity detection (VAD). VAD-based segmentation can reduce irrelevant non-speech input and computational load, but segmentation errors may truncate low-energy speech or distort boundaries, increasing deletions and substitutions and ultimately worsening Word Error Rate (WER). The joint impact of noise conditions, enhancement, segmentation, and recognition is therefore difficult to infer from component-level evaluation and is best assessed at the pipeline level.

In addition to recognition accuracy, computational efficiency is often critical in real-time and resource-limited deployments. Real-Time Factor (RTF) is adopted as an efficiency metric, defined as the ratio between processing time and audio duration, enabling direct comparison of different pipeline configurations in terms of real-time feasibility. Model initialization and startup latency are excluded from RTF to focus on steady-state inference performance.

A fixed three-stage pipeline is evaluated, in which noise suppression is followed by voice activity detection and speech-to-text transcription. The noise

suppression stage is instantiated as either no suppression or DeepFilterNet; the VAD stage as no VAD, Silero, or WebRTC; and the ASR stage as wav2vec2 or faster-whisper. Clean and noise-degraded conditions are evaluated under controlled in-vehicle noise, including a clean baseline processed by the same pipeline. Experiments are conducted on three speech datasets representing scripted speech, meeting conversations, and spontaneous dialogue. Recognition accuracy (WER) and computational efficiency (RTF) are reported for all method combinations, enabling comparison of accuracy–efficiency trade-offs under realistic noise conditions. Analyses are performed separately per dataset and language.

2 RELATED WORK

Robust automatic speech recognition (ASR) under noisy and mismatched acoustic conditions has been an active area of research for decades. Early studies addressed robustness primarily through signal processing techniques, noise estimation, and statistical acoustic modeling, often relying on accurate voice activity detection (VAD) to separate speech from background noise [1]. These works highlighted that errors in speech segmentation can significantly affect downstream recognition performance, particularly in non-stationary noise environments where reliable speech–non-speech discrimination is challenging.

Recent advances in deep learning have led to substantial improvements in ASR performance. Large-scale end-to-end and self-supervised models enable strong recognition accuracy across diverse datasets and acoustic conditions [2]. However, such models are typically evaluated on pre-segmented or cleanly prepared audio, while their behavior within realistic processing pipelines that include noisy inputs and imperfect segmentation remains less explored. Recent benchmarking efforts report considerable variability in Word Error Rate (WER) across datasets, noise conditions, and speaker characteristics, highlighting the need for controlled and reproducible evaluation protocols [3].

Noise suppression and speech enhancement are commonly employed as front-end components to improve intelligibility under adverse acoustic conditions. However, prior work has shown that gains in perceptual speech quality or signal-to-noise ratio do not necessarily translate into lower ASR error rates, and may even degrade recognition performance due to enhancement-induced speech distortion and mismatch with acoustic models [4]. These findings suggest that the impact of noise suppression should be evaluated in conjunction with downstream components, rather than in isolation.

Voice Activity Detection remains a critical intermediate stage in speech processing pipelines. Traditional VAD approaches based on energy thresholds or statistical divergence are computationally efficient but often degrade under low-SNR or non-stationary noise [1]. More recent neural VAD models demonstrate improved robustness by learning noise-invariant representations, including architectures designed for real-time or resource-constrained scenarios [5]. In addition, integrating noise suppression prior to VAD has been shown to improve frame-level VAD performance in degraded environments, indicating that segmentation quality itself is sensitive to front-end processing [6].

Despite improvements in VAD accuracy, multiple studies report that frame-level VAD metrics such as accuracy or AUROC correlate weakly with downstream ASR performance, particularly when segmentation boundary errors occur [7]. Analyses of streaming and API-based ASR systems further demonstrate that VAD failures may lead to the complete loss of user utterances, which can be more detrimental than transcription errors alone in interactive settings [8]. These observations highlight the importance of evaluating segmentation and recognition jointly within an integrated pipeline.

Beyond recognition accuracy, real-time and streaming ASR systems must also satisfy constraints on computational efficiency and processing latency. Several studies explicitly compare recognition quality and inference speed, revealing trade-offs between WER and end-to-end processing time [9]. These findings motivate treating runtime and latency as first-class evaluation metrics along with transcription accuracy.

While prior work often focuses on individual components such as noise suppression, VAD, or ASR models under fixed conditions, fewer studies systematically evaluate complete speech processing pipelines using multiple alternative methods at each stage. This work addresses this gap by analyzing a full speech processing pipeline under controlled noise conditions, comparing different combinations of noise characteristics, VAD algorithms, and speech-to-text systems, and evaluating their impact on both recognition accuracy and processing efficiency.

3 EXPERIMENTAL PIPELINE AND SETUP

Figure 1 provides an overview of the speech processing pipeline evaluated in this study. Clean speech recordings are evaluated both in a clean baseline condition and after transformation into noisy signals under controlled acoustic conditions. The resulting audio is then processed by a fixed sequence of components consisting of noise suppression (NS), voice activity detection (VAD), and automatic speech recognition (ASR), followed by quantitative evaluation. All combinations of pipeline components are evaluated.

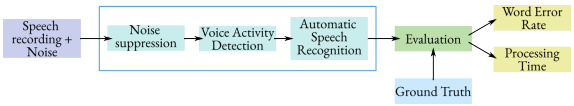


Figure 1: Overview of the evaluated speech processing pipeline.

Evaluated methods:

- Noise Suppression: `none`, `deepfilternet`
- VAD: `none`, `silero`, `webrtc`
- ASR: `wav2vec2`, `faster-whisper`

We selected the evaluated components to represent common, reproducible design choices in practical ASR pipelines while keeping the full-factorial grid compact. For noise suppression, we compare `none` against DeepFilterNet as a widely used, off-the-shelf neural speech enhancement method that can be integrated as a drop-in front end without task-specific training or tuning. This pairing provides a clear control condition and allows us to quantify when enhancement helps or harms downstream recognition under non-stationary in-vehicle noise, including potential distortion-induced mismatches.

For segmentation, we evaluate `none`, WebRTC VAD, and Silero VAD to cover three typical deployment regimes: no explicit segmentation, a lightweight real-time classical VAD (WebRTC) commonly used in streaming/RTC systems, and a popular neural VAD (Silero) often adopted for improved robustness under low SNR and conversational dynamics. For ASR, `faster-whisper` (large-v3) is used as the baseline recognizer due to its strong robustness in diverse acoustic conditions and its practical optimized inference implementation; `wav2vec2` is included as a representative self-supervised ASR alternative with different modeling assumptions and runtime characteristics. Together, these choices span realistic engineering trade-offs in accuracy, segmentation reliability, and computational efficiency while enabling systematic analysis of main effects and non-additive interactions across the $2 \times 3 \times 2$ pipeline grid.

All combinations are evaluated: 2 (denoise) $\times$ 3 (vad) $\times$ 2 (asr) $=$ 12 pipelines. Each pipeline is evaluated at each configured SNR value.

3.1 Datasets

The experimental evaluation was conducted on two speech corpora selected to cover scripted and conversational conditions and different recording setups: a GTA-V derived dataset [10] and the AMI Meeting Corpus [11]. To keep the computational budget comparable, approximately three hours of audio were used from each corpus; the exact file manifests and all data-preparation scripts were released in the accompanying repository.[1] Fixed subsets were curated manually and are fully specified via released manifests.

For the GTA-V corpus, we selected 3,886 subtitle-aligned scripted utterances, corresponding to approximately three hours of audio. The subset includes all non–cut-scene recordings, while files containing Chinese-language speech were excluded in order to maintain language consistency across the experiments. From the AMI Meeting Corpus, we used the headset-mix recordings, which amount to roughly 100 hours of spontaneous multi-speaker meeting audio. This dataset provides naturally occurring conversational speech recorded in realistic meeting scenarios.

Dataset-provided transcripts were normalized and were used as reference text for per-file WER computation; results were reported separately for each dataset (and language) to avoid cross-domain aggregation effects.

3.2 Noise Modeling

To simulate in-vehicle acoustic interference under controlled and reproducible conditions, background noise is added to the speech signals using a single representative in-car noise profile. Noise recordings were collected during car trips, resulting in more than 130 recordings. Candidate recordings were characterized using power spectral density (PSD) representations, and k-means clustering was applied to group acoustically similar recordings. The representative profile was selected from the largest cluster as the recording closest to the cluster centroid in feature space.

The selected profile is illustrated in Fig. 2 using its waveform and log-Mel spectrogram.

For each utterance, the representative noise recording is used as the noise signal $v[n]$. Length

[1] `github.com/AnastasiiaSapeha/ICAIIT_Audio_pipeline`

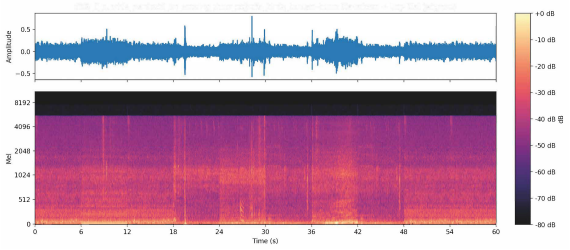


Figure 2: Waveform and log-Mel spectrogram of the selected in-vehicle noise profile.

matching to the clean speech $x[n]$ is achieved deterministically by truncation if the noise recording is longer than the utterance and by cyclic repetition followed by truncation if it is shorter. Additive mixing [12] is performed according to Eq. (1) to obtain the noisy mixture $y[n]$.

$$y[n] = x[n] + \alpha v[n]. \tag{1}$$

The target $\mathrm{SNR}_{\mathrm{dB}}$ is enforced over speech-active samples only to avoid bias from long silent regions. A binary speech-activity mask $m[n] \in \{0,1\}$ is defined for SNR calibration, where $m[n] = 1$ denotes speech-active samples. The mask $m[n]$ is obtained using an active speech level estimate (ITU-T P.56) and is used only for SNR calibration. The mean-square powers P_x and P_v are defined for clean speech and length-matched noise, respectively, and are computed only over speech-active samples by averaging $x[n]^2$ and $v[n]^2$ over indices with $m[n] = 1$. If speech-activity masking is disabled for SNR calibration, $m[n] \equiv 1$ is used and powers are computed over the full utterance.

The noise scaling factor α is computed as in Eq. (2), ensuring that the desired SNR is achieved on speech-active samples; active-speech-level-based scaling prior to additive mixing has been used in related speech enhancement evaluations [13].

$$\alpha = \sqrt{\frac{P_x}{P_v}}\, 10^{-\mathrm{SNR}_{\mathrm{dB}}/20}. \tag{2}$$

To evaluate robustness under varying noise conditions, SNR values (in dB) are selected on a quasi-logarithmic scale: $\{0, 2, 4, 7, 10, 14, 20, 70\}$.

3.3 Noise Suppression

Two noise suppression settings are evaluated: `deepfilternet` and `none`. In the none condition, the input signal is forwarded unchanged. DeepFilterNet is applied using deepfilternet 0.5.6. No task-specific tuning is performed.

By comparing pipelines with and without noise suppression at each SNR level, the study assesses whether front-end enhancement improves downstream segmentation and recognition accuracy, and how it affects computational efficiency.

3.4 Voice Activity Detection

Voice activity detection is applied to segment speech from non-speech regions prior to recognition in order to reduce the amount of irrelevant audio forwarded to the speech-to-text systems and to limit the impact of background noise and silence on downstream recognition performance.

Three VAD settings are evaluated: `none`, `silero`, and `webrtc`. In the none condition, no explicit segmentation is performed and the full input signal is forwarded to the ASR system. Silero VAD is a neural frame-level detector widely used in practical speech processing pipelines. Silero VAD is applied with a threshold of 0.35 and minimum speech/silence durations of 250/100 ms at 16 kHz. The WebRTC voice activity detector is included as a widely used signal-processing-based VAD originally developed for streaming and interactive speech communication systems and optimized for low-latency operation. WebRTC VAD is applied with aggressiveness 2 and 30 ms frames at 16 kHz.

The VAD stage produces time-aligned speech segments that are forwarded to the speech-to-text systems. VAD performance is not evaluated using standalone segmentation metrics; instead, its effect is assessed indirectly through its impact on downstream recognition accuracy and computational efficiency at the pipeline level.

3.5 Automatic Speech Recognition Systems

Speech recognition is performed using `wav2vec2` and `faster-whisper`. Faster-Whisper is applied using faster-whisper 1.1.0 with ctranslate2 4.6.0 and the model faster-whisper-large-v3.

wav2vec2 inference is performed using the Hugging Face stack. For English datasets, wav2vec2-base-960h is used. No additional language-model decoding or task-specific adaptation is applied.

All ASR systems operate on the segments produced by the VAD stage (or on the unsegmented signal when VAD is disabled) and generate textual hypotheses evaluated using WER and RTF.

3.6 Evaluation Metrics

Pipeline performance is evaluated using Word Error Rate (WER) and Real-Time Factor (RTF). WER mea-

sures recognition accuracy by comparing ASR outputs with ground-truth transcripts, while RTF characterizes computational efficiency relative to real-time processing.

Before computing recognition accuracy, reference (ground truth) and hypothesis (ASR output) texts are normalized to reduce the influence of formatting and orthographic variation. The normalization procedure includes conversion to lowercase, removal of punctuation symbols, and whitespace normalization.

After normalization, recognition accuracy is computed using Word Error Rate according to Equation 3, which measures the normalized number of word-level errors between a reference transcription and the corresponding ASR hypothesis.

$$\mathrm{WER} = \frac{S + D + I}{N}, \tag{3}$$

where S denotes substitutions, D deletions, I insertions, and N the number of words in the reference transcription. WER is computed per audio file after normalization using the `jiwer` 3.1.0 library, which performs word-level alignment based on Levenshtein distance.

Computational efficiency is evaluated using the Real-Time Factor for the end-to-end pipeline defined in Equation 4.

$$\mathrm{RTF} = \frac{T_{\text{processing}}}{T_{\text{audio}}}, \tag{4}$$

where $T_{\text{processing}}$ denotes the effective processing time and T_{audio} the duration of the corresponding input audio. RTF values below 1 indicate faster-than-real-time processing, whereas values above 1 indicate slower-than-real-time operation.

RTF is computed based on steady-state inference performance. Model initialization and startup latency are excluded from the reported values, as the evaluation focuses on continuous processing rather than one-time model loading overhead. The audio duration used for RTF computation corresponds to the original input signal before VAD-based segmentation, ensuring consistent comparison across different pipeline configurations.

3.7 Hardware and Execution Environment

All experiments were conducted on a single hardware platform to ensure consistent and reproducible performance measurements across all pipeline configurations. The experimental setup was based on an Intel NUC8i5BEH2 mini-PC equipped with an Intel Core i5-8259U processor, 32 GB of DDR4 memory, and a 512 GB NVMe solid-state drive.

Compute-intensive stages of the pipeline, including neural noise suppression and selected speech-to-text systems, were accelerated using an external NVIDIA GeForce RTX 3090 GPU with 24 GB of memory. All pipeline components were executed locally, and GPU acceleration was applied during inference.

4 RESULTS

ASR accuracy is evaluated under controlled addi-tive noise by sweeping SNR, and the averaged WER is reported for all combinations of NS, VAD, and ASR backend on AMI and GTA.

Fig. 3 shows that WER decreases monotonically with increasing SNR for all configurations. Across the full SNR range, faster-whisper consistently out-performs wav2vec2, with the largest gap at low SNR (0–7dB) and a persistent separation at higher SNR (14–20 dB), indicating that the ASR backend is the dominant factor for WER on AMI. The effect of NS is backend-dependent: for faster-whisper it is largely neutral or slightly detrimental, while wav2vec2 ben-efits more clearly from NS, particularly at low SNR. VAD has a comparatively minor influence on WER, as curves for different VAD variants largely overlap within each ASR/NS group.

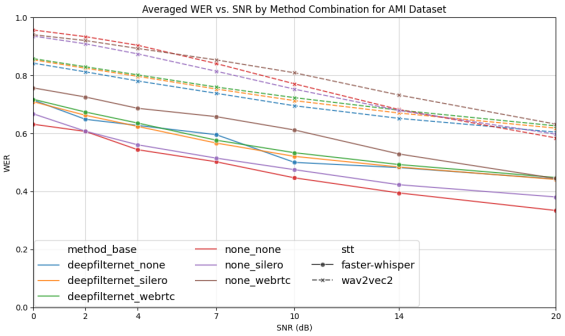


Figure 3: Average WER versus SNR on AMI for all NS/VAD/ASR configurations.

Fig.4 exhibits the same overall trend: WER improves with SNR, with most gains between 0 and 10dB and saturation toward higher SNR. Faster-whisper achieves the lowest WER for all SNR values and shows limited sensitivity to VAD, whereas wav2vec2 remains substantially worse despite improving with SNR. As on AMI, NS primarily benefits wav2vec2 (especially at low SNR), while its effect on faster-whisper is small.

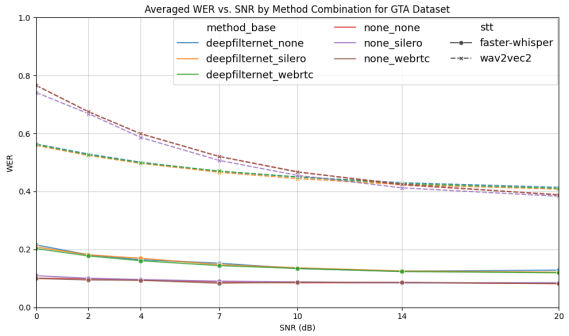


Figure 4: Average WER versus SNR on GTA for all NS/VAD/ASR configurations.

Since the real-time factor (RTF) shows negligible dependence on SNR in our experiments, runtime is analyzed separately using a factorial design over NS/VAD/ASR, without an SNR sweep.

4.1 Statistical Analysis

The main and interaction effects of the pipeline components are quantified using a linear mixed-effects factorial model with a random intercept per utterance. For WER, the following model is fitted:

$$\mathrm{WER}_i = \beta_0 + \mathbf{X}_i\beta + \beta_1\,\mathrm{SNR}_i + \beta_2\,\mathrm{SNR}_i^2 + u_{a(i)} + \varepsilon_i,$$

where X_i encodes the full factorial design over noise suppression (NS), VAD, and ASR backend (including all two- and three-way interactions), and also includes their interactions with SNR_i. The term $u_{a(i)} \sim \mathcal{N}(0,\sigma_u^2)$ is a random intercept for the audio segment a(i) to account for repeated measurements per utterance, and $\varepsilon_i \sim \mathcal{N}(0,\sigma^2)$ is the residual. Runtime is analyzed analogously using $\log(\mathrm{RTF})$ as the dependent variable.

Across both datasets, the mixed-effects factorial analysis shows that WER is dominated by the ASR backend: wav2vec2 yields markedly higher error rates than faster-whisper across configurations. Noise suppression (DeepFilterNet) exhibits a significant but ASR-dependent effect, being neutral to slightly adverse for faster-whisper while mitigating the elevated WER of wav2vec2 (NS$\times$ASR). SNR has a strong nonlinear influence on WER with diminishing gains at higher SNR, and significant ASR$\times$SNR and NS$\times$SNR interactions indicate that wav2vec2 benefits more from improving acoustic conditions, whereas denoising effectiveness decreases as SNR increases. Dataset-specific differences are observed for VAD: on GTA, VAD choice (Silero/WebRTC) does not produce measurable effects on WER, while on AMI VAD effects are statistically detectable (with WebRTC showing the largest degradation and Silero a smaller penalty), partially attenuated when wav2vec2 is used; higher-order interactions (including NS$\times$ASR$\times$VAD) further indicate non-additive coupling among pipeline components.

For runtime, the mixed-effects analysis of $\log(\mathrm{RTF})$ indicates that computational performance is primarily governed by the ASR backend and strong interactions with preprocessing. Noise suppression and VAD generally increase processing time, whereas wav2vec2 is faster in isolation; however, its speed advantage is substantially reduced when combined with DeepFilterNet and VAD, demonstrating that runtime cannot be modeled as an additive sum of module costs and must be assessed at the level of complete pipeline configurations. SNR plays a secondary role for runtime: on GTA it introduces a statistically significant but comparatively small nonlinear modulation with component-dependent interactions, while on AMI it shows no significant main effect and only weak interaction terms. Overall, accuracy is driven by ASR selection and SNR, with backend-dependent denoising effects and dataset-specific VAD sensitivity, whereas runtime is dominated by architectural configuration and its non-additive couplings across stages.

As a secondary confirmation of the mixed-effects factorial analysis, we additionally report partial effect sizes (η_p^2) from a classical factorial ANOVA. The motivation is twofold: (i) η_p^2 provides an intuitive, standardized summary of the relative importance of factors and interactions that is easy to compare across datasets and metrics, and (ii) it serves as a robustness check that the qualitative ranking of influential components is not an artifact of the mixed-effects specification. Since ANOVA assumes independent observations and our design contains repeated measurements per audio segment, this analysis is used for descriptive corroboration only; all inferential conclusions are based on the MixedLM models.

For WER (Fig.5), the effect-size ranking is consistent with the MixedLM findings: the ASR backend explains the largest share of variance for both datasets ($\eta_p^2 \approx 0.95$ on AMI and ≈ 0.99 on GTA). Noise suppression and its interaction with ASR (NS$\times$ASR) also contribute non-negligibly, supporting backend-dependent denoising effects. Dataset-specific differences are visible for VAD: its contribution is negligible on GTA ($\eta_p^2 \approx 0$) but noticeable on AMI ($\eta_p^2 \approx 0.33$), matching the MixedLM result that VAD effects are detectable primarily for AMI.

For runtime (RTF; Fig.6), partial η_p^2 values are uniformly high across ASR, VAD, NS, and their interactions, indicating that runtime is largely configuration-driven and exhibits strong non-additive

coupling between stages. This corroborates the mixed-effects runtime analysis, where large interaction terms imply that computational cost cannot be decomposed into independent module contributions.

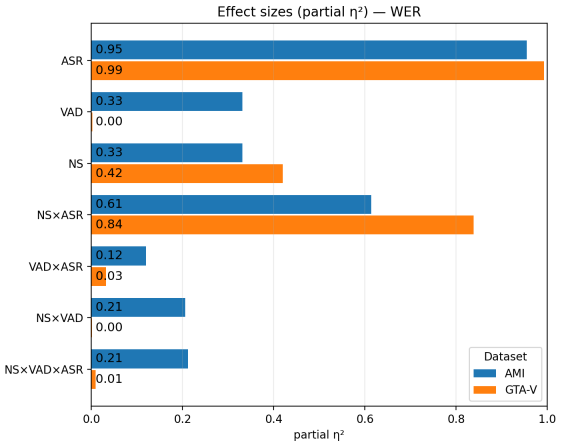


Figure 5: Partial effect sizes (η_p^2) from factorial ANOVA for WER across AMI and GTA-V.

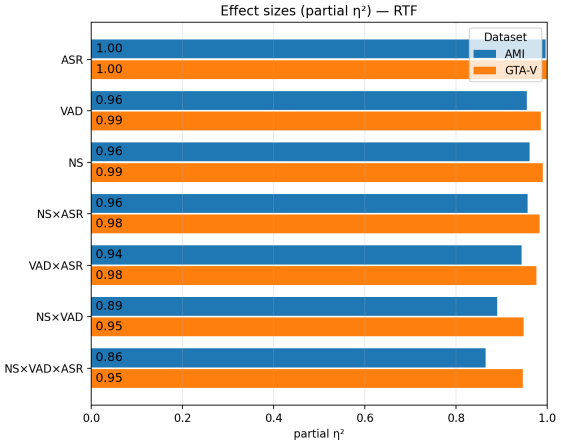


Figure 6: Partial effect sizes (η_p^2) from factorial ANOVA for RTF across AMI and GTA-V.

5 CONCLUSIONS

This paper presented a controlled, pipeline-level evaluation of a three-stage ASR chain comprising noise suppression (none vs. DeepFilterNet), VAD-based segmentation (none, Silero, WebRTC), and ASR backends (wav2vec 2.0, faster-whisper large-v3). Experiments were conducted under additive in-vehicle noise using a representative car-noise profile and speech-active SNR calibration over multiple SNR levels, while reporting both recognition accuracy (WER) and computational efficiency (RTF).

Across both datasets, the principal outcome is that ASR backend choice dominates the accuracy–efficiency trade-off. Faster-whisper consistently achieved lower WER, whereas wav2vec2 incurred higher error rates but offered lower computational cost when used without additional stages. Noise suppression and VAD introduced non-additive, model-dependent effects: denoising tended to benefit wav2vec2 under noisy conditions while being neutral to slightly adverse for faster-whisper, and VAD effects were dataset-dependent (negligible on GTA, but statistically detectable on AMI). The mixed-effects factorial analysis further revealed strong interactions between pipeline stages, demonstrating that preprocessing gains cannot be treated as independent and should be assessed at the level of complete pipeline configurations.

Regarding the evaluation corpora, AMI represents meeting-style conversational speech, whereas GTA provides scripted, studio-recorded utterances. Despite these differences, we observed consistent qualitative dynamics across datasets: WER improved nonlinearly with SNR, the relative ordering of ASR backends remained stable, and backend-dependent denoising effects were reproduced. At the same time, the strength of VAD-related effects differed across datasets. Overall, this cross-dataset agreement supports using GTA-style scripted, high-quality recordings as a practical benchmark for controlled pipeline studies, provided that additive noise is calibrated in a reproducible manner.

As a practical guideline, the best accuracy-first configuration is obtained by selecting faster-whisper and avoiding unnecessary preprocessing stages: DeepFilterNet and VAD do not yield consistent WER gains for faster-whisper and may increase runtime. Conversely, the best efficiency-first configuration is obtained by using wav2vec2 with minimal preprocessing, while enabling DeepFilterNet when operating at low SNR, where it provides the most consistent accuracy improvements for wav2vec2.

Future work. The evaluation uses a single representative in-car noise profile and fixed off-the-shelf component settings; results may vary under different noise types, domains, or hardware constraints. In addition, the number of long-form recordings is limited for some datasets, which constrains statistical stability. Future work will extend the benchmark to multiple noise profiles and environments, include additional recognizers and streaming-oriented latency metrics, and analyze segmentation quality with dedicated boundary/error measures alongside WER and RTF.

6 ACKNOWLEDGMENTS

This work was supported by the European Regional Development Fund (ERDF/EFRE) and the State of Saxony-Anhalt within the programme *Sachsen-Anhalt WISSENSCHAFT Forschung und Innovation (EFRE) 2021–2027*, project ReSeDiUm (grant no. ZS/2023/12/182669).

We acknowledge support by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) and the Open Access Publishing Fund of Anhalt University of Applied Sciences.

REFERENCES

[1] J. Ramírez, J. C. Segura, C. Benítez, Á. de la Torre, and A. Rubio, "A new voice activity detector using subband order-statistics filters for robust speech recognition," in ICASSP, 2004.

[2] A. Baevski, H. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A framework for self-supervised learning of speech representations," in NeurIPS, 2020.

[3] K. Kuhn, V. Kersken, B. Reuter, N. Egger, and G. Zimmermann, "Measuring the accuracy of automatic speech recognition solutions," ACM Transactions on Accessible Computing, vol. 16, 12 2023.

[4] S. Braun and H. Gamper, "Effect of noise suppression losses on speech distortion and asr performance," in ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022, pp. 996–1000.

[5] J. H. Ko, J. Fromm, M. Philipose, I. Tashev, and S. Zarar, "Limiting numerical precision of neural networks to achieve real-time voice activity detection," in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 2236–2240.

[6] M. R. Prasad, S. B. Gowda, M. B. Talawar, and N. Jagadisha, "Integrated noise suppression techniques for enhancing voice activity detection in degraded environments," International Journal of Speech Technology, vol. 27, pp. 987–995, 2024.

[7] S. Tong, N. Chen, Y. Qian, and K. Yu, "Evaluating vad for automatic speech recognition," in 2014 12th International Conference on Signal Processing (ICSP), 2014, pp. 2308–2314.

[8] K. Yamamoto, R. Takeda, and K. Komatani, "Analysis of voice activity detection errors in API-based streaming ASR for human-robot dialogue," in Proceedings of the 15th International Workshop on Spoken Dialogue Systems Technology. Bilbao, Spain: Association for Computational Linguistics, May 2025, pp. 245–253. [Online]. Available: https://aclanthology.org/2025.iwsds-1.26/

[9] C. Arriaga, A. Pozo, J. Conde, and A. Alonso, "Assessing latency in asr systems: A methodological perspective for real-time use," 2025. [Online]. Available: https://arxiv.org/abs/2409.05674

[10] A. Sapeha, E. Sariiev, M. Sapeha, I. Kovan, S. Rajanayagam, K. Karpov, M. Gering, D. Kachan, and E. Siemens, "GTA-NarrativeTraj: Language-aware trajectory prediction from GPS and dialogue in an open-world simulator," in Proc. Int. Conf. Appl. Innov. IT, vol. 13, no. 5, pp. 193–199, doi: 10.25673/122853.

[11] J. Carletta, "Unleashing the killer corpus: experiences in creating the multi-everything ami meeting corpus," Language Resources and Evaluation, vol. 41, no. 2, pp. 181–190, 2007.

[12] D. Orel and H. A. Varol, "Noise-robust automatic speech recognition for industrial and urban environments," in IECON 2023- 49th Annual Conference of the IEEE Industrial Electronics Society, 2023, pp. 1–6.

[13] M. Kolbæk, Z.-H. Tan, S. H. Jensen, and J. Jensen, "On loss functions for supervised monaural time-domain speech enhancement," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 825–838, 2020.

Investigating Multi-Camera Tracking and Re-Identification Pipelines Based on YOLOv8

Boshen Ruan, Beifei Wei and Stefan Twieg
Institut für Medizin und Technik e.V., Bernburger Str. 57, 06366 Koethen, Germany
Boshen.Ruan@student.hs-anhalt.de, Beifei.Wei@student.hs-anhalt.de, Stefan.Twieg@hs-anhalt.de

Keywords: Deep Learning, YOLO, ByteTrack, BoT-SORT, Object Detection, Multi-Camera Tracking, Re-Identification, Machine Learning, Animal.

Abstract: Multi-camera identification and tracking of individual animals is a challenging task in complex monitoring environments due to frequent occlusions, viewpoint changes, and visual similarity between individuals. This challenge is particularly relevant for future animal monitoring scenarios, such as goose tracking in agricultural settings. However, collecting large-scale, well-annotated multi-camera animal datasets remains difficult. In this work, we investigate a conceptual multi-camera tracking and re-identification (ReID) framework based on modern detection, tracking, and appearance-based re-identification methods. The proposed pipeline integrates YOLOv8 for object detection with ByteTrack and BoT-SORT for single-camera multi-object tracking, generating stable tracklets as the basis for cross-camera identity association. To evaluate the feasibility of the re-identification component, experiments are conducted on publicly available human benchmark datasets, which are widely used for appearance-based ReID research. The results demonstrate that the proposed modular pipeline can reliably perform detection, tracking, and identity embedding extraction in multi-camera scenarios. Furthermore, we discuss the transferability of the framework to animal monitoring applications and analyze the limitations arising from domain gaps between human and animal data. The presented study serves as a proof-of-concept and provides a practical foundation for future deployment in real-world goose monitoring scenarios.

1 INTRODUCTION

Continuous monitoring of individuals in complex environments is an important task in applications such as smart farming and automated surveillance [1]. Multi-camera systems are commonly used to cover large areas and reduce blind spots that cannot be handled by single-camera setups [2]. However, maintaining consistent identities over time and across different camera views remains challenging due to occlusions, viewpoint changes, and visually similar objects [3].

In animal monitoring scenarios, including future goose husbandry applications [4], these challenges are further intensified. Animals often exhibit similar appearances and frequent interactions, making reliable long-term identity tracking difficult. Manual monitoring is inefficient, while single-camera systems easily lose identities when objects leave the field of view.

To address these issues, this paper investigates a conceptual multi-camera tracking and re-identification framework based on modern deep learning methods [5]. The proposed pipeline combines YOLOv8 for object detection with ByteTrack [6], [7] and BoT-SORT [8] for single-camera multi-object tracking, producing stable tracklets for further identity association [9].

Due to the limited availability of annotated multi-camera animal datasets, the re-identification component is evaluated using well-known human benchmark datasets, including Market1501 and DukeMTMC [10], [11]. This allows a controlled validation of the framework before its application to animal monitoring. In this study, the term conceptual framework refers to a modular pipeline architecture that integrates object detection, single-camera tracking, tracklet generation, and multi-camera identity association. The primary objective of this work is to establish a practical and modular framework that can be adapted to real-world goose tracking scenarios in future studies.

Therefore, the main contribution of this work is not the development of a fundamentally new tracking

algorithm, but the engineering integration and experimental analysis of a modular multi-camera tracking pipeline. By evaluating YOLOv8-based detection, ByteTrack- and BoT-SORT-based tracking, and geometry- and appearance-based cross-camera identity association under a unified workflow, this study provides practical insights into the strengths and limitations of current multi-camera tracking and re-identification methods in dense scenes.

2 THEORETICAL FOUNDATIONS

2.1 Object Detection in Agriculture

Deep learning-based object detection has been widely applied in agricultural and livestock monitoring scenarios to automate the observation of animals and farming environments. In recent years, YOLO-based models have gained significant attention due to their real-time performance and competitive detection accuracy [1]. Several studies have demonstrated the effectiveness of YOLO variants for detecting livestock such as cattle, pigs, and poultry in complex farm environments, making them suitable candidates for practical monitoring systems [12]. While two-stage detectors (e.g., Faster R-CNN, Mask R-CNN), anchor-free detectors (e.g., FCOS, CenterNet), and transformer-based models (e.g., DETR, Deformable DETR) are also applicable to agricultural object detection tasks, YOLOv8 is selected in this study. Although YOLOv8 may experience performance degradation in scenarios with heavy object overlap or when distinguishing visually similar animals, it provides superior real-time performance, seamless integration with tracking algorithms such as ByteTrack and BoT-SORT, and a stable baseline supported by a large and active community.

2.2 Animal Re-Identification

Animal re-identification aims at recognizing individual animals across time and different viewpoints [13]. Existing approaches can be broadly categorized into biometric-based methods, which rely on natural appearance features, and marker-based methods, which utilize artificial identifiers such as tags or collars. While biometric approaches have shown promising results in controlled settings, their performance often degrades in real-world livestock scenarios due to strong appearance similarity between individuals. Marker-based identification, when available, can provide more reliable identity cues but requires additional infrastructure and data acquisition [13]. Alternative approaches for individual recognition, including trajectory-based features, gait-based features, spatio-temporal behavior signatures, and multi-modal sensing, have demonstrated promising potential. However, these methods are not considered in the present study due to the absence of specialized annotated datasets for animal gait analysis and the high system complexity associated with multi-modal hardware deployment. For visual feature extraction, established software libraries, namely FastReID and the OpenAI implementation of CLIP, are utilized.

2.3 Multi-Object Tracking in Livestock Monitoring

Multi-object tracking (MOT) plays a crucial role in maintaining consistent identities over time [14]. Recent tracking-by-detection methods, such as ByteTrack [6] and BoT-SORT [3], have achieved strong performance in crowded scenes by effectively associating detection results into stable trajectories. Although these methods were primarily developed and evaluated on human-centered datasets, their robustness and modular design make them attractive candidates for exploratory studies in animal monitoring and livestock tracking scenarios [15]. ByteTrack and BoT-SORT are selected instead of alternatives such as DeepSORT or StrongSORT. The latter relies heavily on appearance features for data association, which can be unreliable in livestock monitoring scenarios due to the high visual similarity between individuals. In contrast, ByteTrack and BoT-SORT emphasize motion consistency and high-confidence detection results, enabling more stable and continuous trajectories.

Table 1: Comparison of evaluated multi-camera identity association strategies.

Strategy	Features	Limitation
Geometry-based	Ground-plane coords	Calibration, density
Appearance-based	Visual embeddings	Viewpoint, similarity
Semantic-based	CLIP embeddings	Noise in dense scenes

Table 1 summarizes the key characteristics of the evaluated multi-camera identity association strategies under a unified detection and tracking setup, highlighting the trade-offs between geometry-based,

appearance-based, and semantic-based approaches in dense livestock monitoring scenarios.

2.4 Proposed Conceptual Pipeline for Multi-Camera Identity Tracking

The proposed multi-camera tracking and identity association pipeline follows a modular structure, as illustrated in Figure 1. The process starts with synchronized multi-camera video input, in which each camera stream is processed independently. Single-camera object detection is first performed using YOLOv8, followed by single-camera tracking with ByteTrack or BoT-SORT to link detections over time and generate continuous trajectories within each view. These trajectories are then temporally aggregated into tracklets, which provide more stable units for subsequent cross-camera analysis.

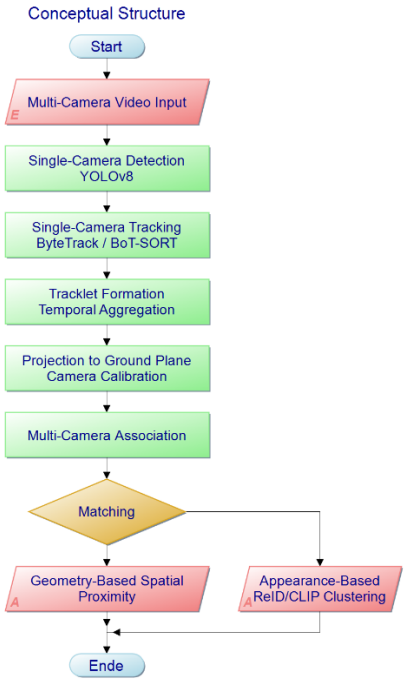


Figure 1: Conceptual pipeline for multi-camera object tracking and identity association.

The resulting tracklets are projected onto a shared ground-plane coordinate system using camera calibration to enable spatial alignment across different camera views. Based on these aligned representations, multi-camera identity association is performed by matching tracklets across cameras, either through geometry-based spatial proximity or through appearance-based clustering using ReID or CLIP embeddings.

3 PRACTICAL IMPLEMENTATION AND RESULTS

3.1 Dataset and Experimental Setup

To validate the complete multi-camera tracking and identification pipeline, the Wildtrack dataset is employed as a proof-of-concept benchmark. Wildtrack provides synchronized multi-camera videos together with precise ground-truth annotations and calibrated camera parameters, making it suitable for evaluating multi-camera tracking workflows.

In our experiments, videos from Cameras 1, 3, and 5 are selected to represent different viewpoints and spatial configurations. The dataset is primarily used to verify the feasibility of the proposed pipeline rather than to achieve state-of-the-art performance.

3.2 Detection and Single-Camera Tracking

In the first stage, the original annotations provided by the Wildtrack dataset are used to train a YOLOv8-based object detector [7]. The training process converges steadily, and the detector achieves high detection performance on the validation set. As shown in Figure 2, the mAP@0.5 increases rapidly during training and stabilizes at a high value, indicating reliable localization accuracy for pedestrian detection in crowded scenes.

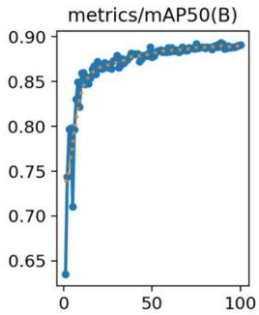


Figure 2: Evolution of mAP50 on the validation set during training.

In addition to quantitative metrics, qualitative detection results further confirm the robustness of the trained detector. Figure 3 present representative detection outputs on Wildtrack video frames, where pedestrians are accurately localized despite strong perspective variations and high object density. These results demonstrate that the trained YOLOv8 model

provides sufficiently reliable detections as input for subsequent tracking and identity association stages.

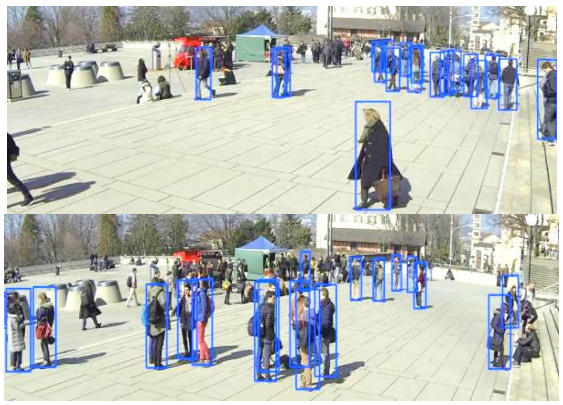


Figure 3: Qualitative detection results of the trained YOLOv8 detector on the Wildtrack dataset.

Subsequently, the trained detector is applied to reprocess the Wildtrack videos from Camera 1, Camera 3, and Camera 5. Detection outputs are integrated with a tracking-by-detection framework to generate videos with persistent track IDs and corresponding label files [6]. Visual inspection of the tracking results shows that identities remain consistent over consecutive frames within a single camera view. As illustrated in Figure 4, tracked individuals preserve stable track IDs even under moderate viewpoint changes and partial occlusions.



Figure 4: Single-camera tracking results on the Wildtrack dataset.

Overall, both quantitative detection metrics and qualitative visualization results confirm that the proposed detection and single-camera tracking pipeline operates reliably. This stable performance provides a solid foundation for the subsequent investigation of cross-camera identity association and re-identification.

3.3 Geometry-Based Multi-Camera Association

To explore multi-camera identity association without appearance-based features, the generated label files are first converted from text format to CSV files. Using the intrinsic and extrinsic camera parameters provided by Wildtrack, detections are projected onto a common ground-plane coordinate system.

Based on this representation, identities across different cameras are associated by comparing the spatial proximity of individuals within a manually defined temporal window. Additional constraints, including mutual nearest-neighbor matching and multi-frame consistency, are introduced to reduce incorrect associations.

Although the number of overlapping frame pairs is abundant (over 540,710 pairs with $\geqslant$10 frames overlap, see Table 2), no tracklets were actually merged during the process. The final number of global IDs equals 29,478, almost identical to the input tracklets (Table 2). This indicates that geometric information alone is insufficient to resolve identities in dense and occluded scenarios, as present in the Wildtrack dataset.

Table 2: Results of geometry-based multi-camera association.

Metric	Value
Number of valid tracklets	27,512
Number of candidates tracklets for multi-camera association	27,512
Number of merge operations (union)	0
Total number of tracklet pairs across cameras	242,326,032
Tracklet pairs with ≥1 frame overlap	789,754
Tracklet pairs with ≥3 frames overlap	763,743
Tracklet pairs with ≥10 frames overlap	540,710
Final number of global IDs	29,478

3.4 Re-Identification and Embedding-Based Association

To further improve cross-camera identity association, appearance-based re-identification (ReID) is integrated into the pipeline. ReID models are trained using Market1501 and DukeMTMC to learn discriminative feature embeddings. Using the YOLOv8 detection results, video frames are cropped to generate person image patches. For each track, embeddings are extracted using a trained FastReID model and aggregated via track-level mean pooling. Subsequently, pairwise cosine distances between track embeddings are computed and the DBSCAN

clustering algorithm is applied to assign global identities across cameras.

The intrinsic capability of the ReID embeddings is first validated using standard ReID evaluation metrics. As shown in Figure 5, the model trained/evaluated on Market1501 achieves consistently high retrieval performance (approximately Rank-1 ≈ 95%, Rank-5 ≈ 99%, Rank-10 ≈ 100%), indicating strong top-k identification accuracy on that benchmark. In contrast, the DukeMTMC setting shows noticeably lower performance (approximately Rank-1 ≈ 40%, Rank-5 ≈ 55%, Rank-10 ≈ 62%), suggesting that appearance-based discrimination becomes significantly harder under that dataset's camera conditions.

Beyond rank-based metrics, mAP and mINP provide additional insight into overall ranking quality and hard-case robustness. In Figure 5, Market1501 maintains a high mAP (≈ 90%) and a relatively strong mINP (≈ 65%), which implies that correct matches tend to appear early in the ranked list even for more challenging identities. Conversely, DukeMTMC exhibits much lower mAP (≈ 25%) and near-zero mINP (≈ 5%), indicating frequent failure cases where true matches are ranked far down the list. This gap suggests substantial sensitivity of appearance embeddings to viewpoint, illumination, and background changes.

In practice, only 6 merge operations occurred (Table 3), and the final number of global IDs was 29,472, almost equal to the input tracklets. The average cosine similarity of 0.755 indicates high overlap between embeddings of different individuals, resulting in poor clustering performance. This demonstrates that ReID embeddings alone are insufficient for robust identity matching in dense, occluded scenarios.

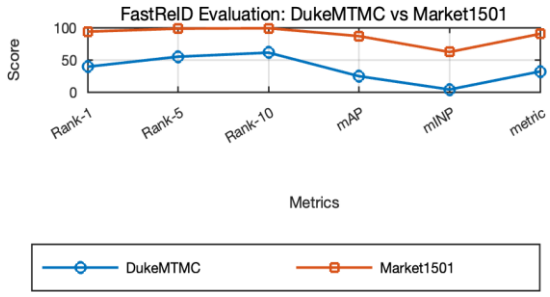


Figure 5: Comparison of ReID models trained on DukeMTMC and Market1501.

Table 3: Results of ReID- and embedding-based association.

Metric	Value
Original number of tracklets	29,478
Number of valid tracklets (sufficient frames + embedding)	27,512
Tracklets with ground coordinates	27,512
Number of candidate tracklets for multi-camera association	27,512
Number of merge operations (union)	6
Cosine similarity range	0.629 – 0.904
Average cosine similarity	0.755
Final number of global IDs	29,472

3.5 Appearance-Based Identity Matching using CLIP

CLIP-based association [16] uses pre-trained visual embeddings for each tracked individual and assigns global IDs via DBSCAN clustering [17]. Track-level embeddings are computed by averaging frame-wise embeddings, and pairwise cosine distances are used for clustering. The final number of global IDs is 245, with 3,739 tracklets marked as noise (Table 4). Although CLIP provides a stronger visual representation than ReID, clustering remains unstable in dense and occluded scenarios [3], [18]. This suggests that relying solely on appearance features is insufficient. Compared to geometric and ReID methods, CLIP shows some improvement but still requires additional strategies such as trajectory-based association, motion consistency, or gait features for robust multi-camera tracking [19], [20].

Table 4: Results of CLIP-based visual identity matching.

Metric	Value
Total number of tracklets	26,283
Final number of global IDs	245
Tracklets marked as noise (-1)	3,739
DBSCAN parameter eps	0.035
DBSCAN parameter min_samples	6

3.6 Comparative Analysis and Observations

The results obtained from both FastReID-based and CLIP-based appearance matching approaches reveal similar limitations in dense multi-camera scenarios. In both cases, the global identity assignment is highly sensitive to clustering parameters, resulting either in identity fragmentation-where the same individual is assigned multiple global identities-or identity merging, where multiple individuals are incorrectly grouped [3].

Despite the stronger representation capability of CLIP, no consistent improvement over traditional ReID embeddings is observed [16]. These findings indicate that appearance-based features alone are insufficient to reliably resolve cross-camera identity association under heavy occlusion, strong viewpoint variation, and crowded conditions, as exemplified by the Wildtrack dataset [18].

These findings highlight the need to investigate alternative multi-camera collaboration strategies, such as trajectory-based association, motion consistency, or gait-related features [19], which may offer better robustness than appearance-based methods in animal monitoring contexts.

It is important to note that Table 3, which reports tracklet-level statistics for ReID, and Table 4, which summarizes clustering outcomes for CLIP, are based on fundamentally different evaluation protocols. Specifically, FastReID is assessed using retrieval-based metrics, including Rank-1/5/10 accuracy and mAP, whereas CLIP is evaluated using clustering statistics. Consequently, these tables should be interpreted as diagnostic indicators rather than as results intended for direct quantitative comparison. In the absence of unified multi-object tracking metrics (e.g., IDF1 or MOTA), the analysis is therefore limited to a qualitative assessment of the feasibility of the respective methods.

4 DISCUSSION AND FUTURE WORK

4.1 Environment-Related Challenges in Real-World Animal Monitoring

The experimental results clearly reveal both the strengths and limitations of the proposed multi-camera tracking and identity association framework. While object detection and single-camera tracking achieve stable performance, cross-camera identity association remains a major challenge, especially in dense scenes [9]. Experiments on the Wildtrack dataset demonstrate that under conditions of high target density and frequent occlusions, both geometry-based methods and appearance-based re-identification struggle to maintain consistent identities.

A key challenge arises from occlusion handling. When multiple animals overlap or move in proximity, identity switches frequently occur, and simple spatial or appearance-based cues are insufficient to preserve identity continuity [3]. Addressing such cases requires more robust occlusion-aware strategies, including long-term trajectory reasoning, temporal consistency constraints, and identity recovery mechanisms after occlusion events.

A further limitation is the domain gap between human-centered benchmark datasets and animal monitoring scenarios. The Wildtrack dataset contains pedestrians with upright body structures, relatively stable shapes, and more distinguishable clothing-based appearance cues. In contrast, geese and other livestock animals often have highly similar visual appearances, non-rigid body poses, smaller visible body regions, and frequent physical interactions. In addition, animal monitoring environments are affected by dirt, changing illumination, partial occlusions, and camera viewpoints close to the ground. As a result, ReID models trained or validated on human datasets cannot be directly transferred to animal tracking without further adaptation.

In addition, real-world animal monitoring environments introduce challenges that are not fully represented in existing human-centered datasets. Factors such as environmental contamination, uneven ground surfaces, and unfavorable lighting conditions inside barns or livestock facilities can significantly degrade detection and tracking performance. These environmental factors further amplify identity ambiguities and reduce the reliability of appearance-based representations. Overcoming these challenges will require dedicated animal-specific data collection and the incorporation of illumination-robust feature representations and adaptive preprocessing techniques.

4.2 Beyond Appearance-Based Identity Cues

Given the limitations observed in appearance-based re-identification, it is necessary to explore alternative identity cues beyond visual appearance. In particular, trajectory-based and gait-based information represent promising directions for multi-camera identity association in animal monitoring scenarios.

Compared to appearance features, motion-related cues rely less on texture, color, and fine-grained visual details, which are often unstable in real-world animal monitoring environments. For animals such as geese, appearance-based ReID is especially challenging due to the high visual similarity between individuals and the limited availability of distinctive visual traits. Environmental factors such as dirt, mud attachment, and lighting variations further reduce the reliability of appearance-based representations.

In contrast, trajectory-based features, including movement direction, speed patterns, and temporal-spatial consistency, can provide more stable identity cues across multiple camera views [3], [21]. Such features capture behavioral regularities over time rather than instantaneous visual similarity. Similarly, gait-based features focus on periodic motion patterns instead of static appearance, making them potentially more robust to occlusion and viewpoint changes.

Although gait recognition has been extensively studied in human identification, its application in animal monitoring remains limited and largely unexplored. For geese, relatively consistent walking rhythms and motion dynamics may form implicit identity signatures when observed over extended time periods. Therefore, integrating gait- and trajectory-based cues with existing detection and tracking pipelines represents an important direction for future research.

Overall, these findings suggest that relying solely on appearance-based ReID is insufficient for reliable identity association in animal monitoring. Future work should focus on combining motion-based cues with detection and tracking frameworks, and, where applicable, fusing weak appearance information to achieve more robust multi-camera identity tracking in livestock environments.

For future adaptation to goose monitoring, several concrete steps are required. First, animal-specific multi-camera datasets should be collected and manually annotated under different lighting conditions, camera viewpoints, and density levels. Second, the YOLOv8 detector should be fine-tuned on goose images to improve the detection of small, occluded, and overexposed animals. Third, cross-camera association should combine multiple cues, including calibrated ground-plane positions, temporal movement consistency, trajectory direction, gait-related motion patterns, and, where available, external identity markers such as neck collars. Fourth, appearance-based ReID should not be used as the only identity cue, but rather as a weak complementary feature together with motion-based and scene-geometric information. These steps would allow the proposed pipeline to move from a proof-of-concept on human benchmark data toward a practical multi-camera animal monitoring system.

5 CONCLUSIONS

This paper presents a conceptual multi-camera tracking and identification framework and evaluates its feasibility using the Wildtrack dataset. By integrating YOLOv8-based detection with modern tracking methods and cross-camera association strategies, the complete processing pipeline is successfully implemented and validated in practice.

Experimental results show that object detection and single-camera tracking achieve stable performance in crowded scenes. However, the study also demonstrates that both geometry-based association and appearance-based re-identification face significant limitations in dense multi-camera environments. These findings highlight the challenges of maintaining consistent identities under occlusions and complex scene conditions.

Overall, this work provides valuable insights into the strengths and weaknesses of current multi-camera collaboration approaches. Rather than delivering a finalized identification system, the presented framework serves as a practical baseline and reference for future research. The conclusions drawn from this study will guide further development toward more robust multi-camera animal monitoring systems, including potential applications in goose tracking and smart farming.

ACKNOWLEDGMENTS

The research project underlying this contribution (NEXON) was funded by the European Regional Development Fund (EFRE). The responsibility for the content of this publication lies with the author.

REFERENCES

[1] D. Wang, W. Cao, F. Zhang, Z. Li, S. Xu, and X. Wu, “A Review of Deep Learning in Multiscale Agricultural Sensing,” Remote Sens., vol. 14, no. 3, p. 559, Jan. 2022, [Online]. Available: https://doi.org/10.3390/rs14030559.

[2] T. I. Amosa et al., “Multi-camera multi-object tracking: A review of current trends and future advances,” Neurocomputing, vol. 552, p. 126558, Oct. 2023, [Online]. Available: https://doi.org/10.1016/j.neucom.2023.126558.

[3] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, and T.-K. Kim, “Multiple object tracking: A literature review,” Artif. Intell., vol. 293, p. 103448, Apr. 2021, [Online]. Available: https://doi.org/10.1016/j.artint.2020.103448.

[4] R. A. Rajagukguk et al., “Deep learning for visual animal monitoring (detection, tracking, pose estimation, and behavior classification): A comprehensive review,” Smart Agric. Technol., vol. 12, p. 101539, Dec. 2025, [Online]. Available: https://doi.org/10.1016/j.atech.2025.101539.

[5] M. A. Ingrisch, R. M. Schilling, I. Chmielewski, and S. Twieg, “Determining the Optimal T-Value for the Temperature Scaling Calibration Method Using the Open-Vocabulary Detection Model YOLO-World,” Appl. Sci., vol. 15, no. 22, p. 12062, Jan. 2025, [Online]. Available: https://doi.org/10.3390/app152212062.

[6] Y. Zhang et al., “ByteTrack: Multi-object Tracking by Associating Every Detection Box,” in Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part XXII, Berlin, Heidelberg: Springer-Verlag, Oct. 2022, pp. 1-21, [Online]. Available: https://doi.org/10.1007/978-3-031-20047-2_1.

[7] L. You, Y. Chen, C. Xiao, C. Sun, and R. Li, “Multi-Object Vehicle Detection and Tracking Algorithm Based on Improved YOLOv8 and ByteTrack,” Electronics, vol. 13, no. 15, p. 3033, Jan. 2024, [Online]. Available: https://doi.org/10.3390/electronics13153033.

[8] J. Zhao and J. Chen, “YOLOv8 Detection and Improved BOT-SORT Tracking Algorithm for Iron Ladles,” in Proceedings of the 2024 7th International Conference on Image and Graphics Processing, in ICIGP ’24, New York, NY, USA: Association for Computing Machinery, May 2024, pp. 409-415, [Online]. Available: https://doi.org/10.1145/3647649.3647713.

[9] E. Ristani and C. Tomasi, “Features for Multi-target Multi-camera Tracking and Re-identification,” in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT: IEEE, Jun. 2018, pp. 6036-6046, [Online]. Available: https://doi.org/10.1109/CVPR.2018.00632.

[10] Z. Tian, S. Chen, D.-H. Wang, and J. Lu, “A Survey of Person Re-identification Based on Deep Learning,” in Proceedings of the 2021 10th International Conference on Computing and Pattern Recognition, in ICCPR ’21, New York, NY, USA: Association for Computing Machinery, Feb. 2022, pp. 36-42, [Online]. Available: https://doi.org/10.1145/3497623.3497630.

[11] Z. Hao, H. Ge, and J. Huang, “Research on an unsupervised person re-identification based on image quality enhancement method,” Eng. Appl. Artif. Intell., vol. 123, p. 106392, Aug. 2023, [Online]. Available: https://doi.org/10.1016/j.engappai.2023.106392.

[12] Z. Zheng, J. Li, and L. Qin, “YOLO-BYTE: An efficient multi-object tracking algorithm for automatic monitoring of dairy cows,” Comput. Electron. Agric., vol. 209, p. 107857, Jun. 2023, [Online]. Available: https://doi.org/10.1016/j.compag.2023.107857.

[13] H. Meng et al., “Livestock Biometrics Identification Using Computer Vision Approaches: A Review,” Agriculture, vol. 15, no. 1, p. 102, Jan. 2025, [Online]. Available: https://doi.org/10.3390/agriculture15010102.

[14] M. Galinski, I. Jatz, and L. Soltes, “Head-On Collision Avoidance Using V2X Communication,” in 2023 International Symposium ELMAR, Sep. 2023, pp. 37-40, [Online]. Available: https://doi.org/10.1109/ELMAR59410.2023.10253932.

[15] J. Yang, K. Wang, R. Li, Z. Qin, and P. Perner, “A novel fast combine-and-conquer object detector based on only one-level feature map,” Comput. Vis. Image Underst., vol. 224, p. 103561, Nov. 2022, [Online]. Available: https://doi.org/10.1016/j.cviu.2022.103561.

[16] S. Li, L. Sun, and Q. Li, “CLIP-ReID: Exploiting Vision-Language Model for Image Re-identification without Concrete Text Labels,” Proc. AAAI Conf. Artif. Intell., vol. 37, no. 1, pp. 1405-1413, Jun. 2023, [Online]. Available: https://doi.org/10.1609/aaai.v37i1.25225.

[17] J. Kim and J. Cho, “DBSCAN-Based Tracklet Association Annealer for Advanced Multi-Object Tracking,” Sensors, vol. 21, no. 17, p. 5715, Jan. 2021, [Online]. Available: https://doi.org/10.3390/s21175715.

[18] T. Chavdarova et al., “WILDTRACK: A Multi-camera HD Dataset for Dense Unscripted Pedestrian Detection,” in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT: IEEE, Jun. 2018, pp. 5030-5039, [Online]. Available: https://doi.org/10.1109/CVPR.2018.00528.

[19] A. Sepas-Moghaddam and A. Etemad, “Deep Gait Recognition: A Survey,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 1, pp. 264-284, Jan. 2023, [Online]. Available: https://doi.org/10.1109/TPAMI.2022.3151865.

[20] Y. Li, L. Wu, Y. Chen, X. Wang, G. Yin, and Z. Wang, “Motion estimation and multi-stage association for tracking-by-detection,” Complex Intell. Syst., vol. 10, no. 2, pp. 2445-2458, Apr. 2024, [Online]. Available: https://doi.org/10.1007/s40747-023-01273-3.

[21] R. Menghani, S. Rajanayagam, and S. Twieg, “Optimized Indoor and Outdoor SLAM for a Quadruped-Based Robot Walking Partner to Safely Promote Elderly Mobility and Support Nursing Staff,” Jun. 2025, [Online]. Available: https://doi.org/10.1109/ISSC67739.2025.11291245.

Image Detection Method Based on CNN and Pre-Trained Models

Oleg Pursky[1], Tetyana Filimonova[1], Anna Selivanova[1], Tatiana Dubovyk[1], Victor Kovtunenko[2] and Iryna Buchatska[1]

[1]*State University of Trade and Economics, Kioto Str. 19, 02156 Kyiv, Ukraine*
[2]*Cherkasy State Technological University, Shevchenka Blvd. 460, 18006 Cherkasy, Ukraine*
pursky_o@ukr.net, t.filimonova@knute.edu.ua, a.selivanova@knute.edu.ua, t.dubovyk@knute.edu.ua, v.kovtunenko@chdtu.edu.ua, i.buchatska@knute.edu.ua

Keywords: Image Detection, CNN, Transfer Learning, Deep Learning, Computer Vision, Neural Networks.

Abstract: This work presents a comprehensive study of the image detection method using CNN and pre-trained models. An object detection method has been developed based on a combination of modern convolutional neural network (CNN) architectures and a transfer learning strategy. Special attention in the developed method is paid to the algorithm for geometric transformation of the input image, which performs scaling while preserving the proportions (aspect ratio) and adding the minimum necessary fields (padding). To improve the interpretation of the neural network output data, a visualization module has been developed that imposes bounding boxes and class labels on the output image. The proposed approach uses the weights of pre-trained models. The results demonstrate an increase in network convergence speed and detection accuracy compared to training models "from scratch", implemented in Python using the TensorFlow/Keras, OpenCV, and NumPy libraries. The practical value of the work lies in creating a full cycle (pipeline) of developing a detection system: from data preparation to a ready-made inference module. The developed toolkit is universal and can be quickly adapted to solve applied tasks (video surveillance, quality control, traffic monitoring) by replacing the dataset without changing the software architecture.

1 INTRODUCTION

Computer vision and image detection technologies play a key role in many areas, such as robotics, video surveillance, medicine, and unmanned transportation. The use of convolutional neural networks (CNN) is the de facto standard for solving these problems, but training such networks "from scratch" requires enormous computational resources. The use of pre-trained models allows you to significantly increase the efficiency of development, reduce training time, and achieve high accuracy even on limited data sets. The development of detection methods based on this approach is relevant, as it provides a balance between speed and accuracy, which is critical for real-time systems and devices with limited computing power.

The aim of the research is to develop an image detection method using CNN and pre-trained models to create an effective object recognition system, optimize the training process, and increase detection accuracy by adapting existing architectures.

The work uses Deep Learning methods to build and configure convolutional neural networks, Transfer Learning to adapt pre-trained models, digital image processing methods to prepare input data, and statistical analysis methods to assess the metrics of model performance (accuracy, completeness, mAP).

The novelty of the results lies in the study of the effectiveness of using different strategies for adapting pre-trained CNN models to solve the problem of object detection in specific conditions. A comparative analysis of architectures was conducted and optimal pre-training parameters were determined, which allow achieving high accuracy of object classification and localization while minimizing computational costs compared to training networks without prior initialization of weights. The developed methodology and software implementation can be used to create applied computer vision systems where fast and accurate object detection is required.

2 LITERATURE REVIEW

A significant number of works are devoted to the study of computer vision and pattern recognition

issues [1]-[11]. The principles of the modern paradigm of machine vision, according to which high detection efficiency can be achieved by using convolutional neural networks (CNN) and ensuring hierarchical feature extraction, were substantiated by J. Lecoun [12].

J. Redmon [11] improved the detection methodology by proposing the "You Only Look Once" (YOLO) approach, which considers detection as a single regression problem, significantly accelerating the processing process.

In the studies of K. He [10], the prospects for the development of deep networks based on residual learning (ResNet), which allowed solving the problem of vanishing gradient when training ultra-deep models.

It should be noted that when considering the functions of computer vision systems, the authors of modern reviews [13] focus on the creation of complex intelligent systems, that is, they analyze the process of processing visual information as the formation of multi-level abstractions.

The scientific and methodological principles of using pre-trained models (Transfer Learning), which, taking into account the peculiarities of learning on limited samples, characterize the possibility of transferring knowledge between domains, are discussed in depth in the scientific works of S. Rader [14]. In [15], the theoretical and methodological principles of adapting CNN architectures for specific detection tasks are defined.

Currently, the main goal of the development of computer vision technologies is to bring the capabilities of artificial systems closer to human perception. It is no coincidence that in the strategies for the development of the digital economy of many countries tasks are defined that are aimed at creating autonomous navigation, monitoring and situation analysis systems, and ultimately - at increasing the safety and efficiency of human activity. The implementation of the tasks is facilitated by using as the goals of the development of the detection method, first of all, such as increasing the accuracy of recognition in difficult weather conditions, reducing the number of false positives, optimizing work in conditions of partial overlap of objects.

In order to study and understand the essence of visual images in computer vision, various interdependent specific research methods are used, the totality of which forms the methodology of digital image processing (Digital Image Processing) [16].

There are several approaches to building effective object detectors [17], [18]. One of them (Two-Stage Detectors, e.g., R-CNN) consists in generating candidate regions (Region Proposals) that potentially contain objects, with the subsequent classification of each region. The disadvantage of this approach is the division of the process into stages and, as a result, the low speed of operation, which makes it impossible to use in real time. Another approach (One-Stage Detectors, e.g., YOLO, SSD) consists in simultaneously predicting the coordinates of frames and object classes through a single neural network. A variation of this approach is the use of anchor frames (Anchors), relative to which displacements are predicted. In this case, the values of the network outputs correspond to the coordinates and probabilities in the image grid. The disadvantages of early versions of this approach include difficulties with detecting small objects, as well as the fact that among the large number of predicted frames, many are false (the problem of imbalance of background and object classes).

Fine-tuning mechanisms of pre-trained models depend on the nature of the new data, the similarity of domains, the required accuracy and speed, and both full unfreezing of weights and training only the last layers can be used. To assess the quality of the detection method, it is necessary to determine its validation metrics - the degree of reliability of the obtained predictions. The basis of the theory of assessing the quality of models is the analysis of Precision-Recall curves [19]. Also, quite often in modern research, data augmentation methods are used. Augmentation is usually carried out by geometric transformations (rotations, scaling) or changing color characteristics.

To determine the quality of detection in situations that are difficult to characterize with a single number (for example, visual quality of segmentation, work in fog or rain), methods of visual inspection of results on test video sequences are used [20]. This approach, although subjective and does not provide an accurate mathematical estimate, allows you to identify specific model errors that may be critical for practical application.

3 METHODOLOGY

This study uses the PASCAL Visual Object Classes (VOC) 2012 dataset [21]. The dataset contains a variety of natural images with many objects of different classes, differing in scale, position, illumination, and degree of occlusion. The dataset has been modified to ensure compatibility with the YOLO family of models. For this purpose, the original annotations in XML format (VOC) were

converted, and the object markup was converted to YOLO format using official scripts provided by the Ultralytics library. As a result, each image corresponds to a text file with bounding box coordinates and class identifiers normalized to the image size. The dataset contains 23.1 thousand files, of which 11.5 thousand are images in **.jpg* format, 11.5 thousand are annotation files in **.txt* format, and one configuration file in **.yaml*. The dataset is distributed under the Creative Commons Attribution-ShareAlike 4.0 (CC BY-SA 4.0) license.

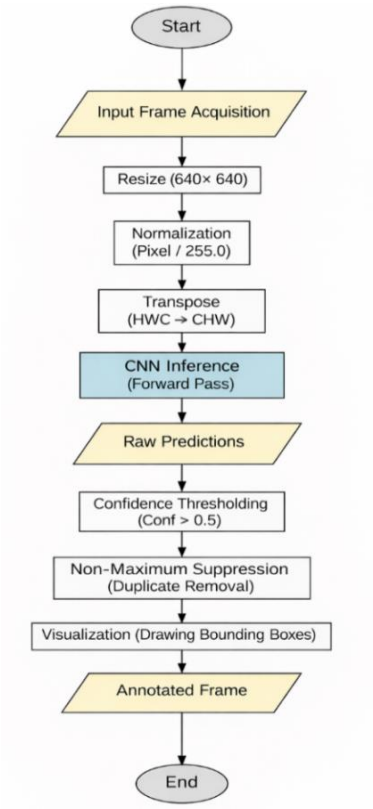


Figure 1: General structure of the algorithm for processing one frame of a video stream or a static image.

A general algorithm for the detection system has been developed. The logic of the system is decomposed into three sequential stages: preprocessing of input data (Preprocessing), inference of the neural network (Inference) and post-analysis of the results (Post-processing). At the Input Frame Acquisition stage, the input image or video frame is read, which is then transmitted to the preprocessing block.

Special attention in the developed method is paid to the algorithm for geometric transformation of the input image (Fig. 1). Most modern CNNs, including YOLO, have a fixed input tensor size (for example, 640×640). Simple bilinear scaling (resize) of images with an aspect ratio other than 1:1 (for example, 1920×1080) leads to geometric distortions of objects, which negatively affects the accuracy of detection. To solve this problem, the Letterbox Resizing algorithm is used, which performs scaling while preserving the proportions (aspect ratio) and adding the minimum necessary margins (padding).

The software complex was developed in Python 3.10 using the Google Colab cloud computing environment, which provides access to graphics accelerators (NVIDIA Tesla T4 GPU). The dataset split includes 5,717 training images and 5,823 validation images, resulting in 11,540 images in total, which corresponds to approximately 11.5 thousand images mentioned above. These numbers refer to images, not to annotated objects. The class distribution is characterized by the expected dominance of the person class, as well as a high representation of the dog, chair, and car classes. The class diagram is presented in Figure 2.

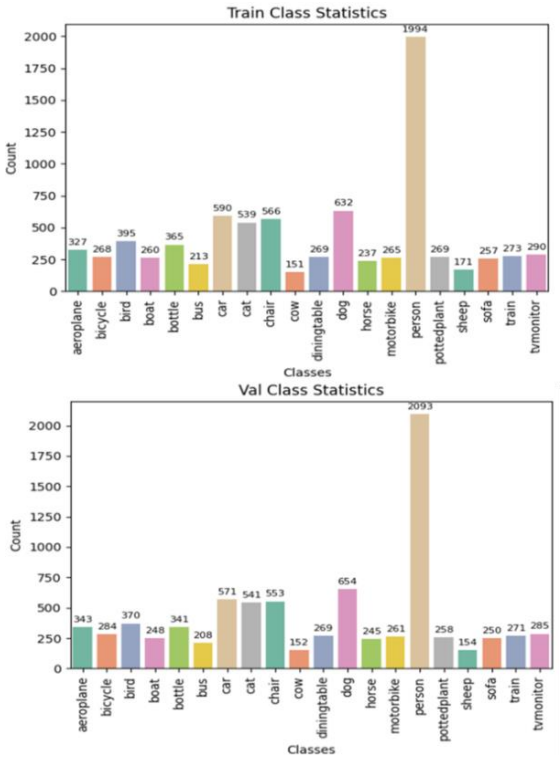


Figure 2: Class diagram of the PASCAL VOC 2012 dataset.

To start training, a *data.yaml* configuration file was programmatically generated, containing all the necessary information about the data set and the training process. The process of initialization and starting training was implemented through the

Ultralytics library API. Note that the training was carried out on the full set of training data, the number of classes used for training was 20. YOLOv8n with pre-trained weights was used as the basic architecture. The model was trained for 50 epochs.

The image before being fed to the model was read using OpenCV and scaled to a fixed size of 640×640 pixels. After resizing, its basic properties after resizing are determined, including width, height, number of channels (1 for grayscale or 3 for RGB), and pixel data type.

After training, the model was exported to ONNX format, which ensures compatibility with various deployment environments and the possibility of further use in autonomous inference systems.

To test the system on arbitrary images (Inference), the *show_prediction* function was developed. Thus, the software implementation represents a full cycle of ML system development: from data preparation and format conversion to model training and creation of tools for user testing. All intermediate results and weights of the trained model are automatically stored in the Google Colab cloud storage.

4 EXPERIMENTS AND RESULTS

A series of experiments were conducted to verify the effectiveness of the developed method. The dynamics of minimizing the loss function (Loss) and increasing accuracy (mAP) are presented in Figure 3. Based on the analysis of the learning dynamics, it was found that the model demonstrates stable and balanced convergence. The values of localization losses (box loss), classification (cls loss) and probability distribution (DFL loss) consistently decrease both on the training and validation samples. At the same time, there is an increase in the precision, recall indicators, as well as the mAP@0.5 and mAP@0.5:0.95 metrics, which confirms the gradual improvement of the detection quality throughout all training epochs. Quantitative indicators of the model quality were obtained by evaluation on the PASCAL VOC validation subset. The final values are given in Table 1.

As a result of testing the developed system, the mAP@0.5 value was 0.7152, while the mAP@0.5:0.95 value was 0.5210. The Precision value of 0.7563 is higher than the Recall value of 0.6458, which indicates that the model tends to reduce the number of false-positive detections. At the same time, the lower Recall value shows that some objects may still be missed in complex scenes. The lower value of mAP@0.5:0.95 compared with mAP@0.5 indicates that detection quality decreases under stricter IoU thresholds. Therefore, the obtained results should be interpreted as a practical trade-off between detection accuracy and real-time inference speed for a lightweight YOLOv8n-based model, rather than as state-of-the-art performance.

Table 1: Summary model performance metrics.

Metrics	Value	Description
Precision	0.7563	Detection accuracy (correct detection rate)
Recall	0.6458	Completeness (fraction of objects found)
mAP@0.5	0.7152	Average accuracy at 50% overlap threshold
mAP@0.5:0.95	0.5210	Integral accuracy metric (COCO metric)

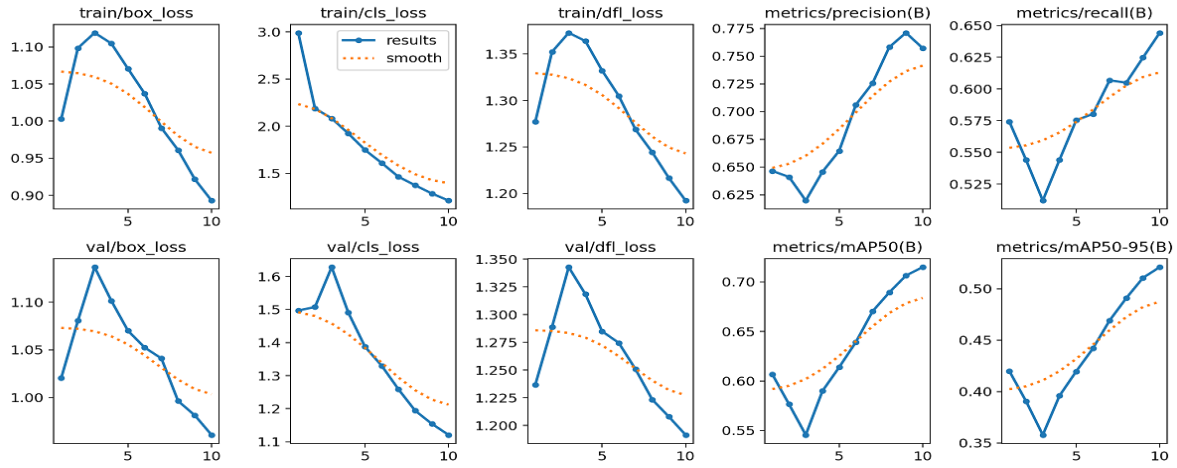


Figure 3: Graphs of changes in metrics (Precision, Recall, mAP) and loss functions during training.

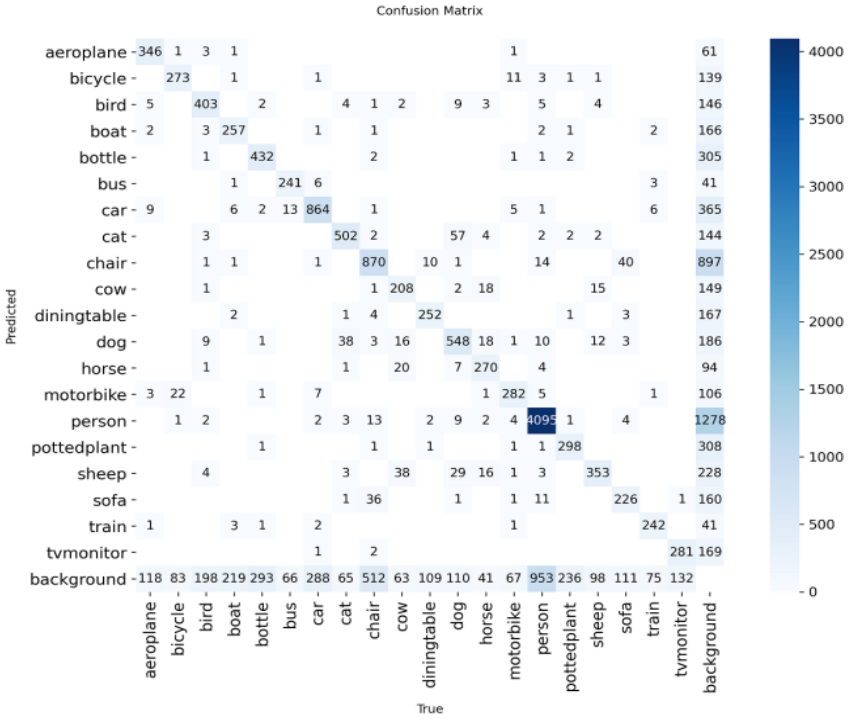


Figure 4: Confusion matrix of detection results on the validation sample.

For a detailed analysis of classification errors, a confusion matrix was constructed, shown in Figure 4.

Inference time was measured for different input image resolutions. The performance analysis was performed on the hardware platform of the Google Colab cloud service (GPU NVIDIA Tesla T4).

Table 2: Dependence of working speed on image size.

Input image size (px)	Inference time (ms)	FPS (frames/ sec)	Note
320 x 320	6.84	146	High speed, reduced accuracy
640 x 640	10.43	96	Optimal balance
1280 x 1280	18.83	53	High precision, low speed

As can be seen from Table 2, at the standard resolution of 640×640 pixels, the model provides an inference time of 10.43 ms, which corresponds to 96 FPS. This value significantly exceeds the real-time processing threshold of 24-30 FPS. When the resolution is increased to 1280×1280 pixels, the inference time increases to 18.83 ms, while the processing speed decreases to 53 FPS. Therefore, the 640×640 configuration can be considered the most balanced mode in terms of speed and detection quality.

To visually assess the adequacy of the system, testing was carried out on real images that did not participate in training (Fig. 5a, b and 6 a, b).

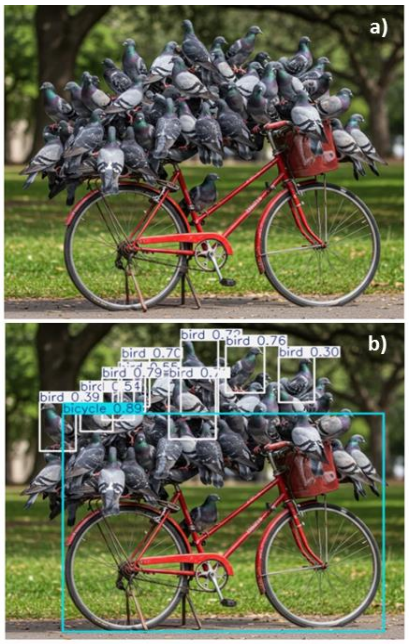


Figure 5: Example of object detection results: original image a), annotated image with detected objects and confidence scores; b) obtained using the YOLOv8n model.

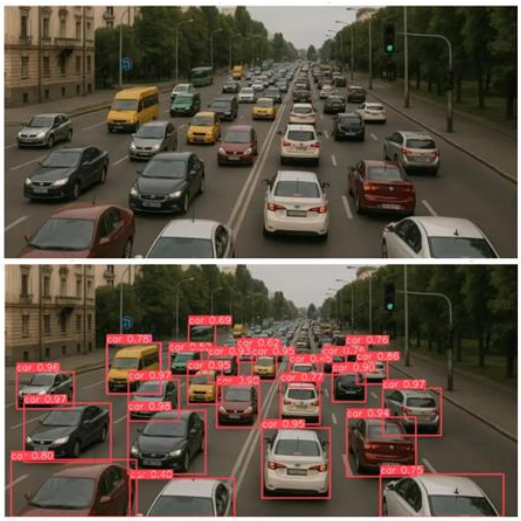


Figure 6: Original photo: a) Vehicle detection in the evening; b) demonstration of low-light resistance.

It should be noted that to improve the interpretation of the neural network output data, a visualization module was developed that imposes bounding boxes and class labels on the output image. The key problem in the visualization of multi-class detection is ensuring perceptual distinguishable objects. The use of random color generation is undesirable, since during video stream processing the color of the frame of the same object may change from frame to frame, which leads to the appearance of visual noise and a flickering effect.

To solve this problem, a deterministic color generation algorithm based on class name hashing is implemented. For each class i, the hash value H is calculated from the class name according to (1):

$$H = Hash(Name_i). \quad (1)$$

The RGB color components are then obtained from the hash value using (2)-(4):

$$R = (H >> 16) \ \& \ 0xFF. \quad (2)$$

$$G = (H >> 8) \ \& \ 0xFF. \quad (3)$$

$$B = H \ \& \ 0xFF. \quad (4)$$

This approach ensures that all objects of the same class (e.g., Person) will always be marked with the same unique color regardless of the system startup time or the order of frame processing.

5 CONCLUSIONS

This paper addresses the problem of improving the efficiency of computer vision systems by proposing and evaluating an image object detection method using convolutional neural networks (CNN) and pre-trained models. The application of Transfer Learning methodology allows reducing the network training time and achieving model convergence in just 10-50 epochs even on limited datasets. A model adaptation method has been developed, which includes the use of pre-trained weights (on the COCO dataset) and a fine-tuning strategy for the specifics of new classes.

For practical implementation, the YOLOv8n (Nano) architecture was selected, as it provides the best balance between inference speed and detection accuracy, making it suitable for real-time applications. The PASCAL VOC 2012 reference dataset was used as an experimental base. A statistical analysis of the class distribution was performed. The speed of image processing allows the system to be used in real time.

The software package was implemented in the Google Colab cloud environment using Python with GPU support. It includes modules for automated data loading, dataset format conversion, model training using the Ultralytics library, evaluation of detection results, inference speed measurement, and visualization of the obtained detections. The effectiveness of the proposed approach was experimentally evaluated on the PASCAL VOC validation subset using standard object detection metrics.

The model achieved Precision = 0.7563, Recall = 0.6458, mAP@0.5 = 0.7152, and mAP@0.5:0.95 = 0.5210. The higher Precision value compared with Recall indicates that the model tends to reduce the number of false-positive detections, although some objects may still be missed in complex scenes. The obtained mAP@0.5 value should not be interpreted as state-of-the-art performance, but rather as a practical trade-off between detection quality and real-time inference speed for a lightweight model, which achieved 95.9 FPS at an input resolution of 640 × 640 pixels.

The practical value of the work lies in creating a full cycle (pipeline) of developing a detection system: from data preparation to a ready-made inference module. The developed toolkit is universal and can be quickly adapted to solve applied tasks (video surveillance, quality control, traffic monitoring) by replacing the dataset without changing the software architecture.

Prospects for further research are seen in optimizing the model for work on mobile devices (Edge Devices) using quantization methods, as well as in expanding the functionality for tracking objects in the video stream.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436-444, 2015, [Online]. Available: https://doi.org/10.1038/nature14539.

[2] R. Szeliski, Computer Vision: Algorithms and Applications, Cham: Springer, 947 p., 2023, [Online]. Available: https://doi.org/10.1007/978-3-030-34372-9.

[3] S. K. Rao, Artificial Neural Networks: Advanced Applications, New Delhi: Discovery Publishing House PVT LTD, 248 p., 2024.

[4] A. B. Lozynskyy, I. M. Romanyshyn, and B. P. Rusyn, "Intensity estimation of noise-like signal in presence of uncorrelated pulse interferences," Radioelectronics and Communications Systems, vol. 62, no. 5, pp. 214-222, 2019, [Online]. Available: https://doi.org/10.3103/S0735272719050030.

[5] C. Wöhler, 3D Computer Vision: Efficient Methods and Applications, London: Springer, 400 p., 2014.

[6] O. Alonso and R. Baeza-Yates, Information Retrieval: Advanced Topics and Techniques, New York: Association for Computing Machinery, 836 p., 2024, [Online]. Available: https://dl.acm.org/doi/book/10.1145/3674127.

[7] T. Filimonova, O. Pursky, V. Babenko, A. Nechepourenko, V. Shvets, and V. Gamaliy, "Text Sentiment Analysis using Different Types of Recurrent Neural Networks," in 2024 5th International Conference on Image Processing and Capsule Networks (ICIPCN), Jul. 2024, [Online]. Available: https://doi.org/10.1109/ICIPCN63822.2024.00068.

[8] J. F. Peters, Foundations of Computer Vision: Computational Geometry, Visual Image Structures and Object Shape Detection, Cham: Springer, 744 p., 2017, [Online]. Available: https://doi.org/10.1007/978-3-319-52483-2.

[9] Y. P. Putyatin, Image Processing in Computer Vision Systems, Kharkiv: Prom-Art Publishing, 288 p., 2019.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2016, [Online]. Available: https://doi.org/10.1109/CVPR.2016.90.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2016, [Online]. Available: https://doi.org/10.1109/CVPR.2016.91.

[12] J. Fan, S. Deng, X. Song, J. Liu, and Y. Sun, "A gradient-based lightweight network automated design method for facial expression recognition," Expert Systems with Applications, vol. 296, p. 129130, 2026, [Online]. Available: https://doi.org/10.1016/j.eswa.2025.129130.

[13] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," International Journal of Computer Vision, vol. 128, pp. 261-318, 2020, [Online]. Available: https://doi.org/10.1007/s11263-019-01247-4.

[14] S. Ruder, "An overview of multi-task learning in deep neural networks," arXiv, 2017, [Online]. Available: https://doi.org/10.48550/arXiv.1706.05098.

[15] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2020, [Online]. Available: https://doi.org/10.1109/CVPR42600.2020.01079.

[16] R. C. Gonzalez and R. E. Woods, Digital Image Processing, London: Pearson, 1168 p., 2018.

[17] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2014, [Online]. Available: https://doi.org/10.1109/CVPR.2014.81.

[18] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 2017, [Online]. Available: https://doi.org/10.1109/TPAMI.2016.2577031.

[19] M. Everingham, S. M. Eslami, L. Gool, C. K. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes Challenge: A Retrospective," International Journal of Computer Vision, vol. 111, no. 1, pp. 98-136, 2015, [Online]. Available: https://doi.org/10.1007/s11263-014-0733-5.

[20] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," International Journal of Computer Vision, vol. 115, no. 3, pp. 211-252, 2015, [Online]. Available: https://doi.org/10.1007/s11263-015-0816-y.

[21] "PASCAL VOC 2012. A Benchmark Dataset for Multi-Class Object Detection," [Online]. Available: https://www.kaggle.com/datasets/banuprasadb/pascal-voc-2012, [Accessed: Jan. 14, 2026].

Architecture of Ontology-Driven Multidimensional Analytical Systems Based on Formal Ontologies

Oleksandr Stryzhak[1], Andrii Yaremenko[2] and Viacheslav Gorborukov[1,3]
[1]National Center "Junior Academy of Sciences of Ukraine", Dekhtyarivska Str. 38-44, 04119 Kyiv, Ukraine
[2]National University "Yuri Kondratyuk Poltava Polytechnic", Pershotravneva Avenue 24, 36011 Poltava, Ukraine
[3]Department of Multimedia Systems, National University of Kyiv-Mohyla Academy, Skovorody Str. 2, 04655 Kyiv, Ukraine
sae953@gmail.com, y_aa@ukr.net, slavon07@gmail.com

Keywords: Formal Ontology, Multidimensional Analytics, Ontology Engineering, Ontology-Driven Analytical Systems, Semantic Normalization, Multicriteria Analysis, Inductive Construction of Analytical Structures, Intelligent Analytics.

Abstract: The article presents a formally grounded approach to the development of ontology-driven systems for multidimensional analytics. In this approach, a formal domain ontology is treated as the primary and invariant knowledge model that provides a sufficient basis for the inductive construction of analytical structures. Unlike traditional solutions, where ontologies are typically applied only as auxiliary semantic or integration layers, the proposed framework derives analytical dimensions, hierarchies, indicators, and permissible aggregation operations directly from the logical organization of the ontological model. The study formalizes several key processes required for building such systems. These include the structuring and semantic alignment of heterogeneous data sources, the construction of a formal domain ontology, and the subsequent use of this ontology to support multicriteria evaluation and the generation of multidimensional analytical representations. Within this framework, a mathematical definition of an ontology-analytic mapping operator is introduced. This operator ensures a rigorous transformation from the ontological model to a multidimensional analytical structure while preserving type information, semantic constraints, and interpretive properties. The results demonstrate that analytical facts can be induced from ontological elements, which enables the correct handling of context-dependent and partially defined dimensions. In addition, the article proposes a conceptual architecture for an ontology-driven analytical system organized into functional and conceptual layers. This layered architecture supports methodological consistency, reproducibility of analytical procedures, and scalability of the proposed approach. The presented formalization is general in nature and can be applied to the development of intelligent analytical systems in various domains. In particular, it is well suited for systems designed to analyze and evaluate educational and intellectual achievements, where consistency of semantics, criteria, and analytical procedures is critically important. Overall, the obtained results provide a formal foundation for constructing coherent, semantically consistent, and explainable analytical systems across diverse application areas.

1 INTRODUCTION

1.1 Motivation

The development of information and analytical systems is accompanied by a continuous increase in the complexity of domain environments, the diversity of data sources, and the requirements for semantic consistency, reliability, and interpretability of analytical results. In contemporary applications, data are generated from heterogeneous sources, including documents, registries, tabular datasets, and textual materials. These data often differ in their level of structure and semantic clarity and are used in contexts characterized by multiple layers of interpretation and strong contextual dependence. Under such conditions, analytical systems are increasingly viewed not merely as tools for data processing or aggregation but as instruments for constructing a formally consistent analytical space that enables comparison, generalization, and interpretation of results across different application contexts.

This situation creates a demand for approaches that combine rigorous formal modeling with the

ability to support semantic interpretation, traceability, and structural extensibility. Ontology-based approaches play a particularly important role in addressing these challenges, as they allow the formal specification of domain structures, including key concepts, relationships, and interpretative constraints. The use of ontologies facilitates the transition from fragmented and heterogeneous datasets to coherent analytical models capable of supporting semantic consistency, reproducibility of model construction, and interpretability of analytical outcomes.

1.2 State of the Art

Modern information and analytical systems are traditionally built on data warehousing, OLAP, and multidimensional data model technologies, within which analytical facts are described through a combination of dimensions, hierarchies, measures, and aggregation operations [1], [2]. In classical OLAP systems, the main focus is on cube design, finding relevant analytical views, query optimization, and user interaction with multidimensional structures. Djiroun et al. address the problem of discovering and constructing OLAP cubes based on the analysis of user queries, demonstrating the importance of adaptive formation of analytical views [1].

At the same time, such approaches remain predominantly focused on existing collections of multidimensional cubes or on the construction of additional cubes in response to user queries, whereas the semantic nature of the subject domain typically does not serve as a formal basis for deriving the analytical structure itself.

A separate area of research is related to the transformation of traditional data warehouses under the influence of Big Data, NoSQL databases, and unstructured or semi-structured data sources. Martinez-Mosquera et al. systematize approaches to integrating OLAP with NoSQL databases and demonstrate that the transition to Big Data and NoSQL environments requires adapting OLAP cube construction models, warehouse types, modeling methods, processing modes, and performance criteria [3]. Ptiček et al. emphasize that for the integration of NoSQL data and traditional data warehouses, a mere technical combination of sources is insufficient, as the problem must be solved at a higher level of abstraction, specifically at the level of semantic modeling and design [4]. This points to the need for models capable not only of integrating heterogeneous data but also of ensuring its semantic interpretation in the analytical process.

The fundamental principles of this level of formalization are established in the works of Gruber, Guarino, Uschold, and Gruninger, as well as in subsequent generalizations by Staab and Studer [5]-[8]. Gruber defines an ontology as an explicit specification of conceptualization, oriented toward the sharing and reuse of formalized knowledge [5]. Guarino emphasizes the importance of formal ontology for separating domain knowledge from the specific procedures for its use [6]. Uschold and Gruninger view ontologies as a means of achieving shared understanding, interoperability, reusability, and reliability of software systems [7]. Within the scope of this study, an ontology is treated not as a reference dictionary or an external metadata system, but as a formal model of a subject domain capable of defining valid interpretations of objects, relations, attributes, values, and semantic constraints [8].

Semantic technologies and ontologies are actively used in semantic data warehouses, semantic ETL, and exploratory OLAP. Abelló et al. demonstrate that Semantic Web technologies can support the discovery, integration, semantic annotation, and analytical processing of external sources in exploratory OLAP scenarios [2]. Deb Nath et al. propose high-level RDF-based ETL constructs for defining, mapping, transforming, integrating, and loading semantic multidimensional data [9]. Antunes et al. demonstrate the application of Semantic Web techniques to link strategic management, the Balanced Scorecard ontology, and the BI environment [10]. In practice-oriented BI literature, the semantic layer is also viewed as a business-oriented abstraction between the data warehouse and user analytical applications, which improves data accessibility, terminology consistency, governance, and visualization [11]. However, in a significant portion of these approaches, ontologies or the semantic layer are used primarily for integration, annotation, contextualization, traceability, semantic ETL, or to support predefined or semi-generated analytical structures. At the same time, the rules for deriving a complete multidimensional analytical model from a unified ontological basis, including dimensions, hierarchies, measures, admissible aggregation operations, and analytical facts, remain only partially formalized.

The work closest to the subject matter of this study is that of Romero and Abelló, dedicated to the automation of multidimensional design based on ontologies [12]. It proposes an approach to identifying multidimensional concepts from a domain ontology and using these concepts for designing data warehouses. This work is particularly important for

positioning the present study because it directly addresses the transition from an ontological description of a domain to multidimensional design.

At the same time, the present article is positioned within the line of ontology-based multidimensional design, but shifts the focus from the detection of multidimensional schema elements to the formal role of the ontology in analytical model construction. In Romero and Abelló's approach, the ontology serves as the input for identifying multidimensional design elements - facts, dimensions, measures, bases, and dimension hierarchies - and for constructing conceptual data warehouse schemas. In the present article, the ontology is treated as a formal domain model whose analytical function is defined after semantic normalization and interpretive specification of domain indicators. It is then used not only to identify multidimensional schema elements, but to derive a multidimensional analytical model under explicit semantic, functional, interpretive, and aggregation constraints.

Thus, the proposed ontology-analytical mapping operator formalizes a controlled transition from a formal domain ontology to a logical analytical cube, in which dimensions, hierarchies, measures, admissible aggregation operations, and analytical facts are systematically derived.

In the context of intelligent, educational, and explainable systems, ontologies are also viewed as a means of bridging the gap between formal knowledge, practical application, and the explainability of results. Mizoguchi and Bourdeau demonstrate that ontology engineering provides a shared vocabulary, supports consistent domain modeling, and facilitates the construction of theory-compliant systems in educational environments [13]. Futia and Vetrò, as well as Confalonieri and Guizzardi, link knowledge graphs and ontologies to the problem of explainable and neuro-symbolic AI, where explicit knowledge is used to enhance the comprehensibility, justifiability, and interpretability of results [14], [15]. For analytical systems, this means that an ontological model can perform not only an integrative but also an explanatory function, specifying the semantic origin of indicators, criteria, aggregations, and analytical conclusions.

Ontology-driven knowledge structuring and decision support systems constitute a separate line of application. Nadutenko et al. consider ontology-driven lexicographic systems as one example of applying the ontological approach to the organization, structuring, and interpretation of complex knowledge resources [16]. Stryzhak et al. propose a decision-making system based on the ontology of the choice problem, in which the task of rational choice is formalized through an ontological representation of alternatives, criteria, and ranking indicators [17]. Taken together, these works demonstrate that ontological models can support both the organization of complex information resources and decision-oriented analytical procedures. At the same time, this article shifts the focus from a specific choice problem or lexicographic structuring to the broader problem of inductively forming a multidimensional analytical model from a formal domain ontology.

Thus, an analysis of existing approaches allows us to identify three main limitations of the current state of research. First, in classical OLAP and DW approaches, analytical schemas are predominantly designed in advance or reconstructed according to the user's query, but their connection to the formal semantics of the subject domain is usually not defined as a separate, mathematically specified transition from ontology to analytical model [1], [2]. Second, in semantic OLAP, semantic DW, and semantic ETL, ontologies are mostly used for integration, annotation, contextualization, traceability, or ETL support, while their role as a generative basis for constructing multidimensional structures remains limited [2], [4], [9]-[11]. Third, even in approaches directly aimed at automating multidimensional design from ontologies [12], the main emphasis is placed on identifying multidimensional schema elements and supporting the construction of conceptual DW structures, whereas the present article focuses on defining a unified ontology-analytical mapping operator that specifies how dimensions, hierarchies, measures, admissible aggregation operators, and analytical facts are derived from the formal ontology.

Accordingly, the proposed approach builds upon existing work in ontology engineering, semantic OLAP, semantic data warehouses, semantic ETL, ontology-based multidimensional design, and ontology-based decision support, but shifts the focus from using ontology as an auxiliary semantic layer or a source of individual multidimensional design patterns to using it as a formal generative basis for multidimensional analytics. Within this approach, a multidimensional analytical structure is not simply projected from a predefined schema or user query, but is derived under explicitly defined mapping rules from concepts, relations, functions, types, and semantic constraints represented in the domain ontology.

1.3 Objectives and Tasks

Despite the widespread adoption of ontological models in modern information and analytical systems, the problem of formally aligning domain semantic models with analytical structures remains unresolved. This issue is particularly evident in relation to multidimensional representations, indicators, hierarchies, and aggregation operations.

In many existing implementations, analytical structures are still developed separately from the ontological model, which weakens the formal connection between domain semantics and analytical operations. This separation complicates the reproducibility of analytical models, limits their transferability across different domains, and weakens the ability to ensure the semantic correctness of analytical operations.

Consequently, there is a scientific need to develop a formal approach for ontology-driven construction of multidimensional analytical representations. Such an approach should support the inductive derivation of analytical dimensions, indicators, hierarchies, and permissible aggregation operations directly from the domain ontology while preserving its semantic constraints and interpretative properties. Addressing this problem is essential for the development of coherent, scalable, and methodologically consistent analytical systems capable of operating with heterogeneous data across a wide range of applied tasks.

The aim of this paper is to develop and demonstrate a formal ontology-driven framework in which a domain ontology serves as a generative formal basis for deriving multidimensional analytical structures, including dimensions, hierarchies, analytical measures, domain-specific indicators, admissible aggregation operations, and analytical facts.

To achieve the stated objective, the following tasks must be addressed:

- To formalize the ontological model of the domain as a structured system of entities, relationships, and semantic constraints.
- To establish principles for the inductive derivation of analytical dimensions, hierarchies, and analytical measures, including domain-specific indicators, from the components of the ontology.
- To develop a formal operator that maps the ontological structure to a multidimensional analytical model.
- To illustrate how the proposed formal constructs are instantiated in the domain of intellectual competition analytics.
- To substantiate the semantic consistency, traceability, and reproducibility conditions of the proposed approach.

The article is organized as follows. Section 2 presents the formal ontology-analytical model, including the formalization of input data, the construction of the domain ontology, the definition of the multidimensional analytical model, and the specification of the ontology-to-analytics mapping operator. Section 3 describes the conceptual multilayer architecture of the ontology-driven analytical system and its functional components. Section 4 demonstrates the application of the proposed approach to the problem of analyzing and ranking intellectual achievements within the context of competitive events. The final section summarizes the conclusions and outlines directions for future research.

2 MATERIALS AND METHODS

2.1 Formal Ontology-Analytical Model

The formal model introduced in this section describes the sequence of transformations from heterogeneous input sources to a normalized domain structure, then to a formal ontology, and subsequently to an enriched ontology that can be used for multidimensional analytical construction. The definitions below specify the main structural components of the proposed approach and provide the basis for the ontology-analytical mapping operator introduced in the next subsection.

The input data of the domain may originate from heterogeneous documentary sources, including various electronic forms, questionnaires, reports, registries, and textual descriptions. To ensure a unified approach to the further processing of such sources, a set of input documents is introduced:

$$S = \{s_1, s_2, \dots, s_n\}, \quad (1)$$

where S - denotes the set of all documents obtained directly from digital sources, s_i - represents an individual document within the domain.

Because input documents may differ significantly in their internal organization, level of structure, and degree of semantic completeness, a structural transformation operator is introduced. This operator maps the set of input documents to a unified

normalized data structure N (which is suitable for further formal analysis and semantic interpretation):

$$\Phi_{\text{norm}}: S \to N. \quad (2)$$

Where:

- S - the set of input documents;
- Φ_{norm} - the operator of structural transformation and alignment;
- N- a unified normalized data structure.

The result of this transformation stage is described by the structure:

$$N = \left(X',\ A',\ V_A',\ F',\ C',\ \tau_X'\right). \quad (3)$$

Where:

- X'-the set of domain objects;
- A'- a set of normalized and semantically consistent attributes;
- V_A'- a set of normalized attribute values;
- $F' \subseteq X' \times A' \times V_A'$- a ternary relation of the «object-attribute-value type»;
- C'- a set of object types (classes);
- $\tau_X': X' \to C'$ - an object typing function.

Based on the normalized structure, a formal ontology of the domain is constructed that provides a semantic interpretation of objects, attributes, their values, and the relationships between them. The construction of the ontology is described by an ontological mapping operator:

$$\Phi_{\text{ont}}: N \to O. \quad (4)$$

Where:

- N- is the normalized data structure;
- O- is the formal ontology of the domain;
- Φ_{ont}- is the ontological mapping operator that performs the construction of the ontological model.

After the ontology has been constructed, multicriteria evaluation becomes possible, followed by the extension of the ontological model with rating values.

$$\Phi_{MR}: O \to MR,\ \Omega_{MR}: O \times MR \to O^*. \quad (5)$$

Where:

- O- is the formal ontology of the domain;
- MR - is the formal model of the multicriteria evaluation and ranking problem;
- Φ_{MR} - is the operator for constructing the ranking model based on the ontology;
- Ω_{MR} - is the operator for applying the model MR to the ontology O with the formation and integration of evaluation results;
- O^* - is the extended ontology with integrated ranking results.

Thus, the first part of the model defines the transformation chain:

$$S \xrightarrow{\Phi_{olap}} N \xrightarrow{\Phi_{ont}} O \xrightarrow{\Omega_{MR}} O^*.$$

This chain is important for the architecture proposed in Section 3: the system does not transform raw data directly into an analytical cube. Instead, heterogeneous sources are first normalized, then interpreted ontologically, then enriched with evaluation results, and only after that used as the basis for ontology-driven multidimensional analytical construction.

The ontology-driven construction of a multidimensional analytical representation is based on the principle that a sufficiently specified formal ontological model provides an internally consistent semantic basis for analytical model construction. Such a description supports the derivation of analytical structures under explicitly defined mapping rules and reduces the need for ad hoc non-semantic assumptions. Within this framework, the ontology functions as a formal source of analytical interpretation. Formally, the ontology is defined as an ordered triple:

$$O = (X_o, R_o, F_o), \quad (6)$$

where: $X_o = X_C \cup X_I$ - denotes the set of ontological elements, comprising both abstract concepts X_C, and concrete domain objects X_I, R_o - denotes the set of binary relations between ontology elements, and F_o - denotes the set of ontological functions that define the interpretation of attributes and values of ontology elements.

Each element $x \in X_o$ is typed according to the ontological typing function:

$$\tau_X: X_o \to C, \quad (7)$$

where τ_X- denotes the ontological typing function, and C- denotes the universal set of ontological types and categories.

Typing makes it possible to distinguish concepts of classificatory, contextual, and factual nature, as well as individual objects considered as instances of the corresponding concepts. Ontological relations R_o are defined as a subset of the Cartesian product of the set of ontological elements:

$$R_o \subseteq X_o \times X_o, \quad (8)$$

where $(x_i, x_j) \in R_o$ indicates the existence of a semantically interpreted relation between elements x_i

and x_j. Such relations may represent generalization, inclusion, composition, instance-class membership, or other structural dependencies that form the semantic framework of the ontology.

The set of ontological functions is defined as:

$$F_o = \{f \mid f: X_o \to V\}, \quad (9)$$

where f - an ontological function that interprets an attribute or property of an ontology element, V - the set of admissible values, which may include numerical, symbolic, or ordered domains.

The presence of functions with numerical or ordered value domains creates formal prerequisites for the subsequent analytical use of the ontology, in particular for the construction of indicators and aggregation operations within a multidimensional representation.

Ontological-analytical transformation consists in the formal mapping of the ontological model into a multidimensional analytical structure while preserving the semantic constraints, typing, and interpretations specified in the ontology. Unlike the procedural construction of analytical schemas, this transformation is considered a mathematically defined operator acting on formal structures.

For further formalization of ontology-driven analytical mapping, it is necessary to provide a rigorous formal description of the multidimensional analytical model. Such a representation is treated as an abstract mathematical structure independent of any specific implementation of data storage or computation. The multidimensional analytical model is defined as an ordered quintuple:

$$M_{OLAP} = (D, H, M, A, \mu). \quad (10)$$

Where:

- D - is a finite set of analytical dimensions;
- H - is the set of hierarchies defined over the dimensions, M - is the set of analytical measures;
- A - is the set of admissible aggregation operators;
- μ - is a multidimensional analytical mapping function.

Each dimension $D_i \in D$ is interpreted as an ordered set of discrete values or levels of generalization corresponding to a particular aspect of the subject domain. Formally, the set of admissible values for each analytical dimension is defined by the following mapping:

$$\mathrm{dom}: D \to 2^{V_D}, \quad (11)$$

for each $D_i \in D$ let $dom(D_i) \subseteq V_D$ - denote the set of admissible values of dimension D_i, while V_D - represents the universal set of all possible values of analytical dimensions, and 2^{V_D} - denotes the power set ofV_D. In this context $D_i$denotes the dimension itself (i.e., the axis), where as $d_i \in dom(D_i)$ - represents a specific value of that dimension (i.e., the coordinate of a fact).

The hierarchies H define admissible transitions between levels of granularity within a single dimension or across related dimensions. The set of measures M consists of numerical or ordered functions whose values can be aggregated along the dimensions. For each measure, a subset of admissible aggregation operators is defined to ensure the correctness of generalization and interpretation.

The multidimensional mapping functionμ formalizes the relationship between dimension values and measures, thereby determining the structure of facts in the multidimensional model:

$$\mu: dom(D_1) \times \cdots \times dom(D_k) \times M \to V', \quad (12)$$

Where:

- $D_i \in D$ - denotes an individual analytical dimension, $dom(D_i)$ - is the set of admissible values of the i -th analytical dimension,
- k - is the number of dimensions involved in the analytical mapping,
- M - is the set of measures, V′ is the set of admissible measure values (numerical or ordered domains),
- μ - is a multidimensional function that maps a combination of dimension values and a selected measure to its corresponding value.

As a result, each element of the multidimensional model can be represented as an ordered tuple:

$$f_{OLAP} = (d_1, d_2, \ldots, d_k, m, v), \quad (13)$$

where $d_i \in dom(D_i)$ -is a specific value of the corresponding dimension, $m \in M$ - is the measure associated with the given combination of dimensions, and $v \in V'$ - denotes the value of that measure.

Thus, the multidimensional analytical model provides the target structure for ontology-driven construction. The next subsection specifies how its components - dimensions, hierarchies, measures, admissible aggregation operations, and analytical facts - are derived from the formal ontology rather than externally imposed as a predefined OLAP schema.

2.2 Formal Operator of Ontology-Analytical Transformation

A key element of ontology-driven construction of multidimensional analytical representations is the ontology-analytical mapping operator, which provides a formal transition from an ontological model to the structure of multidimensional analytics. Unlike procedural or heuristic approaches to the construction of OLAP schemas, this operator is defined as a strict mapping between formal structures that preserves the semantic properties of the source ontology.

Formally, the ontology-analytical mapping operator is defined as:

$$\Phi_{olap}: O \to M_{OLAP}. \quad (14)$$

Where:

- Φ_{olap} - denotes the ontology-analytical mapping operator;
- O - is the formal ontological model;
- M_{OLAP} - is the multidimensional analytical model.

The operator Φ_{olap} interprets ontological elements according to their semantic roles, which are determined by their typing and structural relations. In particular, concepts classified as categorical or contextual types may be mapped to the set of analytical dimensions; hierarchical relations between concepts may be transformed into dimension generalization structures; and ontological functions with numerical value domains may be mapped to the set of analytical measures.

For each component of the ontology, the operator Φ_{olap}defines a set of partial mappings:

$$\Phi_{olap} = (\Phi_D, \Phi_H, \Phi_M, \Phi_A, \Phi_\mu). \quad (15)$$

Where:

- $\Phi_D: X_o \to D$- denotes the mapping of concepts to dimensions;
- $\Phi_H: R_o \to H$- represents the mapping of ontological relations to dimension hierarchies;
- $\Phi_M: F_o \to M$- denotes the mapping of ontological functions to analytical measures;
- $\Phi_A: M \to 2^A$ - represents the mapping of measures to the set of admissible aggregation operator;
- Φ_μ - denotes the mapping of ontology facts to the multidimensional function μ.

Thus, the ontology-analytical mapping operator formalizes the principle according to which the multidimensional analytical model is considered a derived structure of the ontology, rather than an independently designed data schema. The formation of analytical dimensions constitutes the first stage in the specification of the multidimensional analytical model. At this stage, ontological concepts are interpreted as axes of the multidimensional space. Within the ontology-driven approach, dimensions are not predefined; instead, they are inductively derived from the ontology based on the typing and semantic roles of concepts.

The set of analytical dimensions is defined as a subset of the set of ontological concepts:

$$D \subseteq X_C, \quad (16)$$

where D - denotes the set of analytical dimensions and X_C - denotes the set of ontological concepts.

The criterion for including a concept $x \in X_C$ in the set of dimensions is its membership in types that are admissible for analytical decomposition. Formally, this condition is defined as follows:

$$x \in D \quad \Leftrightarrow \quad (x \in X_C \ \wedge \ \tau_X(x) \in C_{dim}), \quad (17)$$

Where:

- $\tau_X(x)$ - denotes the type of the ontological concept x;
- $C_{dim} \subseteq C$- is the subset of ontological types permitted as analytical dimensions;
- C - is the universal set of ontological types.

The set C_{dim} typically includes concepts of a classification or contextual nature that determine the way facts are grouped, segmented, or compared. Concepts of a factual or interpretative nature are not included in the set of dimensions because they do not define independent axes of the analytical space.

In general, the set of analytical dimensions is formed as follows:

$$D = \{x \in X_C \mid \tau_X(x) \in C_{dim}\}. \quad (18)$$

Where:

- D - is the set of analytical dimensions;
- X_C - is the set of ontological concepts;
- τ_X- is the typing function;
- C_{dim} - is the set of types admissible for analytical dimensions.

Thus, the formation of the set of analytical dimensions directly depends on the ontological typing and semantic structure of the domain, which ensures the correctness and semantic consistency of the analytical axes.

The hierarchical organization of dimensions is a necessary condition for supporting aggregation and drill-down operations in multidimensional analytics. Within the ontology-driven approach, dimension hierarchies are not constructed independently of the domain but are instead derived from ontological

relations that already capture semantic dependencies between concepts. However, not every ontological relation can be interpreted as hierarchical in the analytical sense. Such a relation must define a direction of hierarchy from more detailed levels to more generalized ones and satisfy the properties of transitivity and acyclicity, which ensure the correct formation of hierarchical structures.

Formally, the set of dimension hierarchies can be represented as:

$$H = \{(x_i, x_j) \in R_o \mid x_i \prec x_j \wedge x_i, x_j \in D\}. \quad (19)$$

Where:

- H - is the set of dimension hierarchies,
- R_o - is the set of ontological relations,
- D - is the set of analytical dimensions,
- $\prec$ - denotes the binary generalization relation corresponding to the roll-up operation in multidimensional analytics.

Thus, the hierarchies of dimensions in the multidimensional analytical model constitute a direct mapping of the semantic hierarchies specified in the ontology. This ensures that analytical aggregation operations remain consistent with the conceptual structure of the domain and prevents the emergence of artificial or semantically inconsistent hierarchies.

Measures in the multidimensional analytical model represent quantitative or ordered characteristics of the domain whose values are subject to aggregation and comparative analysis. Within the ontology-driven approach, the set of measures is defined as the class of ontological functions that permit analytical operations:

$$M = \{f \in F_o \mid \forall x \in X_o : \ f(x) \in V'\}, \quad (20)$$

where M - is the set of analytical measures, F_o - is the set of ontological functions, X_o - is the set of ontological elements (concepts and/or objects) for which functions from F_o are defined, V' - is the set of admissible numerical or ordered value domains (i.e., domains that allow aggregation and/or comparison), and f - is an ontological function interpreted as a measure when the condition $f: X_o \to V'$.

The value of a measure $m \in M$ for an ontological element $x \in X_o$ is determined by applying the corresponding function:

$$m(x) = v, \ v \in V', \quad (21)$$

Where:

- m - is an analytical measure (a function from the set M);
- x - is the ontological element for which the measure value is computed;
- v - is the numerical or ordered value of the measure;
- V' - is the domain of admissible measure values.

Thus, measures are derived directly from ontological interpretations of values, ensuring semantic unambiguity of the measures and the correctness of their subsequent aggregation within analytical views.

Aggregation is a fundamental mechanism of multidimensional analysis; however, the admissibility of aggregation operations is determined not only by the data type but primarily by the semantics of the measures. In the ontology-driven approach, aggregation operations are aligned with measures at the level of their ontological interpretation. This prevents semantically incorrect generalizations and ensures the meaningful interpretability of analytical results.

The set of basic aggregation operators is defined as:

$$A = \{sum, avg, count, min, max\}, \quad (22)$$

Where:

- A - is the set of aggregation operators;
- sum - denotes the summation operator;
- avg - denotes the averaging operator;
- $count$ - represents the counting operator;
- min and max - denote the minimum and maximum operators, respectively.

The admissibility of aggregation is a property not of the operator itself but of a particular measure. Therefore, for each measure $m \in M$, the set of admissible aggregation operators is defined directly as a semantically determined subset of the set A:

$$A_m = \{a \in A \mid (m, a) \in Compat\}, \quad (23)$$

Where:

- A_m - is the set of aggregation operators admissible for the measure m;
- m - is an analytical measure;
- A - is the base set of aggregation operators;
- $Compat \subseteq M \times A$ - denotes the relation of semantic admissibility of the aggregation operator a for the measure m.

Thus, the correspondence between measures and admissible aggregation operations can be formalized by the mapping:

$$\Phi_A: M \to 2^A, \quad (24)$$

where Φ_A - is the function that semantically aligns measures with aggregation operators, and 2^A - denotes the power set of the set of aggregation operators.

The definition of the set A_m is based on the ontological properties of the measure. Additive measures allow summation and averaging operations; intensive measures allow averaging and extreme-value operations; whereas ordinal or nominal measures are typically restricted to minimum and maximum operations. Such semantic alignment ensures the correctness of aggregation and prevents the application of operations that contradict the semantic meaning of the measures.

The ontology-driven construction of a multidimensional analytical representation is based on identifying ontological elements that serve as sources of analytical facts. This role is not an inherent property of a concept and is not determined solely by its type; rather, it is induced by its ontological interpretation, particularly by the presence of ontological functions whose values permit analytical aggregation or comparison.

Fact-generating ontological elements are defined by the following condition:

$$X_F = \{x \in X_o \mid \exists f \in F_o\colon f(x) \in V'\}, \qquad (25)$$

Where:

- X_F - is the subset of ontological elements that serve as sources of analytical facts;
- X_o - is the set of ontological elements of the domain;
- F_o - is the set of ontological functions;
- $f\colon X_o \to V$ - is an ontological interpretation function;
- $V' \subseteq V$ - is the domain of values admissible for analytical interpretation (numerical or ordered).

Thus, fact generation is a role induced by ontological interpretation, rather than a consequence of membership in a predefined class. An ontological fact is represented as an ordered triple:

$$f_O = \big(x, f, f(x)\big), \qquad (26)$$

Where:

- $x \in X_F$ - is a fact-generating ontological element;
- $f \in F_o$ - is an ontological function;
- $f(x) \in V'$ - is an analytically significant value.

An ontological fact is semantically defined; however, it does not yet possess coordinates within the multidimensional analytical space. The coordinates of an analytical fact are formed through the induction of relevant analytical dimensions from the ontological context of the fact-generating element.

Let D - denote the set of analytical dimensions previously defined as a subset of the set of ontological concepts. For each $x \in X_F$ the set of dimensions that are semantically relevant for constructing the corresponding analytical fact is defined as:

$$D_{axes}(x) = \left\{ y \in D \;\middle|\; \begin{matrix}(x,y) \in R_o \vee \\ (y,x) \in R_o\end{matrix} \right\}. \qquad (27)$$

Where:

- $D_{axes}(x) \subseteq D$ - is the set of analytical dimensions (ontological axes induced for the fact-generating element x);
- $x \in X_F$ - is the fact-generating ontological element;
- X_o - is the set of ontological elements of the domain;
- $R_o \subseteq X_o \times X_o$ - is the set of ontological relations;
- $y \in D$ - is an ontological element considered as a candidate dimension in the context of the fact induced by the element x.

The set D defines the complete space of admissible analytical dimensions of the domain, whereas $D_{axes}(x)$ specifies only those dimensions that are semantically relevant in the context of a particular fact. This makes it possible to formally describe situations in which not all dimensions participate in the formation of each analytical fact, without introducing artificial or semantically empty values.

Because a multidimensional analytical model operates not with dimensions themselves but with their values, a mapping is introduced for each fact-forming element that induces the values of the corresponding dimensions within the context of a given fact:

$$\delta\colon \{(x,y) \mid x \in X_F,\ \ y \in D_{axes}(x)\} \to V_D. \qquad (28)$$

Where:

- X_F - denotes the set of fact-forming ontological elements;
- V_D - is the set of admissible dimension values;
- $\delta(x,y)$ - is the value of dimension $y \in D_{axes}(x)$, induced in the context of the fact associated with element x.

The function δ therefore provides a formal transition from ontological concepts to the coordinates of the analytical space. Based on this, the ontologically induced multidimensional mapping function is defined as a specialization of μ, restricted to the set of ontological facts:

$$\mu_F\colon (x,f) \mapsto (d_1, d_2, \dots, d_k, m, v),$$
$$d_i = \delta(x, y_i) \in V_D, \quad i = 1, \dots, k. \qquad (29)$$

Where:

- μ_F - denotes the ontologically induced multidimensional mapping function;
- $x \in X_F$ - is a fact-forming ontological element;
- $f \in F_o$ - represents the ontological function;
- $D_{axes}(x) \subseteq D$ - the set of dimensions relevant to the fact x, $\langle y_1, \dots, y_k \rangle = ord\big(D_{axes}(x)\big)$;
- $k = \mid D_{axes}(x) \mid$ - an ordered list of relevant dimensions and their cardinality;
- δ - a mapping defined by formula (28), that induces the values of the dimensions for $y \in D_{axes}(x)$;
- $m = \Phi_M(f) \in M$ - an indicator corresponding to the ontological function,;
- $v = f(x) \in V'$ - the value of the indicator.

The resulting tuple $(d_1, \dots, d_k, m, v)$ constitutes an analytical fact of the multidimensional model and corresponds to a single row of the fact table. The coordinates of the fact are determined by the values of semantically relevant dimensions, whereas the measure is given by the value of the ontologically interpreted indicator.

The set of analytical facts is formed as the image of the set of ontological facts under the mapping μ_F:

$$F_{OLAP} = \left\{ \mu_F(x, f) \middle| \begin{matrix} x \in X_F, f \in F_o, \\ f(x) \in V' \end{matrix} \right\}, \qquad (30)$$

where F_{OLAP} denotes the set of analytical facts of the multidimensional model, μ_F - is the ontologically induced fact mapping, X_F - denotes the set of fact-forming ontological elements, F_o is the set of ontological functions, and V' - is the set of admissible indicator values.

Thus, the multidimensional analytical structure emerges as a formally induced derivative of the domain ontology, in which the fact table is constructed directly from ontological elements, relations, and functions. The logical model of analytical facts is based on the semantic relevance of dimensions and does not require filling missing coordinates with zero or null values. This property ensures correct aggregation, consistency of interpretation, and independence from the specific characteristics of the physical data storage implementation.

Consequently, the induced analytical facts are semantically traceable to their ontological origin. Each fact is determined by a fact-generating element, the relevant dimensional context, and an ontologically interpreted measure. The practical instantiation of these constructs is presented in Section 4.

3 ONTOLOGY-DRIVEN ANALYTICAL SYSTEM ARCHITECTURE

3.1 Conceptual Layered Structure

The proposed architecture of the ontology-driven analytical system is designed as a layered model that supports the transformation of heterogeneous domain data into semantically grounded multidimensional analytical representations. Its purpose is to organize the interaction between data structuring, ontology construction, multicriteria enrichment, analytical transformation, logical cube construction, and interpretation of results within a unified formal framework. The general layered architecture of the proposed system is shown in Figure 1.

The architecture reflects a sequence of ontology-analytical transformations, within which primary information sources are reduced to a normalized structure, interpreted as a formal ontology, enriched with evaluation results, and transformed into a multidimensional analytical model. In this framework, the ontology serves as an invariant semantic core from which dimensions, hierarchies, indicators, admissible aggregation operations, analytical facts, and a logical multidimensional cube are inductively derived.

Conceptually, the system consists of four functional-conceptual layers:

1) Data Structuring and Ontological Preparation Layer.
2) Ontological Formation, Management, and Cognitive Interpretation Layer.
3) Multicriteria Enrichment and Analytical Transformation Layer.
4) Interpretation, Navigation, and Visualization Layer.

These layers do not represent independent subsystems. Rather, they specify successive functional roles in the implementation of the ontology-analytical transformation. The first layer prepares semantically aligned data, the second constructs and maintains the ontology, the third enriches the ontology and constructs the multidimensional analytical model, and the fourth provides interpretation of the induced analytical structures through projections, rankings, filters, and visual representations.

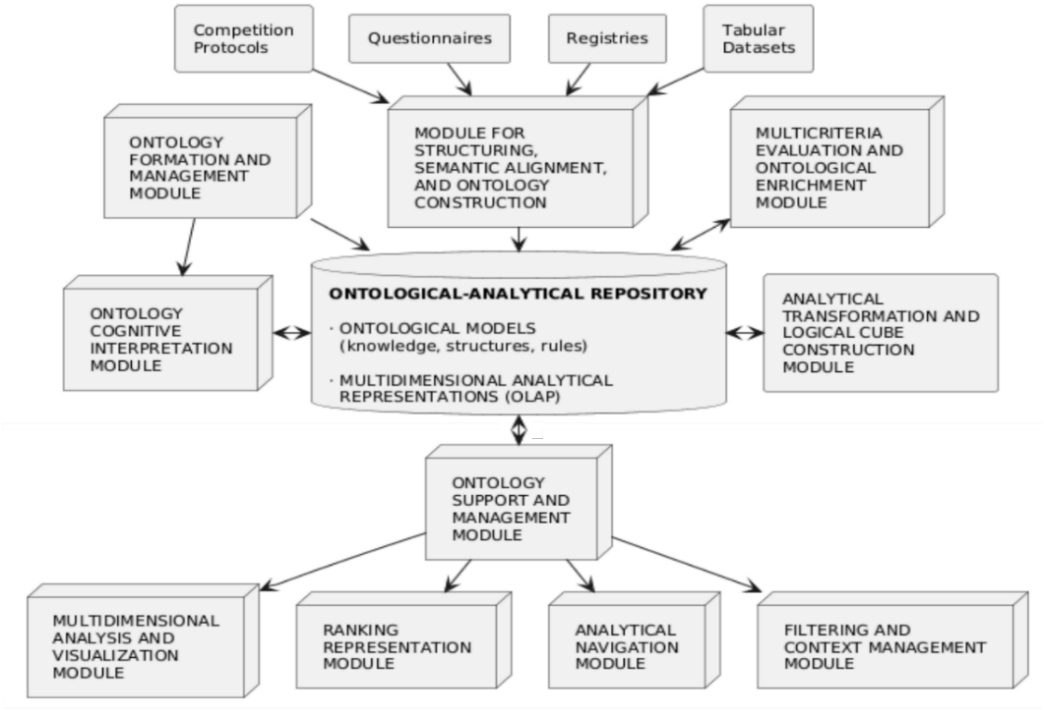


Figure 1: Conceptual architecture of the ontology-driven analytical system.

Table 1: Correspondence between architectural layers and ontology-analytical transformations.

Architectural layer	Formal component	Input	Output	Functional role
Data Structuring and Ontological Preparation	Structural transformation and semantic alignment operator - (1)-(3)	Documents, protocols, tables, registries, textual descriptions	Normalized object-attribute-value structure	Formation of a semantically aligned representation of source data
Ontological Formation, Management and Cognitive Interpretation	Ontological mapping operator and formal ontology model - (4), (6)-(9)	Normalized domain structure	Formal domain ontology with typed elements, relations and interpretation functions	Construction and maintenance of the invariant semantic model
Multicriteria Enrichment and Analytical Transformation - stage 1	Ranking model construction and ontology enrichment operators - (5)	Formal ontology	Enriched ontology with integrated evaluative or ranking values	Integration of multicriteria evaluation results into the ontology as semantically interpreted properties
Multicriteria Enrichment and Analytical Transformation - stage 2	Analytical transformation model, ontology-analytical mapping operator and multidimensional model - (10)-(30)	Enriched ontology	Multidimensional analytical model / logical analytical cube with dimensions, hierarchies, measures, aggregation rules and analytical facts	Construction of the ontology-induced logical cube from semantic structures and fact-generating elements
Interpretation, Navigation and Visualization	Analytical projections and multidimensional interpretation mechanisms - (12), (13), (29), (30)	Analytical facts, indicators, rankings and multidimensional structures	Rankings, analytical slices, filters, visual representations	Practical interpretation and exploration of ontology-induced analytical results

Within this architecture, the third layer is particularly important, as it implements the transition from formal ontology to a multidimensional analytical structure. This transition corresponds to the formal operators introduced in Section 2 and links the extended ontology O^* to the multidimensional analytical model M_{OLAP}.

The first stage of the analytical transformation forms an extended ontology with integrated evaluation values; the second stage transforms this ontology into a multidimensional model containing dimensions, hierarchies, indicators, aggregation rules, and analytical facts. The transformation implemented by the architecture can be represented as the following sequence:

$$S \xrightarrow{\Phi_{olap}} N \xrightarrow{\Phi_{ont}} O \xrightarrow{\Omega_{MR}} O^* \xrightarrow{\Phi_{olap}} M_{OLAP}.$$

This sequence shows that the multidimensional analytical model is not introduced independently of the ontology. It is constructed as a derived structure based on the enriched ontology and the ontology-analytical mapping operator. This correspondence is summarized in Table 1.

3.2 Data Structuring and Ontological Preparation Layer

The first layer transforms heterogeneous sources into a normalized representation suitable for further ontological interpretation. Its input may include electronic forms, competition protocols, registries, tabular datasets, questionnaires, reports and textual descriptions. These sources may differ in structure, terminology, completeness and level of formalization; therefore, the primary task of the layer is to eliminate structural, lexical and semantic heterogeneity.

At this level, the transition from source documents to a normalized object-attribute structure takes place. Such a structure captures the entities of the domain, their attributes, values, types, and basic relationships between them. It is not yet an ontology and does not contain an analytical model. Its function is to prepare a formally consistent representation for subsequent ontological mapping.

This layer includes the Module for Data Structuring and Semantic Alignment. It performs ingestion and consolidation of source data, normalizes terms and attributes, aligns equivalent or related values, and produces a coherent representation that can be used as input for ontology construction.

No comparison, ranking, aggregation or cube construction is performed at this stage. The output of the layer is a normalized structure that preserves the relevant domain content while eliminating inconsistencies that would prevent further semantic interpretation.

3.3 Ontological Formation, Management and Cognitive Interpretation Layer

The second layer transforms the normalized domain structure into a formal ontology and maintains this ontology as the central semantic mechanism of the system. The ontology defines concepts, individual objects, relations, typing rules and interpretation functions that determine how domain data may be consistently understood and further processed.

Within this layer, normalized records are given an ontological interpretation. Values that may have been simple fields at the input structure level can, after ontological mapping, become distinct entities, relationships, or functions. Contextual, classificatory, spatial, temporal, organizational, and personal characteristics are interpreted as ontological elements associated with relevant events or objects in the domain. Numerical or ordered values can be interpreted as ontological functions if they have domain-specific, evaluative, or analytical semantics.

The Ontology Formation and Management Module supports the structure and consistency of the ontology as the domain evolves. It controls changes in concepts, relations, types and interpretation functions without requiring manual redesign of analytical tables.

The Ontology Cognitive Interpretation Module uses the ontology as a source of semantic constraints for subsequent evaluation and analytical transformation. This layer does not calculate rankings, does not perform aggregation and does not construct the multidimensional cube. It defines the semantic conditions under which such operations become valid.

Thus, the ontology formed at this layer serves as the invariant semantic model of the domain. It specifies which elements can later function as contextual dimensions, which relations may define hierarchies, which functions may become measures, and which elements may generate analytical facts.

3.4 Multicriteria Enrichment and Analytical Transformation Layer

The third layer implements the analytical core of the system. At this level, the formal ontology is used not only as a semantic description of the domain but also

as the basis for multicriteria enrichment, analytical transformation and construction of the multidimensional analytical model.

This layer includes two functionally distinct modules: the Multicriteria Evaluation and Ontological Enrichment Module and the Analytical Transformation and Logical Cube Construction Module.

The Multicriteria Evaluation and Ontological Enrichment Module defines evaluation criteria, weights and rules of interpretation according to the ontology of the domain. Primary evaluative properties are interpreted within the domain model and transformed into semantically defined evaluative values. The resulting values are integrated into the ontology as properties of the corresponding ontological elements rather than as external technical fields. At this stage, an enriched ontology is formed in which the primary evaluative characteristics of the subject domain may be supplemented with integral rating values, normalized scores, or other interpreted indicators.

The Analytical Transformation and Logical Cube Construction Module implements the transition from the enriched ontology to the multidimensional analytical model. This is the stage at which the logical analytical cube is constructed. Contextual and classificatory ontology elements are mapped to dimensions, ontological relations define hierarchies, analytically meaningful functions are mapped to measures, admissible aggregation rules are assigned to measures, and fact-generating ontology elements produce analytical facts.

This module is responsible for actually building the multidimensional cube. Here, the cube is not necessarily understood as a physical OLAP structure in a specific storage system, but rather as a logical multidimensional analytical model. It includes a set of dimensions, hierarchies, measures, admissible aggregation operations, and analytical facts. It is this model that is subsequently used to build rankings, slices, projections, filters, and visualizations.

As a result, the multidimensional model is not manually imposed on the data. It is derived from the logical organization of the ontology and from the enriched semantic representation of the domain. The analytical cube therefore preserves the connection between domain semantics and analytical operations.

For the domain of intellectual competitions, such a transformation may derive dimensions such as year, region, institution, competition, section, participant and mentor. Measures may include ranking score, achievement indicators and aggregated counts. The admissibility of aggregation depends on the semantic role of the measure: counts may be summed, ranking scores may be averaged or compared, and ordinal characteristics require restricted interpretation.

Thus, the third layer is responsible not only for ranking or evaluation, but also for the formal construction of the ontology-induced multidimensional analytical cube.

3.5 Interpretation, Navigation and Visualization Layer

The fourth layer provides practical access to the results of ontology-driven analysis. It operates on analytical facts, indicators, rankings and multidimensional projections generated by the previous layer. It does not modify the ontology and does not construct a new analytical cube. Instead, it supports interpretation, filtering, visualization and navigation across levels of detail.

At this level, the user gains access to various projections of a single ontologically induced analytical structure. These may include ranking views, aggregated snapshots, comparative projections, filters by dimensions, as well as transitions between aggregated measures and detailed analytical facts.

The layer includes Semantic Consistency Support, Multidimensional Visualization, Ranking Representation, Analytical Navigation and Contextual Filtering. These functions allow users to explore analytical slices, restrict results by selected dimensions, compare aggregated indicators and move between levels of granularity without changing the underlying ontology or the induced multidimensional model.

The fourth layer therefore works with the already constructed logical cube. It produces user-oriented representations of the analytical model but does not change the model itself. This separation prevents visualization, filtering or navigation procedures from altering the semantic basis of the analysis.

Thus, the proposed architecture establishes a coherent interaction between semantic modeling, multicriteria enrichment, analytical transformation, logical cube construction and interpretation. The ontology remains the invariant foundation of the system, while the multidimensional analytical model and user-oriented representations are derived from it in a controlled and semantically consistent manner.

4 RESULTS AND DEMONSTRATION OF THE PROPOSED MECHANISM

4.1 Demonstration Scenario and Input Records

To demonstrate the formal instantiation of the proposed mechanism, we consider the problem of multidimensional analysis of the results of intellectual competitions. This subject area is representative of the proposed approach, as it combines diverse data sources, the organizational structure of competitions, personal relationships between participants and mentors, the spatial distribution of educational institutions, as well as evaluation characteristics that require a coordinated analytical representation.

The system receives a set of documents defined in the formal model (1). Each document may correspond to a competition protocol, a participant registry, a results table, a list of sections, a list of educational institutions, or a description of a competition event. The demonstration fragment examines four records of student participation in competitions. A sample of the input records is presented in Table 2.

Each entry in Table 2 describes not an isolated numerical result, but an instance of a specific student's participation in a particular competitive context.

Therefore, the entry s_i is interpreted as a source for constructing an individual ontological element Participation:

$$s_i \to p_i,\ \tau_X(p_i) = Participation.$$

In the example snippet, the set of elements looks like this:

$$\mathcal{P} = \{p_1, p_2, p_3, p_4\}.$$

The Participation element is interpreted as an ontological representation of a participation event, rather than as the technical equivalent of a table row. It integrates personal, organizational, temporal, spatial, classificatory, and evaluative contexts, on the basis of which analytical facts will subsequently be derived.

4.2 From Input Fields to Ontological Interpretation

At the first stage, the structural normalization operator Φ_{norm} defined in (2) is applied. Its result is a normalized object-attribute structure N formally specified in (3).

For the corresponding record s_i, the normalized object p_1 can be represented as a structured description of Student 1's participation in the National research competition in 2025:

$$p_i = \begin{pmatrix} 2025, NRC, CS, AI, KyivRegion, \\ Lyceum_A, Student_1, Mentor_1, 92, I \end{pmatrix}.$$

At the level of the normalized structure, the fields competition, student, region, institution, section, and mentor may be represented as attributes of the object p_i. However, after ontological mapping, they do not remain simple attribute values. The corresponding values are interpreted as separate ontological entities connected with Participation through semantic relations. Table 3 summarizes how the main input fields are interpreted in the ontological model.

Table 2: Fragment of input competition records.

Record	Year	Competition	Field	Section / discipline	Region	Institution	Student	Mentor	Protocol score	Place
s_1	2025	National research competition	Computer Science	Artificial intelligence systems	Kyiv region	Lyceum A	Student 1	Mentor 1	92	I
s_2	2025	National research competition	Computer Science	Artificial intelligence systems	Lviv region	Lyceum B	Student 2	Mentor 2	84	II
s_3	2025	Informatics olympiad	Informatics	Programming	Kyiv region	Lyceum A	Student 3	Mentor 1	78	III
s_4	2024	National research competition	Applied Mathematics	Mathematical modeling	Kharkiv region	Gymnasium C	Student 4	Mentor 3	88	II

Table 3: Interpretation of input fields in the ontological model

Input field	Normalized representation	Ontological interpretation
competition	value of competition attribute in p_i	individual of class Competition, connected with p_i by *hasCompetition*
student	value of student attribute in p_i	individual of class Student, connected with p_i by *hasStudent*
mentor	value of mentor attribute in p_i	individual of class Mentor, connected with p_i by *hasMentor*
institution	value of institution attribute in p_i	individual of class Institution, connected with p_i by *hasInstitution*
region	value of region attribute in p_i	individual of class Region, connected with p_i by *hasRegion*
field	value of field attribute in p_i	individual of class Field, connected with p_i by hasField
section / discipline	value of section or discipline attribute in p_i	individual of class Section or Discipline, connected with p_i by *hasSection* or *hasDiscipline*
year	temporal value in p_i	contextual value connected with p_i by *hasYear*
protocol score	numerical value in the protocol	source evaluative function $ProtocolScore(p_i)$
place	ordered competition result	ordinal evaluative function $Place(p_i)$

This distinction is essential for the correct transition from a tabular representation to an ontological model. Not every field of a normalized record is an attribute in the narrow ontological sense, and not every numerical value automatically becomes an analytical measure. For example, the year value is numerical, but it functions as a temporal coordinate. By contrast, the protocol score and award place have evaluative semantics because they characterize the result of a specific participation.

4.3 Ontological Fragment Centered on Participation

At the second stage, the normalized structure is interpreted as a formal ontology using the operator Φ_{ont} defined in (4). The structure of the formal ontology $O$is defined in (6), while its typing, relations, and functions are specified according to (7)-(9).

Within the demonstration domain, the ontology fragment is constructed around the Participation element, which serves as the central fact-forming node (Fig. 2).

In this fragment, the relations between *Participation* and the entities *Student*, *Mentor*, *Competition*, *Year*, *Section*, *Institution*, and *Region* belong to the set of ontological relations defined in (8). They specify the semantic context of participation. The evaluative characteristics *ProtocolScore*, *Place*, and *RankingScore* belong to the set of ontological functions F_o defined in (9), since they are defined as functions on *Participation*.

For the corresponding Participation instance p_1, the relations are instantiated as follows:

- $hasStudent(p_1, Student\ 1)$;
- $hasMentor(p_1, Mentor\ 1)$;
- $hasCompetition(p_1, NationalResearchComp)$;
- $hasYear(p_1, 2025)$;
- $hasField(p_1, ComputerScience)$;
- $hasSection(p_1, ArtificialIntelligenceSystems)$;
- $hasInstitution(p_1, Lyceum\ A)$;
- $hasRegion(p_1, KyivRegion)$;

The primary evaluative characteristics are defined as functions on the set of participations:

$$ProtocolScore: \mathcal{P} \to V_{score},\ Place: \mathcal{P} \to V_{place}.$$

For the first participation:

$$ProtocolScore(p_1) = 92,\quad Place(p_1) = I.$$

4.4 Multicriteria Enrichment and Logical Cube Construction

After constructing the ontology, the multicriteria evaluation model is applied MR, and the evaluation results are integrated into the ontology by the operator Ω_{MR}defined in (5). In the demonstration example, the model converts the primary evaluative characteristics of a participation event - the protocol score and award place - into an integral rating value.

The following function is used to illustrate this:
$R(p_i) = 0.7 \cdot ProtocolScore(p_i) + 30 \cdot W_{place}(p_i)$,
where: $W_I = 1.00,\ W_{II} = 0.85,\ W_{III} = 0.70$.
This function is a demonstration instance of the model. In a complete system, the evaluation model may take into account the level of the competition, the weight of the section, the type of competition, scale normalization rules, criterion weights, and other parameters defined by the domain ontology. The results of enriching the Participation elements are presented in Table 4.

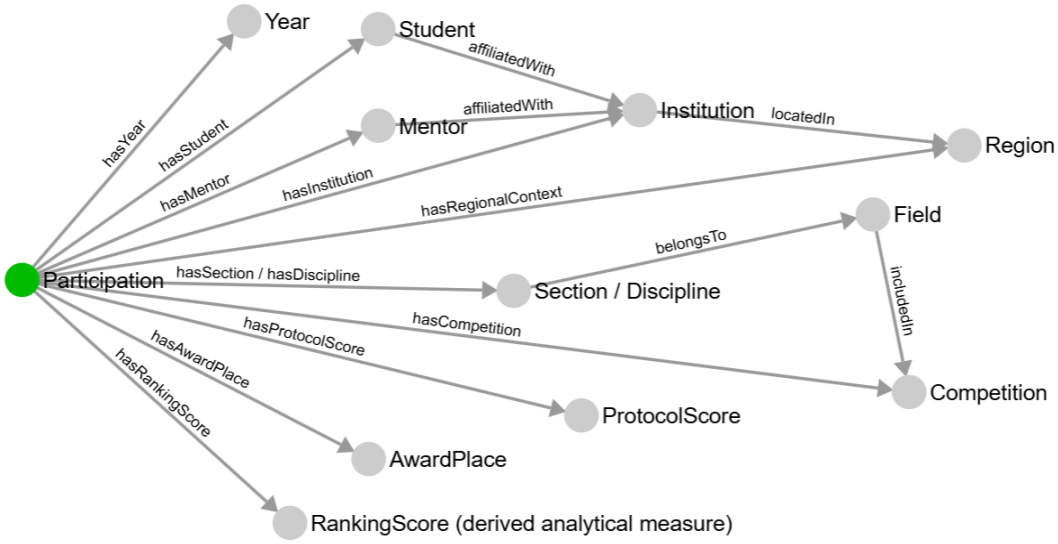


Figure 2: Fragment of the domain ontology centered on participation.

Table 4: Multicriteria enrichment of participation elements.

Participation	*Protocol score*	*Place*	W_{place}	*RankingScore*
p_1	92	I	1.00	94.4
p_2	84	II	0.85	84.3
p_3	78	III	0.70	75.6
p_4	88	II	0.85	87.1

After applying the evaluation model Ω_{MR}, the ontology transitions to an enriched state O^* in which the rating values are interpreted as ontological functions of participation:

$$RankingScore(p_1) = 94.4;$$
$$\mathrm{RankingScore}(\mathrm{p}_2) = 84.3;$$
$$\mathrm{RankingScore}(\mathrm{p}_3) = 75.6;$$
$$RankingScore(p_4) = 87.1.$$

Next, the ontology-analytical mapping operator Φ_{olap} defined in (14) is applied to map the extended ontology O^* to the multidimensional analytical model M_{OLAP} defined in (10). In this section, M_{OLAP} is viewed as a logical analytical cube, that is, as a structure containing dimensions, hierarchies, measures, admissible aggregation operations, and analytical facts.

In the example provided, the dimensions are inductively derived from ontological elements related to *Participation*:

$$D = \begin{Bmatrix} Year, Region, Institution, Competition, \\ Field, Section, Discipline, Student, Mentor \end{Bmatrix}.$$

Measures are not derived from all numerical attributes, but only from those ontological functions or derived values that have analytical semantics with respect to the fact-forming element. For example:

$$M = \{RankingScore, AchievementIndicator\}.$$

The achievement indicator is defined for individual participationas:

$$AchievementInd(p_i) = \begin{cases} 1, & Place(p_i) \in \{I, II, III\}, \\ 0, & otherwise. \end{cases}$$

Measures for the number of participations or achievements are generated at the level of an aggregated analytical group G, for example, a region, educational institution, section, or supervisor:

$$ParticipationCount(G) = |\mathcal{P}_G|,$$
$$AchievementCount(G) = \sum_{p_i \in \mathcal{P}_G} AchievementInd(p_i),$$

where $\mathcal{P}_G$ - is the set of participations belonging to the group G. The set of basic aggregation operations is defined in (22), and for each indicator, a semantically admissible subset of operations is determined according to (23). The semantic roles of indicators and admissible aggregation operations are summarized in Table 5.

Whether aggregation is permissible depends not only on the type of value but also on its subject-specific semantics. Therefore, participation counts can be summed, achievement indicators can be aggregated, and ranking scores can be summed, averaged or compared, but ordinal characteristics such as rank should not be arbitrarily summed.

The hierarchies of the logical cube are derived from ontological relations R_o (8). The following links are relevant in the demo domain:

$$Student \rightarrow Institution \rightarrow Region;$$
$$Mentor \rightarrow Institution \rightarrow Region;$$
$$Section \rightarrow Field \rightarrow Competition;$$
$$Institution \rightarrow Region \rightarrow Country.$$

Thus, the logical analytical cube is derived from an extended ontology rather than being a pre-designed data schema.

4.5 Ontologically Induced Analytical Facts

The fact-forming ontological elements are defined in the formal model (25). In the demonstration domain, the set of instances is a subset of these elements: $\mathcal{P} \subseteq X_F$, since for each entry $p_i \in \mathcal{P}$ there are functions that have an evaluative or analytical interpretation:

$$ProtocolScore(p_i) \in V_{score}, \quad Place(p_i) \in V_{place}, \quad RankingScore(p_i) \in V_{rank}$$

For each instance p_i a set of relevant dimensions is determined D_{p_i}, which corresponds to the formal description of context-dependent axes of analysis in (27). For the participation instance p_1:

$$D_{p_1} = \begin{Bmatrix} Year, Region, Institution, Competition, \\ Field, Section, Student, Mentor \end{Bmatrix}.$$

The values of these dimensions in the context of a specific instance are determined by the function δ (28). For the participation instance p_1:

$$\delta_{p_1}(Year) = 2025, \quad \delta_{p_1}(Region) = KyivRegion,$$
$$\delta_{p_1}(Institution) = Lyceum\ A,$$
$$\delta_{p_1}(Competition) = NationalResearchCompetition,$$
$$\delta_{p_1}(Field) = ComputerScience,$$
$$\delta_{p_1}(Section) = ArtificialIntelligenceSystems,$$
$$\delta_{p_1}(Student) = Student\ 1, \delta_{p_1}(Mentor) = Mentor\ 1$$

The formulation of an analytical fact follows the general structure f_{OLAP} (13), and a specialized ontologically induced mapping μ_F, (29). For the participation instance p_1, based on the RankingScore measure, we obtain the following analytical fact:

$$f_{OLAP}(p_1, RScore) = \begin{pmatrix} 2025, KyivRegion, Lyceum\ A, \\ NRC, CS, AI, Student\ 1, \\ Mentor\ 1, RankingScore, 94.4 \end{pmatrix}$$

Table 5: Measures and admissible aggregation operations.

Ontological function and Semantic role	Analytical use	Admissible aggregation
$ProtocolScore(p_i)$, source evaluative function	input for ranking model; optional analytical variable	avg, min, max, if analytically allowed
$Place(p_i)$, ordinal evaluative function	input for ranking model	comparison, ordering
$RankingScore(p_i)$, derived analytical measure	ranking and aggregation	sum, avg, min, max
$AchievementInd(p_i)$, *derived analytical measure*	basis for counting achievements	sum, count
$Year(p_i)$, contextual value	dimension value	not a measure

Table 6: Analytical facts induced from Participation elements.

Fact	Year	Region	Institution	Competition	Field	Section / discipline	Student	Mentor	Measure	Value
f_1	2025	Kyiv region	Lyceum A	National research competition	Computer Science	Artificial intelligence systems	Student 1	Mentor 1	RankingScore	94.4
f_2	2025	Lviv region	Lyceum B	National research competition	Computer Science	Artificial intelligence systems	Student 2	Mentor 2	RankingScore	84.3
f_3	2025	Kyiv region	Lyceum A	Informatics olympiad	Informatics	Programming	Student 3	Mentor 1	RankingScore	75.6
f_4	2024	Kharkiv region	Gymnasium C	National research competition	Applied Mathematics	Mathematical modeling	Student 4	Mentor 3	RankingScore	87.1

The analytical facts induced from the demonstration recordsare presented in Table 6.

For the same participation instance, other analytical facts may be generated when the indicator changes. For example, for the achievement indicator:

$$f_{OLAP}(p_1, AchInd) = \begin{pmatrix} 2025, KyivRegion, Lyceum\ A, \\ NRC, CS, AI, Student\ 1, \\ Mentor\ 1, AchInd, 1 \end{pmatrix}$$

Thus, a single Participation element can generate several analytical facts if it is associated with multiple analytically significant functions or derived indicators. The set of such facts f_{OLAP} is formally defined in (30).

4.6 Multidimensional Aggregation and Analytical Projections

Once the analytical facts have been formed, the multidimensional function defined in (12) provides indicator values for various combinations of dimensions. For example, a regional projection is formed by aggregating facts by dimension $Region$.

For each regional group G, the number of participations is determined as:

$$ParticipationCount(G) = |\mathcal{P}_G|.$$

The sum of the rating values:

$$TotalRankingScore(G) = \sum_{p_i \in \mathcal{P}_G} RankingScore(p_i),$$

and the average rating:

$$AverageRankingScore(G) = \frac{TotalRankingScore(G)}{ParticipationCount(G)}.$$

The regional analytical projection is presented in Table 7.

Table 7: Regional analytical projection.

Region	Number of participations	Sum of ranking scores	Average ranking score
Kyiv region	2	170.0	85.0
Lviv region	1	84.3	84.3
Kharkiv region	1	87.1	87.1

Similar projections can be constructed based on other dimensions, such as educational institution, supervisor, section, field, or competition. These results are not distinct independent models, but rather different projections of a single multidimensional structure M_{OLAP} induced from the ontology.

Changing the level of analysis does not require redesigning the data schema. It involves selecting a different subset of dimensions and applying semantically valid aggregation operations to the corresponding measures. The key result is that the analytical structure is formed not through prior manual design of an OLAP schema, but through an ontological interpretation of participation as an event to which personal, organizational, spatial, temporal, classificatory, and evaluative characteristics are attached. In such a model, the participation result is not an isolated numerical field, but a function defined on the ontological element Participation.

The resulting multidimensional model supports analytical views at the level of the student, the administrator, the educational institution, the region, the section, the field, or the competition. All of these views are different projections of a single ontologically induced structure. This allows for maintaining the semantic integrity of the data, controlling the validity of aggregation, and correctly working with context-dependent dimensions.

5 CONCLUSIONS

The article develops a formally grounded approach to the construction of ontology-driven information and analytical systems, in which the formal domain ontology serves as the primary invariant knowledge model, used for the inductive formation of analytical structures, indicators, hierarchies, and admissible aggregation operations. The proposed approach is oriented toward the consistent integration of the semantic model of the domain and multidimensional analytics within a unified formal framework, while minimizing ad hoc assumptions through explicitly defined ontology-analytical mapping rules.

A key result of the study is the formalization of the ontology-analytical mapping operator, which defines a formal transition from the ontological model to a multidimensional analytical representation, taking into account the typing of ontological elements, semantic relations, and interpretative functions. Such an operator supports the derivation of analytical dimensions, hierarchies, and indicators directly from the logical organization of the domain, while preserving semantic constraints at the stages considered in the proposed transformation model.

Within the proposed formalization, it is shown that analytical facts can be induced from ontological elements and their interpretations without the prior specification of fixed analytical schemas. This makes it possible to formally describe analytical situations with context-dependent or partially defined dimensions, support the semantic consistency of aggregation, and reduce the dependence of the logical analytical model on specific implementations of physical data storage or computation.

The proposed architecture of the ontology-driven analytical system formalizes the interaction of semantic, analytical, and interpretative components

in the form of functional-conceptual layers with clearly defined roles. Such an architecture separates data structuring, ontology construction, multicriteria enrichment, analytical transformation, logical cube construction, and user-oriented interpretation, while preserving the connection between the semantic model of the domain and the induced analytical structures.

The developed approach is general in nature and can be applied to various classes of information and analytical tasks, including the analysis and evaluation of intellectual and educational achievements as a specific applied case. The demonstration presented in the article illustrates how input records can be transformed into Participation-centered ontological elements, enriched with evaluation values, and further used to induce dimensions, measures, aggregation rules, and analytical facts.

The obtained results provide a formal foundation for further research on ontology-driven analytical systems, including the automated derivation of analytical models, the integration of multicriteria evaluation with semantic structures, and the development of semantically traceable analytical and ranking mechanisms in complex information and analytical environments.

ACKNOWLEDGMENTS

The authors express their gratitude to colleagues and experts who contributed to the discussion of the conceptual provisions of the study.

REFERENCES

[1] R. Djiroun, K. Boukhalfa and Z. Alimazighi, “Data cubes retrieval and design in OLAP systems: From query analysis to visualisation tool,” International Journal of Business Intelligence and Data Mining, vol. 14, nos. 1/2, pp. 267-298, 2019, doi: 10.1504/IJBIDM.2019.096813.

[2] A. Abelló et al., “Using semantic web technologies for exploratory OLAP: A survey,” IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 2, pp. 571-588, 2015, doi: 10.1109/TKDE.2014.2330822.

[3] D. Martinez-Mosquera, R. Navarrete, S. Luján-Mora, L. Recalde and A. Andrade-Cabrera, “Integrating OLAP with NoSQL databases in big data environments: Systematic mapping,” Big Data and Cognitive Computing, vol. 8, no. 6, art. 64, 2024, doi: 10.3390/bdcc8060064.

[4] M. Ptiček, B. Vrdoljak and M. Gulić, “The potential of semantic paradigm in warehousing of big data,” Automatika, vol. 60, no. 4, pp. 393-403, 2019, doi: 10.1080/00051144.2019.1630582.

[5] T. R. Gruber, “A translation approach to portable ontology specifications,” Knowledge Acquisition, vol. 5, no. 2, pp. 199-220, 1993, doi: 10.1006/knac.1993.1008.

[6] N. Guarino, “Understanding, building and using ontologies,” International Journal of Human-Computer Studies, vol. 46, nos. 2-3, pp. 293-310, 1997, doi: 10.1006/ijhc.1996.0091.

[7] M. Uschold and M. Gruninger, “Ontologies: Principles, methods and applications,” Knowledge Engineering Review, vol. 11, no. 2, pp. 93-136, 1996, doi: 10.1017/S0269888900007797.

[8] S. Staab and R. Studer, Eds., Handbook on Ontologies. Berlin, Heidelberg: Springer, 2009, doi: 10.1007/978-3-540-92673-3.

[9] R. P. Deb Nath, O. Romero, T. B. Pedersen and K. Hose, “High-level ETL for semantic data warehouses,” Semantic Web, vol. 13, no. 1, pp. 85-132, 2022, doi: 10.3233/SW-210429.

[10] A. L. Antunes, J. Barateiro and E. Cardoso, “Application of semantic web techniques in DW/BI systems for strategic management,” IEEE Access, vol. 13, pp. 163103-163119, 2025, doi: 10.1109/ACCESS.2025.3610577.

[11] H. P. Bomma, “Enhancing business analytics with a semantic layer,” International Journal of Business Quantitative Economics and Applied Management Research, vol. 7, no. 9, pp. 93-99, 2023.

[12] O. Romero and A. Abelló, “Automating multidimensional design from ontologies,” in Proceedings of the ACM 10th International Workshop on Data Warehousing and OLAP (DOLAP ’07), Lisbon, Portugal, 2007, pp. 1-8, doi: 10.1145/1317331.1317333.

[13] R. Mizoguchi and J. Bourdeau, “Using ontological engineering to overcome AI-ED problems: Contribution, impact and perspectives,” International Journal of Artificial Intelligence in Education, vol. 26, no. 1, pp. 91-106, 2016, doi: 10.1007/s40593-015-0077-5.

[14] G. Futia and A. Vetrò, “On the integration of knowledge graphs into deep learning models for a more comprehensible AI: Three challenges for future research,” Information, vol. 11, no. 2, art. 122, 2020, doi: 10.3390/info11020122.

[15] R. Confalonieri and G. Guizzardi, “On the multiple roles of ontologies in explanations for neuro-symbolic AI,” Neurosymbolic Artificial Intelligence, vol. 1, pp. 1-14, 2025, doi: 10.3233/NAI-240754.

[16] M. Nadutenko, V. Prykhodniuk, V. Shyrokov and O. Stryzhak, “Ontology-driven lexicographic systems,” in Advances in Information and Communication. FICC 2022. Lecture Notes in Networks and Systems, vol. 438. Cham, Switzerland: Springer, 2022, pp. 204-215, doi: 10.1007/978-3-030-98012-2_16.

[17] O. Stryzhak, V. Horborukov, V. Prychodniuk, O. Franchuk and R. Chepkov, “Decision-making system based on the ontology of the choice problem,” Journal of Physics: Conference Series, vol. 1828, no. 1, art. 012007, 2021, doi: 10.1088/1742-6596/1828/1/012007.

Software Implementation of Application for Correlation Analysis

Vitalii Bazurin, Oleg Pursky, Ihor Tyschenko, Andrey Nechepourenko and Alexander Vishnevsky
Department of Computer Science and Information Systems, State University of Trade and Economics, Kyoto Str. 19, 02156 Kyiv, Ukraine
v.bazurin@knute.edu.ua, o.pursky@knute.edu.ua, i.tyshchenko@knute.edu.ua, a.nechepurenko@knute.edu.ua, o.vyshnivskyy@knute.edu.ua

Keywords: Correlation, Statistics, Software Library, Pearson Correlation, Spearman Correlation, Kendall Correlation, Algorithm, Application.

Abstract: The article contains the results of research on the development of an application for determining the correlation between two data samples. Such an application is relevant for correlation analysis. The central component of the applications the Correlation library, which contains classes and methods for calculating the Pearson, Spearman, and Kendall correlation coefficients. The library is implemented in Python. The created Correlation library utilizes built-in Python data types and does not rely on third-party libraries. The article describes the algorithms used to calculate these coefficients, as well as the Python implementations for calculating the specified correlation coefficients. For the practical implementation of the library, the PyCorrelation application has been developed in two versions. The interface of the first version of the application is implemented using the PyQT 5 library; the interface of the second version is implemented using the free Custom Tkinter library. The results of calculating the Pearson, Spearman, and Kendall correlation coefficients are verified using MS Excel spreadsheets. The developed Correlation library has significant prospects for modernization and use as a component of software tools for statistical processing of experimental data. It can also be integrated into future versions of the StatCriterion software tool, previously developed by the authors of the article.

1 INTRODUCTION

In the process of conducting many studies, including pedagogical ones, there is a need to establish the existence of a relationship between two phenomena. Quantitative characteristics of such phenomena are usually expressed in numerical arrays. For example, the relationship between voltage usage and current fluctuations, algebra grades and chemistry grades, algebra test results and physics test results, etc. To determine such relationships, there are methods of correlation analysis: Pearson, Spearman, and Kendall. According to R. Taylor, correlation analysis is of great importance in scientific research. All these methods differ in algorithm, but provide high reliability of the calculated results. The only drawback of these methods, in our opinion, is that these calculations are cumbersome, and it takes a lot of time to perform them manually. That is why it is advisable to develop a software tool that will automate correlation calculations and draw conclusions from the calculations performed. In the case of a component-oriented architecture of such a software tool, one of its main components will be a library of functions for calculating correlation coefficients. Such libraries for the Python language already exist: NumPy and Pandas. Using these libraries, you can create a program to determine the existence of a relationship between two data sets. However, here too, problems may arise in the further use of such a program. There is a possibility, albeit insignificant, that the release of the NumPy and Pandas libraries for new versions of Python may be discontinued. That is why, to develop and support software, it is advisable to first develop your own modules in the appropriate programming languages. The purpose of the article is to reveal the features of building algorithms and software code for a Python library for calculating Pearson, Spearman, and Kendall correlation coefficients.

2 LITERATURE REVIEW

The problem of applying correlation in the practice of scientific research has two aspects. The first is to develop correlation methods that will unambiguously show the presence or absence of a relationship between two data sets. R. Taylor reveals the role of correlation analysis in scientific research, the features of calculating Pearson's correlation [1]. The Pearson correlation coefficient varies within [-1;1] and characterizes the presence or absence of a relationship between two data sets. B. Ratner, in his article, presents an algorithm and basic formulas for calculating the Pearson correlation coefficient [2]. The article provides examples of calculating the Pearson correlation coefficient in the form of tables. D. Chicco and G. Jurman in their article [3] reveal the meaning of the Matthews correlation coefficient and give examples of using this coefficient to determine whether there is a relationship between two sets of features (even of different sizes). The Matthews coefficient takes into account true positives and negatives, false positives and false negatives. J.Jiang, X.Zhang, and Zhong Yuan in their article demonstrate an example of the application of Spearman correlation in the field of machine learning [4]. S.Chatterjee in his article states that the most popular classical correlation coefficients are Pearson, Spearman and Kendall coefficients [5]. S.Chatterjee developed his own correlation coefficient, which was called the Chatterjee coefficient. X.Su, S.Shang, Z.Xu, H.Quian and X.Pan conducted a study of risk factors using Pearson correlation [6]. In particular, they used Pearson correlation to determine the type and strength of the relationship between two factors of productivity formation. H.Yu and A.D.Hutson in their study check the validity of Spearman correlation for different samples [7]. Scientists prove that the test they developed to check the validity of Spearman correlation is valid for small and large data sets. F.Han and Z.Huang adapted Azadkia-Chatterjee's correlation coefficient for different data [8]. In their article, scientists prove the theorem and 11 lemmas. C.A.Metler, R.A.Vannatta and K.N.LaVenia in their book reveal the features of algorithms for calculating univariate and multivariate correlation [9].

The second direction of research is to create software tools that can automate the process of calculating the correlation coefficient and formulating a conclusion about the presence or absence of a relationship between two data sets. For example, V. Mikhalchuk in his study gives examples of creating pseudocode for calculating the t-test and p-value for two samples [10]. But to calculate the p-value, V. Mikhalchuk uses the function of the corresponding statistical library. H. Kang in his study uses the G* Power software tool to perform statistical data processing, including calculating the Pearson correlation coefficient [11]. Using the Pearson correlation, the presence or absence of a relationship between two samples is determined. The author recommends the G* Power software tool for performing statistical calculations. R.Bivand, G.Millo and G.Piras analyze the software tools for statistical data processing available in the R system [12]. Scientists have developed appropriate tests for evaluating statistical criteria, including correlation coefficients. L.Qi, W.Lin, X.Zhang, W.Dou, X.Xu and J.Chen investigate the issue of using the correlation graph in order to select and recommend web APIs for mobile applications [13]. S.Mguil-Touchal, F.Morestin and M.Brunei in their article demonstrate examples of creating experimental software for calculating correlation [14]. The article provides examples of using software tools for calculating the correlation coefficient in the process of studying the mechanical properties of solids. S.D.Rahmawati, A.F. Ihsan, F.Priadi and U.W.Siagian in their article [15] show the results of developing a software tool for studying the physical properties of hydrocarbon liquids. V.M. Bazurin, O.I. Pursky and O.S. Chashechnikova in their article demonstrate the structure and interface of an application for statistical processing of experimental results [16]. The developed application compares two samples of values using the Pearson, Whitney-Mann, and Student criteria. It would be advisable to supplement this application with correlation functions.

The analysis of research on the creation, justification of correlation coefficients and their use in practice shows that the problem of developing correlation methods is relevant and continues to be studied by many scientists.

3 OBJECTIVES AND METHODS

(RQ1) How to build algorithms for calculating Pearson, Spearman, and Kendall correlation coefficients and write the code as a software library in Python?

(RQ2) How to integrate the developed library into a Python application to provide a convenient interface and the necessary functionality?

Problem: How to build algorithms for calculating Pearson, Spearman, and Kendall correlation

coefficients and present them as a software library integrated into the application?

In the process of research, we used the following methods:

- general scientific analytical method - for analyzing the state of the problem, analyzing literary sources;
- modeling method - for developing a model for calculating correlation coefficients;
- algorithmic method - for developing algorithms for calculating correlation coefficients;
- testing method.

4 RESULTS

When starting to create a software library, we first define the functional and non-functional requirements for this library:

- the software library must be autonomous and operate without references to third-party modules (for example, NumPY);
- the software library must use an object-oriented programming paradigm. This will provide a clearer structure of the library and simplify further modernization of the library;
- the software library must use built-in Python data types and not use data types from third-party libraries. This will facilitate the autonomous operation of the library regardless of access to third-party modules;
- the software library must contain error prevention tools;
- the software library methods must return the value of the selected correlation coefficient and a conclusion about the similarity of two data sets;
- the software library must be quickly and easily integrated into the application.

The software library can be developed in different programming languages. We consider it advisable to write a software library in the Python language. Python language tools provide:

- the use of simple and complex data types when calculating correlation coefficients;
- ease of implementation of basic algorithmic structures;
- various platforms for the functioning of applications developed in Python;
- comparative ease of generating a report on correlation analysis.

Pearson's correlation coefficient is used to determine the linear relationship between two data sets. The value of the Pearson's correlation coefficient is within [-1;1]. The example of Pearson correlation calculation is shown in Table 1.

Table 1: Calculation of Pearson correlation coefficient.

Table	x	y	x_i^2	y_i^2	$x_i y_i$
1	3	5	9	25	15
2	4	2	16	4	8
3	5	8	25	64	40
4	7	9	49	81	63
5	2	4	4	16	8
6	6	9	36	81	54
7	4	8	16	64	32
Summ	31	45	155	335	220

The value of the Pearson's correlation coefficient is calculated by (1). The Pearson correlation coefficient is calculated using the formula [17]:

$$r_{xy} = \frac{n\cdot\sum_i(x_i\cdot y_i)-(\sum_i x_i)\cdot(\sum_i y_i)}{\sqrt{(n\cdot\sum_i x_i^2-(\sum_i x_i)^2)\cdot(n\cdot\sum_i y_i^2-(\sum_i y_i)^2)}}. \quad (1)$$

For two data sets (Table 1), the Pearson correlation coefficient is 0.728. This indicates the existence of a direct strong relationship between the samples. The algorithm for calculating the Pearson correlation coefficient is shown in the diagram (Fig.1).

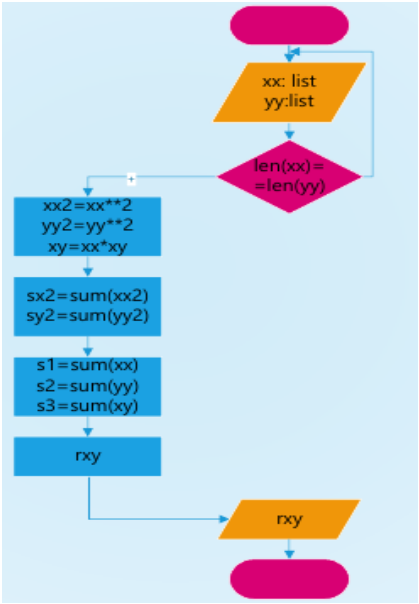


Figure 1: Schematic of the algorithm for calculating the Pearson correlation coefficient.

The algorithm for calculating the Pearson correlation coefficient is implemented as a Pearson

class. This class has the following fields: N - size of each array; xx - first list of values; uu - second list of values; xy - list of pairwise products of the elements of the first list by the elements of the second list; xx2 - list of squares of the numbers of the first list; yy2 - list of squares of the numbers of the second list; s1 - sum of the elements of the first list; s2 - sum of the elements of the second list; s3 - sum of the elements of the list xy; sx2 - sum of the squares of the elements of the first list; sy2 - sum of the squares of the elements of the second list, correl - Pearson correlation coefficient.

Methods of the Pearson class: init() - constructor; correlation - returns the Pearson correlation coefficient; concl() - returns a conclusion about the degree of connection between the initial lists of data. The generated PDF report contains the value of the correlation coefficient, a conclusion about the nature of the relationship between the samples, and a plot of the distribution of the samples being compared. The results of the Pearson class testing revealed that the algorithm was O(n) in complexity. The processing time for one pair of values did not change.

Spearman's correlation coefficient is used to determine the relationship in cases where it is not possible to use the Pearson correlation coefficient. Ranks are used when calculating the Spearman correlation coefficient. The advantage of the Spearman coefficient is that it is more sensitive to the relationship. The limitation of the Spearman coefficient is that each sample must contain at least 5 values. Algorithm for calculating the Spearman correlation:

1) record both samples in the table (Table 2);
2) assigning a rank to each sample value (relative to the elements of this sample);
3) calculating the difference in ranks for each pair of values;
4) calculating the square of each difference in ranks;
5) calculating the sum of the ranks of each sample and the sum of the squares of the differences;
6) calculating the Spearman correlation coefficient according to formula [18]:

$$r_s = 1 - \frac{6 \cdot \sum (R(x)_i - R(y)_i)^2}{n \cdot (n^2 - 1)}. \quad (2)$$

Based on the data in Table 2, the value rs = 0.8214 was obtained. The algorithm for the Spearman correlation coefficient calculating is shown in Figure 2.

To perform the calculations, the corrValue (numerical value) and Spearman (calculation of the correlation coefficient itself and drawing a conclusion) classes were designed and implemented.

The corrValue class describes the numerical value used in the calculations and has the following fields: value - numerical value, rank - rank, num - ordinal number of the value in the list. Class methods: init - constructor, setRank() - sets the rank, getRank() - returns the rank, setValue() - sets the value, getValue() - returns the value, setNum() - sets the ordinal number, getNum() - returns the ordinal number.

Table 2: Calculation of spearman correlation coefficient.

Num	*x*	*y*	*R(x)*	***R(y)***	$(R(x)-R(y))^2$
1	3	5	2	3	1
2	4	2	3	1	4
3	5	8	5	4	1
4	7	9	7	6	1
5	2	4	1	2	1
6	6	9	6	7	1
7	4	8	3	4	1
Sum	31	45	27	27	10

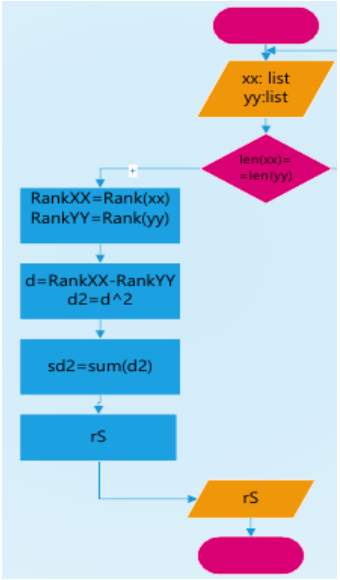


Figure 2: Schematic of the algorithm for calculating Spearman's correlation.

The Spearman class performs calculations and draws a conclusion about the existence and nature of the relationship. Class fields: xx - first sample, uu - second sample, N - number of values in each sample, RankXX - list of objects of class corrValue (first sample), RankYY - list of objects of class corrValue (second sample), RankXXX - list of objects of class corrValue (first sample, used to find rank), RankYYY - list of objects of class corrValue (second sample, used to find rank), correl - value of correlation

coefficient, Conclusion - conclusion about existence of connection.

The Spearman class has the following methods: init() - constructor, correlation () - calculation and return of Spearman's correlation coefficient, concl() - formulation and return of conclusion about existence of connection between samples and nature of this connection. The generated PDF report contains the value of the correlation coefficient, a conclusion about the nature of the relationship between the samples, and a plot of the distribution of the samples being compared. For the Spearman class, performance testing results revealed an algorithm complexity of $O(n^2)$. The time to process one pair of values increased proportionally to the sample size.

The τ-Kendall coefficient is used as an alternative to determine the relationship in the same cases as Spearman's rank correlation coefficient, but has a different calculation algorithm. Let's consider this algorithm in more detail. Let there be two samples of the same size. These values are written in the table (Table 3).

Table 3: Initial data for calculation the Kendall correlation coefficient.

Num	*x*	*y*
1	7	5
2	1	4
3	9	1
4	3	9
5	5	8
6	7	3
7	6	6
8	2	11
9	10	12

The algorithm for calculating the τ-Kedall coefficient is as follows:

1) sort the records by one of the characteristics (in our example - X) in ascending order;
2) calculate the ranks of the values for each column separately and write them next to each other (Table 4);
3) calculate the number of matches: for each record, count how many times its rank in Y is less than the ranks of the records below it and enter this number in column P;
4) calculate the number of inversions: for each record, count how many times its rank in Y is greater than the rank of the records below it;
5) calculate the τ-Kendall coefficient using formula [19]:

$$\tau = \frac{\sum P - \sum Q}{n \cdot (n-1)/2}\ , \qquad (3)$$

where ΣP is the number of matches; ΣQ is the number of inversions; n is the total number of records.

The scheme of algorithm is shown in Figure 3.

To implement the developed algorithm, the Record and Kendall classes were created. The Record class contains all the data about one row of the table of values and has the following fields: num - ordinal number of the record, x - value of the first sample, y - value of the second sample, rank1 - rank of the value of the first sample, rank2 - rank of the value of the second sample. The Record class has the following methods: setRank1() - sets the rank of the element of the first sample, setRank2() - sets the rank of the element of the first sample, setX() - sets the value of the element of the first sample, setY() - sets the value of the element of the second sample.

Table 4: Calculation of Kendall correlation coefficient.

Num	*x*	*R(x)*	*y*	$\boldsymbol{R(y)}$	Coincidence (P)	Inversion (Q)
1	1	1	4	3	6	2
2	2	2	11	8	1	6
3	3	3	9	7	1	5
4	5	4	8	6	1	4
5	6	5	6	5	1	3
6	7	6	3	2	2	1
7	7	7	5	4	1	1
8	9	8	1	1	1	0
9	10	9	12	9	0	0

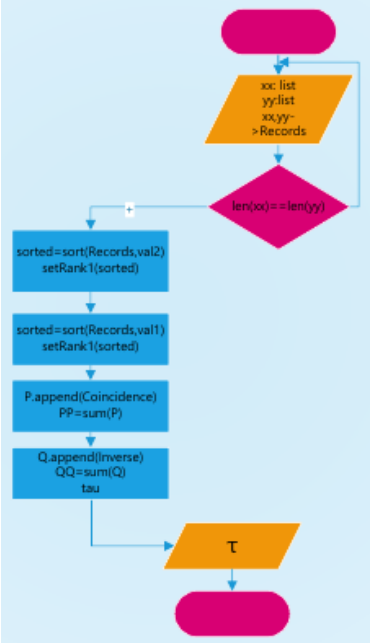


Figure 3: Schematic of the algorithm for calculating the Kendall correlation coefficient.

The Kendall class contains all the samples, initial, intermediate and final data from calculating the Kendall correlation coefficient. The Kendall class has the following fields: x1 is the value of the first sample, y1 is the value of the second sample, Records is a list of Record class objects, N is the number of values in each sample, sorted is a sorted list of Record class objects, P is a list of matches, Q is a list of inversions, PP is the number of matches, QQ is the number of inversions, tau is the value of the Kendall correlation coefficient. The Kendall class has the following methods: init() is a constructor, correlation() is the calculation and return of the correlation coefficient, concl() is the formulation and return of the conclusion. The generated PDF report contains the value of the correlation coefficient, a conclusion about the nature of the relationship between the samples, and a plot of the distribution of the samples being compared. The results of testing the algorithm for calculating Kendall's correlation showed that the algorithm had a complexity of $O(n^2)$. The time to process one pair of values increased proportionally to the increase in sample size.

The developed Correlation library contains the following classes (Fig. 4).

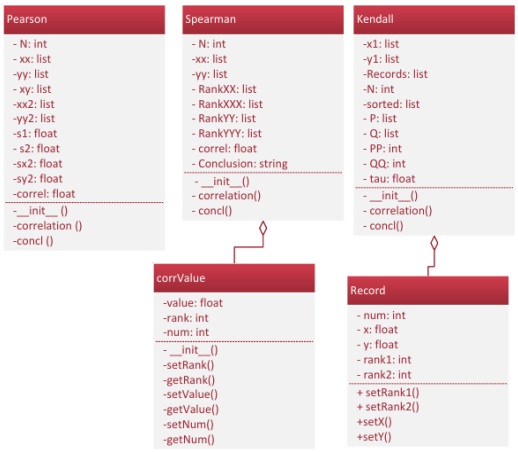


Figure 4: Class diagram of the correlation software library.

The created software library contains 3 main classes that calculate Pearson, Spearman, and Kendall correlation values, as well as additional classes used in the Spearman and Kendall classes.

The created software libraries were integrated into the application, which has a graphical user interface implemented in two versions: one based on the PyQt 5 library and the other based on the Custom Tkinter library. Their use is because the standard Tkinter library lacks tools for creating menus. The interface of the application based on the PyQT 5 library is shown in Figure 5.

The graphical interface of the Correlation Checker application contains a menu consisting of 3 sections: Input data, Data, and Save results. The developed Correlation Checker 1.0 application has the following options for entering initial data: from a text file, from a SQLite3 database, from a table on a form. After entering the initial data, they can be seen in the form of a table. To do this, you need to select the Data menu option. Using the Save results menu, you can generate a report on the correlation study in the form of a PDF document. The main application window contains radio buttons with which you can select the type of correlation: Pearson, Spearman, or Kendall. The Correlation Checker 2.0 application (Fig. 6) was also created based on the freely distributed Custom Tkinter library. The main menu options are the same as in the previous version.

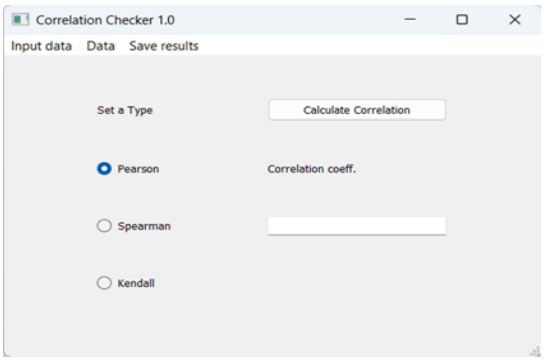


Figure 5: Application interface based on the PyQT 5 library.

The use of the Custom Tkinter library is because the PyQT 5 library is paid, and the Custom Tkinter library is free, although it has fewer widgets. The main window of the application also contains a field for entering initial data (Input data arrays). Further, the entered arrays are displayed in the DataArray1 and DataArray2 fields. This version of the application calculates all 3 types of correlation: Pearson, Spearman, and Kendall. Using the Report menu, you can generate a correlation report in the form of a PDF document. Therefore, applications based on the PyQT 5 and Custom Tkinter libraries have the same functionality and can be used to study the relationship between two data sets.

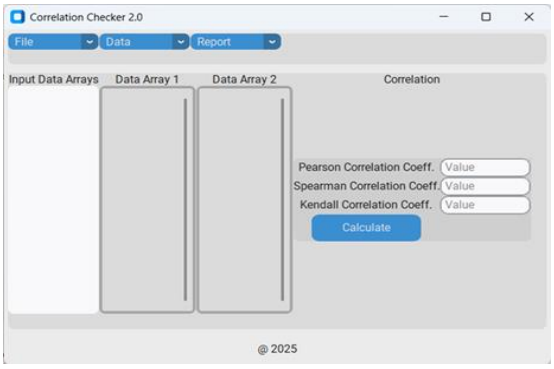


Figure 6: Interface of the application based on the Custom Tkinter library.

Since both versions of the Correlation Checker application use the same libraries, the results they calculate will be the same. Therefore, we used MS Excel spreadsheets for verification. This software tool contains built-in functions for calculating Pearson, Spearman, and Kendall correlation coefficients. For the same input data (Tables 1, 2, 4), MS Excel produced the following values: Pearson=0.640411, Spearman=0.643, and Kendall=0.64286. The program we developed with the same input parameters (Tables 1, 2, 4) displayed the following values: Pearson=0.64041, Spearman=0.64321, and Kendall=0.64286. Therefore, based on the verification of the results obtained, we can conclude that the developed program calculates the correct result.

5 CONCLUSIONS

The developed Correlation Checker application is based on the Correlation library we developed, the classes and methods of which provide the calculation of the Pearson, Spearman, and Kendall correlation coefficients. This application is intended for data analysis, as well as for calculating the specified correlation coefficients during the study of the academic discipline “Fundamentals of Scientific Research”. The developed application has the following advantages: open source; ease of modernization and development (the application is written in Python); modular architecture of the application; use of only built-in Python data types in the Correlation library. Thanks to this, the application can be upgraded in the future, even if support for libraries such as PyQT and Custom Tkinter is discontinued. It will only be necessary to rewrite the interface using other libraries. The generated PDF report does not contain instructions for interpretation, but a conclusion about the existence of a relationship between the two samples.

REFERENCES

[1] R. Taylor, “Interpretation of the correlation coefficient: a basic review,” Journal of Diagnostic Medical Sonography, vol. 6, no. 1, pp. 35-39, 1990.

[2] B. Ratner, “The correlation coefficient: Its values range between +1/-1, or do they?” Journal of Targeting, Measurement and Analysis for Marketing, vol. 17, no. 2, pp. 139-142, 2009.

[3] D. Chicco and G. Jurman, “The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification,” BioData Mining, vol. 16, no. 1, p. 4, 2023.

[4] J. Jiang, X. Zhang, and Z. Yuan, “Feature selection for classification with Spearman’s rank correlation coefficient-based self-information in divergence-based fuzzy rough sets,” Expert Systems with Applications, vol. 249, p. 123633, 2024.

[5] S. Chatterjee, “A new coefficient of correlation,” Journal of the American Statistical Association, vol. 116, no. 536, pp. 2009-2022, 2021.

[6] X. Su, S. Shang, Z. Xu, H. Qian, and X. Pan, "Assessment of Dependent Performance Shaping Factors in SPAR-H Based on Pearson Correlation Coefficient," CMES-Computer Modeling in Engineering & Sciences, vol. 140, no. 2, pp. 23-26, 2024.

[7] H. Yu and A. D. Hutson, "A robust Spearman correlation coefficient permutation test," Communications in Statistics - Theory and Methods, vol. 53, no. 6, pp. 2141-2153, 2024.

[8] F. Han and Z. Huang, "Azadkia-Chatterjee's correlation coefficient adapts to manifold data," The Annals of Applied Probability, vol. 34, no. 6, pp. 5172-5210, 2024.

[9] C. A. Mertler, R. A. Vannatta, and K. N. LaVenia, Advanced and Multivariate Statistical Methods: Practical Application and Interpretation, Abingdon-on-Thames: Routledge, 2021.

[10] V. Mykhalchuk, "Assessment of the objectiveness of cognitive prejudice mitigation in a neuroeducational strategy with the highlighting of vulnerable indicators," Measuring and Computing Devices in Technological Processes, vol. 1, pp. 52-58, 2025.

[11] H. Kang, "Sample size determination and power analysis using the G*Power software," Journal of Educational Evaluation for Health Professions, vol. 18, pp. 52-58, 2021.

[12] R. Bivand, G. Millo, and G. Piras, "A review of software for spatial econometrics in R," Mathematics, vol. 9, no. 11, p. 1276, 2021.

[13] L. Qi, W. Lin, X. Zhang, W. Dou, X. Xu, and J. Chen, "A correlation graph based approach for personalized and compatible web APIs recommendation in mobile app development," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 6, pp. 5444-5457, 2022.

[14] S. Mguil-Touchal, F. Morestin, and M. Brunei, "Various experimental applications of digital image correlation method," in WIT Transactions on Modelling and Simulation, vol. 17, pp. 1227-1236, 2024.

[15] S. D. Rahmawati, A. F. Ihsan, D. Darmadi, F. Priadi, and U. W. R. Siagian, "Web-Based Application for Calculation of Physical Properties of Hydrocarbon Fluids Using Compositional Approach," Journal IATMI, vol. 17, pp. 1-12, 2022.

[16] V. M. Bazurin, O. I. Pursky, and O. S. Chashechnikova, "Application for Statistical Processing of Pedagogical Experiment Results: A Component-Based Approach," in 2025 International Conference on Inventive Computation Technologies (ICICT), Apr. 2025, [Online]. Available: https://doi.org/10.1109/ICICT64420.2025.11004909.

[17] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in Noise Reduction in Speech Processing, Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 1-4.

[18] J. C. De Winter, S. D. Gosling, and J. Potter, "Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data," Psychological Methods, vol. 21, no. 3, p. 273, 2016.

[19] H. Abdi, "The Kendall rank correlation coefficient," in Encyclopedia of Measurement and Statistics, vol. 2, 2007, pp. 508-510.

LLM-Based Synthetic Ground Truth Generation for Audio-Based Emotion Classification via In-Context Learning

Qing Huang[1,2], Pooja Pol[1,2] and Jianing Zhang[1]
[1]School of Business, Technical University of Applied Sciences Augsburg, Friedberger Str. 4, 86161 Augsburg, Germany
[2]Data Science und Autonome Systeme Technologietransferzentrum (TTZ), An der Schmiede 19, 86899 Landsberg am Lech, Germany
qing.huang@tha.de, pooja.pol@tha.de, jianing.zhang@tha.de

Keywords: Large Language Models (LLMs), In-Context Learning (ICL), Synthetic Ground Truth, Affective Computing, Data-Driven Decision Support, Virtual Reality (VR).

Abstract: Understanding human states and interaction dynamics is a core goal of human-computer interaction (HCI). As interaction paradigms become more immersive, virtual reality (VR) has emerged as a powerful platform for studying collaborative work. In such settings, evaluating team collaboration states, including team performance and team resilience, requires continuous and reliable inference of latent team-level cognitive and affective states from multi-modal sensor data, such as speech signals. However, generating ground truth labels for these latent states remains challenging due to sensor-induced noise, contextual variability, and sparse expert annotations. Traditional self-reporting approaches provide only static and delayed measurements and are therefore insufficient for capturing dynamic team processes reflected in continuous speech data. In this work, we propose a large language model (LLM)-driven, agentic inference workflow for automated emotion-related synthetic ground truth generation from streaming speech data in multi-user VR environments. Leveraging the generalization capabilities of LLMs, we use In-Context Learning (ICL) with few-shot demonstrations of paired audio-based samples and their corresponding transcriptions. ICL tends to achieve task adaptation comparable to model fine-tuning while circumventing the computational overhead of parameter updates. To construct informative and robust in-context prompts, we adopt a retrieval-based selection strategy that dynamically identifies relevant audio demonstrations based on similarity in the acoustic feature space. Specifically, retrieval is guided by prosodic descriptors derived from speech signals, whereas transcript-based semantic information is incorporated directly within the prompt and interpreted by the LLM through its reasoning capabilities. This modality-aware strategy promotes consistent affective labeling and improved generalization across previously unseen speech segments. Evaluated on multi-player VR audio recordings, our methodology demonstrates potential as a scalable, data-efficient component for data-driven team-based decision support. By integrating acoustic similarity-based retrieval with LLM-based semantic reasoning, this work contributes to emerging interdisciplinary methodologies at the intersection of scientific machine learning, multi-modal systems, and AI-driven decision-making.

1 INTRODUCTION

Virtual reality (VR) provides an immersive platform for analyzing human behavior through multimodal sensing, including speech, motion, and physiological signals [1]. In collaborative VR tasks, speech plays a central role, reflecting coordination, intent, and affective expression. However, establishing reliable ground truth annotations for affective states remains challenging. Self-report measures are subjective and temporally misaligned [2], while sensor data are noisy and large-scale, making manual annotation costly. These constraints motivate automated and data-efficient annotation strategies.

We focus on speech sentiment annotation in collaborative VR environments. Recent advances in large language models (LLMs) enable affective reasoning via in-context learning (ICL) without task-specific parameter updates, often approaching fine-tuned performance [3], [4]. This makes LLMs particularly suitable for scalable annotation scenarios where labeled data are limited and repeated manual labeling across sessions is costly.

In collaborative VR settings, sessions involve changing player constellations while task contexts and discussion structures remain comparable. Although each session includes only 5-7 participants, affective expressions are influenced not only by individual speaker characteristics but also by shared interactional dynamics emerging from common tasks and conversational context. For consistent annotation across sessions, it is therefore beneficial to align segments at the level of comparable affective interaction patterns that arise under similar task conditions. To support such cross-session consistency, our approach retrieves acoustically comparable segments using generic prosodic descriptors (e.g., pitch, loudness, intensity, and speaking rate). These descriptors capture expressive characteristics that remain relatively stable across different speaker constellations and thereby provide a session-independent basis for demonstration selection.

Building on this motivation, we propose an LLM-based framework for synthetic speech sentiment annotation that integrates speech-oriented LLMs with a modality-aware retrieval-guided ICL strategy grounded in acoustic similarity. While ICL and retrieval-based prompting are established techniques, the novelty of our approach lies in positioning acoustic similarity as the primary mechanism for selecting affectively comparable interaction segments across sessions, thereby enabling context-level alignment rather than session-specific adaptation. Retrieval is conducted in the acoustic feature space to identify segments that share comparable expressive patterns, while transcript-based content is incorporated within the prompt and interpreted by the LLM through its semantic reasoning capabilities. Rather than introducing a new similarity metric per se, we reconceptualize retrieval-guided ICL as a structured mechanism for cross-session affective alignment. This design mitigates over-adaptation to session-specific characteristics and instead supports data-efficient domain adaptation based on shared interactional dynamics. Because retrieval relies on generic acoustic descriptors and semantic interpretation is handled by the LLM, the framework generalizes beyond a particular team composition and can be transferred to other collaborative or conversational settings without model fine-tuning.

2 DATA ACQUISITION

This study is based on data collected from a multi-player VR game in which participants collaborate to achieve a shared objective. The VR scenario is designed to elicit natural team interaction and communication under controlled yet immersive conditions, enabling analysis of individual- and team-level behavioral dynamics.

Data were collected across two VR sessions. The first session served as a familiarization phase, during which participants explored the virtual environment and learned the task mechanics, while the second session focused on task execution, with participants collaboratively completing the assigned task. This session-based design enables the capture of multi-modal sensor data under different interaction conditions and supports analysis of condition-dependent variations in cognitive and affective states.

Each session involved approximately five to seven participants equipped with VR head-mounted displays (Meta Quest Pro), allowing continuous recording of multi-modal sensor data. Speech audio was recorded in a single-channel (mono) format via the headset-mounted microphone array, which provides a downmixed audio signal, and encoded on device at a sampling rate of 48 kHz. Each participant's audio recording has an average duration of approximately 25 minutes.

In this work, we focus on speech signals recorded during team interactions. The collected audio data capture spontaneous verbal communication and are predominantly in German, reflecting realistic collaborative settings. While most recordings were conducted in German, the proposed analysis and annotation framework is not inherently language-specific and can, in principle, be adapted to other languages.

For downstream analysis, the audio streams were processed using an automatic speech recognition (ASR) model based on Whisper [5] to generate textual transcriptions with preserved temporal alignment between audio segments and text. This alignment reduces potential temporal mismatches during affective annotation. The resulting paired audio-text data form the basis for synthetic sentiment annotation and further analysis.

For lexicon-based methods relying on the NRC-VAD resource, which is originally constructed for English, German transcripts were automatically translated into English prior to lexicon lookup. The translation step enables compatibility with the NRC-based affective scoring pipeline while preserving segment-level temporal alignment. In our analysis, cross-linguistic and cultural differences introduced by the translation step were observed to have only a limited impact on the overall annotation results. However, in our framework, lexicon-based

annotations serve primarily as a lightweight reference signal rather than as the sole source of ground truth. Moreover, the retrieval-guided LLM-based annotation process operates directly on original transcripts and acoustic descriptors, thereby reducing dependence on language-specific lexical mappings. This design mitigates potential cultural bias introduced by cross-lingual lexicon transfer while maintaining methodological consistency.

3 METHODOLOGY

To enable automatic synthetic ground truth generation for speech sentiment annotation, we propose an LLM-based methodology built upon the publicly available Voxtral model [6], a speech-oriented large language model capable of jointly processing audio and textual inputs. All experiments are conducted without task-specific fine-tuning. Inference is performed with fixed decoding parameters to reduce stochastic variability during generation. While different LLM backbones or decoding strategies may lead to quantitative differences in output, the proposed framework is designed in a modular manner and can, in principle, be instantiated with alternative speech-capable LLMs. A systematic comparison of different LLM backbones and their inference variability is beyond the scope of the present study and will be investigated in future work.

We first established a vanilla ICL baseline using few-shot prompting with labeled audio examples. We then extend this baseline with a retrieval-guided ICL strategy that integrates acoustically derived representations for demonstration selection while incorporating transcript-based semantic information directly within the prompt for multimodal reasoning. Importantly, retrieval operates exclusively on precomputed acoustic descriptors and relies on similarity-based ranking without random sampling or approximate search. For a fixed ICL bank and identical input, the same set of in-context demonstrations is deterministically retrieved. As a result, retrieval stability is guaranteed by design, and any potential variability is confined to the generative stage of the LLM rather than the demonstration selection process.

3.1 Vanilla ICL with Few-shot Prompting

As a baseline approach for synthetic ground truth generation, we adopt the strategy demonstrated in Figure 1. In this setup, the model is prompted using a small number of labeled audio examples without any task-specific parameter updates. Each shot consists of a triplet of audio snippets, one per sentiment class (positive, negative, and neutral), paired with their corresponding sentiment labels provided in textual form. For each inference instance, a small number of such shots are randomly sampled and included in the prompt context to expose the model to representative sentiment variations. Following these few-shot demonstrations [4], an unseen audio segment is appended to the prompt, and the model is instructed to infer its sentiment based on the contextual examples. This vanilla formulation enables the LLM to adapt to the task at inference time solely through contextual conditioning.

3.2 Acoustically Informed LLM-Based Synthetic Annotation via ICL

Recent progress in self-supervised speech representation learning has demonstrated that pretrained acoustic models can effectively capture affective cues directly from raw audio signals. Transformer-based architectures such as wav2vec 2.0 [7] and HuBERT [7] have been shown to predict continuous emotional dimensions, including valence, arousal, and dominance [8], without relying on textual input. These properties make acoustic representations particularly valuable in spontaneous and conversational settings, where transcripts may be incomplete or noisy.

In parallel, lexicon-based approaches provide an interpretable and lightweight means of extracting affective information from text. Resources such as the NRC word-emotion association [9] and NRC Valence, Arousal, and Dominance (NRC-VAD) lexicons [10] offer linguistically grounded affective cues that complement acoustically derived features and serve as useful reference annotations.

Motivated by these observations, the proposed method integrates acoustic and semantic information into the ICL framework. Specifically, the acoustic information is represented using low- and mid-level speech descriptors, including pitch, loudness, intensity, and speaking rate measured in both words per second and syllables per second. As shown in Figure 1, a small set of few-shot demonstrations is enriched with both transcript-based content and acoustically derived descriptors, enabling LLM inference to be grounded in multimodal context. By combining the generalization capability of LLMs with acoustically informed ICL, the framework aims to improve the robustness and consistency of

synthetic sentiment annotations, particularly in emotionally ambiguous conversational scenarios.

To systematically select informative demonstrations, we replace random sampling with similarity-based retrieval over an annotated segment pool $\mathcal{D}$.

For each segment u_i, we construct an acoustic descriptor vector

$$a_i = E_{ac}(x_i) \in \mathbb{R}^{d_a},$$

where $E_{\text{ac}}(\cdot)$ extracts normalized prosodic features (e.g., pitch statistics, loudness/intensity statistics, and speaking-rate measures) from the raw audio signal. Prior to retrieval, acoustic vectors are standardized to ensure comparability across feature dimensions.

Given a target segment u^* with acoustic representation a^*, the acoustic distance is defined using Euclidean distance:

$$d_{ac}(u^*, u_i) = \|\tilde{a}^* - \tilde{a}_i\|_2,$$

where $\tilde{a}$ denotes the standardized acoustic vector. Since smaller distances indicate higher similarity of acoustic features, the in-context demonstration set is obtained by selecting the top-K segments with the smallest acoustic distance:

$$\mathcal{S}_{retr} = TopK_{(u_i,y_i)\in\mathcal{D}}(-d_{ac}(u^*, u_i)), K \in \mathbb{N}.$$

Importantly, transcript-based semantic information is not used for retrieval ranking. Instead, transcripts are directly incorporated into the ICL prompt together with the corresponding audio segments and acoustic descriptors. This design ensures that demonstration selection is grounded in affective acoustic similarity, while semantic reasoning is handled by the LLM during inference.

3 AI-DRIVEN WORKFLOWS

The individual components introduced in Sections 3.1 and 3.2 are integrated into a unified AI-driven workflow for multimodal sentiment analysis in immersive collaborative environments. The workflow is designed to provide a coherent transition from data acquisition and representation learning to synthetic annotation and evaluation.

At the system level, collaborative interactions take place within simulated environments, such as virtual reality setups, which serve as controlled yet ecologically valid experimental platforms. During these interactions, speech signals are recorded in multi-speaker settings, forming the primary data modality considered in this study.

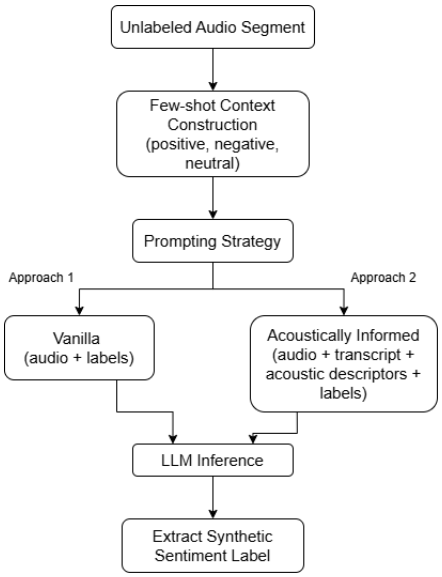


Figure 1: Flowchart illustrating the vanilla and acoustically informed ICL strategies.

The recorded audio data are subsequently processed to obtain structured representations, including automatic speech transcription, temporal segmentation, and the extraction of acoustic features such as pitch, loudness, intensity, and speaking rate. Pretrained acoustic models and lexicon-based methods are used to derive preliminary affective cues, providing complementary continuous and categorical representations.

These intermediate representations are then incorporated into the LLM-based annotation stage through acoustically informed ICL. By conditioning inference on a small set of enriched few-shot examples selected through acoustic similarity, the system generates synthetic sentiment labels at the segment level. This design enables scalable annotation while reducing reliance on manual labeling.

4 RESULTS

This section presents a comparative evaluation of the proposed acoustically informed ICL framework for synthetic sentiment annotation. We first analyze the impact of prompt construction by comparing vanilla few-shot ICL with a retrieval-guided, acoustically informed ICL strategy. We then examine how the proposed ICL approach interacts with different annotation paradigms, including audio-based, lexicon-based, and pretrained text-based sentiment models.

To ensure a fair and consistent comparison, all methods are evaluated on the same set of 794 segment-level samples, comprising all transcribed speech segments produced by all players within a single VR session. Performance is assessed using normalized confusion matrices and quantitative evaluation metrics, as reported in Tables 1-3 and illustrated in Figures 2-5.

4.1 Few-shot Vanilla ICL vs Acoustically Informed ICL

We first examine the effect of prompt construction by comparing vanilla few-shot ICL, which uses randomly sampled demonstrations, with an acoustically informed ICL variant that retrieves a small set of demonstrations using ground-truth-based similarity (Fig. 2). The results show that prompt design strongly determines whether the LLM can recognize non-neutral affective states.

In the vanilla ICL setting (Fig. 2, left), predictions collapse almost entirely into the neutral class. The model correctly identifies neutral segments (89.3%), but fails to recognize the positive class and misclassifies all negative segments as neutral (100.0% neutral for actual negative). This behavior indicates that randomly selected few-shot prompts do not provide sufficient, discriminative context for the model to separate minority affective states from neutral speech.

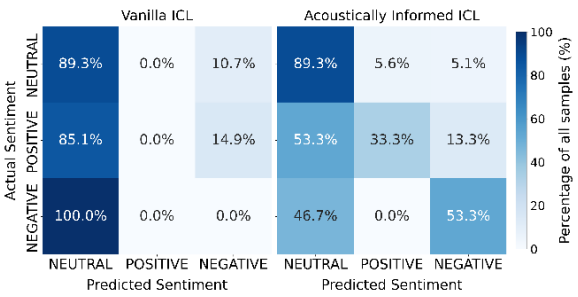


Figure 2: Confusion matrices comparing ground truth sentiment labels with predictions from (left) vanilla few-shot ICL and (right) the retrieval-guided acoustically informed ICL using ground truth-based similarity.

Table 1: Performance comparison between vanilla few-shot ICL and acoustically informed ICL.

	Acc.	Prec.	Recall	F1
$ICL^{Vanilla}$	0.82	0.31	0.30	0.30
ICL^{GT}	0.85	0.46	0.59	0.49

In contrast, the retrieval-guided acoustically informed ICL (ground-truth-based retrieval) substantially improves discrimination of non-neutral classes (Fig. 2, right). For positive segments, the correct prediction rate increases to 33.3%, and for negative segments the model achieves 53.3% correct negative predictions, reducing the systematic "all-neutral" failure mode. This improvement is also reflected in the summary metrics: macro-F1 increases from 0.30 to 0.49, while recall rises from 0.30 to 0.59 (Table 1). Although overall accuracy changes only slightly (0.82 to 0.85), the macro metrics confirm that the gain comes from improved minority-class recognition rather than from the dominant neutral class.

Overall, this experiment demonstrates that retrieval-guided prompt construction—combined with richer, acoustically grounded context—can fundamentally change the LLM's behavior from neutral-dominant collapse to meaningful separation of positive and negative affect.

4.2 Retrieval-Guided Acoustically Informed ICL

4.2.1 Baseline Annotation Performance

We next examine three baseline annotation paradigms: an audio-based model (wav2vec 2.0 [7]), a lexicon-based text method (NRC-VAD [10]), and a pretrained text-based classifier (XLM-Roberta [11]). Table 2 summarizes their overall performance, and Table 3 reports per-class results for the positive and negative classes.

A key methodological note is that both wav2vec 2.0 and NRC-VAD originally output continuous affective representations (valence, arousal, dominance), i.e., numerical scores in a three-dimensional space. To enable direct comparison with categorical sentiment outputs, these continuous representations are mapped into a simplified three-class scheme (positive, neutral, negative) using predefined mapping rules. While this discretization necessarily removes fine-grained affect information, it ensures that all paradigms are evaluated under the same categorical benchmark.

Across baselines, overall accuracy is largely driven by the dominance of neutral samples and therefore provides a limited view of model quality. Consistent with the confusion matrices (Fig. 3-5), all baselines exhibit recurring confusion between neutral and non-neutral classes. Among them, wav2vec and NRC-lexicon achieve comparable macro-F1 scores (0.35 and 0.32 in Table 2, respectively) but show very weak sensitivity to negative affect, with near-zero F1 and recall for the negative class (Table 3),

highlighting their limitations in detecting rare emotional expressions in conversational speech. XLM-Roberta performs slightly better overall (macro-F1 = 0.39) and shows improved balance on minority classes, yet still suffers from substantial neutral-positive confusion, likely due to domain mismatch between written-text training data and spoken team interaction language.

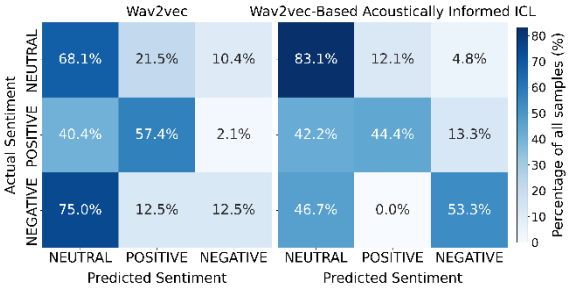


Figure 3: Confusion matrices comparing ground truth sentiment labels with predictions from (left) the wav2vec-based baseline and (right) the retrieval-guided acoustically informed ICL using wav2vec-based similarity.

4.2.2 Effect of Acoustically Informed In-Context Learning

Applying retrieval-guided acoustically informed ICL consistently improves performance across all three annotation paradigms (Tables 2-3), with the largest gains occurring in minority-class recognition, particularly for negative sentiment.

For the wav2vec-based setup, macro-F1 increases from 0.35 to 0.47 (Table 2). The primary improvement is negative-class sensitivity: F1(N) rises from 0.04 to 0.25 and Rec.(N) increases from 0.13 to 0.53 (Table 3). This trend is also visible in Figure 3, where the ICL-enhanced variant reduces the baseline tendency to absorb negative segments into neutral and improves correct negative predictions.

For the NRC-VAD-based setup, macro-F1 improves from 0.32 to 0.45 (Table 2). Negative-class recognition increases markedly (F1(N): 0.07 to 0.24; Rec.(N): 0.06 to 0.47 in Table 3). These gains indicate that retrieval-guided ICL can compensate for the absence of prosodic information in lexicon-driven annotation by providing multimodal context through retrieved examples.

For the XLM-Roberta-based setup, macro-F1 increases from 0.39 to 0.48 (Table 2). Negative-class recall rises from 0.44 to 0.60 and F1(N) improves from 0.22 to 0.27 (Table 3), consistent with the improved diagonal structure visible in Figures 4-5. Given that the text classifier baseline is already relatively strong, the improvements suggest that retrieval-guided ICL acts as a form of contextual adaptation to the characteristics of spoken interaction data.

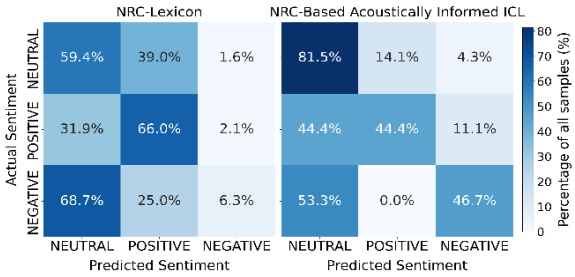


Figure 4: Confusion matrices comparing ground truth sentiment labels with predictions from (left) the NRC-VAD lexicon-based baseline and (right) the retrieval-guided acoustically informed ICL using NRC-VAD-based similarity.

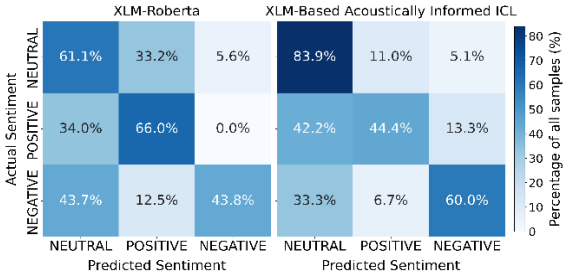


Figure 5: Confusion matrices comparing ground truth sentiment labels with predictions from (left) the XLM-Roberta-based baseline and (right) the retrieval-guided acoustically informed ICL using XLM-Roberta-based similarity.

A consistent trade-off across paradigms is that positive recall decreases (from 0.57/0.66/0.66 to 0.44 across wav2vec/NRC/XLM in Table 3), while F1(P) increases. This indicates that the ICL-enhanced system becomes more conservative in predicting positive sentiment, reducing false positives at the cost of missing some positive instances. In contrast, improvements for the negative class are both large and consistent, which is particularly important because negative affect is the most challenging and sparsely represented class.

Table 2: Overall performance comparison between baseline methods and acoustically informed ICL.

	Acc.	Prec.	Recall	F1
wav2vec	0.66	0.37	0.46	0.35
ICL^{w2v}	0.80	0.44	0.60	0.47
NRC	0.59	0.37	0.44	0.32
ICL^{NRC}	0.79	0.43	0.58	0.45
XLM	0.61	0.40	0.57	0.39
ICL^{XLM}	0.81	0.45	0.63	0.48

Table 3: Per-class performance comparison between baseline methods and acoustically informed ICL.

	F1 (P)	Rec. (P)	F1 (N)	Rec. (N)
wav2vec	0.23	0.57	0.04	0.13
ICL^{w2v}	0.26	0.44	0.25	0.53
NRC	0.17	0.66	0.07	0.06
ICL^{NRC}	0.24	0.44	0.24	0.47
XLM	0.19	0.66	0.22	0.44
ICL^{XLM}	0.28	0.44	0.27	0.60

Taken together, the results demonstrate that retrieval-guided acoustically informed ICL is a robust enhancement mechanism across fundamentally different annotation paradigms. The largest relative improvements are observed for wav2vec and NRC-VAD, where the baselines struggle most with minority affective states. For XLM-Roberta, improvements remain consistent but are smaller in magnitude, reflecting its stronger baseline capacity.

Importantly, the gains are best reflected by macro-level metrics and per-class negative performance rather than accuracy. This confirms that the proposed workflow improves sensitivity to non-neutral affective deviations and reduces class-imbalance-driven distortions, making it a practical and scalable approach for synthetic sentiment annotation in immersive team-based VR interaction settings.

5 CONCLUSIONS

This work presents a data-efficient and scalable framework for synthetic ground truth generation in speech-based sentiment annotation within immersive virtual reality environments. Experimental results show that a vanilla few-shot ICL approach is insufficient for reliable affective labeling, exhibiting a strong bias toward neutral predictions and limited sensitivity to minority emotional states. By incorporating acoustic information and structured feature representations into a retrieval-guided ICL paradigm, the proposed approach substantially improves annotation robustness and mitigates the effects of class imbalance.

Beyond individual model comparisons, we introduce an end-to-end AI-driven workflow that automates the sentiment annotation process, spanning representation learning, example retrieval, in-context inference, and evaluation. For controlled and interpretable comparison, this study focuses on a three-class sentiment setting comprising neutral, positive, and negative affective states. While this simplified label space facilitates systematic evaluation, the proposed framework is not inherently restricted to it and can be extended to more fine-grained emotional taxonomies through its retrieval-based and modality-aware design.

Across audio-based, lexicon-based, and pretrained text-based annotation paradigms, the proposed framework consistently outperforms vanilla ICL, demonstrating its general applicability. Although the generated annotations are not intended to replace expert-labeled ground truth, they reliably capture deviations from neutral affective states and are therefore well suited for large-scale analysis scenarios where manual annotation is impractical.

Finally, the proposed methodology is not limited to speech data. Its underlying principles—retrieval-guided ICL, modality-specific grounding, and data-efficient annotation—are transferable to other modalities used in affective computing and psychology, including physiological signals, behavioral measures, and multimodal interaction data. This work thus provides a foundation for unified synthetic ground truth generation to support multi-modal, team-level affective analysis in complex interactive systems.

ACKNOWLEDGMENTS

Support by the Technology Transfer Center for Data Science and Autonomous Systems in Landsberg am Lech and the Bavarian Collaborative Research Program (BayVFP), funding line Digitalization, for the project DaTeam-VR is kindly acknowledged.

REFERENCES

[1] I. A. Castiblanco Jimenez, E. C. Olivetti, E. Vezzetti, S. Moos, A. Celeghin, and F. Marcolin, “Effective affective EEG-based indicators in emotion-evoking VR environments: An evidence from machine learning,” Neural Computing and Applications, vol. 36, pp. 22245-22263, 2024.

[2] N. Gao, M. S. Rahaman, W. Shao, and F. D. Salim, “Investigating the reliability of self-report data in the wild: The quest for ground truth,” in Proc. ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp Adjunct), pp. 237-242, 2021.

[3] M. Ihori, T. Yamane, N. Kawata, N. Makishima, T. Tanaka, S. Suzuki, S. Orihashi, and R. Masumura, “Few-shot personalization via in-context learning for speech emotion recognition based on speech-language model,” arXiv preprint arXiv:2509.08344, 2025.

[4] M. Mosbach, T. Pimentel, S. Ravfogel, D. Klakow, and Y. Elazar, “Few-shot fine-tuning vs. in-context learning: A fair comparison and evaluation,” in Findings of the Association for Computational

Linguistics: ACL 2023, Toronto, ON, Canada, pp. 12284-12314, 2023.

[5] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," in Proc. 40th International Conference on Machine Learning (ICML 2023), Honolulu, HI, USA, pp. 28492-28518, 2023, [Online]. Available: https://doi.org/10.48550/arXiv.2212.04356.

[6] A. H. Liu, A. Ehrenberg, L. A. Lo, C. Denoix, C. Barreau, G. Lample, J.-M. Delignon, K. R. Chandu, P. von Platen, and P. R. Muddireddy, "Voxtral," arXiv preprint arXiv:2507.13264, 2025.

[7] J. Wagner, A. Triantafyllopoulos, H. Wierstorf, M. Schmitt, F. Burkhardt, F. Eyben, and B. W. Schuller, "Dawn of the transformer era in speech emotion recognition: Closing the valence gap," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 9, pp. 10745-10759, 2023.

[8] S. M. Mohammad, "Sentiment analysis: Automatically detecting valence, emotions, and other affectual states from text," in Emotion Measurement, 2nd ed., H. L. Meiselman, Ed. Cambridge, U.K.: Woodhead Publishing, 2021, pp. 323-379.

[9] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word-emotion association lexicon," Computational Intelligence, vol. 29, no. 3, pp. 436-465, 2012.

[10] S. M. Mohammad, "NRC VAD lexicon v2: Norms for valence, arousal, and dominance for over 55k English terms," 2025, doi: 10.48550/arXiv.2503.23547.

[11] C. Lalk, K. Targan, T. Steinbrenner, J. Schaffrath, S. Eberhardt, B. Schwartz, A. Vehlen, W. Lutz, and J. Rubel, "Employing large language models for emotion detection in psychotherapy transcripts," Frontiers in Psychiatry, vol. 16, Art. no. 1504306, 2025.

A Comparative Analysis of AI-Assisted Development Tools in Software Development

Marija Miloshevska[1] and Marija Kalendar[2]
[1]*MIR! Software Solutions, 1000 Skopje, North Macedonia*
[2]*Department of Computer Technologies and Engineering, Faculty of Electrical Engineering and Information Technologies, SS. Cyril and Methodius University in Skopje, Rugjer Boshkovikj Str. 18, 1000 Skopje, North Macedonia*
milosevskamarija99@yahoo.com, marijaka@feit.ukim.edu.mk

Keywords: Artificial Intelligence (AI), AI-Assisted Programming, Software Development, Frontend Development, Code Generation, Software Quality, Developer Productivity, Tool Comparison.

Abstract: Artificial Intelligence (AI) tools are increasingly integrated into modern software development workflows, assisting developers in tasks such as code generation, testing, and debugging. This study presents a comparative evaluation of three AI-assisted development tools - Cursor, GitHub Copilot, and Gemini - with a focus on their effectiveness in frontend development. The comparison was conducted through practical experimentation within a small-scale Vue 3 project, where the tools were assessed across representative tasks including component implementation, utility function development, and test creation. The findings indicate that AI-based tools can significantly improve productivity. However, the generated outputs frequently require manual review, correction, and refinement to ensure correctness and alignment with project requirements. Each tool demonstrated distinct strengths, but no single tool consistently outperformed the others across all evaluated tasks. The results suggest that AI tools are most effective when used as supportive assistants rather than autonomous replacements for developers. Future research may extend this evaluation to larger-scale projects, different frameworks, and more complex development scenarios.

1 INTRODUCTION

Software development has evolved rapidly over the past few decades, largely because businesses and consumers expect more advanced, reliable, and feature-rich software. As projects become increasingly complex, developers often struggle with tight deadlines, limited resources, and the constant need to maintain high code quality. To keep up with these demands, the industry has started embracing new and innovative approaches, and Artificial Intelligence (AI) has become one of the most influential drivers of this change.

AI technologies such as machine learning (ML), natural language processing (NLP), are changing how developers work across different stages of the software development lifecycle. Many tasks that once required significant manual effort can now be supported or automated by AI: code generation, testing, debugging, and deployment. Tools like OpenAI Codex and GPT-4 can turn natural language instructions into working code, suggest improvements, and even spot potential bugs. This not only saves time but also helps reduce human error, leading to more stable and dependable software.

As AI becomes increasingly integrated into software development workflows, its role is expanding beyond simple productivity gains. It has the potential to reshape how efficiency, accuracy, and innovation are defined in software engineering. Rather than replacing developers, AI is increasingly becoming a powerful assistant that helps them focus on more creative and high-level problem solving.

In this context, the study presented in this paper examines the practical effectiveness of contemporary AI-assisted development tools in frontend software engineering. Specifically, it provides a comparative evaluation of Cursor, GitHub Copilot, and Gemini within a controlled Vue 3 project environment. The tools were assessed through real development tasks, including component creation, utility function implementation, and test generation, in order to analyze their impact on productivity, code quality, contextual awareness, and overall usability. By systematically comparing their strengths and limitations, this paper aims to contribute empirical

insights into the role of AI as a collaborative support mechanism in modern software development workflows.

2 RELATED WORK

Artificial intelligence is being increasingly used in software development recently. Consequently, the topic has attracted a lot of research in the last few years. In the early days, AI was mostly used for automation or data analysis, but more recently the industry, as well as the research community, have increasingly focused on integrating generative AI and coding assistants directly into development environments. As a result, researchers are working on understanding the performance of these tools in real-world environments, and what are their limitations.

There have already been a number of research studies comparing AI coding assistants and evaluating how they perform. Results usually show that these tools do not behave the same in every situation. Some work better for structured programming tasks, while others are more useful when the developer interacts with the tool step by step. Experiments with programming problems show that AI can often handle simple tasks, but performance drops when tasks become more complex or require deeper reasoning [1] - [3]. This suggests that AI is becoming more and more helpful, but not something developers can fully rely on yet.

There are also studies that research generative AI in software engineering in a more general way. Generative AI is being used for testing, bug prediction, and code suggestions. Reviews that summarize research from recent years show that development and testing are the areas where AI appears most often [4], [5]. These works usually mention time savings and reduced manual effort as benefits. At the same time, they also point out issues like unclear model decisions, data quality problems, and difficulty fitting AI tools into existing workflows [6] - [8]. So, while the potential is clear, there are still some open questions.

Another topic that appears in the research literature is how developers feel about using AI tools. The verdict is that “trust is not automatic”. Developers tend to trust suggestions more when they are accurate and match the task they are working on. If suggestions are wrong or strange, confidence drops quickly [9]. This indicates that adoption depends not only on performance but also on user experience.

Some research work also focuses on how these tools are built and used in practice. For example, local code completion systems show that AI support does not always need cloud services to be useful [10]. This can matter for privacy and response time. At the same time, AI-based software brings new challenges, since models depend on training data and may change over time, which is different from traditional software behavior [11].

Overall, existing research suggests that AI-assisted development tools are steadily advancing and demonstrate high potential to support developers. However, a substantial portion of prior studies relies on benchmark evaluations or controlled experimental settings, which may not fully capture the complexity and variability of real-world software development environments. In practice, development workflows often involve changing requirements, contextual dependencies, and imperfect conditions that are difficult to replicate in laboratory settings. Consequently, there remains a need for empirical investigations conducted within realistic development scenarios. This study builds on prior work by presenting a practical, task-oriented comparison of AI tools within a frontend development context, thereby offering insights from within applied software engineering practice.

3 AI TRANSFORMATIVE ROLE IN SOFTWARE DEVELOPMENT

With recent progress in machine learning, natural language processing (NLP), and large language models (LLMs), AI technology has become increasingly integrated into modern development environments. Its role is not limited to improving productivity; it is also influencing how development teams collaborate and how software quality is evaluated. AI is integrated in several stages of the software development lifecycle, assisting with tasks related to coding, testing, design, and maintenance.

One of the areas where this influence can be observed is code generation and auto-completion. AI-supported tools are able to provide context-aware suggestions while developers write code. These suggestions can reduce the amount of repetitive work and, in some cases, speed up implementation, since they allow developers to focus on application logic. ML models contribute to this process by identifying common coding patterns, proposing small improvements, and pointing out possible mistakes. Although these suggestions are not always correct, they can still serve as a helpful starting point.

Machine learning techniques are increasingly applied in software testing activities. Several ML approaches focus on predicting which components or areas of a codebase are more likely to contain defects, enabling teams to focus testing efforts more effectively. This is particularly valuable in large-scale projects in order to achieve as vast test coverage as possible. By identifying high-risk modules, predictive models can enable more efficient allocation of testing resources and improve overall quality assurance processes.

In addition to predictive testing, AI-based tools are widely used to support debugging and test generation, identifying irregular patterns in code, suggesting potential fixes, and automatically generating preliminary test cases. By highlighting anomalies and suspicious code structures, these tools assist developers in detecting issues earlier in the development lifecycle. Early detection is generally associated with reduced correction costs and shorter feedback cycles; however, the overall effectiveness of these AI-driven solutions depends significantly on the quality of their outputs and the extent of human oversight during review and refinement.

Natural Language Processing has introduced new possibilities in communication and documentation. Tools such as OpenAI Codex and GPT-4 can convert natural language descriptions into code examples. This is often useful during early prototyping, when ideas are still being explored. In addition, NLP-based tools can produce summaries or explanations of existing code. Such features may support knowledge sharing within teams and make it easier for new members to become familiar with a codebase.

At a broader level, AI can provide some support in software design and architecture. For example, certain tools attempt to recommend design patterns or highlight parts of the code that could be refactored. Others can simulate aspects of system behavior to inform design choices. Reinforcement learning (RL) has also been studied in the context of testing, where models adjust test scenarios based on prior results in order to explore edge cases.

AI tools are also present in deployment and operational practices. AI-supported DevOps tools can help automate parts of continuous integration and deployment pipelines. Some systems monitor performance metrics and attempt to detect unusual behavior in production environments. In certain cases, predictive analytics are used to estimate the likelihood of failures so that preventative actions can be taken. While these approaches show promise, their success depends heavily on data quality and proper configuration.

4 CHALLENGES OF USING AI TOOLS IN SOFTWARE DEVELOPMENT

While AI tools offer significant benefits to software development, their integration and use come with a range of challenges that developers and organizations must address to fully harness their potential.

Accuracy and reliability. AI-driven tools, especially those using LLMs, can occasionally generate inaccurate or incomplete code, which can lead to bugs, security vulnerabilities, or inefficient implementations. Thus, AI-generated suggestions need to be carefully reviewed so they align with best practices and project-specific requirements.

Over-reliance on AI. Excessive dependence on AI for tasks like code generation, debugging, and documentation may reduce developers' problem-solving skills and deep technical understanding. Over time, this could lead to a decline in critical thinking, where engineers accept AI suggestions without fully evaluating their correctness or efficiency. Maintaining a balance between AI assistance and human expertise is crucial.

Ethical and security risks. AI tools may inadvertently introduce security vulnerabilities or use code that violates intellectual property rights. For instance, AI models can generate code that resembles open-source or proprietary codebases, leading to copyright issues. There is also the risk of embedding biases into the AI's decision-making, which could negatively affect the software's functionality and inclusivity.

Data privacy concerns. AI models trained on large datasets often require access to substantial amounts of code or proprietary information. This raises concerns about data privacy, especially in industries dealing with sensitive or regulated data. Unauthorized data exposure or improper model training could lead to legal and reputational risks.

Bias and inclusivity**.** AI models can inherit biases from the datasets they are trained on. If these datasets include biased examples, the AI tools may maintain or amplify those biases in generated code, recommendations or algorithms. This is particularly concerning in applications related to hiring, healthcare, and finance, where fairness and impartiality are crucial.

Integration and compatibility**.** While AI tools are powerful, integrating them into existing development workflows, systems, and tools can be a complex process. Compatibility with legacy systems or specific development environments can be an issue,

requiring additional time, effort, and resources to adapt.

Learning curve. Many AI-powered development tools come with a steep learning curve. Developers may struggle to understand how to effectively use these tools, which could reduce their overall impact. Additionally, tools that require prompt engineering or deep customization in order to work effectively, can be challenging for new users.

Lack of human context. AI tools are trained on vast datasets, but may lack the contextual understanding of a project that a human developer brings. This can lead to suggestions and code generation that may technically work but are not the most efficient solutions for the problem at hand.

5 COMPARATIVE ANALYSIS OF AI TOOLS

This paper compares three popular AI tools: Cursor, GitHub Copilot, and Gemini. We chose these three as representative examples for our empirical comparative analysis since they represent a spectrum of AI-assisted coding tools, from widely adopted solutions to emerging, innovative systems. Their differing architectures, interaction paradigms, and levels of context awareness allow for a comprehensive evaluation of code suggestion accuracy, developer productivity, and workflow integration. By including both established and cutting-edge tools, the study provides meaningful insights into the strengths and limitations of current AI coding assistants, based on their integration, learning curve, model flexibility, effectiveness in code development and applicability across different stages of development, with a primary focus on frontend development.

5.1 Integration Ease

Cursor, GitHub Copilot, and Gemini each offer different levels of integration with development tools, particularly benefiting frontend workflows.

Cursor functions as an AI-enhanced code editor based on VS Code, providing seamless compatibility with a wide range of extensions such as linters (ESLint), formatters (Prettier), and frontend frameworks like React and Vue, while also supporting native Git and GitHub integration.

GitHub Copilot offers deep integration with popular editors like VS Code and JetBrains IDEs, providing real-time code completion, frontend scaffolding, and CSS assistance, and now extends to GitHub PR workflows with Copilot Chat for interactive code reviews.

Gemini primarily integrates via APIs, Google Cloud services, and certain IDE extensions, making it especially useful for large-scale code analysis, optimization, and documentation in frontend projects, though it is less directly embedded into coding environments compared to Cursor or Copilot.

5.2 Learning Curve

The learning curve varies across Cursor, GitHub Copilot, and Gemini, depending on how naturally they fit into existing development workflows.

Cursor has a low to moderate learning curve for developers already familiar with VS Code, since it is built on top of it. Most standard coding habits transfer directly, but exploring its multi-model management and prompt engineering features can require some initial learning, especially to fully take advantage of different LLMs and custom API setups.

GitHub Copilot has a very low learning curve. It embeds directly into the typing flow of IDEs, such as: VS Code or JetBrains, offering inline suggestions and code completions. Frontend developers can start benefiting from it almost immediately, with minimal adjustment to their coding habits. Copilot Chat (for interactive conversations) may require a bit more exploration but remains intuitive.

Gemini tends to have a moderate learning curve, mainly because it is not as tightly embedded into coding IDEs. Developers often interact with Gemini via APIs, web UIs, or cloud environments (like Vertex AI Studio). To fully leverage its capabilities (large code analysis, summarization, architecture suggestions) users need to learn how to properly structure prompts and sometimes adjust their workflow to accommodate out-of-IDE interactions.

In summary, GitHub Copilot is the easiest to adopt quickly, Cursor is user-friendly for VS Code users, but requires some effort to master its multi-model capabilities, Gemini demands more adjustment due to its external/cloud-based integration.

5.3 Code Generation and Quality

The impact on code development and quality varies across Cursor, GitHub Copilot, and Gemini, depending on how each tool supports coding workflows and promotes best practices.

Cursor significantly enhances code development by supplying intelligent code completions, refactor suggestions, and in-editor AI chat assistance. Thanks to its flexible access to top-performing LLMs like

GPT-4o, Claude-3, and Gemini 2.0 Flash, Cursor generates high-quality, context-aware code, helps structure frontend components more cleanly, and encourages better coding practices. It also assists with documentation, code reviews, even generates tests, improving overall project quality.

GitHub Copilot accelerates code development speed dramatically by suggesting real-time code completions, snippets, and even complex frontend structures (React hooks, TailwindCSS layouts, Vue composables). While Copilot improves developer productivity, the quality of its suggestions varies. It sometimes produces code that works but is not always optimized or fully aligned with best practices. Developers need to actively review and refine Copilot's outputs to maintain high code quality. Therefore, while Copilot enhances productivity, the focus is on developers to ensure optimal code quality.

Gemini contributes more at a macro level, focusing on large-scale code understanding, architectural refactoring, and documentation. It is particularly strong at identifying redundant patterns, suggesting code optimizations, and helping restructure large frontend codebases. While Gemini does not directly generate small snippets during typing, its analyses can lead to cleaner, and better-architected frontend systems.

5.4 Model Flexibility

Cursor, GitHub Copilot, and Gemini each differ in the AI models they utilize and how they expose model capabilities to developers. Table 1 presents the findings.

Cursor offers extensive multi-model support, allowing users to select from a wide range of models: o1, o1-mini, GPT-4o, Claude-3.5-Sonnet, Claude-3-Opus, Gemini 2.0 Flash Exp. Additionally, Cursor allows users to integrate custom API keys from various providers, giving developers flexibility to use their own subscriptions instead of paying for Cursor's bundled service. This approach provides an advantage in cost control and model selection, especially for developers working on frontend projects requiring specific strengths from different models (e.g., faster models for autocompletion, stronger models for architectural planning).

GitHub Copilot is primarily powered by OpenAI's Codex (derived from GPT-3) and incorporates GPT-4 models, particularly in Copilot X. These models are fine-tuned for code completion, offering excellent real-time assistance in frontend development tasks such as building React/Vue components, generating CSS, and refactoring.

Gemini, through its Gemini 1.5 and Gemini 1.5 Flash models, delivers large-context AI capabilities, supporting advanced codebase analysis, summarization, and optimization. While Gemini is less tightly integrated into coding environments compared to Cursor or Copilot, its ability to handle large and complex frontend codebases makes it highly valuable for high-level refactoring and architecture design.

6 USAGE OF AI TOOLS IN VUE FRONTEND DEVELOPMENT

To objectively evaluate the effectiveness of AI-assisted coding tools in frontend development, this paper adopts a hands-on, task-based methodology. Rather than relying solely on theoretical comparisons, each tool Cursor (version 1.2), GitHub Copilot (VS Code version 1.101, powered by GPT-4), and Gemini (web interface using Gemini 2.5 Pro available at https://gemini.google.com,) was evaluated through practical implementation within a Vue 3 project scaffolded using Vite.

The selected tasks reflect common frontend challenges that developers frequently encounter: component creation, utility function abstraction, and unit testing. These tasks help assess the tools' code quality, contextual awareness, and speed of execution in a real-world development setting.

Table 1: Comparative analysis.

Tool	Integration	Models	Learning Curve	Best For
Cursor	VS Code-based	GPT-4o, Claude 3.5, Gemini 2.0, custom APIs	Low to Moderate	Flexible, high-quality code generation, multi-model workflows, larger apps
GitHub Copilot	VS Code, JetBrains, and Neovim	Codex, GPT-4	Very Low	Fast autocomplete while coding components/utilities
Gemini	API & external tools	Gemini 1.5 / Flash	Moderate	Codebase analysis, refactoring, debugging, feature-spec generation

6.1 Create a Vue Component

Objective: Create a UserCard Vue component that accepts a user prop (name, email, and avatar) and emits an onClick event when the card is clicked. The used prompt in all tools: "I am building a Vue 3 application using the Composition API. Please create a *UserCard* Vue component that accepts a user prop, containing name, email and avatar. It should emit an *onClick* event when the card is clicked. The component should follow Vue 3 best practices." The results are summarized in Table 2.

6.2 Create and Use a Utility Function

Objective: Write a reusable JavaScript function to capitalize each word in a full name string, then use it in UserCard to format the displayed name. The used prompt in all tools: "Write a reusable JavaScript utility function called *capitalizeWords* that takes a full name string and capitalizes the first letter of each word. The function should be placed in src/utils/stringUtils.js and imported into the *UserCard* component to format the displayed name." The results are summarized in Table 3.

6.3 Write Unit Tests

Objective: Write unit tests for the *UserCard* component, verifying prop rendering, event emission, and formatted name output using Jest. The used prompt in all tools: "Write unit tests for the *UserCard* Vue 3 component using Jest. The tests should verify: 1) that the name and email props render correctly, 2) that the *onClick* event is emitted when the card is clicked, and 3) that the *capitalizeWords* utility correctly formats the displayed name." The results are summarized in Table 4.

Table 2: Results from Task 1 - creating a Vue component.

Criteria	Cursor	Copilot	Gemini
Syntax Used	Vue 3 (Composition API)	Vue 2 (Options API)	Vue 2-like syntax
Functionality	Fully functional with prop and emit logic	Functional with minor Vue 3 fixes	Functional, similar to Copilot
Context Awareness	High-created files, registered components properly	Medium - generated code in isolation	Low - unaware of integration context
Time to Generate	Fast	Medium	Slow
Comments/ Explanation	Minimal - clean, working code	Minimal	Extensive code comments
Required Edits	None	Minor syntax corrections	Structural revisions and syntax updates
Code Quality	Clean, idiomatic Vue 3 Functional	Clean, idiomatic Vue 3 Functional but outdated syntax	Verbose and not Vue 3 optimized

Table 3: Results from Task 2 - working with a utility function.

Criteria	Cursor	Copilot	Gemini
Function Output	Optimized and maintainable	Optimized and maintainable	Overly complex implementation
Code Location	src/utils/stringUtils.js - follows best folder structure	utils/stringUtils.js - placed at root level	src/utils/stringUtils.js - follows best folder structure
Code Complexity	Simple, clean, and easy to follow	Simple and readable	Verbose and overengineered
Time to Generate	Fast (~10 seconds)	Fast (~10-15 seconds)	Slow (>1 minute)
Maintainability	High - easily testable and reusable	High - clear and reusable	Low - unnecessarily detailed logic
Context Awareness	High - respected existing project structure	Medium - ignored folder hierarchy	Low - limited awareness of structure
Code Quality	5/5	4/5	2/5

Table 4: Results from Task 3 -writing unit tests.

Criteria	Cursor	Copilot	Gemini
Test Coverage	Comprehensive - props, event emit, and formatted name rendering	Comprehensive and accurate	Covered main behavior but tests were overly verbose
Tooling Awareness	High - updated package.json, included install and config instructions	Medium - assumed Jest was preconfigured	Medium - included install steps but lacked integration context
Code Simplicity	Clean and readable test code	Very concise and readable	Complex and harder to follow
Time to Generate	Fast (~10 seconds)	Fast (~10-15 seconds)	Slow (>1 minute)
Maintainability	High	High	Low
Code Quality	5/5	4/5	2/5

These three small tasks showed that simple frontend features in Vue can be built more quickly with the help of modern AI tools, and they reduce the amount of manual coding needed. However, some level of developer input and correction was still required.

Among the tested tools, Cursor showed the strongest overall performance in this evaluation. It generated accurate code, was aware of the project context, and often included useful setup suggestions. Copilot was sometimes faster in producing inline suggestions, and Gemini was helpful for explanations and understanding the code. Cursor generally provided the most complete results for the chosen tasks.

7 CONCLUSIONS

This paper focused on how AI tools like Cursor, GitHub Copilot, and Gemini can help in frontend development. The goal was to compare them in real tasks and see how useful they actually are when building a Vue application.

The results indicate that these tools can save time and reduce some repetitive coding. They are especially helpful for generating components, utility functions, and test examples. However, they do not always provide high-quality results. Sometimes the code needs fixing, adjusting, or rewriting. Because of this, developers still need to check and understand what the AI produces.

Each tool showed different strengths. Cursor performed effectively when project context mattered. Copilot was fast and convenient while typing. Gemini was more useful for explanations and bigger-picture suggestions. None of them was the best in every situation. These observations are consistent with prior research, which also reports that AI coding tools provide useful assistance but still require human oversight and validation [1], [2].

This study also showed that relying too much on AI is not a particularly good idea. AI can make mistakes, and if developers trust it without any consideration, problems can appear later, even in production environments. It works best when it is used as support, not as a replacement for thinking and coding skills.

There are also some limits to this research. Only a few tasks were tested, and only in Vue. Bigger projects or other frameworks might show different results. AI tools are also improving quickly, so their performance can change over time.

Overall, AI tools are becoming part of frontend development. They can help developers work faster, but they still need human control. The best results come when developers use AI carefully and combine it with their own knowledge. AI should be viewed as a supportive tool rather than a replacement for developers.

REFERENCES

[1] S. Sarkar, “Comparative Analysis of AI-Powered Coding Assistants for Software Engineering Practices: An Evaluation of Claude Sonnet, ChatGPT, Google Gemini, and Competing Platforms,” SSRN Electronic Journal, Sep. 2025, , doi: 10.2139/ssrn.5465115.

[2] M. S. I. Ovi, N. Anjum, T. H. Bithe, M. M. Rahman and M. S. A. Smrity, “Benchmarking ChatGPT, Codeium, and GitHub Copilot: A Comparative Study of AI-Driven Programming and Debugging Assistants,” arXiv preprint arXiv:2409.19922, 2024, , doi: 10.48550/arXiv.2409.19922.

[3] E. Holloway, C. Chu, J. Tenenbaum, S. Puranic and J. Wei, “Benchmarking Code Generation Models with BIG-Bench Lite CodeGen: An Evaluation of Instruction-Tuned Large Language Models,” arXiv preprint arXiv:2304.10778, 2023, doi: 10.48550/arXiv.2304.10778.

[4] G. Shekhar, “The Impact of AI and Automation on Software Development: A Deep Dive,” ESP International Journal of Advancements in Computational Technology (ESP-IJACT), vol. 2, no. 1, pp. 162-174, 2024.

[5] U. K. Durrani, M. Akpinar, M. F. Adak, A. T. Kabakus, M. M. Öztürk and M. Saleh, “A Decade of Progress: A Systematic Literature Review on the Integration of AI in Software Engineering Phases and Activities (2013-2023),” IEEE Access, vol. 12, pp. 171185-171204, 2024, doi: 10.1109/ACCESS.2024.3488904.

[6] R. A. Husein, H. Aburajouh and C. Catal, “Large Language Models for Code Completion: A Systematic Literature Review,” Computer Standards & Interfaces, vol. 92, p. 103917, 2025, doi: 10.1016/j.csi.2024.103917.

[7] M. Barenkamp, J. Rebstadt and O. Thomas, “Applications of AI in Classical Software Engineering,” AI Perspectives, vol. 2, no. 1, 2020, , doi: 10.1186/s42467-020-00005-4.

[8] M. Adil, M. Sharif and A. Masood, “The Role of Artificial Intelligence in Software Development: A Literature Review,” Science, Engineering and Technology Review (SETR), vol. 9, no. 11, pp. 214-226, Nov. 2024.

[9] A. Brown, S. D’Angelo, A. Murillo, C. Jaspan and C. Green, “Identifying the Factors That Influence Trust in AI Code Completion,” in Proceedings of the 1st ACM International Conference on AI-Powered Software (AIware), Porto de Galinhas, Brazil, Jul. 2024, , doi: 10.1145/3664646.3664757.

[10] A. Semenkin, V. Bibaev, Y. Sokolov, K. Krylov, A. Kalina, A. Khannanova, D. Savenkov, D. Rovdo, I. Davidenko, K. Karnaukhov, M. Vakhrushev, M. Kostyukov, M. Podvitskii, P. Surkov, Y. Golubev, N. Povarov and T. Bryksin, “Full Line Code Completion: Bringing AI to Desktop,” arXiv preprint arXiv:2405.08704, 2024, revised Jan. 8, 2025, , doi: 10.48550/arXiv.2405.08704.

[11] S. Martínez-Fernández, J. Bogner, X. Franch, M. Oriol, J. Siebert, A. Trendowicz, A. M. Vollmer and S. Wagner, “Software Engineering for AI-Based Systems: A Survey,” ACM Transactions on Software Engineering and Methodology (TOSEM), vol. 31, no. 2, art. 37e, pp. 1-59, 2022, , doi: 10.1145/3487043.

Adaptive Governance Architecture for Digital Innovation Ecosystems Based on Psychophysiological User Profiling

Olena Ronzhes
Department of Applied Psychology, Kharkiv National University, Svobody Square 4, 61022 Kharkiv, Ukraine
olena.ronzhes@gmail.com

Keywords: Adaptive Governance, Digital Innovation Ecosystems, Psychophysiological Profiling, User Adaptation, Industry 4.0, Psychological Safety.

Abstract: Digital innovation ecosystems operating in Industry 4.0 environments increasingly encounter resistance and systemic instability driven not by technological limitations but by user adaptation stress under conditions of rapid change. Although Quadruple Helix governance models structure institutional collaboration among academia, industry, government, and civil society, they do not incorporate empirically validated human adaptation parameters into ecosystem design. The aim of this study is to develop and empirically ground a four-level adaptive governance architecture integrating psychophysiological user profiling into digital ecosystem management. The research is based on a multi-stage empirical study (N = 479; EEG subsample n = 50) combining psychophysiological measurement and subjective state assessment. Latent Profile Analysis identified four stable adaptation profiles (entropy = 0.93). Experimental findings demonstrate a significant stress amplification effect under conditions of novelty combined with time pressure, reflected in increased mental workload and reduced relaxation levels. The proposed architecture translates adaptation profiles into operational governance parameters across four layers: operational interaction modulation, innovation rollout sequencing, conceptual governance integration, and socio-cultural stabilization. The framework enables integration with wearable stress indicators, behavioral analytics, and AI-driven interface adaptation. The results provide both theoretical advancement by bridging behavioral adaptation research with digital ecosystem governance and practical value through a scalable decision-support model for adaptive innovation management.

1 INTRODUCTION

Digital transformation within Industry 4.0 has led to the rapid proliferation of interconnected digital innovation ecosystems integrating cyber-physical systems, artificial intelligence, and data-driven services [1].

These ecosystems are increasingly structured through institutional collaboration models such as the Triple and Quadruple Helix frameworks, which emphasize interaction between academia, industry, government, and civil society [2], [3]. While such models effectively describe macro-level innovation dynamics, they insufficiently address micro-level human adaptation processes within digital environments.

In the context of this study, user adaptation is examined within specific Industry 4.0 environments, including AI-driven digital platforms, smart manufacturing interfaces, augmented reality-based training systems, and data-intensive decision-support environments. These contexts are characterized by high levels of cognitive load, continuous system updates, and interaction with adaptive algorithms, which directly influence user experience and adaptation dynamics.

In practice, technological innovation often encounters resistance not due to architectural flaws or functional limitations, but because of elevated cognitive load, technostress, and reduced perceived stability among users [4]. Rapid system updates and compressed innovation cycles create conditions of novelty combined with time pressure, which have been shown to amplify mental workload and decrease relaxation indicators in digital task environments [5].

Current ecosystem governance models rarely incorporate empirically validated psychophysiological indicators of user adaptation.

User behavior is typically assessed through adoption metrics or acceptance models such as TAM and UTAUT [6], while measurable cognitive and stress-related parameters remain under-integrated into system design.

At the same time, psychological safety has been identified as a critical factor in innovation acceptance and learning performance [7], yet it is seldom operationalized within digital ecosystem governance.

To address this gap, this study proposes a four-level adaptive governance architecture based on psychophysiological user profiling derived from latent profile analysis of EEG-based indicators and subjective state assessment [5], [8]. By mapping empirically identified adaptation profiles onto governance layers aligned with Quadruple Helix innovation systems [3], the architecture transforms human adaptive variability into structured decision-support parameters.

2 THEORETICAL BACKGROUND AND CONCEPTUAL FOUNDATIONS

2.1 Digital Innovation Ecosystems and Multi-Helix Governance

Digital innovation ecosystems are complex adaptive systems integrating technological infrastructure, institutional actors, and user communities. The Quadruple Helix model extends earlier Triple Helix frameworks by incorporating civil society as a structural innovation driver [3], [9].

This extension recognizes that innovation diffusion depends not only on institutional collaboration but also on societal acceptance and participatory legitimacy.

However, while Helix models successfully describe institutional interaction, they operate predominantly at the macro-structural level. They do not incorporate measurable cognitive or psychophysiological adaptation variables into governance mechanisms.

As digital ecosystems become increasingly human-centric within Industry 4.0 and Industry 5.0 paradigms [1], the absence of adaptation-sensitive governance parameters becomes a structural limitation.

2.2 VUCA Conditions and Technostress Dynamics

Digital ecosystems operate under conditions characterized as volatile, uncertain, complex, and ambiguous (VUCA) [10]. Continuous system updates, interface modifications, and rapid innovation cycles increase cognitive demands on users. Empirical studies indicate that technostress and digital overload negatively affect system acceptance, learning performance, and long-term engagement [11] - [13].

Novelty combined with time pressure represents a particularly stress-inducing condition in digital task environments. Such conditions amplify mental workload and destabilize cognitive equilibrium, potentially reducing ecosystem resilience at scale.

2.3 Psychological Safety, Agency, and Adaptation Profiling

Psychological safety has been identified as a critical factor influencing innovation participation and error tolerance in organizational contexts [7]. In digital environments, perceived agency and cognitive stability mediate adaptation processes and trust formation [6], [14].

Recent advances in psychophysiological monitoring allow objective measurement of cognitive load and stress-related parameters in real time [5], [8]. Yet these measurable indicators remain largely disconnected from innovation governance models.

This theoretical gap motivates the integration of empirically validated adaptation profiling into ecosystem-level governance architecture.

3 METHODOLOGY

3.1 Research Design

To empirically investigate adaptation variability within digital innovation ecosystems, the study employed a multi-stage mixed-method research design integrating large-scale survey assessment with controlled psychophysiological experimentation.

The research was conducted between 2020 and 2024 and aimed to quantify measurable cognitive and stress-related adaptation patterns in digital task environments.

The methodological framework consisted of two complementary stages: (1) macro-level assessment of digital adaptation indicators (N = 479), and (2) controlled EEG-based experimental validation under stress-inducing innovation conditions (n = 50). This structure enabled both large-scale profiling of adaptation variability and in-depth physiological measurement of stress amplification effects.

3.2 Participants

The total sample included 479 participants aged 16–75 who actively used digital technologies. Data for the macro-level assessment were collected through a combination of online surveys and in-person participation in Cologne for the experimental part.

The experimental subsample (n = 50; age 30–50) participated in controlled digital task sessions designed to simulate innovation-intensive interaction scenarios.

All participants performed tasks under four conditions:

1) familiar environment without time constraint;
2) familiar environment with time constraint;
3) novel digital environment with time constraint;
4) novel digital environment without time constraint. The third condition represented the highest innovation-stress configuration.

3.3 Psychophysiological Data Acquisition

EEG signals were recorded using a portable non-medical neurointerface (Neeuro SenzeBand 2). Raw signals were processed into real-time indicators: attention index, relaxation index, mental workload, alpha/Beta activity ratio, delta differences relative to baseline condition.

In Industry 4.0 interaction scenarios, different technologies generate distinct cognitive and stress-related responses measurable through EEG-based indicators.

For example, augmented reality interfaces may increase cognitive load and attention demands due to multimodal information processing, while AI-driven adaptive systems may reduce or redistribute mental workload depending on interface predictability and personalization level. These variations are reflected in changes in attention, relaxation, and alpha/beta activity ratios.

In addition, participants provided subjective state assessments after task completion.

These variables were selected to capture cognitive load modulation and stress-related adaptation dynamics in digital environments.

3.4 Statistical Modeling

Latent Profile Analysis (LPA) was applied to nine psychophysiological and subjective indicators to identify homogeneous adaptation clusters [15], [16].

Models with 2 to 6 classes were evaluated using Bayesian Information Criterion (BIC), Entropy and Classification quality indicators. The four-class solution demonstrated optimal model fit: BIC = −1238.721, Entropy = 0.93. High entropy indicates strong classification precision and profile separation. Analysis of variance (ANOVA) was conducted to compare stress responses across experimental conditions.

3.5 Governance Mapping Procedure

To translate empirical findings into ecosystem-level governance implications, a thematic mapping procedure was applied. Identified adaptation profiles were mapped onto functional governance layers aligned with Quadruple Helix architecture.

This procedure transformed psychophysiological clusters into structured governance parameters, enabling integration into adaptive digital ecosystem design.

4 RESULTS

4.1 Stress Amplification Effect

Analysis of variance (ANOVA) revealed a statistically significant interaction effect between environmental novelty and time pressure. The condition combining a novel digital interface with execution under time constraint produced the highest mental workload scores, the lowest relaxation indices, reduced subjective state stability.

Compared to baseline conditions (familiar environment without time constraint), this configuration demonstrated a pronounced amplification of cognitive strain.

These findings confirm that rapid innovation exposure under temporal pressure constitutes a measurable stress-inducing configuration in digital task environments. At ecosystem scale, such configurations may accumulate into systemic resistance to innovation rollout.

4.2 Latent Adaptation Profiles

Latent Profile Analysis identified four stable psychophysiological adaptation clusters (Table 1).

Table 1: Adaptation profiles and corresponding governance modulation strategies.

Profile	Cognitive Pattern	Stress Response	Governance Action
Adaptive Leaders	High stability	Low amplification	Early rollout
Balanced Adapters	Moderate regulation	Controlled load	Standard sequencing
Adaptive Novices	Elevated workload	Moderate sensitivity	Guided onboarding
Pressure-Vulnerable	Low regulation	High amplification	Protective staging

The first cluster (Adaptive Leaders) demonstrated high attention levels, stable relaxation, low mental workload, high subjective state stability. This group demonstrates efficient cognitive regulation under stress-intensive digital conditions.

The second cluster (Balanced Adapters) demonstrated moderate attention and workload, stable self-assessment, gradual adaptation under novelty. Represents stable but not stress-optimized adaptation.

The third cluster (Adaptive Novices) demonstrated increased cognitive effort, moderate stress sensitivity, elevated workload during novelty exposure. Indicates higher resource expenditure during adaptation.

The fourth cluster (Pressure-Vulnerable Users) demonstrated low attention stability, high mental workload, reduced relaxation. Pronounced stress amplification under time pressure. This cluster represents the highest risk group for innovation resistance within digital ecosystems.

The four adaptation profiles identified through latent classification are summarized in Table 1.

Entropy value (0.93) confirms strong separation between profiles, supporting classification reliability.

4.3 Governance-Relevant Interpretation

The distribution of adaptation profiles suggests that digital ecosystems contain heterogeneous cognitive-regulatory capacities. Uniform innovation rollout strategies therefore generate asymmetric stress impacts across user groups.

From a governance perspective, the presence of pressure-vulnerable clusters indicates that unmodulated innovation cycles may reduce overall ecosystem resilience. Conversely, adaptive leaders may function as stabilizing agents in staged rollout configurations.

These results justify the transformation of psychophysiological profiling into structured governance parameters, forming the basis for adaptive ecosystem architecture.

4.4 Ecosystem-Level Implications of Adaptation Heterogeneity

The identified heterogeneity of adaptation profiles has systemic implications beyond individual task performance. In digital innovation ecosystems structured through Quadruple Helix interaction, users represent the civil society helix and directly influence innovation diffusion dynamics.

The presence of pressure-vulnerable clusters indicates that innovation-intensive environments may generate asymmetric stress distribution across ecosystem participants. Without adaptive modulation, such asymmetry can accumulate into resistance patterns, reduced participation, and lowered institutional trust.

Conversely, adaptive leaders may act as early adopters and stabilizing agents during innovation rollout. This suggests that ecosystem governance should not assume homogeneous adaptation capacity but instead incorporate profile-sensitive modulation mechanisms. From a systemic perspective, psychophysiological heterogeneity functions as a hidden governance variable affecting ecosystem resilience, innovation velocity, and long-term sustainability.

These findings support the need for transforming adaptation profiling into a structural governance component rather than treating it as a descriptive psychological classification.

5 ADAPTIVE GOVERNANCE ARCHITECTURE FOR DIGITAL INNOVATION ECOSYSTEMS

The empirical findings demonstrate that adaptation to innovation-intensive digital environments is structurally heterogeneous and systematically

distributed across identifiable psychophysiological clusters. This heterogeneity should not be interpreted as incidental variation but rather as a latent regulatory factor influencing ecosystem stability. Within digital innovation ecosystems structured according to Quadruple Helix logic, user communities represent the civil society helix and therefore constitute a central element in innovation diffusion dynamics. If adaptation variability is ignored, governance decisions remain detached from actual cognitive capacity distributions across the ecosystem.

To address this limitation, the proposed architecture translates psychophysiological adaptation profiling into a structured governance mechanism. The model follows a layered logic in which measurable cognitive and stress-related indicators inform operational, strategic, conceptual, and socio-cultural dimensions of ecosystem management. In this framework, adaptation profiles function as dynamic regulatory signals that shape how innovation is introduced, scaled, and socially integrated.

From a functional perspective, the proposed system operates as an adaptive decision-support mechanism that processes psychophysiological and behavioral input data and translates them into governance actions. Specifically, the system performs several core management functions. First, it classifies users into adaptation profiles based on measurable cognitive and stress-related indicators. Second, it dynamically adjusts interaction parameters, including interface complexity, pacing, and feature exposure, in order to maintain cognitive stability. Third, it regulates innovation rollout by sequencing feature introduction according to aggregate adaptation capacity within the ecosystem. Finally, it generates aggregated adaptation metrics that serve as governance indicators for monitoring ecosystem resilience and guiding strategic decision-making.

At the operational level (Level A), the architecture focuses on real-time modulation of user-system interaction. Instead of assuming homogeneous cognitive capacity, the system dynamically adjusts interaction complexity based on identified adaptation profiles. Users with higher stress sensitivity may be exposed to gradual feature introduction, simplified interface modes, or contextual microlearning support. Conversely, users demonstrating stable cognitive regulation under novelty conditions may be offered expanded functionality earlier in the adoption cycle. In this way, psychophysiological profiling becomes an instrument for reducing cognitive overload and stabilizing short-term interaction dynamics.

The second layer (Level B) operates at the strategic innovation management level. Here, profile distribution across the ecosystem informs the sequencing of innovation rollout. Rather than implementing uniform release strategies, governance mechanisms calibrate innovation velocity according to aggregate adaptation capacity. Early adoption phases may prioritize cognitively resilient clusters, while optional activation mechanisms protect stress-vulnerable groups from excessive overload. This adaptive sequencing reduces systemic resistance and enhances diffusion stability, particularly in volatile and rapidly evolving technological environments.

At the conceptual governance level (Level C), adaptation sensitivity is embedded into institutional design standards. Psychological safety and cognitive stability are reframed as measurable system parameters rather than abstract organizational values. Technical specifications may include stress-modulation criteria, while interdisciplinary governance structures integrate behavioral expertise into engineering decision processes. In this configuration, psychophysiological variability becomes a formal design constraint within digital ecosystem architecture, contributing to long-term resilience and sustainable innovation cycles.

The highest layer (Level D) addresses socio-cultural stabilization within the broader Helix ecosystem. Innovation sustainability depends not only on technical efficiency but also on perceived legitimacy, trust, and user agency. Transparent signaling of collaboration between academia, industry, government, and civil society strengthens institutional visibility and contextual trust formation. At this level, Helix-based macro-nudging operates as a structural framing mechanism: the perceived presence of multiple institutional actors influences how innovation is interpreted before direct interaction occurs. Ethical signaling and participatory framing enhance perceived autonomy and long-term ecosystem engagement.

Taken together, the four layers form a modular adaptive governance architecture in which psychophysiological input data are processed through latent profile classification and mapped onto ecosystem-level management strategies. The resulting system enables scalable integration with wearable sensing technologies, behavioral analytics infrastructures, and AI-driven interface modulation engines. By embedding measurable adaptation parameters into governance logic, the architecture aligns human cognitive variability with innovation

management processes and enhances digital ecosystem resilience under Industry 4.0 conditions.

6 TECHNOLOGICAL IMPLICATIONS AND IMPLEMENTATION SCENARIOS

The proposed adaptive governance architecture is designed to be technically scalable and compatible with existing digital ecosystem infrastructures. Its implementation does not require radical restructuring of platform architecture but rather the integration of an adaptive profiling and modulation layer within existing data-processing pipelines. The multi-layer governance structure is illustrated in Figure 1.

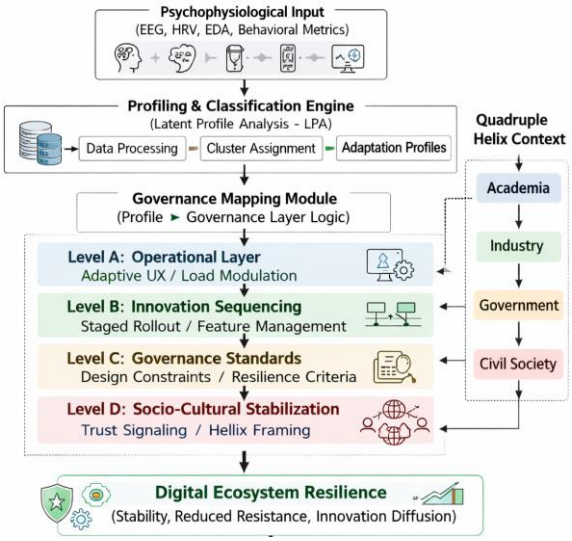


Figure 1: Multi-layer adaptive governance architecture integrating psychophysiological profiling into digital innovation ecosystems.

In practical Industry 4.0 applications, such adaptation-sensitive modulation can be implemented in environments such as smart manufacturing systems, where operator interfaces dynamically adjust information density, or in AI-driven digital platforms, where recommendation and interaction logic adapt to user cognitive states. In augmented reality-based training systems, real-time stress indicators may be used to regulate task complexity and pacing, thereby preventing cognitive overload and improving learning efficiency.

At the input level, psychophysiological indicators may be obtained through wearable devices measuring heart rate variability (HRV), electrodermal activity (EDA), or sleep-based stress proxies, as well as through EEG-based devices in controlled environments. In large-scale digital ecosystems, behavioral analytics can function as indirect cognitive-load proxies, capturing response latency, error frequency, task abandonment rates, or interaction volatility. These data streams can be integrated into a profiling engine based on latent classification models similar to those validated in the present study.

The profiling engine operates as an intermediate decision layer that dynamically assigns users to adaptation clusters based on real-time or periodically updated indicators. At the system level, this decision layer enables continuous feedback loops in which user interaction data are used to refine adaptation profiles and update governance parameters. As a result, the system evolves over time, improving its ability to predict stress responses and optimize innovation deployment strategies.

This classification does not require medical precision; rather, it functions as a probabilistic modulation mechanism aimed at reducing cognitive overload and preventing stress accumulation. Machine learning models may continuously refine cluster assignment through feedback loops incorporating behavioral performance and subjective micro-assessments.

At the operational level, adaptive interface engines can modulate complexity, pacing, and feature exposure. For example, interface density, notification frequency, or workflow branching can be dynamically calibrated according to detected adaptation capacity. In innovation-intensive environments, staged release algorithms may prioritize exposure for cognitively resilient clusters while introducing optional activation paths for stress-sensitive users. From a governance analytics perspective, aggregated profile distributions can serve as ecosystem health indicators. A growing proportion of stress-vulnerable clusters may signal excessive innovation velocity, prompting strategic recalibration.

Conversely, stable cognitive regulation across clusters may support accelerated rollout cycles. In this configuration, psychophysiological profiling becomes a systemic feedback variable influencing innovation management decisions.

Importantly, implementation must adhere to strict data privacy and ethical transparency standards. Adaptation-sensitive systems should operate with

informed consent, anonymized aggregation, and clear communication regarding profiling logic. Transparency strengthens perceived agency and prevents adaptive governance from being interpreted as manipulative control.

Overall, the architecture functions as an adaptive overlay compatible with AI-driven personalization engines, cloud-based analytics dashboards, and Industry 4.0 cyber-physical infrastructures. By integrating measurable human adaptation parameters into innovation governance, digital ecosystems can shift from reactive stress management to proactive resilience modulation.

7 CONCLUSIONS

Digital innovation ecosystems increasingly operate under conditions of accelerated technological change, where systemic instability often arises not from technological insufficiency but from misalignment between innovation velocity and human adaptive capacity. While Quadruple Helix models provide a macro-structural understanding of institutional collaboration, they do not incorporate measurable cognitive and stress-related parameters into governance logic.

This study introduced a four-level adaptive governance architecture that integrates psychophysiological user profiling into digital ecosystem management. Based on empirical identification of four stable adaptation profiles and a validated stress amplification effect under novelty and time pressure conditions, the proposed model transforms human adaptation heterogeneity into a structured governance variable. By mapping adaptation profiles onto operational, strategic, conceptual, and socio-cultural layers, the architecture aligns user cognitive variability with innovation management processes.

The framework contributes theoretically by bridging behavioral adaptation research with digital ecosystem governance and extending Helix-based innovation models through measurable human-system integration. Practically, it provides a scalable decision-support structure compatible with wearable sensing technologies, behavioral analytics systems, and AI-driven interface modulation engines.

Future research may focus on cross-cultural validation of adaptation distributions, integration with real-time machine learning classifiers, and longitudinal assessment of ecosystem resilience under adaptive governance conditions. Embedding measurable adaptation parameters into digital infrastructure design represents a shift from reactive stress mitigation toward proactive resilience engineering in Industry 4.0 ecosystems.

REFERENCES

[1] M. Ghobakhloo, H. A. Mahdiraji, M. Iranmanesh and V. Jafari-Sadeghi, “From industry 4.0 digital manufacturing to industry 5.0 digital society: A roadmap toward human-centric, sustainable, and resilient production,” Information Systems Frontiers, 2024, doi: 10.1007/s10796-024-10476-z.

[2] R. R. Kunwar and N. Ulak, “Extension of the triple helix to quadruple to quintuple helix model,” Journal of APF Command Staff College, vol. 7, no. 1, pp. 241-280, Jun. 2024, doi: 10.3126/japfcsc.v7i1.67006.

[3] E. G. Carayannis and D. F. J. Campbell, “Triple helix, quadruple helix and quintuple helix and how do knowledge, innovation and the environment relate to each other?,” International Journal of Social Ecology and Sustainable Development, vol. 1, no. 1, pp. 41-69, 2010, doi: 10.4018/jsesd.2010010105.

[4] M. Weinmann, C. Schneider and J. vom Brocke, “Digital nudging,” Business & Information Systems Engineering, vol. 58, no. 6, pp. 433-436, 2016, doi: 10.1007/s12599-016-0453-1.

[5] O. Ronzhes, “Psychological, psychophysiological and social adaptation of a person during the period of active transition to digital technologies,” Ph.D. dissertation, V. N. Karazin Kharkiv National University, Kharkiv, Ukraine, 2025.

[6] A. T. Lee, R. K. Ramasamy and A. Subbarao, “Understanding psychosocial barriers to healthcare technology adoption: A review of TAM and UTAUT frameworks,” Healthcare, vol. 13, no. 3, p. 250, 2025, doi: 10.3390/healthcare13030250.

[7] A. C. Edmondson, The Fearless Organization: Creating Psychological Safety in the Workplace for Learning, Innovation, and Growth. Hoboken, NJ, USA: Wiley, 2018.

[8] O. Ronzhes and I. V. Kryazh, “Latent profiles of human adaptation indicators to new digital environments,” Problems of Modern Transformations. Series: Pedagogy, no. 7, 2025, doi: 10.54929/2786-9199-2025-7-10-03.

[9] E. G. Carayannis and D. F. J. Campbell, Mode 3 Knowledge Production in Quadruple Helix Innovation Systems. New York, NY, USA: Springer, 2012.

[10] M. Salun and K. Zaslavska, “Strategies for resilience in a dynamic world: From VUCA to BANI,” in Socratic Lectures 10. University of Ljubljana Press, 2024, doi: 10.55295/psl.2024.i23.

[11] R. Tarafdar, C. Cooper and J. Stich, “The technostress trifecta: Techno eustress, techno distress and design,” Information Systems Journal, vol. 29, no. 1, pp. 6-42, 2019, doi: 10.1111/isj.12169.

[12] S. Ayyagari, V. Grover and R. Purvis, “Technostress: Technological antecedents and implications,” MIS Quarterly, vol. 35, no. 4, pp. 831-858, 2011.

[13] R. M. Ferrari, A. A. dos Santos, G. Salomão and M. A. Terlizzi, “Replication of technostress: Technological antecedents and implications in the post-pandemic context,” Communications of the

Association for Information Systems, vol. 57, no. 1, pp. 1111-1133, 2025, doi: 10.17705/1cais.05748.

[14] R. M. Ryan and E. L. Deci, "Self-Determination theory," in Encyclopedia of Quality of Life and Well-Being Research. Cham, Switzerland: Springer International Publishing, 2023, pp. 6229-6235, doi: 10.1007/978-3-031-17299-1_2630.

[15] K. L. Nylund-Gibson and A. Y. Choi, "Ten frequently asked questions about latent class analysis," Translational Issues in Psychological Science, vol. 4, no. 4, pp. 440-461, 2018, doi: 10.1037/tps0000176.

[16] B. S. Everitt and T. Hothorn, An Introduction to Applied Multivariate Analysis with R. New York, NY, USA: Springer, 2011.

Spectral Solution of Fractional Delay Equations via Fermat Polynomials

Abdullah Hussein[1] and Ali Khalaf Hussain Al-Hachami[2]
Department of Mathematics, College of Education for Pure Science, Wasit University, 52001 Wasit, Iraq
std.2024205.a.khalee@uowasit.edu.iq, alhachamia@uowasit.edu.iq

Keywords: Fermat Polynomials, Fractional a Delay Equation, Fermat Operational Matrix, Galerkin Method.

Abstract: This paper presents an efficient and accurate numerical approach for solving fractional differential equations (FDEs), particularly those with delay terms, using a truncated series of Fermat polynomials. The developed method converts the fractional differential equation and its associated initial conditions into a system of algebraic equations by employing the Galerkin spectral technique combined with the operational matrix of fractional-order derivatives in the Caputo sense. Fermat polynomials are adopted as basis functions due to their advantageous analytical properties and recursive structure, making them suitable for constructing spectral approximations. The operational matrix formulation significantly reduces computational complexity and facilitates efficient implementation. The approximate solution is obtained with high accuracy using only a limited number of basis functions, ensuring rapid convergence. To demonstrate the effectiveness of this approach, several numerical examples, including fractional delay differential equations, are investigated. The results show excellent agreement with exact solutions, even for non-polynomial problems, and comparisons with existing methods confirm the superiority, stability, and computational efficiency of the proposed technique.

1 INTRODUCTION

The capacity of fractional differential equations (FDEs) to simulate complicated processes that display memory effects and cumulative behaviors in a variety of disciplines, including as physics, engineering, biological systems, and control theory, has drawn a lot of interest in recent years. A particularly important category within this field is fractional differential equations with delay, when a system's development is not only dependent on its present condition, but also on its previous states. The presence of fractional derivatives with delay increases the complexity of these equations, making obtaining analytical solutions difficult, and therefore numerical methods are often used. Spectral methods are very effective numerical techniques for solving FDEs, offering quick convergence and great accuracy in comparison to conventional techniques. These methods approximate the solution using a series of basis functions, with the expansion coefficients determined by spectroscopic techniques such as (Collocation), (Tau), and (Galerkin) methods. Among them, Galerkin's method is widely used, approximating the solution using a set of base functions while imposing the condition of orthogonality to the remainder with respect to the chosen domain of functions. For more details about spectral methods, you can refer to [1]-[4]. Fermat polynomials have also received attention in numerical analysis because of their mathematical properties and iterative relationships, which make them suitable for constructing spectral approximations and operational matrices of fractional derivatives. It is a special case of generalized Fibonacci polynomials of the type (p, q) [5]. Several studies [6]-[8] have shown that the use of Fermat polynomials within spectral methods leads to accurate and efficient numerical solutions to a variety of fractional equations. In order to solve FDEs with delay, we provide a spectral approach based on Fermat polynomials in this paper. Fermat polynomials in a finite series are used to approximate the answer, a proposed Galerkin-type formulation is employed to convert the equation and its associated conditions into a set of algebraic equations that can be used to compute the expansion coefficients. Several numerical examples were given to illustrate the effectiveness of the suggested approach, comparing the results with other existing methods, showing high accuracy and rapid convergence using relatively few

basis functions. The structure of this research is as the paper is organized as follows. First, the basic concepts of fractional differentiation are reviewed. Next, the definitions of Fermat polynomials are introduced. Then, the operational matrices for fractional derivatives in the Caputo sense are developed. After that, the proposed numerical method is described, followed by various numerical examples. Finally, concluding remarks are presented.

2 PRELIMINARIES

The principles of fractional calculus theory, which will be applied in this investigation, are reviewed in this part.

Definition 1: [9] The following is the definition of the Riemann-Liouville (R._L.) fractional integrals of order β operator I^{β}:

$$I^{\beta}\varphi(x)=\begin{cases}\frac{1}{\Gamma(\beta)}\int_a^x \frac{\varphi(t)}{(x-t)^{1-\beta}}\,dt, & x>a,\\ \varphi(x) & \beta=0\end{cases}. \quad (1)$$

These are the characteristics of the (R._L.) fractional integral [10]:

1) if $\beta>0$ & $\delta>0$ $I^{\beta}I^{\delta}\varphi(x)=I^{\beta+\delta}\varphi(x)$;
2) let $n,m\in R$: $I^{\beta}\big(n\varphi(x)+m\psi(x)\big)=nI^{\beta}\varphi(x)+mI^{\beta}\psi(x)$;
3) for $k>0$: $I^{\beta}x^{k}=\frac{\Gamma(k+1)}{\Gamma(k+1+\beta)}x^{k+\beta}$;
4) C is constant: $I^{\beta}C=\frac{C}{\Gamma(\beta+1)}x^{\beta}$.

Definition 2: [11] The following is the definition of the Riemann-Liouville (R._L.) fractional differential of order β operator D^{β}:

$$D^{\beta}\varphi(x)=\frac{1}{\Gamma(n-\beta)}\frac{d^{n}}{dx^{n}}\int_a^x\frac{\varphi(t)}{(x-t)^{\beta+1-n}}dt. \quad (2)$$

Where n is an integer and $n-1<\beta\le n$

Definition 3: [12] The operator representing the Caputo fractional derivative of order β can be formulated as

$${}^{C}D^{\beta}\varphi(x)=\frac{1}{\Gamma(n-\beta)}\int_a^x\frac{\varphi^{(n)}(t)}{(x-t)^{\beta+1-n}}dt. \quad (3)$$

Where n is an integer and $n-1<\beta\le n$

Some Properties of Fractional Derivative:

1) ${}^{C}D^{\beta}I^{\beta}\varphi(x)=\varphi(x)$.
2) $I^{\beta}\ {}^{C}D^{\beta}\varphi(x)=\varphi(x)-\sum_{k=0}^{n-1}\frac{\varphi^{k}(0)}{k!}x^{k}, x>0$
3) $D^{\beta}x^{k}=\frac{\Gamma(k+1)}{\Gamma(k+1-\beta)}x^{k-\beta}, k\in\mathbb{N}, k\ge\lceil\beta\rceil$.
4) ${}^{C}D^{\beta}\varphi(x)=D^{\beta}\varphi(x)-\sum_{k=0}^{n-1}\frac{(x-t)^{k-\beta}}{\Gamma(k+1-\beta)}D^{K}\varphi(t)$.

3 A REVIEW OF FERMAT POLYNOMIALS

If we examine the (p,q)-Fibonacci polynomials' recurrence relation

$$u_n(x)=p(x)u_{n-1}(x)+q(x)u_{n-2}(x).$$

Considering the initial circumstances $u_0(x)=0$ and $u_1(x)=1$ by substituting $p(x)=3x$ and $q(x)=-2$, we obtain the Fermat polynomials, which are considered A particular instance of the Fibonacci polynomials [13]. We can define the Fermat polynomials by means of the difference equation [14].

$$F_i(x)=3xF_{i-1}(x)-2F_{i-2}(x), \quad (3)$$
$$F_0(x)=0, F_1(x)=1, \quad i\ge 2,$$

we can write the Binet formula for the Fermat polynomials. First, we find the two roots:

$$\alpha(x)=\frac{3x+\sqrt{9x^2-8}}{2}, \qquad \beta(x)=\frac{3x-\sqrt{9x^2-8}}{2}.$$

$$F_i=\frac{\alpha^{i}(x)-\beta^{i}(x)}{\alpha(x)-\beta(x)}. \quad (4)$$

A Fermat polynomial can be written explicitly over the domain $x\in(0,1)$ as follows:

$$F_k(x)=\begin{cases}\frac{\left(3x+\sqrt{9x^2-8}\right)^{k}-\left(3x-\sqrt{9x^2-8}\right)^{k}}{2^{k}\sqrt{9x^2-8}}, & x\ne\frac{2}{3}\\ 2^{\frac{k}{2}}sin\left(\frac{\pi}{4}k\right), & x=\frac{2}{3}\end{cases} \quad k=1,2,. \quad (5)$$

The polynomial $F_k(x)$, of degree $k-1$, is characterized by integer coefficients, which is an important point

The analytic version of Fermat polynomials can be expressed as.$F_{i+1}(x), i\ge 1$ as follows:

$$F_{i+1}=\sum_{m=0}^{\left\lfloor\frac{i}{2}\right\rfloor}(-2)^{m}\,3^{i-2m}\binom{i-m}{m}x^{i-2m}. \quad (6)$$

Theorem 1.[6] The inversion formula presented below is applicable for every nonnegative integer k

$$x^{k}=\frac{1}{3^{k}}\sum_{i=0}^{\left\lfloor\frac{k}{2}\right\rfloor}\frac{2^{i}(k+1-2i)}{k+1-i}\binom{k}{i}F_{k+1-2i}(x). \quad (7)$$

Theorem 2.[6] The formula below establishes the connection between Fermat polynomials and their first derivatives:

$$DF_{i+1}(x)=3\sum_{k=0}^{\left\lfloor\frac{i-1}{2}\right\rfloor}2^{k}\,(i-2k)F_{i-2k}(x). \quad (8)$$

4 THE CAPUTO FRACTIONAL DERIVATIVE'S FERMAT OPERATIONAL MATRIX

4.1 Preliminary Formulation of Fermat Operational Matrices

Let $\varphi(x)$ denote a function on $(0,1)$ that is square-integrable under the Lebesgue measure. It is assumed that $\varphi(x)$ can be expanded in terms of a collection of linearly independent Fermat polynomials as follows:

$$\varphi(x) = \sum_{m=1}^{\infty} r_j F_j(x).$$

The following is applicable if the series is shortened.

$$\varphi(x) \approx \varphi_N(x) = \sum_{m=1}^{N+1} r_m F_m(x) = \mathrm{R}^T \Phi(x). \quad (9)$$

Where

$$\mathrm{R}^T = [r_1, r_2, r_3, \ldots\ldots, r_{N+1}], \quad (10)$$

and

$$\Phi(x) = [F_1(x), F_2(x), F_3(x), \ldots\ldots, F_{N+1}(x)].^T \quad (11)$$

Assuming $\frac{d\Phi(x)}{dx}$, we may write it as:

$$\frac{d\Phi(x)}{dx} = \mathrm{G}^1 \Phi(x), \quad (12)$$

$\varphi(x) = R^T \mathrm{G}^1 \Phi(x)$ where the non-zero entries of the matrix G^1,can be obtained directly from relation (8). which is the $(N+1) \times (N+1)$ (OMD).

Remark 1. The overall structure of the (OMD) where $\mathrm{G}^1 = g_{ij}^{(1)}$ then can be written as follows:

$$g_{ij}^{(1)} = \begin{cases} 3(j+1)2^{\frac{i-j-1}{2}} & \text{if } i > j, (i+j)odd \\ 0 & ,otherwise \end{cases} \quad (13)$$

4.2 Building of Fermat OMFD

In this section, we aim to formulate the operational matrix of fractional derivatives as an extension of the corresponding matrix for integer-order derivatives. It follows directly from Relation (8) that for any positive integer:

$$\frac{d^s \Phi(x)}{dx^s} = \mathrm{G}^{(s)} \Phi(x) = \left(\mathrm{G}^{(1)}\right)^s \Phi(x). \quad (14)$$

Theorem 3. [6] Let the vector be $\Phi(x)$ of Fermat polynomials described by Equation (11). For $x \in (0,1)$ and any $\beta > 0$, one has

$$D^\beta \Phi(x) = x^{-\beta} \mathrm{G}^{(\beta)} \Phi(x), \quad (15)$$

where $\mathrm{G}^{(\beta)} = g_{i,j}^{(\beta)}$, which is explicitly written as follows, represents the $(N+1) \times (N+1)$ Fermat (OMFE) of order β in the Caputo sense.

$$G^{(\beta)} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \gamma_\beta(\lceil\beta\rceil, 1) & \gamma_\beta(\lceil\beta\rceil, \lceil\beta\rceil) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \gamma_\beta(i,1) & \cdots & \gamma_\beta(i,i) & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \gamma_\beta(n+1,1) & \gamma_\beta(n+1,2) & \gamma_\beta(n+1,3) & \cdots & \gamma_\beta(N+1,N+1) \end{pmatrix}. \quad (17)$$

You may write the entries $g_{i,j}^{(\beta)}$ like this:

$$g_{ij}^{(\beta)} = \begin{cases} \gamma_\beta(i,j) & \text{if } i \geq \lceil\beta\rceil, i \geq j \\ 0 & ,otherwise \end{cases}. \quad (16)$$

Where:

$$\gamma_\beta(i,j) = j \sum_{\substack{m=\lceil\beta\rceil \\ (i+j)odd,(j+m)odd}}^{i} \frac{m!\,(-1)^{\frac{1}{2}(i-2j+m+1)} 2^{\frac{i-j}{2}} \left(\frac{i+m-1}{2}\right)!}{\left(\frac{i-m-1}{2}\right)! \left(\frac{m-j+1}{2}\right)! \left(\frac{m+j+1}{2}\right)! \Gamma(-\beta+m+1)} \quad (19)$$

For instance, when $N = 5$, the operational matrix $\mathrm{G}^{\frac{3}{2}}$ is defined as:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{8}{\sqrt{\pi}} & 0 & \frac{4}{\sqrt{\pi}} & 0 & 0 & 0 \\ 0 & \frac{32}{\sqrt{\pi}} & 0 & \frac{8}{\sqrt{\pi}} & 0 & 0 \\ \frac{272}{5\sqrt{\pi}} & 0 & \frac{264}{5\sqrt{\pi}} & 0 & \frac{64}{5\sqrt{\pi}} & 0 \\ 0 & \frac{768}{7\sqrt{\pi}} & 0 & \frac{576}{7\sqrt{\pi}} & 0 & \frac{128}{7\sqrt{\pi}} \end{pmatrix}.$$

5 METHOD OF SOLUTION

The recommended approach for creating a numerical solution to the Delay equation is described in this section.

Examine the fractional delay differential equation that follows:

$$D^s \varphi(x) + D^\beta \varphi(x) = H[\varphi(x), \varphi(dx-\tau), z(x)] \quad x \in (0,1) \quad (20)$$

with initial conditions

$$\varphi^{(i)}(0) = v_i, \quad i = 0,1,\ldots,m \quad (17)$$

where $m < \beta \leq m+1$, β Largest fractional power $d \in (0,1]$, $\tau \in (0,1]$ $H\big(\varphi(x), \varphi(dx-\tau), z(x)\big)$ is a

polynomial in $\varphi(x), \varphi(dx), z(x)$, $D^s\varphi(x)$ is the derivative of integers,$D^\beta\varphi(x)$ is the Caputo derivative.

Now we can approximate $\varphi(x)$ and $\varphi(dx-\tau)$

$$\varphi(x) \approx \varphi_N(x) = \mathrm{R}^T\Phi(x),$$
$$\varphi(dx-\tau) \approx \varphi_N(dx-\tau) = \mathrm{R}^T\Phi(dx-\tau), \quad (18)$$
$$D^s\varphi(x) \approx \mathrm{R}^T\mathrm{G}^{(s)}\Phi(x). \quad (19)$$

The following approximation can be produced by virtue of Theorem 3:

$$D^\beta\Phi(x) \approx x^{-\beta}\mathrm{R}^T\mathrm{G}^{(\beta)}\Phi(x). \quad (20)$$

The formula for the residual of **Помилка! Джерело посилання не знайдено.** is obtained by applying the approximations in (19) and (20). (25)

$$x^\beta R(x) = x^\beta \mathrm{R}^T\mathrm{G}^{(s)}\Phi(x) + \mathrm{R}^T\mathrm{G}^{(\beta)}\Phi(x) - x^\beta H[\varphi(x), \varphi(dx), z(x)]$$

and hence the application of Our proposed Galerkin-type approach:

$$\langle R(x), F_j(x)\rangle = \int_0^1 x^\beta\, R(x)F_j(x)dx = 0 \quad , \quad (26)$$
$$j = 1,2,3,\ldots,N-m.$$

Also, by inserting (18) into (17),we get:

$$\begin{aligned} \varphi(0) &= R\hat{}T\,\Phi(0) = v_0 \\ \varphi^1(0) &= R^T G^1\,\Phi(0) = v_1 \\ &\vdots \\ \varphi^m(0) &= R^T G^m\,\Phi(0) = v_m. \end{aligned} \quad (27)$$

Finally, using (26) and (27) of dimension $(N+1)$, a system of algebraic equations is obtained. It is possible to solve these linear eq. given the unknown expansion factors.$\mathrm{r_i}$. As a result, (22) can estimate $\varphi(x)$.

6 NUMERICAL EXAMPLES

In this part, we use the Fermat Galerkin operational matrix (FGM) method to numerically solve fractional differential equations.

Example 1: Consider the following FDD equation [15]-[17]:

$$D^\beta\varphi(x) + \varphi(x) - \varphi(x-\tau) = \frac{\Gamma(3)}{\Gamma(3-\beta)}x^{2-\beta} - \frac{\Gamma(2)}{\Gamma(2-\beta)}x^{1-\beta} - \tau^2 + 2\tau x - \tau \quad (28)$$
$$\varphi(0) = 0, \quad x \in (0,1), \quad \tau \in (0,1], \quad \beta \in (0,1]$$

The precise solution to (28) is $\varphi(x) = x^2 - x$ at $\beta = 1$ by applying the method from Section 5, Figure 1 shows the estimated answer and the corresponding precise real solution. The findings of the Krawtchouk wavelet technique (KWM) described in reference [15] and the spectral approach with non-integer order Taylor basis reported in reference [16] are contrasted with the achievable absolute error produced by the provided FGM for $N = 2$ in Table 1. It is evident that the introduced method is more accurate because we obtain the exact solution for $\beta = 1$ and for any value of τ. As an example, we used three terms for FGM when $\tau = 0.01$ in Table 1, whereas they only obtained approximate solutions using Taylor functions with eight terms of fractional order and nine terms of the Krawtchouk wavelet method. and also, better than the methods used in [18]-[20].

Table 1: The absolute errors at $\tau = 0.01$ and $\beta = 1$ are compared for Example 1.

x	Results of [16]	KWM	P. method
0.2	3.47×10^{-16}	0	0
0.4	2.36×10^{-16}	5.55×10^{-17}	0
0.6	3.47×10^{-16}	5.55×10^{-17}	0
0.8	9.69×10^{-16}	5.55×10^{-17}	0
1	0	7.81×10^{-17}	0

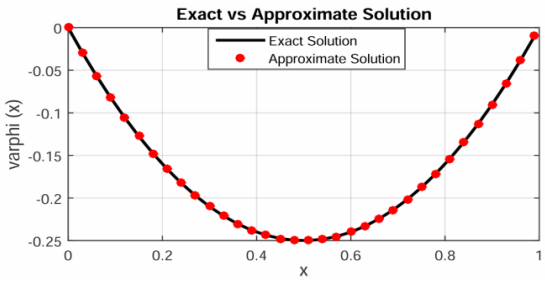


Figure 1: Graphical representations of the exact and approximate solutions for Example 1

Example 2: Consider the following FDD equation [21], [22]:

$$D^{0.2}\varphi(x) + \varphi(x-1) = \frac{\Gamma(3)}{\Gamma(0.2)}x^{1.8} - \frac{\Gamma(2)}{\Gamma(1.8)}x^{0.8} + x^2 - 3x + 1 \quad (29)$$
$$\varphi(0) = -1 \qquad x \in (0,1).$$

The precise solution to equation **Помилка! Джерело посилання не знайдено.** is $\varphi(x) = x^2 - x - 1$ by applying the method from Section 5, The estimated solution is displayed beside the corresponding precise real solution and the absolute error in Figure 2, Figure 3. In Table 2, the results of the fractional-order Chelyshkov functions (FCHFs) given in reference [21] and the results of the Legendre collocation technique reported in reference

[22] have been compared with the achievable absolute error acquired by the provided FGM for $N = 3$.

Table 2. The findings concerning the approximation solutions absolute errors Example 2.

x	FCHFs	Results of [16]	P. method
0.2	5.2×10^{-10}	8.9×10^{-16}	2.6×10^{-17}
0.4	2.4×10^{-10}	1.6×10^{-15}	5.8×10^{-17}
0.6	1.1×10^{-10}	1.6×10^{-15}	9.5×10^{-17}
0.8	9.7×10^{-12}	1.3×10^{-15}	1.4×10^{-16}
1	7.3×10^{-11}	7.8×10^{-16}	1.8×10^{-16}

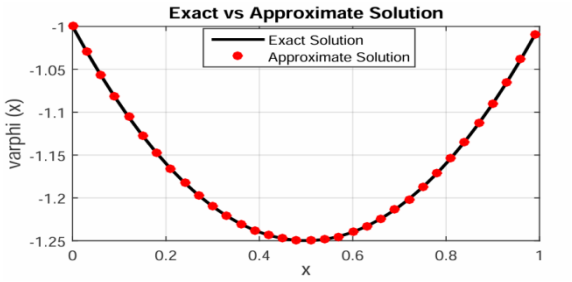


Figure 2: Graphical representations of the exact and approximate solutions for Example 2.

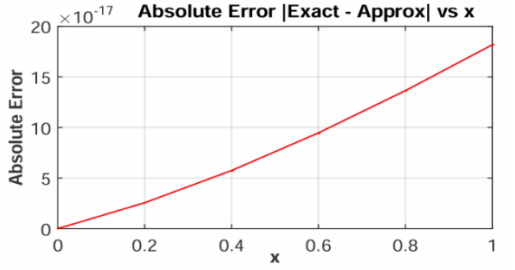


Figure 3: Absolute errors of Example 2.

Example 3: Consider the following FDD equation [20], [21], [23]:

$$D^{\beta}\varphi(x) + \varphi(x) = \frac{\tau}{2}\varphi(\tau x) - \frac{\tau}{2}e^{-\tau x}. \qquad (21)$$
$$\varphi(0) = 1, \quad x \in (0,1), \quad \tau \in (0,1], \quad \beta \in (0,1]$$

The precise solution to equation (21) is $\varphi(x) = e^{-x}$, at $\beta = 1$ by applying the method from Section 5,In Figure 4 We show how the approximate solution behaves, and Figure 5 shows the absolute errors obtained using the current approach. In Table 3, We have compared the findings of the fractional-order Chelyshkov functions (FCHFs) published in reference [21] with the attained absolute error acquired by the provided FGM for $N = 10$,The comparison indicates that the maximum absolute error produced by the proposed FGM is of order (4.57×10^{-14}), while the FCHFs method yields an error of order (1.49×10^{-12}), highlighting the effectiveness and high precision of the proposed method.

Table 3: The absolute errors at $\tau = 0.01$ and $\beta = 1$ are compared for Example 3.

x	FCHFs	P. method
0	2.16×10^{-14}	0
0.2	4.66×10^{-15}	4.57×10^{-14}
0.4	6.22×10^{-15}	3.11×10^{-14}
0.6	8.09×10^{-14}	1.84×10^{-14}
0.8	3.84×10^{-13}	2.10×10^{-14}
1	1.49×10^{-12}	1.80×10^{-14}

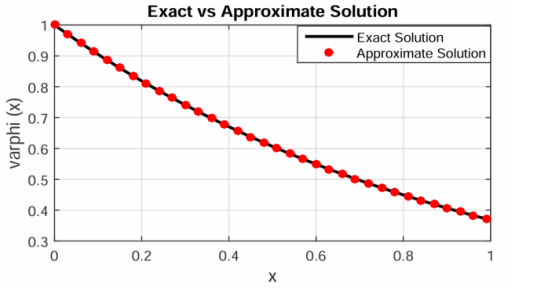


Figure 4: The visual representations of the exact and approximate solutions for Example 3.

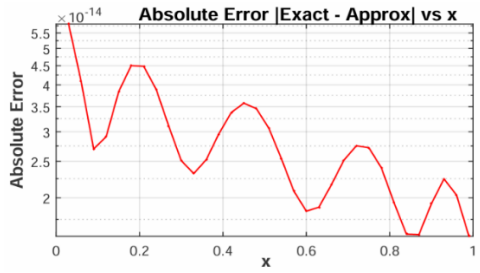


Figure 5: Absolute errors of Example 3.

Example 4: Consider the following FDD equation [24]-[26]

$$D^{0.3}\varphi(x) + \varphi(x) = \varphi(x-1) + 3x^2 - 3x + 1 + \frac{6}{\Gamma(4-0.3)}x^{2.7} \qquad (31)$$
$$\varphi(0) = 0 \qquad x \in (0,1).$$

Table 4: The findings concerning the approximation solutions' absolute errors Example 4.

x	$N = 6$
0	0
0.2	1.33×10^{-18}
0.4	9.01×10^{-18}
0.6	2.70×10^{-17}
0.8	5.98×10^{-17}
1	1.12×10^{-16}

The precise solution to **Помилка! Джерело посилання не знайдено.** is $\varphi(x) = x^3$, by applying the method from Section 5.In Table 4 presents the absolute errors at $N = 6$, The greatest absolute error of the suggested technique (FGM) and the collocation method that uses the Haar wavelet collocation method [25] and the Chebyshev cardinal functions (CCFs) mentioned in reference [24] are compared in Table 5 for different values of N. Figure 6, where the approximate answer is given in close agreement with the precise solution, demonstrates the efficacy and precision of the suggested strategy. Additionally, Figure 7 shows the method's absolute accuracy. Notably, the proposed technique attains high accuracy despite employing fewer terms compared to Methods [24] and [25].

Table 5: The largest absolute error calculated for Example 4.

N	FCHFs	N	Results of [16]	N	P.method
4	5.2×10^{-10}	4	8.9×10^{-16}	4	2.6×10^{-17}
8	2.4×10^{-10}	6	1.6×10^{-15}	5	5.8×10^{-17}
16	1.1×10^{-10}	8	1.6×10^{-15}	6	9.5×10^{-17}
32	9.7×10^{-12}	10	1.3×10^{-15}	7	1.4×10^{-16}

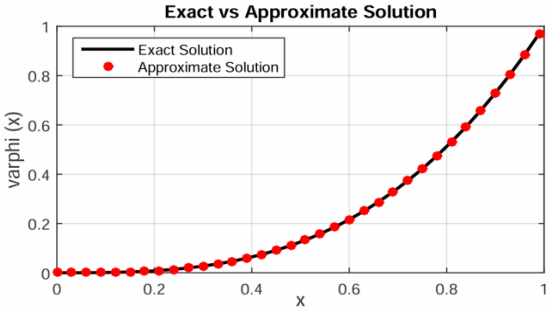


Figure 6: Graphical representations of the exact and approximate solutions for Example 4.

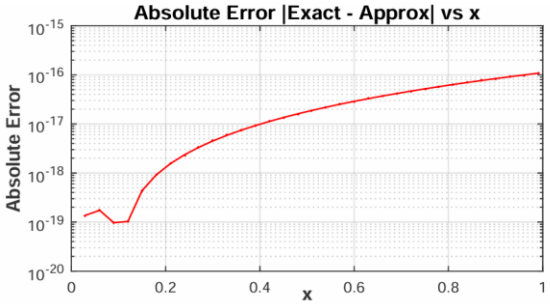


Figure 7: Absolute errors of Example 4.

Example 5: Consider the following FDD equation [27], [28].

$$D^{\beta}\varphi(x) + a\varphi(x) + b\varphi(\tau x) = (a-1)sinx + bsin(\tau x) \quad (32)$$

$$\varphi(0) = 1, \quad \varphi'^{(0)} = 1,$$
$$x \in (0,1), \quad \tau \in (0,1] \qquad \beta \in (1,2]$$

The precise solution to **Помилка! Джерело посилання не знайдено.** is $\varphi(x) = sinx$, at $\beta = 2$ for any $a, b \in R$ by applying the method from Section 5.We solve this problem when $\tau = 0.25$. The greatest absolute error of the suggested technique (FGM) and the explicit Chebyshev tau (ECT) approach described in reference [27] are compared in Table 6 for a range of N values. In Table 7 presents the absolute errors for different values of N. Figure 8 and Figure 9 show the approximate answer and the absolute accuracy that our technique produced, respectively.

Table 6: The largest absolute error calculated for Example 5.

x	ECT	P. method
4	2.2×10^{-3}	2.1×10^{-3}
8	4.6×10^{-7}	1.3×10^{-8}
12	2.8×10^{-10}	7.4×10^{-15}
16	4.4×10^{-16}	0

Table 7: The estimated solutions' absolute errors for Example 5 at various N values.

x	$N = 8$	$N = 10$	$N = 12$
0.2	2.8×10^{-9}	2.7×10^{-12}	1.7×10^{-15}
0.4	5.6×10^{-9}	5.4×10^{-12}	3.3×10^{-15}
0.6	8.3×10^{-9}	7.9×10^{-12}	4.9×10^{-15}
0.8	1.1×10^{-8}	1.0×10^{-11}	6.2×10^{-15}
1	1.3×10^{-8}	1.2×10^{-11}	7.4×10^{-15}

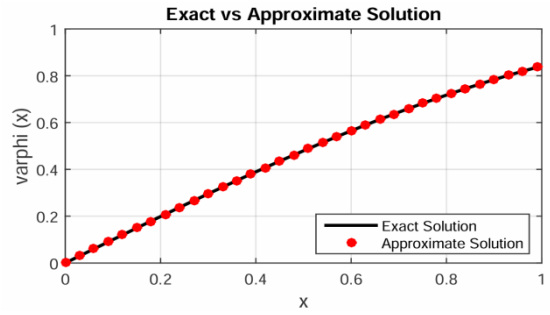


Figure 8: Graphical representations of the exact and approximate solutions for Example 5.

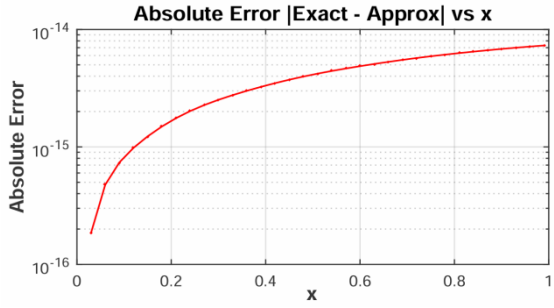


Figure 9: Absolute errors of Example 5.

7 CONCLUSIONS

In this paper, an efficient spectral Galerkin method based on Fermat polynomials and the operational matrix of fractional derivatives in the Caputo sense was proposed for solving fractional delay differential equations. The developed approach transforms the original problem into a system of algebraic equations, which can be solved efficiently with reduced computational effort.

The proposed Fermat Galerkin method demonstrated high accuracy and rapid convergence while requiring only a small number of basis functions. Several numerical examples were investigated to validate the effectiveness of the method. The obtained numerical solutions showed excellent agreement with the corresponding exact solutions and produced smaller absolute errors compared with several existing numerical techniques reported in the literature.

The results confirm that the proposed method is stable, reliable, and computationally efficient for solving fractional delay differential equations. Therefore, the presented approach can serve as an effective numerical tool for a broad class of fractional problems arising in applied mathematics, engineering, and physical sciences. Future work may focus on extending the method to nonlinear systems, multi-dimensional fractional models, and variable-order fractional differential equations.

REFERENCES

[1] J. S. Hesthaven, S. Gottlieb, and D. Gottlieb, Spectral Methods for Time-Dependent Problems, vol. 21, Cambridge Monographs on Applied and Computational Mathematics. Cambridge, U.K.: Cambridge University Press, 2007.

[2] J. P. Boyd, Chebyshev and Fourier Spectral Methods. Mineola, NY, USA: Dover Publications, 2001.

[3] L. N. Trefethen, Spectral Methods in MATLAB, vol. 10, Software, Environments, and Tools. Philadelphia, PA, USA: SIAM, 2000.

[4] H. Bhrawy, T. M. Taha, and J. A. T. Machado, "A review of operational matrices and spectral techniques for fractional calculus," Nonlinear Dynamics, vol. 81, pp. 1023-1052, 2015, [Online]. Available: https://doi.org/10.1007/s11071-015-2087-0.

[5] G. Lee and M. Asci, "Some properties of the (p, q)-Fibonacci and (p, q)-Lucas polynomials," Journal of Applied Mathematics, vol. 2012, Art. no. 264842, 18 pages, 2012, [Online]. Available: https://doi.org/10.1155/2012/264842.

[6] Y. H. Youssri, "A new operational matrix of Caputo fractional derivatives of Fermat polynomials: An application for solving the Bagley-Torvik equation," Advances in Difference Equations, vol. 2017, Art. no. 73, 2017, [Online]. Available: https://doi.org/10.1186/s13662-017-1123-4.

[7] W. M. Abd-Elhameed, O. M. Alqubori, N. M. A. Alsafri, A. K. Amin, and A. G. Atta, "A matrix approach by convolved Fermat polynomials for solving the fractional Burgers' equation," Mathematics, vol. 13, Art. no. 1135, 2025, [Online]. Available: https://doi.org/10.3390/math13071135.

[8] M. A. Taema, M. D. Dagher, and Y. H. Youssri, "Spectral collocation method via Fermat polynomials for Fredholm-Volterra integral equations with singular kernels and fractional differential equations," Palestine Journal of Mathematics, vol. 14, no. 2, pp. 481-492, 2025.

[9] R. Gorenflo and F. Mainardi, "Fractional Calculus," in A. Carpinteri and F. Mainardi (eds.), Fractals and Fractional Calculus in Continuum Mechanics, International Centre for Mechanical Sciences, vol. 378, Springer, Vienna, 1997, [Online]. Available: https://doi.org/10.1007/978-3-7091-2664-6_5.

[10] A. Kilbas, H. M. Srivastava, and J. J. Trujillo, "Theory and Applications of Fractional Differential Equations," North-Holland Mathematics Studies, vol. 204, pp. 1-523, Elsevier, 2006.

[11] M. Abu-Shady and M. K. A. Kaabar, "A generalized definition of the fractional derivative with applications," Mathematical Problems in Engineering, vol. 2021, Art. no. 9444803, 9 pages, 2021, [Online]. Available: https://doi.org/10.1155/2021/9444803.

[12] Li, F. Zeng, and F. Liu, "Spectral approximations to the fractional integral and derivative," Fractional Calculus and Applied Analysis, vol. 15, no. 3, pp. 383-

406, 2012, [Online]. Available: https://doi.org/10.2478/s13540-012-0028-x.

[13] J. Wang, "Some new results for the (p, q)-Fibonacci and Lucas polynomials," Advances in Difference Equations, vol. 2014, Art. no. 64, 2014, [Online]. Available: https://doi.org/10.1186/1687-1847-2014-64.

[14] W. Wang and H. Wang, "Some results on convolved (p, q)-Fibonacci polynomials," Integral Transforms and Special Functions, vol. 26, no. 5, pp. 340-356, 2015, [Online]. Available: https://doi.org/10.1080/10652469.2015.1007502.

[15] H. Mohammadi, H. Saeedi, and M. Izadi, "A novel design of Krawtchouk wavelet based projection approach to solve the nonlinear fractional delay differential equations," Journal of Nonlinear Mathematical Physics, vol. 32, no. 1, 2025, [Online]. Available: https://doi.org/10.1007/s44198-024-00256-3.

[16] İ. Avcı, "Numerical simulation of fractional delay differential equations using the operational matrix of fractional integration for fractional-order Taylor basis," Fractal and Fractional, vol. 6, no. 1, Art. no. 10, 2022, [Online]. Available: https://doi.org/10.3390/fractalfract6010010.

[17] M. Kumar, "An efficient numerical scheme for solving a fractional-order system of delay differential equations," International Journal of Applied and Computational Mathematics, vol. 8, Art. no. 262, 2022, [Online]. Available: https://doi.org/10.1007/s40819-022-01466-3.

[18] H. Singh, "Numerical simulation for fractional delay differential equations," International Journal of Dynamics and Control, vol. 9, pp. 463-474, 2021, [Online]. Available: https://doi.org/10.1007/s40435-020-00671-6.

[19] K. K. Ali, M. A. A. El-Salam, and E. M. Mohamed, "Chebyshev operational matrix for solving fractional order delay-differential equations using spectral collocation method," Arab Journal of Basic and Applied Sciences, vol. 26, no. 1, pp. 342-353, 2019, [Online]. Available: https://doi.org/10.1080/25765299.2019.1629543.

[20] P. Rahimkhani, Y. Ordokhani, and E. Babolian, "A new operational matrix based on Bernoulli wavelets for solving fractional delay differential equations," Numerical Algorithms, vol. 74, pp. 223-245, 2017, [Online]. Available: https://doi.org/10.1007/s11075-016-0146-3.

[21] I. Ahmed and M. S. Al-Sharif, "An effective numerical method for solving fractional delay differential equations using fractional-order Chelyshkov functions," Boundary Value Problems, vol. 2024, Art. no. 107, 2024, [Online]. Available: https://doi.org/10.1186/s13661-024-01913-8.

[22] P. K. Abbassi, M. Fathy, R. A. Elbarkoki, et al., "Semi-analytical solution of linear and nonlinear fractional delay differential equations via the Legendre collocation method," Journal of Nonlinear Mathematical Physics, vol. 32, no. 46, 2025, [Online]. Available: https://doi.org/10.1007/s44198-025-00285-6.

[23] Y. Muroya, E. Ishiwata, and H. Brunner, "On the attainable order of collocation methods for pantograph integro-differential equations," Journal of Computational and Applied Mathematics, vol. 152, nos. 1-2, pp. 347-366, 2003, [Online]. Available: https://doi.org/10.1016/S0377-0427(02)00716-1.

[24] H. B. Jebreen and I. Dassios, "The collocation method based on the new Chebyshev cardinal functions for solving fractional delay differential equations," Mathematics, vol. 12, Art. no. 3388, 2024, [Online]. Available: https://doi.org/10.3390/math12213388.

[25] R. Amin, K. Shah, M. Asif, and I. Khan, "A computational algorithm for the numerical solution of fractional order delay differential equations," Applied Mathematics and Computation, vol. 402, no. 125863, 2021, [Online]. Available: https://doi.org/10.1016/j.amc.2020.125863.

[26] M. L. Morgado, N. J. Ford, and P. M. Lima, "Analysis and numerical methods for fractional differential equations with delay," Journal of Computational and Applied Mathematics, vol. 252, pp. 159-168, 2013, [Online]. Available: https://doi.org/10.1016/j.cam.2012.06.034.

[27] W. M. Abd-Elhameed, J. A. T. Machado, and Y. H. Youssri, "Hypergeometric fractional derivatives formula of shifted Chebyshev polynomials: tau algorithm for a type of fractional delay differential equations," International Journal of Nonlinear Sciences and Numerical Simulation, vol. 23, nos. 7-8, pp. 1253-1268, 2022, [Online]. Available: https://doi.org/10.1515/ijnsns-2020-0124.

[28] H. Brunner, Q. Huang, and H. Xie, "Discontinuous Galerkin methods for delay differential equations of pantograph type," SIAM Journal on Numerical Analysis, vol. 48, no. 5, pp. 1944-1967, 2010, [Online]. Available: https://doi.org/10.1137/090771922.

Hochshule Anhalt

EDITION

ICAIIT

CONTROL AND AUTOMATION

Section 4

APRIL 2026



Proceedings of the International Conference on Applied Innovations in IT

Automation of a Pneumatic Substrate Mixing System in a Biogas Plant

Sergey Pochinok[1], Alexey Lavrentyev[1], Alaibek Obozov[2], Raya Sultanalieva[1] and Veronika Kirova[3]
[1]*Kyrgyz State Technical University named after I. Razzakov, Ch. Aitmatov Avenue 66, 720044 Bishkek, Kyrgyz Republic*
[2]*Laboratory of Renewable Energy Sources, National Academy of Sciences of the Kyrgyz Republic, Skribina Str. 23, 720055 Bishkek, Kyrgyz Republic*
[3]*Anhalt University of Applied Sciences, Bernburger Str. 55, 06366 Koethen, Germany*
info@incomcompany.kz, teolidz@mail.ru, alaibek.obozov@gmail.com, raia-ktu@mail.ru

Keywords: Biogas Plant, Anaerobic Digestion, Pneumatic Mixing, Barbotage, Intermittent Mixing, Automation, Solenoid Valve, Time Relay, Duty Cycle, Small-Scale Biogas Systems.

Abstract: Effective substrate mixing is a critical factor in the stable operation of anaerobic digesters, directly influencing mass transfer, temperature distribution, and overall biogas productivity. This paper reviews the engineering principles and practical implementation of pneumatic (barbotage) mixing in biogas reactors, with emphasis on intermittent operation controlled by standard industrial automation components. The system utilizes internally produced biogas as the working medium for cyclic gas injection, eliminating the need for mechanical agitators. The paper formalizes the duty-cycle concept for intermittent barbotage through two timing parameters - valve opening time (T_on) and pause duration (T_off) - and discusses their practical selection based on reactor geometry, substrate rheology, and operational requirements. A structured comparison of manual and automated mixing modes is presented, highlighting the limitations of operator-dependent control and the engineering advantages of timer-based actuation. The reviewed literature consistently indicates that intermittent gas mixing reduces energy demand while maintaining acceptable substrate homogeneity in small and medium-scale anaerobic digesters. The described approach demonstrates that low-cost automation based on a time relay and a solenoid valve can substantially improve operational repeatability in small-scale biogas installations without requiring complex programmable control systems. Practical timing guidelines and component specifications are provided to support engineering implementation.

1 INTRODUCTION

Biogas plants are widely used for the conversion of organic wastes into biogas and digestate that can be utilized as fertilizer [1], [2]. A key unit of a biogas plant is the anaerobic reactor, where methane fermentation occurs. To maintain favorable conditions for methanogenic microorganisms and to mitigate medium non-uniformity, the substrate in the reactor must be mixed. Among available mixing strategies, pneumatic (barbotage) mixing by biogas jets offers a mechanically simple alternative to conventional impeller-based agitation. The biogas generated during anaerobic digestion is supplied to a gas distribution system and used as the driving medium for mixing. This approach enables substrate agitation without a mechanical agitator and can improve process stability by maintaining more homogeneous conditions in the reactor [3]. A major drawback of manual mixing control is its dependence on the operator: both the switching frequency and the duration of gas supply may deviate from the intended schedule, thereby reducing the repeatability of the operating regime [4]. To eliminate the human factor, an automatic mode based on a solenoid valve actuated by a time relay according to a predefined program is considered a practical and cost-effective solution [5]. The objective of this paper is to review the system configuration and the control principles that enable automated barbotage mixing in small-scale biogas plants. In contrast to general descriptions of gas mixing in digesters, this work provides a structured engineering description of intermittent pneumatic mixing as a timer-controlled duty-cycle process. The main contribution is the formalization of duty-cycle parameters (T_on and T_off) and the practical control sequence enabling repeatable cyclic barbotage operation using standard industrial components. This

makes it possible to compare operating modes, identify the engineering benefits of automation, and define practical timing parameters for small-scale biogas plants without using complex programmable control systems [6].

2 BACKGROUND AND ANALYSIS OF MIXING APPROACHES

Mixing in anaerobic reactors is commonly regarded as an essential operation that affects the spatial distribution of temperature, substrate concentration, microbial populations, and overall mass-transfer conditions. Insufficient mixing may lead to stratification, local accumulation of solids, and the formation of zones with different physicochemical conditions, which can negatively influence process stability. Conversely, excessive mixing intensity may increase energy demand and may adversely affect the anaerobic microbial consortia, particularly when the hydrodynamic shear becomes significant. Therefore, the selection of a mixing strategy remains an important engineering task and is widely discussed in the literature, especially in the context of balancing process stability and energy efficiency [2], [3].

Several studies have emphasized that mixing performance should not be interpreted solely in terms of achieving complete homogeneity. In practical anaerobic digestion systems, the purpose of mixing is often to provide sufficient contact between microorganisms and substrate, to improve heat distribution, and to prevent the formation of stagnant zones, while avoiding unnecessary energy consumption. Consequently, researchers consider the choice between continuous and intermittent mixing regimes and propose that intermittent operation can be advantageous, particularly for full-scale plants where mixing accounts for a noticeable portion of total auxiliary power consumption [2], [3].

A number of publications indicate that intermittent mixing can serve as an effective alternative to continuous mixing. By providing periodic “on/off” operation, intermittent regimes may ensure acceptable homogeneity at reduced energy consumption and allow the optimization of mixing duty cycles depending on digester loading and operating conditions [4] - [6]. In addition, the intermittent approach provides a technical basis for automation, since the mixing program can be explicitly expressed through time parameters defining the activation duration and the pause interval. This feature is important for real plants, where operational simplicity and reproducibility of control actions are essential for stable long-term operation.

Recent review papers underline that the mixing strategy is determined not only by hydrodynamics but also by the overall energy balance of the plant, equipment configuration, maintenance requirements, and operating expenditures [2], [3], [7] - [9]. For this reason, an “optimal” mixing strategy is not universal: it is influenced by reactor geometry, substrate rheology, the presence of gas-liquid-solid phases, and the constraints of the available mixing equipment. Accordingly, there is continuous interest in alternative mixing concepts that reduce mechanical complexity and enable reliable control of the mixing regime under industrial conditions.

Barbotage (gas mixing) is commonly defined as the motion of a multitude of gas bubbles through a liquid medium [10]. In anaerobic digestion systems, gas-driven mixing is attractive because it may eliminate the need for mechanical agitators and may utilize part of the produced biogas to generate the mixing flow. From an engineering perspective, this approach shifts the primary challenges from mechanical design to the design of the gas line, distribution elements, and control devices that ensure stable and repeatable biogas delivery to spargers.

Based on the above considerations, gas-driven (barbotage) mixing combined with intermittent operation appears to be a promising approach for anaerobic reactors, as it reduces mechanical complexity while enabling flexible timing control. In this context, the implementation of an automatic control algorithm for biogas supply becomes crucial for achieving repeatable operating regimes and minimizing operator influence. Therefore, the following section describes the configuration of the proposed pneumatic mixing system and its key functional elements, which form the basis for automated cyclic barbotage mixing in a biogas plant.

3 DESCRIPTION OF THE PNEUMATIC MIXING SYSTEM

The pneumatic mixing system utilizes the biogas released during anaerobic digestion as the working medium for substrate agitation. During normal operation, biogas accumulates in the reactor headspace at relatively low pressure. A compressor withdraws the biogas from the headspace and directs it through the gas line. The gas then passes through a

water seal, which serves as a protective element of the pneumatic circuit, and enters a receiver, where it is accumulated under pressure. Subsequently, the biogas is supplied to the collector inlet chamber and distributed through pipelines to the spargers installed in the reactor (Fig. 1). This configuration enables the use of internally generated biogas as an energy carrier for mixing, thereby reducing the need for mechanical agitation devices.

The spargers are equipped with perforated nozzles. During operation, biogas jets pass through the perforations, rise through the substrate volume, and form a gas–liquid flow structure that promotes mixing. This mechanism facilitates redistribution of the substrate and contributes to mitigating stagnant zones within the reactor. A cross-shaped sparger configuration is known to provide more uniform spatial distribution of gas jets over the reactor working area, preventing void formation and local thermal gradients [3] (Fig. 2).

The control arrangement includes two valves integrated into the biogas supply line: a shut-off valve and a solenoid valve. In the manual mode, the solenoid valve remains closed, and the operator periodically opens and closes the shut-off valve, thereby controlling the mixing cycles. In the automatic mode, the shut-off valve is closed, and the solenoid valve is opened and closed according to commands from a time relay in accordance with a predefined program. Such an approach enables barbotage mixing without a mechanical agitator and provides regular cyclic operation with a specified time interval, improving repeatability of the mixing regime and reducing operator dependency.

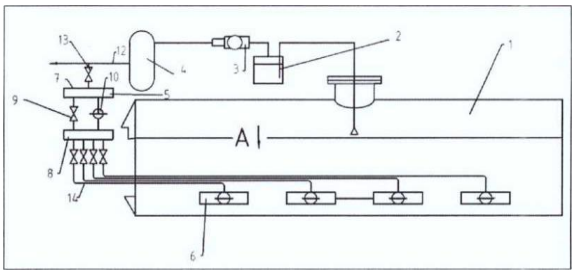


Figure 1: Structural schematic of the biogas plant (BEM-100): 1-reactor; 2-water seal; 3-compressor; 4-receiver; 5-collector; 6-spargers; 7-inlet chamber; 8-distribution chamber; 9-shut-off valve; 10-solenoid valve; 11-pipeline; 12-valve.

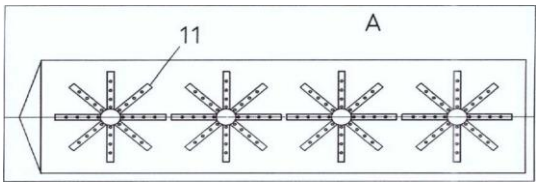


Figure 2: Arrangement of the mixing system inside the biochemical reactor: 11-perforated nozzles.

Figures 1-2 provide the structural overview and the in-reactor configuration of the sparger-based mixing system; the following section describes the timing control implementation and the relay settings used to realize the cyclic barbotage program.

4 CONTROL ALGORITHM AND OPERATING MODES

The biogas supply to the spargers can be controlled in two operating modes: manual and automatic. Such dual-mode implementation is practically important, since manual operation allows maintenance, commissioning, and emergency intervention, whereas automatic operation ensures stable and repeatable long-term mixing regimes.

4.1 Manual Mode

In the manual mode, the solenoid valve remains closed and does not participate in the mixing control. Mixing cycles are created by the operator through periodic opening and closing of a mechanical shut-off valve. This approach requires no additional control logic and may be acceptable for temporary operation and adjustment procedures. However, manual control is inherently sensitive to human factors. Deviations in switching time, inconsistent valve-opening duration, and irregular cycle spacing are typical and lead to low repeatability of the mixing pattern. From the process standpoint, such variability may affect hydrodynamic conditions in the digester, resulting in unstable circulation intensity between cycles and non-uniform disturbance of settled layers. Therefore, manual mode is mainly suitable for short-term operation or as a backup mode.

4.2 Automatic Mode

In the automatic mode, the mechanical shut-off valve is closed and cyclic mixing is achieved by the solenoid valve actuated by a time relay. In this implementation, the control strategy is based on a predefined switching schedule and corresponds to open-loop time-based control. The mixing algorithm can be formally represented as a cyclic sequence of discrete states: waiting → valve opening → gas supply (barbotage) → valve closing → waiting (Fig. 3).

Each cycle initiates gas injection through the spargers, forming a bubble flow that induces circulation of the liquid–sludge mixture and promotes periodic resuspension of settled solids.

The automatic approach provides several advantages relevant to practical digester operation:

1) High repeatability of the duty cycle and stable mixing intensity over time;
2) Reduced dependence on personnel, improving process stability and safety;
3) Consistency of operational documentation, enabling comparable results across operating periods;
4) Simplicity and robustness, since timing control can be implemented using standard industrial components.

For unambiguous specification and practical adjustment of the operating program, two timing parameters are introduced: T_{on} and T_{off}. The parameter T_{on} denotes the solenoid valve open duration, i.e., the time during which biogas is supplied to the spargers and active barbotage occurs. The parameter T_{off} denotes the pause duration between consecutive barbotage events, during which the valve remains closed. Accordingly, the duration of one complete operating cycle, T_{cycle}, is given by (1):

$$T_{cycle} = T_{on} + T_{off}. \quad (1)$$

To characterize the intermittent operating regime more explicitly, the duty-cycle coefficient D can be introduced:

$$D = T_{on} / (T_{on} + T_{off}), \quad (2)$$

where D is the fraction of the cycle during which active gas injection occurs. This parameter can be used as a simple engineering indicator of mixing intensity and auxiliary energy demand. A lower value of D corresponds to shorter active gas supply periods and longer pauses, whereas a higher value of D indicates a more intensive barbotage regime. Thus, the intermittent pneumatic mixing process can be adjusted not only by changing the absolute values of T_{on} and T_{off}, but also by controlling the duty-cycle ratio.

For practical implementation, representative timing programs (T_{on} and T_{off} settings) used for intermittent pneumatic mixing are summarized in Table 1.

The proposed ranges are intended as practical engineering guidelines for commissioning and adjustment. The final settings should be selected based on reactor geometry, sparger configuration, substrate rheology, and the required level of hydrodynamic disturbance.

The values of T_{on} and T_{off} should be selected based on the process requirements and the digester configuration, considering factors such as reactor volume, sparger geometry, substrate properties, and the acceptable level of hydrodynamic disturbance. In practice, periodic rather than continuous gas mixing is often preferred because it reduces energy consumption and avoids excessive shear, while still preventing long-term sedimentation and stratification.

For reliable operation and reproducibility, the selected settings should be recorded in the operating documentation (relay configuration sheets, maintenance logs, or standard operating procedures). This enables consistent operation, simplifies maintenance, and improves transparency of the experimental and industrial results obtained from the reactor.

The wiring of the automation actuators (the time relay and the solenoid valve) is shown in Figure 4. A generalized connection diagram of the time relay and the solenoid valve is presented in Figure 5.

Table 1: Example timing programs for intermittent pneumatic mixing.

Program	T_{on} (gas injection)	T_{off} (pause)	Application note
A (commissioning)	10–30 s	10–30 min	commissioning / checking spargers
B (standard intermittent)	30–90 s	1–3 h	routine operation, low energy consumption
C (enhanced mixing)	60–180 s	30–90 min	increased solids tendency / stratification risk
D (literature-based)	60–120 s	4–6 h	recommended in [6] for similar configurations

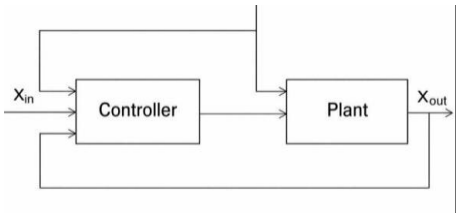
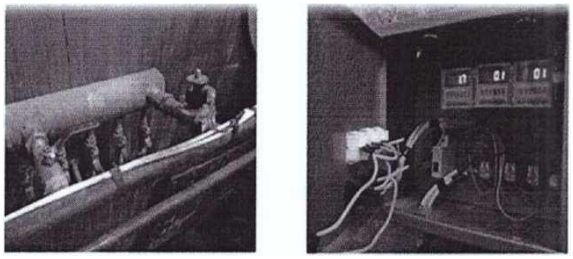


Figure 3: Automatic control scheme.



a) b)

Figure 4: Connection of the solenoid valve and the time relay for substrate mixing in the biochemical reactor: a) general view of the installation; b) relay control panel with wiring.

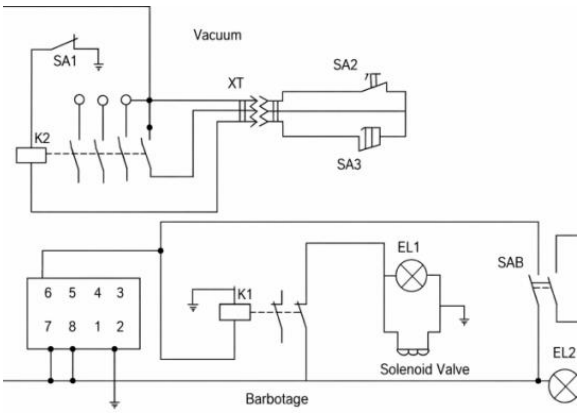


Figure 5: Generalized wiring diagram of the time relay and the solenoid valve.

The technical specifications of the DH-48S-S time relay used for implementing the timing control are summarized in Table 2.

Overall, the specifications in Table 2 confirm that the DH-48S-S relay is suitable for implementing cyclic time-based control of a 220 VAC solenoid valve in the proposed mixing system. In the following subsection, the practical implementation on the BEM-100 setup and the observed operational effects of timer-controlled barbotage are described.

Table 2: Technical specifications of the DH-48S-S time relay.

Specification	Value
Model	DH48S-S
Product type	Repeat cycle time relay
Time setting range	0.01–99.99 s / 0.1–999.9 s / 0.1–99h 59min
Operating mode	Repeat cycle (cyclic on/off)
Setting method	Dial switch
Power supply	220 VAC, 50/60 Hz
Power consumption	≤ 2 W
Contact form	SPDT (1NO + 1NC)
Relay contact rating	5 A @ 250 VAC
Display type	4-digit LED (8-segment)
Number of pins	8
Operating temperature	0…+50 °C
Relative humidity	≤ 95% RH (non-condensing)
Overall dimensions (W×H×D)	48 × 48 × 110 mm
Mechanical life	≥ 1 × 10^6 operations
Weight	≤ 200 g

4.3 Practical Implementation and Observations

The proposed automation scheme was implemented using a DH-48S-S time relay and a standard 220 VAC solenoid valve installed in the biogas supply line. The configuration was tested on the BEM-100 biogas plant setup (see Fig. 1). Commissioning was performed by configuring the relay to generate periodic barbotage cycles defined by the timing parameters T_{on} (valve open duration) and T_{off} (pause duration).

The described automation scheme can be implemented using a DH-48S-S time relay and a standard 220 VAC solenoid valve installed in the biogas supply line, as illustrated by the BEM-100 biogas plant configuration (see Fig. 1). Commissioning is performed by configuring the relay to generate periodic barbotage cycles defined by the timing parameters T_on (valve open duration) and T_off (pause duration). The selection of appropriate timing values depends on reactor volume, sparger geometry, and substrate properties, and should follow the engineering guidelines summarized in Table 1. A structured comparison of the key operational characteristics of manual and automatic mixing modes is presented in Table 3.

Table 3: Comparative characteristics of manual and automatic pneumatic mixing modes.

Parameter	Manual mode	Automatic mode
Control principle	The operator opens and closes the shut-off valve manually	The solenoid valve is actuated by a time relay
Cycle repeatability	Depends on operator actions	Defined by T_on and T_off settings
Operation during unattended periods	Limited or absent	Possible according to a preset switching program
Main limitation	Human-factor variability	No feedback from pressure, gas flow rate, or temperature sensors
Practical effect	Irregular mixing frequency and duration	Stable cyclic barbotage and reduced operator dependency

ompared with manual switching, the automatic mode provided improved repeatability of the mixing duty cycle. Timer-based actuation reduced operator-dependent variability, including inconsistent valve opening time and irregular cycle spacing. Accordingly, cyclic barbotage could be performed according to a predefined schedule, which simplifies routine operation and reduces the need for continuous operator intervention.

The dual-mode (manual/automatic) implementation supports practical maintainability: the manual shut-off valve enables commissioning, adjustment, and emergency operation, whereas the solenoid valve controlled by the time relay provides consistent long-term cyclic actuation. Overall, the implemented configuration demonstrates the feasibility of intermittent pneumatic mixing automation using standard low-cost industrial components without a programmable controller and without the use of feedback sensors.

5 DISCUSSION AND LIMITATIONS

The use of barbotage as a mixing method is consistent with the general concept of gas-driven mixing in anaerobic reactors, where the objective is to maintain sufficient medium homogeneity while minimizing additional energy demand and mechanical complexity. In the proposed system, intermittent operation is implemented by means of a time relay and a solenoid valve, i.e., mixing intensity is indirectly regulated through cyclic gas supply. This approach corresponds to reported conclusions indicating that intermittent mixing regimes can be technically feasible and energetically advantageous in anaerobic digestion applications [4], [5].

It should be noted that the engineering assessment of automation benefits in this review is based on operational and structural arguments rather than direct measurement data. The energy-related advantage of the intermittent mode is discussed in terms of reduced active gas-supply duration and lower duty-cycle operation. Published studies confirm that compressor operating time under intermittent schedules is substantially lower than under continuous agitation, with reported duty-cycle reductions of 60–90% depending on T_{on}/T_{off} combinations [8], [9]. Future work should address direct measurements of compressor energy consumption for varying duty-cycle configurations.

At the same time, the proposed solution is characterized by limitations inherent to time-based actuation. The control program is defined by fixed time intervals and does not employ feedback from process variables (e.g., gas-line pressure, gas flow rate, or reactor temperature). Therefore, variations in operating conditions may require adjustment of the timing parameters to maintain stable and repeatable performance. In addition, because the control scheme is implemented in an open-loop manner, deviations caused by external disturbances cannot be automatically compensated.

From an operational perspective, the reliability of the barbotage regime depends on the integrity of the pneumatic circuit, including gas-line tightness and proper functioning of the compressor and the valves. Furthermore, the long-term stability of mixing performance is influenced by the condition of the perforated nozzles, since clogging or fouling may reduce gas distribution uniformity and alter the effective mixing intensity. Consequently, regular inspection and maintenance of the gas distribution elements is recommended to ensure consistent operation of the pneumatic mixing system.

6 CONCLUSIONS

This paper reviews the engineering principles and practical implementation of automated barbotage

substrate mixing in biogas plants, utilizing biogas generated during anaerobic digestion as the working medium. A pneumatic mixing scheme is described, including biogas withdrawal from the reactor headspace, passage through a water seal, accumulation in a receiver, and supply to a collector with subsequent distribution to spargers equipped with perforated nozzles. The dual-mode (manual/automatic) control architecture is analyzed with respect to repeatability, operational simplicity, and suitability for small-scale installations.

The literature consistently supports that timer-controlled intermittent barbotage provides more repeatable operating conditions compared with manual switching, reduces operator-dependent variability, and lowers compressor duty cycle relative to continuous agitation [9], [5]. The duty-cycle coefficient D, defined as the ratio T_on / (T_on + T_off), is identified as a key engineering indicator linking mixing intensity to energy demand. Practical timing programs (T_on ranging from 10 s to 3 min; T_off from 10 min to 6 h) cover a wide range of operational scenarios from commissioning to routine operation.

The reviewed approach is practically significant for small-scale and farm biogas plants, where complex programmable automation may be economically unjustified. The use of standard industrial components - a time relay and a solenoid valve - makes the system relatively simple, maintainable, and suitable for implementation under field conditions. Future research directions include closed-loop control with process feedback, long-term evaluation of mixing schedules under varying substrate compositions, and integration with digital monitoring systems.

ACKNOWLEDGMENTS

We acknowledge support by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) and the Open Access Publishing Fund of Anhalt University of Applied Sciences.

REFERENCES

[1] P. Weiland, “Biogas production: Current state and perspectives,” Applied Microbiology and Biotechnology, vol. 85, no. 4, pp. 849–860, 2010. [Online]. Available: https://doi.org/10.1007/s00253-009-2246-7.

[2] D. Deublein and A. Steinhauser, Biogas from Waste and Renewable Resources: An Introduction, 2nd ed. Weinheim, Germany: Wiley-VCH, 2011.

[3] B. Singh, Z. Szamosi and Z. Siménfalvi, “Impact of mixing intensity and duration on biogas production in an anaerobic digester: A review,” Critical Reviews in Biotechnology, vol. 40, no. 4, pp. 508–521, 2020. [Online]. Available: https://doi.org/10.1080/07388551.2020.1731413.

[4] I. Gayfullin, M. Murodov, B. Ziganshin and D. Khaliullin, "Results of Research on the Technological Process of Cattle Manure Utilization in a Biogas Plant," Proceedings of International Conference on Applied Innovation in IT, vol. 13, issue 5, pp. 1305–1310, 2025. doi: 10.25673/123173.

[5] J. Lindmark, E. Thorin, R. Bel Fdhila and E. Dahlquist, “Effects of mixing on the result of anaerobic digestion: A review,” Renewable and Sustainable Energy Reviews, vol. 40, pp. 1030–1047, 2014. [Online]. Available: https://doi.org/10.1016/j.rser.2014.07.182.

[6] A. Dzh. Obozov and A. V. Lavrentyev, “Algorithm of software control of the feedstock mixing mode in the methane tank of a biogas plant,” Problems of Automation and Control, no. 2(41), pp. 56-60, 2021.

[7] J. Zhang, L. Loh, D. D. Chen and Y. Dai, “Optimizing mixing strategy to improve the performance of an anaerobic digestion waste-to-energy system for energy recovery from food waste,” Applied Energy, vol. 249, pp. 28–36, 2019. [Online]. Available: https://doi.org/10.1016/j.apenergy.2019.04.142.

[8] U. Bergamo, S. Kujawiak, G. Pecorini, J. Mazurkiewicz and E. C. Rada, “Analysis of anaerobic digester mixing: Comparison of long-shafted paddle mixing vs gas mixing,” Water Science & Technology, vol. 81, no. 7, pp. 1406–1419, 2020. [Online]. Available: https://doi.org/10.2166/wst.2020.248.

[9] S. Dabiri, A. Noorpoor, M. Arfaee, P. Kumar and W. Rauch, “CFD modeling of a stirred anaerobic digestion tank for evaluating energy consumption through mixing,” Water, vol. 13, no. 12, art. 1629, 2021. [Online]. Available: https://doi.org/10.3390/w13121629.

[10] P. Wei, W. Uijttewaal, H. Spanjers, J. B. van Lier and M. de Kreuk, “Optimising flow and mixing in a full-scale gas-mixed anaerobic digester by integrating sludge rheological data using computational fluid dynamics,” Chemical Engineering Journal, vol. 468, art. 143647, 2023. [Online]. Available: https://doi.org/10.1016/j.cej.2023.143647.

Automating Visual Testing in Android Applications

Oleksii Cherkashyn
Blynk Technologies Inc., Brickell Avenue 951, 33131 Miami, FL, USA
oleksii.sciencepapers@gmail.com

Keywords: Android, Android Application Testing, Mobile Test Automation, WebdriverIO, Appium, Pixelmatch.

Abstract: Automated visual testing has become increasingly important in Android application development due to the rapid growth of device fragmentation, diverse screen resolutions, multiple operating system versions, and the widespread adoption of cross-platform frameworks. Ensuring consistent user interface (UI) rendering across different environments is a critical quality attribute, particularly for applications distributed to large and heterogeneous user bases. Traditional functional testing approaches are often insufficient to detect subtle visual regressions that directly affect user experience. Therefore, reliable and scalable visual test automation techniques are essential. This study extends existing research by empirically evaluating pixel-level visual testing across multiple execution environments. The results demonstrate that automated visual testing is feasible and stable in an Android headless emulator environment. Furthermore, the cross-platform nature of mobile applications, including different development frameworks, does not negatively impact the effectiveness of the approach. The findings also confirm that the execution platform-whether a real device, a local emulator, or a cloud-based environment-does not significantly influence test reliability. Finally, the implementation based on the Pixelmatch library shows stable and consistent performance across all experimental conditions. These results highlight the practical applicability and robustness of the proposed visual testing methodology for Android applications.

1 INTRODUCTION

Mobile applications are widely adopted, with millions of users worldwide relying on them for everyday tasks [2]. As of August 2024, approximately 1.7 million Android applications were available on the Google Play Store [1]. With the continuous growth and increasing complexity of mobile applications, ensuring their quality has become a critical challenge. In previous work [3], the importance and growing relevance of mobile application test automation were highlighted.

Mobile applications are inherently GUI-intensive and event-driven, making the Graphical User Interface (GUI) a central component through which most functionality is exposed [4]. Therefore, systematic and reliable testing of GUIs is essential to ensure both functional correctness and a high-quality user experience [5]. However, compared to web applications, mobile applications provide more limited access to UI element attributes and rely on platform-specific view hierarchies, which complicates traditional attribute-based validation approaches. To address these limitations, visual testing has emerged as an important complementary technique, enabling validation of UI elements based on their rendered appearance rather than structural properties. This is particularly relevant for elements such as images, charts, and dynamically rendered text, where correctness depends on visual representation.

This paper presents an approach to automated visual testing of Android mobile applications based on the integration of the WebdriverIO framework [7] and the Appium automation tool [6]. The implementation leverages the Node.js ecosystem, which provides a large set of reusable libraries through the Node Package Manager (npm) [8], enabling flexible and scalable test automation environments. Visual validation is achieved using a pixel-level image comparison strategy implemented with the Pixelmatch library [9] and the Chai assertion framework [10].

Unlike many existing approaches that focus on either attribute-based validation or advanced AI-driven techniques, this study investigates the feasibility and stability of pixel-level visual testing across different execution environments, including

real devices, emulators, and cloud-based platforms. In particular, special attention is given to the use of headless Android emulators, which remain underexplored in the context of visual GUI testing.

To structure the study, the following research questions are defined:

- RQ1: Can pixel-level visual testing be reliably executed on headless Android emulators?
- RQ2: How does the execution environment (real device, local emulator, cloud-based emulator) affect the stability and reproducibility of visual testing results?

The main contributions of this paper are as follows:

- C1: An implementation of a pixel-level visual testing pipeline for Android applications using WebdriverIO, Appium, and Pixelmatch;
- C2: An empirical evaluation of the proposed approach across multiple execution environments, including real devices, local emulators, headless emulators, and cloud-based platforms;
- C3: An analysis of the practical limitations of pixel-based visual testing, including dependency on device-specific baselines and challenges related to dynamic UI content.

2 RELEVANCE

Since only 8% of mobile application development projects currently leverage automated testing practices, the automation of testing represents a timely and significant area of research [12]. Visual test automation as a testing approach in mobile applications is of particular importance because, unlike web applications - where a wide range of test cases can be verified through element attributes, DOM structures, and CSS styles - mobile applications provide limited access to UI element attributes and rely on a more constrained and platform-dependent view hierarchy. As a result, many UI aspects in mobile applications cannot be reliably validated using traditional attribute-based assertions alone. Visual testing becomes essential for verifying UI components whose correctness depends on their visual appearance rather than their structural or semantic properties, such as charts, images, rendered text, graphical indicators, and validation messages. Therefore, automated visual testing serves as a critical complement to functional and attribute-based testing in mobile application quality assurance.

In their scientific work [11], Kraus, Roessler and Sulzmann describe that, based on the principles of golden master testing, they compare the actual GUI state against an expected GUI state. Accuracy is a critical factor in visual testing, as it directly affects the reliability of detected visual differences. Also, according to their scientific work, abstract GUI state (AGS) is:

- Sufficiently expressive to capture the essence of a GUA including visible and non-visible properties;
- Computationally tractable to compare two (expected vs. actual) GUI states and to identify any changes;
- Customizable to let the user decide what changes are important and which are not;
- Platform-independent to achieve a high degree of reusability.

This study is based on the concept of accuracy evaluation, in which the actual rendered image of a UI element is compared with its expected reference image. The comparison is performed with maximum precision, including pixel-by-pixel verification as well as validation of image dimensions such as width, height, and overall size.

The proposed solution is platform-independent and does not rely on a specific Android environment. It operates consistently across different execution contexts, including a local emulator running on a developer's machine, a physical mobile device, or a cloud-based emulator service. Furthermore, when executed on a local emulator, the solution remains independent of the host operating system, whether Microsoft Windows, Linux, or macOS.

This universality ensures broad applicability, reproducibility, and compatibility across heterogeneous development and testing infrastructures.

3 ANALYSIS OF CURRENT RESEARCH

In their academic work [12], Ardito, Coppola, Morisio, and Torchiano highlight that Android applications require distinct interaction patterns from users, and they emphasize that well-designed apps must deliver a high level of usability and work correctly on the broad variety of devices supported by the Android operating system. Because of these unique demands and diversity of platforms, they argue that thorough and systematic testing of the graphical user interfaces (GUIs) is essential to ensure consistent functionality and a positive user

experience. For their experiment, the authors selected two GUI testing tools representing different generations of Android testing. EyeStudio (visual testing) automates GUI interactions using image-based Capture & Replay, evolving from Sikuli and suitable for desktop-emulated apps. Espresso (layout-based testing) interacts with GUI components via properties like IDs and text, using Hamcrest matchers and JUnit for script execution, and is widely adopted in Android development; tests were run on a Nexus 5X emulator.

Yuanchun Li, Ziyue Yang, Yao Guo, and Xiangqun Chen propose Humanoid [13], an automated input generation approach for Android app testing that leverages knowledge learned from real human interaction traces. The system trains a deep learning model offline to predict the most probable user actions in a given GUI state, and then uses these predictions during online testing to guide input generation instead of relying on purely random strategies. By combining the learned interaction model with a UI transition graph, Humanoid prioritizes human-like actions and aims to reach important application states more efficiently and achieve higher test coverage.

Liu, Li, Chen, Wang, Wu, Wang, Hu, and Wang propose VisionDroid [14], a vision-driven automated GUI testing approach that leverages a Multimodal Large Language Model (MLLM) to detect non-crash functional bugs in mobile applications. Their method aligns GUI screenshots with extracted textual and structural information (e.g., widget attributes and app metadata) to enhance semantic understanding of the interface. The approach includes a function-aware explorer that guides deeper and functionality-focused exploration, as well as a logic-aware bug detector that analyzes coherent action sequences to identify inconsistencies between GUI transitions and expected processing logic. Additionally, they employ in-context learning with examples from issue reports and historical test records to further improve bug detection performance.

In their work, Ran, Li, Liu, Wang, Meng, Wu, Jin, Cui, Tang, and Xie propose VTest [15], an industrial-scale visual automated GUI testing framework for mobile applications, primarily focusing on Android and other platforms. In this study, the authors rely entirely on visual information (screenshots) rather than internal API structures, which allows their approach to handle apps with non-standard UI elements. The framework identifies UI elements from images and generates test actions based on the visual context, automatically performing interactions such as taps and swipes.

In their study, Yu, Fang, Tuo, Zhang, Chen, Chen, and Su report that, as of 2024, there are 271 vision-based GUI testing studies [4]. Overall, as the authors note, the majority of the studies focus on testing approaches based on visual analysis.

Mahmood, Mirzaei, and Malek propose EvoDroid [18], an automated Android testing approach that uses evolutionary search-based algorithms to generate UI event sequences in order to systematically explore application behavior and maximize code coverage. Their method combines Android-specific program segmentation with evolutionary test generation to efficiently guide interaction sequences through different app components and GUI states. However, the approach is not focused on visual testing or UI appearance comparison, as it does not perform screenshot-based or pixel-level visual regression analysis.

Yuanhong Lan, Yifei Lu, Zhong Li, Minxue Pan, Wenhua Yang, Tian Zhang, and Xuandong Li propose DQT (Deep Q-network Testing) [19], an approach that applies deep reinforcement learning to automatically explore Android application interfaces and generate test actions. The method encodes GUI states using graph-based representations that preserve structural and semantic UI information, improving state understanding compared to traditional tabular RL approaches. By learning from runtime interactions, the system prioritizes more informative exploration paths and increases test coverage while improving fault detection in complex Android apps.

Tianxiao Gu, Chengnian Sun, Xiaoxing Ma, Chun Cao, Chang Xu, Yuan Yao, Qirun Zhang, Jian Lu, and Zhendong Su propose a model-based automated GUI testing approach (APE) [20] for Android applications that dynamically constructs and refines a GUI state model during execution. Their method continuously updates the abstraction of the application's interface using runtime exploration feedback, which allows more precise modeling compared to static GUI models. By leveraging this adaptive model refinement, the approach systematically generates test events that improve exploration coverage and increase the detection of crashes in real-world Android applications.

Overall, it can be concluded that the majority of existing studies focus on the visual verification of UI elements. This paper contributes to the advancement of visual testing by further exploring the feasibility of automated visual testing using a headless Android emulator [3], a testing approach that was previously described. In addition, this study contributes to the field of visual testing through the application of the Pixelmatch library.

4 HARDWARE AND SOFTWARE CONFIGURATION

4.1 Test Infrastructure Configuration

The experimental setup for Android application automation was based on Node.js version 20.19.4. The implementation employed the WebdriverIO framework with TypeScript integration, along with Appium and the Appium WDIO service. Additionally, Pixelmatch and several supporting libraries were incorporated to facilitate image comparison and enhance the overall testing infrastructure. The complete list of dependencies and libraries is provided in the package.json file, included in Appendix (Listing A1).

At the initial stage of the study, a detailed examination of the application's UI elements was conducted. For this purpose, the Appium Inspector tool was utilized [16], as it facilitates the identification, inspection, and visual analysis of interface elements, including their attributes and hierarchical structure within the application.

Figure 1 illustrates an example of how a UI element appears in Appium Inspector. The top part shows the element in its unselected state, while the bottom part shows it selected. In this specific case, it is a single element, and the screenshot preview of the element is clearly visible for visual inspection. In Appium Inspector, it is also possible to verify the uniqueness of the selected locator. The specific method used to define the locator - such as accessibility ID, XPath, or UIAutomator - is not critical; the key requirement is that the locator must be unique. However, for stability and execution speed, the use of an accessibility ID is recommended.

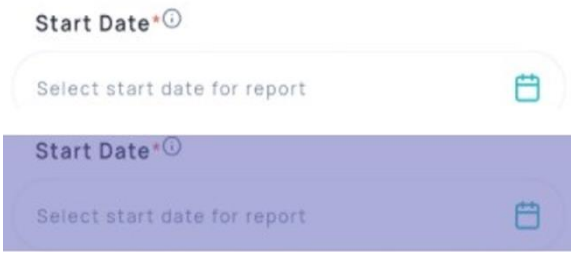


Figure 1: Visual representation of a selected and unselected element in Appium Inspector.

The next stage involves preparing the expected images of the elements. This can be accomplished using the command shown below, or alternatively, by utilizing the previously selected locator.

TypeScript code:

```
//The example uses an accessibility
ID locator.
const checkEl = await $(‘~element’);
await
checkEl.saveScreenshot('./images/ele
mScreenshot.png');
```

4.2 Preparation of the Image Comparison Library

In Appendix (Listing A2), an example of an exported function for image comparison is provided. The function utilizes the Pixelmatch library for pixel-level image comparison, pngjs for parsing PNG images, fs for file system operations, and chai for assertions. It performs a precise visual comparison between two images to verify that a mobile application's UI matches the expected design. The images are first read from files and parsed into pixel data, which includes width, height, and RGBA color values. The function checks that the dimensions of both images are identical, and if they differ, the test fails immediately. Pixel-by-pixel comparison is then performed using pixelmatch, with a small threshold to ignore minor rendering variations. A difference image is generated to visualize any discrepancies. Finally, the function asserts that no pixels differ between the actual and expected images. This approach allows for automated, exact verification of mobile UI rendering across different platforms and devices, ensuring consistency with the intended design.

Compared to a baseline usage of pixelmatch, which typically performs only raw pixel-level comparison and returns a numeric difference metric, the proposed implementation extends the core algorithm with additional layers of validation, including pre-comparison dimension checks, structured assertion handling, automatic generation of diagnostic artifacts, and full pipeline encapsulation. These enhancements improve robustness, interpretability, and usability of pixel-based visual regression testing in automated environments, particularly within continuous integration systems.

5 RESULTS AND DISCUSSIONS

This study is based on a real Samsung Galaxy A34 5G device with a screen resolution of 2340 × 1080, as well as an Android emulator configured as a Pixel 6a

device with a resolution of 2400 × 1080 and the x86_64 architecture.

The study was conducted on eight different cross-platform mobile applications, including Flutter-based, Kotlin-based, and React Native applications. A total of 179 UI element screenshots were captured and verified using automated visual testing. The dataset includes 28 images and photographs, 22 input fields, 18 dropdown menus, 25 buttons, 16 page titles, 24 text elements, and 12 icons. Additionally, 34 graphs of varying complexity and structural types were included to evaluate visual consistency across data-driven and graphical components. All tests demonstrated stable and consistent results across executions. Additionally, minor artificial modifications were introduced into the images to evaluate the sensitivity of the comparison approach. The tests successfully detected even the smallest differences, including discrepancies at the level of a single pixel (1 px), confirming the high precision of the proposed visual verification method.

The study was also conducted on the same emulator operating in headless mode using the same set of applications. The tests similarly demonstrated stable and consistent results. This experiment is particularly significant, as visual test automation studies have not previously been conducted in this mode. The findings confirm that the proposed approach remains reliable even in a headless environment, which is especially relevant for CI/CD pipelines and large-scale automated testing infrastructures. In addition, automated visual tests executed in a headless mode emulator run on average 6 - 7% faster in each test cycle compared to a UI-based emulator, which is consistent with previously conducted research findings [3].

The tests were also executed on cloud-based emulators provided by BrowserStack, as well as on locally configured Android emulators. In both environments, the automated visual tests demonstrated stable and consistent results. These findings further confirm the robustness and reproducibility of the proposed visual testing approach across different infrastructure setups, including cloud-based and local execution environments.

Each test case was executed five times on each mobile device to ensure the reliability and reproducibility of the results. The outcomes demonstrated stable and consistent performance across all environments, regardless of the Android platform, operating system version, or the cross-platform framework used in the applications. These findings confirm the robustness and platform-independence of the proposed visual testing approach.

The results also revealed several important limitations. First, the preparation of baseline (expected) screenshots is a time-consuming process. In addition, identical reference screenshots must be generated separately for each device configuration depending on screen resolution. This implies that a prepared visual test is tightly coupled to a specific device setup, regardless of whether it is executed on a real device or an emulator.

Furthermore, this approach is not suitable for automating UI elements with dynamically changing content, such as animations, video streams, or charts with continuously updating values. Since pixel-level comparison requires a static visual state, any inherent variability in the rendered content leads to test failures, even if the functionality is correct.

The evaluation of pixelmatch demonstrates several measurable advantages compared to visual regression approaches integrated in WebdriverIO via wdio-image-comparison-service and alternative JavaScript libraries including Resemble.js [17], looks-same, blink-diff, and jest-image-snapshot. In terms of execution efficiency, pixelmatch shows reduced computational overhead due to its lightweight architecture and direct manipulation of pixel buffers without framework-level abstractions, resulting in faster execution in batch image comparison scenarios compared to WebdriverIO-based and snapshot-based solutions. Table 1 presents a comparative analysis of the selected libraries. The results demonstrate that pixelmatch is the most lightweight solution among the evaluated tools. The latest version comprises 37 files across 5 directories, with a total size of 732 KB, indicating a minimal dependency footprint and potentially higher execution efficiency. Regarding determinism, pixelmatch produces fully reproducible outputs based on strict pixel-by-pixel comparison, whereas libraries such as Resemble.js and looks-same may introduce variability due to heuristic similarity metrics and configurable tolerance thresholds. In perceptual accuracy, pixelmatch leverages the YIQ color space, which improves alignment with human visual perception and reduces sensitivity to minor color deviations, in contrast to RGB-based approaches used in blink-diff that are more prone to false-positive detections. With respect to robustness, pixelmatch demonstrates improved handling of anti-aliasing artifacts, whereas looks-same and blink-diff exhibit increased noise in UI contexts involving font rendering and graphical smoothing. In terms of integration flexibility, pixelmatch operates as an

independent module without dependency on specific testing frameworks, unlike WebdriverIO visual regression services and jest-image-snapshot which are tightly coupled to their respective ecosystems. Overall, the results indicate that pixelmatch provides a combination of high performance, deterministic behavior, perceptually informed comparison, robustness to rendering artifacts, and framework independence, which makes it particularly suitable for controlled experimental setups and research-oriented visual regression analysis.

Table 1: Comparison of visual regression testing libraries.

Library	Files	Folders	Size
jest-image-snapshot	1,276	29	3.11 MB
looks-same	92	22	420 KB
blink-diff	1,108	188	9.25 MB
resemblejs	681	126	74.6 MB
pixelmatch	37	5	732 KB

6 CONCLUSIONS

Although numerous studies have been devoted to the automation of visual testing for mobile applications, the present research extends existing practices and provides additional empirical evidence. The key findings of this study are as follows:

- Automated visual testing is feasible and stable when executed on an Android headless emulator;
- The cross-platform nature of mobile applications does not affect the effectiveness of visual testing;
- The Android execution environment - whether a real device, a local emulator, or a cloud-based device - does not significantly influence the reliability of visual testing results;
- Automated mobile application testing implemented using the Pixelmatch library demonstrates stable and consistent results across different testing conditions.

REFERENCES

[1] A. Bilal, H. T. Mirza, I. Hussain, and A. Ahmad, “Investigating Influence of Google Play Application Titles on Success,” Big Data Research, vol. 36, 2024, [Online]. Available: https://doi.org/10.1016/j.bdr.2024.100443.

[2] A. Niroshan, S. Seneviratne, and A. Seneviratne, “An Empirical Study of Code Obfuscation Practices in the Google Play Store,” arXiv preprint, arXiv:2502.04636, 2025.

[3] O. Cherkashyn, “Application Test Automation in Headless Android Emulator,” Proceedings of International Conference on Applied Innovation in IT, vol. 13, no. 5, pp. 445-452, 2025, [Online]. Available: https://doi.org/10.25673/123064.

[4] Y. Wu, J. Ling, X. Luo, T. Yang, M. Zhao, C. He, and Y. Wu, “Vision-Based Mobile App GUI Testing: A Survey,” ACM Computing Surveys, vol. 58, no. 6, pp. 1-46, Oct. 2025, [Online]. Available: https://doi.org/10.1145/3773027.

[5] R. Coppola, L. Ardito, and M. Torchiano, “Fragility of layout-based and visual GUI test scripts: An assessment study on a hybrid mobile application,” in Proceedings of the 10th ACM SIGSOFT International Workshop on Automating TEST Case Design, Selection, and Evaluation (A-TEST ’19), New York, NY, USA, Aug. 2019, pp. 28-34, [Online]. Available: https://doi.org/10.1145/3340433.3342824.

[6] S. Adiatma and A. Darmayantie, “Implementation and Comparative Analysis of Test Automation Framework Performance for Functional Testing of Web-Based Applications using the Distance to the Ideal Alternative (DIA) Method,” Widya Teknik, vol. 22, no. 1, 2023, [Online]. Available: https://doi.org/10.33508/wt.v22i1.5027.

[7] S. Godboley, D. Dalei, R. Sadam, and D. P. Mohapatra, “Agile GUI Testing by computing novel Mobile App Coverage Using Appium Tool,” in Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing, pp. 1026-1029, 2023, [Online]. Available: https://doi.org/10.1145/3555776.3577806.

[8] H. Sun, A. Rosà, D. Bonetta, and W. Binder, “Automatically Assessing and Extending Code Coverage for NPM Packages,” in 2021 IEEE/ACM International Conference on Automation of Software Test (AST 2021), pp. 40-49, 2021, [Online]. Available: https://doi.org/10.1109/AST52587.2021.00013.

[9] J. Yang, C. E. Jimenez, A. L. Zhang, K. Lieret, J. Yang, X. Wu, O. Press, N. Muennighoff, G. Synnaeve, K. R. Narasimhan, D. Yang, S. I. Wang, and O. Press, “SWE-bench Multimodal: Do AI Systems Generalize to Visual Software Domains?,” in The Thirteenth International Conference on Learning Representations (ICLR), 2025, [Online]. Availabe: https://doi.org/10.48550/arXiv.2410.03859.

[10] L. Zamprogno, B. Hall, R. Holmes, and J. M. Atlee, “Dynamic Human-in-the-Loop Assertion Generation,” IEEE Transactions on Software Engineering, vol. 49, no. 4, pp. 2337-2351, Apr. 2023, [Online]. Available: https://doi.org/10.1109/TSE.2022.3194038.

[11] L. Ardito, R. Coppola, M. Morisio, and M. Torchiano, “Espresso vs. EyeAutomate: An Experiment for the Comparison of Two Generations of Android GUI Testing,” in Proceedings of the 23rd International Conference on Evaluation and Assessment in Software Engineering (EASE ’19), pp. 13-22, 2019, [Online]. Available: https://doi.org/10.1145/3319008.3319022.

[12] D. Kraus, J. Roessler, and M. Sulzmann, “Visual Testing of GUIs by Abstraction,” arXiv preprint, arXiv:2007.10419, 2020, [Online]. Available: https://arxiv.org/abs/2007.10419.

[13] Y. Li, Z. Yang, Y. Guo, and X. Chen, “Humanoid: A deep learning-based approach to automated black-box Android app testing,” in Proceedings of the 34th IEEE/ACM International Conference on Automated Software Engineering (ASE ’19), pp. 1070-1073, 2019, [Online]. Available: https://doi.org/10.1109/ASE.2019.00104.

[14] Z. Liu, C. Li, C. Chen, J. Wang, B. Wu, Y. Wang, J. Hu, and Q. Wang, “VisionDroid: Vision driven Automated Mobile GUI Testing via Multimodal Large Language Model,” arXiv preprint, arXiv:2407.03037v1, 2024, [Online]. Available: https://arxiv.org/abs/2407.03037v1.

[15] D. Ran, Z. Li, C. Liu, W. Wang, W. Meng, X. Wu, H. Jin, J. Cui, X. Tang, and T. Xie, “Automated Visual Testing for Mobile Apps in an Industrial Setting,” in Proceedings of the 44th International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP ’22), pp. 55-64, 2022, [Online]. Available: https://doi.org/10.1145/3510457.3513027.

[16] R. Mahmood, N. Mirzaei, and S. Malek, “EvoDroid: Segmented Evolutionary Testing of Android Apps,” in Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering (FSE ’14), pp. 599-609, 2014, [Online]. Available: https://doi.org/10.1145/2635868.2635896.

[17] Y. Lan, Y. Lu, Z. Li, M. Pan, W. Yang, T. Zhang, and X. Li, “Deeply Reinforcing Android GUI Testing with Deep Reinforcement Learning,” in Proceedings of the 46th IEEE/ACM International Conference on Software Engineering (ICSE ’24), art. no. 71, pp. 1-13, Feb. 2024, [Online]. Available: https://doi.org/10.1145/3597503.3623344.

[18] T. Gu, C. Sun, X. Ma, C. Cao, C. Xu, Y. Yao, Q. Zhang, J. Lu, and Z. Su, “Practical GUI Testing of Android Applications via Model Abstraction and Refinement,” in Proceedings of the 41st International Conference on Software Engineering (ICSE ’19), pp. 269-280, 2019, [Online]. Available: https://doi.org/10.1109/ICSE.2019.00042.

[19] D. S. Dupakuntla Naga, “Cross-Platform Mobile Testing with Appium: A Framework for High-Accuracy Validation in Healthcare,” International Research Journal of Engineering and Technology (IRJET), vol. 12, no. 10, pp. 450-462, Oct. 2025, [Online].

[20] A. J. Swart and P. E. Hertzog, “Quantifying the Percentage of Shading on a PV Module and Its Subsequent Impact on Its Output Power,” European Chemical Bulletin, vol. 12, special issue 3, pp. 6627-6635, Jul. 2023, [Online]. Available: https://doi.org/10.31838/ecb/2023.12.s3.747.

APPENDIX

The Appendix contains the scripts and sections of programming languages code used in the study.

```
{
  "name": "app_name",
  "type": "module",
  "devDependencies": {
    "@types/chai": "^5.0.1",
    "@types/jimp": "^0.2.28",
    "@types/moment": "^2.13.0",
    "@types/pixelmatch": "^5.2.6",
    "@types/pngjs": "^6.0.5",
    "@wdio/appium-service": "^8.39.1",
    "@wdio/cli": "^8.39.1",
    "@wdio/local-runner": "^8.39.1",
    "@wdio/mocha-framework": "^8.39.0",
    "@wdio/spec-reporter": "^8.39.0",
    "appium": "^2.11.2",
    "chai": "^5.1.0",
    "pixelmatch": "^5.3.0",
    "pngjs": "^7.0.0",
    "ts-node": "^10.9.2",
    "typescript": "^5.5.3",
    "wdio-video-reporter": "^6.1.1"
    "appium-uiautomator2-driver":
"^3.7.2"
  },
  "scripts": {
    "test": "wdio run ./wdio.conf.ts"
  },
}
```

Listing A1: Package.json.

```
 import pixelmatch from 'pixelmatch'
  import { PNG } from 'pngjs'
  import fs from 'fs'
  import { expect } from 'chai';

  export async function
  compareImages(imagePath1: any,
  imagePath2: any, expectImage: any) {
      //Read images from files
      const img1 =
  PNG.sync.read(fs.readFileSync(imageP
  ath1));
      const img2 =
  PNG.sync.read(fs.readFileSync(imageP
  ath2));

      //Checking image sizes
      expect(img1.width, `Images width
  size is different. Expexted image:
  ${expectImage}`).to.equal(img2.width
  );
      expect(img1.height, `Images
  height size is different. Expexted
  image:
  ${expectImage}`).to.equal(img2.heigh
  t);

      //Creating an object to store
  the comparison result
```

```
    const diff = new PNG({ width:
img1.width, height: img1.height });

    //Image comparison
    const numDiffPixels =
pixelmatch(img1.data, img2.data,
diff.data, img1.width, img1.height,
{ threshold: 0.1 });

    diff.pack().pipe(fs.createWriteS
tream('./images/diff/diff.png'));

    // Checking the number of
different pixels
    expect(numDiffPixels, `Images
are different. Expexted image:
${expectImage}`).to.equal(0);
}
```

Listing A2: TypeScript code.

Automation of Water Treatment Facilities with Adaptive KPI Integration within a Sustainable Development Strategy

Nataliia Zaiets[1], Nataliia Lutska[2], Lidiia Vlasenko[3], Volodymyr Shtepa[4] and Stefan Junge[5]
[1]*Department of Electrical Engineering, Mechatronics and Optometry, Berlin University of Applied Sciences, Luxemburger Str. 10, 13353 Berlin, Germany*
[2]*Department of Automation and Computer Technologies of Control Systems, National University of Food Technologies, Volodymyrska Str. 68, 01033 Kyiv, UkraineV*
[3]*Department of Software Engineering and Cyber Security, State University of Trade and Economics, Kyoto Str. 19, 02156 Kyiv, Ukraine*
[4]*International Innovation and Applied Center "Aquatic Arteria", Kharkivska St. 116, 40007 Sumy, Ukraine*
[5]*Department of Packaging Technology, Berlin University of Applied Sciences, Luxemburger Str. 10, 13353 Berlin, Germany*
nataliia.zaiets@bht-berlin.de, lutskanm2017@gmail.com, vlasenko.lidia1@gmail.com, aquartery@ecolog.sumdu.edu.ua, stefan.junge@bht-berlin.de

Keywords: Automated Process Control System, Key Performance Indicators, Sustainable Development, Water Treatment Automation, Digital Maturity, Dynamic Control.

Abstract: This paper is devoted to the development of engineering methods for integrating a Sustainable Development strategy into automated process control systems of industrial enterprises. The relevance of the study is driven by the need to shift from conventional variables stabilization systems to adaptive control models capable of dynamically responding to environmental and energy-related challenges. An original approach is proposed in which Key Performance Indicators (KPIs) are transformed from reporting tools into dynamic setpoints for local control loops. The authors develop a dynamic model for KPI selection and adjustment based on an integral sustainability compliance index, enabling real-time balancing between environmental requirements – such as wastewater treatment quality indicators – and economic efficiency. The adaptive behavior of the model is achieved through time-dependent weighting coefficients, which are automatically adjusted according to the proximity of monitored process variables to regulatory discharge limits, thereby prioritizing environmental safety under critical operating conditions. Special attention is given to an object-oriented approach that ensures the adaptation of strategic sustainability goals to specific technological units of wastewater treatment facilities. The paper also analyzes the impact of an enterprise's digital maturity level on the reliability of KPI calculation and proposes the use of soft sensors for forecasting complex environmental variables under conditions of limited instrumentation. The effectiveness of the proposed solutions is validated through simulation modeling of wastewater treatment processes at a dairy processing plant. The results demonstrate that the implementation of the dynamic control model reduces specific energy consumption by 28.1% while fully complying with established environmental regulations. The practical significance of the study lies in the creation of an algorithmic foundation for modernizing industrial process automation systems, enabling the integration of technological reliability with the requirements of the global green transition.

1 INTRODUCTION

The current stage of industrial automation development is characterized by a transition from conventional systems focused on stabilizing technological variables to intelligent control systems based on global Sustainable Development strategies [1]. Under conditions of the ongoing energy crisis and the tightening of environmental regulations – particularly ESG requirements – industrial facilities such as wastewater treatment systems require not only the maintenance of predefined operating regimes but also real-time dynamic process optimization [2].

KPIs serve as a fundamental tool for assessing the efficiency of such systems. However, from the perspective of automation engineering, KPIs can no

longer be regarded solely as reporting metrics. Instead, they must be transformed into state variables and dynamic setpoints that directly influence the control logic of Programmable Logic Controllers (PLCs) and Supervisory Control and Data Acquisition (SCADA) systems.

Traditional automated process control systems (APCS) are primarily focused on technical reliability and compliance with maximum permissible concentrations (MPC) of pollutants. However, they often lack the flexibility required for instantaneous adaptation to changes in an enterprise's strategic priorities, such as switching to a maximum energy-saving mode under low-load conditions or prioritizing treatment quality during shock discharges. The absence of mechanisms enabling a dynamic linkage between sustainable development objectives and automatic control algorithms creates a gap between business-level planning and the actual technological process.

Industrial wastewater treatment in the food industry is a complex, nonlinear, and high-inertia process with variable influent loads. Key stages – mechanical sedimentation, reagent treatment, and biological degradation – require precise coordination of blowers, pumps, and dosing systems. Maintaining the balance between aeration and pollutant concentration is critical. Consequently, modern facilities demand predictive control, where automation decisions directly align with resource efficiency and environmental safety targets.

The aim of this study is to develop an engineering approach to the design of automation systems that employ a dynamic model for KPI selection and adjustment as the primary factor in shaping control actions. This approach enables the implementation of a "green" development trajectory directly at the level of actuating mechanisms. The object of the research comprises automated control processes of water treatment facilities (with dairy processing plants used as a case study) operating under conditions of variable influent flows and environmental constraints. The subject of the research includes methods and algorithms for integrating dynamic KPIs into automatic control loops in order to realize a sustainable development strategy.

2 LITERATURE REVIEW

Sustainable automation involves the implementation of automated solutions aimed at minimizing environmental impact, ensuring efficient resource utilization, and maintaining long-term ecological stability. In the field of water treatment, this approach is realized through the integration of environmental and energy-related KPIs [3] – in particular, treatment quality indicators and energy intensity metrics [4] – with technological process variables [5]. Given the complexity and nonlinearity of wastewater treatment processes [6], control systems must be based on dynamic models capable of adapting to changing operating conditions and fluctuating load regimes.

Modern computational and intelligent models in water treatment enable KPI monitoring [7], KPI forecasting [8],[9], and effective fault detection both related and unrelated to key performance indicators, thereby transforming raw process data into strategic decision-making tools for improving energy efficiency and process stability. Recent studies focus on control system architectures for wastewater treatment plants that emphasize multi-objective optimization of treatment quality and operational costs. In particular, works [10]–[12] propose multilevel and multi-objective architectures combining benchmarking models, metaheuristic optimization algorithms, and intelligent predictive models to adapt to variable loading conditions. A separate research direction is represented by economically oriented control architectures [13], which integrate predictive control with economic performance criteria to minimize energy consumption while ensuring compliance with regulatory wastewater quality standards.

In wastewater treatment facilities, actual KPIs are typically calculated based on direct measurements, while accurate estimation of complex indicators relies on soft-sensor models [14], [15]. Adaptive soft-sensor architectures enable the simultaneous prediction of multiple hard-to-measure variables, assessment of their reliability, and enhancement of the digital maturity of wastewater treatment control systems. A general review indicates that the evolution of soft sensors from mechanistic models to modern machine learning–based approaches reduces dependence on physical instrumentation and improves the reliability and efficiency of online monitoring [16].

Simulation scenarios demonstrate stable operation under nominal conditions and energy consumption optimization when dynamic KPIs are employed [13], [17]. The application of KPIs as control variables confirms their effectiveness in supporting sustainable development objectives, while the integration of digital twins and soft sensors enhances system adaptability. At the same time, international case studies report significant economic

and environmental benefits associated with such approaches [17], [18].

3 METHODS

To address the problem of efficient control of water treatment facilities within a sustainable development strategy, an approach based on the decomposition of a complex technological process into individual software objects is proposed. The complexity of such facilities lies in their high degree of heterogeneity; therefore, an object-oriented approach enables the adaptation of a global KPI strategy to the specific characteristics of each technological unit [19].

Within the proposed architecture, each unit of the wastewater treatment plant is considered an autonomous software object with its own set of local KPIs, which are integrated into a unified enterprise-level hierarchy:

1) Primary sedimentation unit: this object is oriented toward the KPIs Throughput capacity and Suspended solids removal efficiency, where the control algorithm adapts the influent flow rate according to the current hydraulic load;
2) An eration tank unit: the key performance metrics include Nitrification level and Aeration energy intensity; at this stage, the control algorithm operates with the most dynamic setpoints;
3) Sludge dewatering unit: control is focused on the KPI Minimum flocculant consumption while achieving the specified sludge cake moisture content.

Such decomposition enables the development of flexible control systems: when a single unit is modernized (for example, through the installation of a more accurate ammonium sensor), only the corresponding local object-oriented module requires adjustment, while the overall structure of strategic control remains unchanged. The engineering advantage of this approach lies in the simplification of PLC programming, allowing validated code blocks to be replicated across different process units while preserving a unified sustainable development trajectory for the entire facility.

3.1 KPIs as Dynamic Setpoints in Automated Control Loops

In classical automation systems, the setpoints of local controllers are static or change according to a rigid predefined schedule. However, a sustainable development strategy requires flexibility: the system must be capable of autonomously determining target variable values based on current priorities (environmental performance versus energy efficiency). In this context, KPIs function as dynamic external reference signals for automated process control systems.

To transition from the conceptual foundations of sustainable development to practical automation algorithms, it is necessary to systematize the measurable variables of the technological process. Table 1 presents an extended classification of KPIs, in which each indicator is treated not as a static reporting metric but as an active variable within the control loop of wastewater treatment automation systems.

Table 1: Classification of KPIs as control variables for water treatment automation systems.

Category	KPI (Source, Unit)	Variable Type in APCS	Role in Dynamic Model
1. Technical (Operational)	Flow rate (Flowmeter, m^3/h)	Input variable	System load definition
	System pressure (Pressure sensor, Pa)	Controll variable	Hydraulic regime maintenance
	Tank level (Level sensor, m)	Constraint	Emergency prevention
2. Environmental (Sustainable)	Chemical or biochemical oxygen demand – COD, BOD (Analyzer or Soft sensor, mg/dm^3)	Target setpoint	Effluent quality indicator
	Dissolved oxygen concentration (Oximeter, mg/L)	Intermedi-ate setpoint	Control of aeration intensity
	pH value (pH meter)	Hard constraint	Reagent dosage correction
3. Energy (Economic)	Specific energy consumption (Power-analyzer kWh/m^3)	Objective function	Cost minimization
	Reagent consumption (Dosing pump L/m^3)	Cost variable	Operational optimization

The analysis of Table 1 indicates that the effectiveness of implementing a Sustainable Development strategy directly depends on the ability of automated process control systems to integrate heterogeneous metrics into a unified decision-making framework. In contrast to traditional control approaches, where priority is given exclusively to technical variables, the proposed approach is based on multi-criteria optimization.

The central element of the proposed engineering solution is the calculation of an integral Sustainable Development Compliance Index (I_{sd}). This index enables a quantitative assessment of deviations between the current operating mode of the facility and an ideal state defined by the enterprise's strategic objectives. The mathematical model of the index is expressed as follows:

$$I_{sd}(t) = \sum_{i=1}^{n} w_i(t) \frac{KPI_i^{fact}(t) - KPI_i^{target}(t)}{KPI_i^{max}(t) - KPI_i^{min}(t)}, \quad (1)$$

where:

1) $KPI_i^{fact}(t)$ – denotes the actual value of the i-th indicator obtained from instrumentation and control systems in real time;
2) $KPI_i^{target}(t)$ represents the target (reference) value of the indicator in accordance with the sustainable development strategy;
3) $KPI_i^{max}(t)$, $KPI_i^{min}(t)$ correspond to the maximum and minimum allowable values of the indicator, respectively;
4) $w_i(t)$ are dynamic weighting coefficients that determine the priority of a particular KPI at a given time instant.

The dynamic nature of the model is ensured by the adaptive adjustment of the weighting coefficient vectors $w_i(t)$. For example, when sensors detect a sharp increase in pollutant concentration at the influent (a change in Category 1 indicators), the algorithm automatically increases the weighting coefficients of environmental KPIs (Category 2). This results in an immediate recalculation of setpoints for actuating mechanisms, whereby the priority of effluent water quality exceeds that of energy efficiency indicators (Category 3).

Thus, the proposed model transforms the concept of sustainable development from a declarative framework into a level of direct algorithmic control of technological equipment. This enables the automation system not only to maintain process stability but also to function as a tool for dynamic balancing between economic feasibility and environmental responsibility of an industrial enterprise.

3.2 Development of a KPI-Based Control Architecture

For the practical implementation of a sustainable development strategy, a two-level control architecture is proposed. At the lower level, conventional PID controllers are employed to stabilize process variables, while the upper level is represented by a KPI-oriented supervisory controller that acts as a coordinator of objectives.

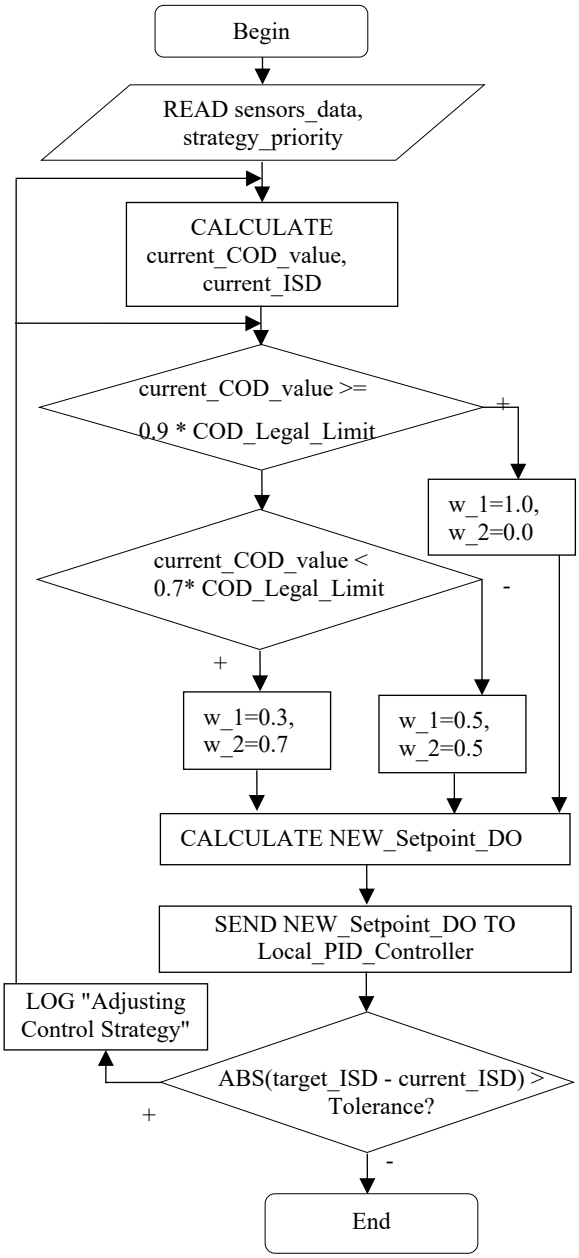


Figure 1: Algorithm of dynamic KPI control.

The key novelty of the approach lies in transforming strategic KPIs into dynamic reference signals for local control loops. This mechanism is illustrated using the example of the biological treatment process in an aeration tank (Figure 1). The proposed dynamic control algorithm is based on a continuous Data-to-Action processing loop implemented at the supervisory level of the automated process control system. The operational logic of the Dynamic KPI Control module consists of the sequential execution of four stages.

First, environmental performance indicators are evaluated using a mathematical soft-sensor model, which enables real-time estimation of COD without the delays inherent to laboratory-based analyses. Second, the algorithm performs adaptive adjustment of the weighting coefficients w_1 and w_2, which define the priority balance between environmental safety and energy conservation depending on the proximity of current process variables to regulatory discharge limits. At the third stage, the integral I_{sd} is calculated and used as the criterion for the optimization model. The final step involves the generation of a dynamic setpoint for the local PID controller of the aeration system, thereby ensuring the immediate translation of the enterprise's strategic objectives into physical operating variables of the equipment.

This approach enables the closure of the control loop based on KPIs, providing system flexibility under conditions of unstable influent loading.

The proposed hierarchical automation architecture (Figure 2) implements vertical integration of sustainable development strategic objectives into control loops through three interconnected levels. At the Field Level, primary data acquisition and physical execution of control commands are performed using the sensor infrastructure and actuating mechanisms, such as variable frequency drives.

The Control Level serves as an intelligent intermediate layer, where indirect estimations of environmental variables are generated using soft sensors and transmitted to the upper level, while PID control is implemented based on the received setpoints.

The highest level, the Management/Cloud Level, provides supervisory control, where the Dynamic KPI Control module dynamically recalculates controller setpoints based on monitoring of the sustainable development strategy, balancing environmental safety and energy efficiency objectives.

This closed-loop structure ensures continuous adaptation of the technological process to enterprise-level global KPIs, transforming the automation system into an active instrument for implementing environmental strategy.

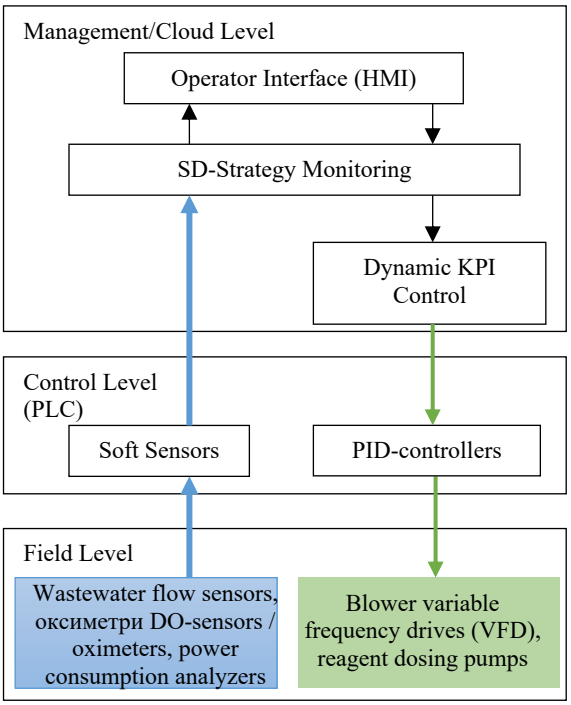


Figure 2: Hierarchical KPI-based control architecture for water.

3.3 Consideration of Digital Maturity of APCS

The effectiveness of the dynamic KPI model is directly constrained by the level of digital maturity of the technological facility. For automation engineers, this primarily refers to the availability, accuracy, and sampling resolution of data acquired from field devices.

Not all sustainable development indicators can be obtained directly from physical sensors in real time. Depending on the instrumentation level of the facility, KPIs can be classified into three levels of availability:

1) Direct KPIs (Direct Measurements) are based on data obtained from standard instrumentation such as flow meters, ammeters, and level sensors. They are used to calculate overall water or energy consumption and represent the most reliable data, forming the foundation of digital maturity (Levels 1–2);
2) Derived KPIs (Derived Metrics) are obtained through mathematical operations applied to direct measurements. An example is specific

energy consumption (kWh/m³). Such calculations require time synchronization of data from power analyzers and flow meters within a single PLC program cycle;

3) Soft sensor–based KPIs (Soft Sensors) rely on mathematical models (regression-based or neural network–based) that predict pollution indicators using indirect variables such as turbidity, electrical conductivity, pH, and aerator power consumption. These KPIs are used for complex environmental indicators (COD, BOD, nitrogen), where the installation of online analyzers is either economically infeasible or technically impractical. At facilities with a high level of digital maturity (Levels 4–5), these models are directly integrated into the SCADA system for dynamic setpoint generation.

To assess the feasibility of implementing the dynamic model at a specific facility, the use of an availability matrix (Table 2) is proposed. This matrix enables automation engineers to quickly identify bottlenecks in the instrumentation and control architecture and to plan the required deployment of virtual sensors.

As shown in Table 2, the environmental KPIs that are most critical for Sustainable Development (COD/BOD) exhibit the highest implementation complexity due to the high cost or technical difficulty of direct measurements. This supports the conclusion that, for enterprises with a medium level of digital maturity, the optimal approach is the deployment of soft sensors. Such solutions enable the acquisition of the data required for the dynamic model by leveraging existing low-complexity instrumentation (flow meters and power sensors), thereby significantly reducing the capital expenditures associated with upgrading automated process control systems.

Measurement accuracy within automation systems is of critical importance for “green” control strategies. In conventional APCS, a sensor error of up to 5% may be acceptable for maintaining tank levels; however, within a sustainable development framework, such an error can lead to incorrect activation of optimization algorithms. In this case, the system may excessively reduce power consumption under the assumption that environmental standards are being met, while in reality polluted wastewater is being discharged. Furthermore, insufficient reliability of energy-related KPIs prevents an objective assessment of the economic return on implemented energy efficiency measures.

To ensure the reliability of sustainable development KPIs, the APCS architecture must include dedicated data validation modules. The dynamic model algorithm should verify input data for value freezing, out-of-range conditions, and anomalous spikes before using them for setpoint recalculation.

Thus, the implementation of dynamic KPIs requires a balance between the available instrumentation infrastructure and the complexity of control algorithms. Facilities with low digital maturity should focus on simple energy-related KPIs, whereas the transition to comprehensive environmentally driven control is feasible only with the deployment of intelligent data analysis tools and soft sensors.

4 RESULTS AND DISCUSSION

To evaluate the effectiveness of the proposed control architecture, simulation modeling of the biological wastewater treatment process was performed based on a typical dairy processing plant. According to the initial data of the study, such facilities are characterized by high influent flow variability (with a coefficient of variation of up to 1.5) and significant concentrations of organic pollutants (COD in the range of 1500–4500 mg/dm³).

Table 2: Availability matrix of sustainable development KPIs in APCS.

Indicator (KPI)	Source / Instrument	Complexity	Engineering Notes
Flow rate	EM / Ultrasonic meter	Low	Baseline variable for most systems
Energy consump-tion	Smart Meter + PLC	Low	Requires integration of digital communication protocols
DO level	Optical / EC sensor	Medium	Needs regular cleaning & calibration
pH value	pH-electrode	Medium	Requires temperature compensation
COD / BOD (online)	Spectro-photometer	High	High cost; complex sample prep
COD (predicted)	Soft sensor (Model)	High	Model based on indirect variables
Nitrogen removal	Ion-selective electrode	High	Interference sensitive; high maintenance

The rationale for the comparative analysis is based on testing the hypothesis that transitioning from static setpoints to dynamic KPI-based control enables the implementation of a sustainable development strategy without major restructuring of the technological scheme. Two simulation scenarios were selected for analysis:

1) Scenario 1 (Baseline). A conventional automated process control system configured to maintain a constant dissolved oxygen concentration (DO = 2.5 mg/L) to ensure reliable oxidation under peak loading conditions;
2) Scenario 2 (Adaptive). Implementation of the dynamic model in which the DO setpoint is adjusted by a supervisory controller based on the calculated Sustainable Development Compliance Index (I_{sd}) and predicted COD values obtained via a soft sensor.

The simulation was conducted over a 24-hour time horizon, covering equipment cleaning cycles and peak production discharges. The results of the comparative analysis of technical and environmental performance indicators are presented in Table 3.

Analysis of the obtained data indicates that Scenario 2 provides a significant reduction in energy consumption (by 28.1%) by eliminating the effect of "excessive aeration" during periods of low load. Although the average COD concentration in the second scenario increased, it consistently remained below the established environmental limit (MPC = 100 mg/dm³), confirming the validity of the selected sustainable development strategy.

Table 3: Comparison of control strategy performance indicators.

Performance Metric	Scenario		Deviation
	1	2	
Average daily energy consumption, kWh	420	302	-28,1%
Specific energy intensity, kWh/m^3	0,92	0,66	-28,3%
Average effluent COD concentration, mg/dm^3	68	84	+23,5%
Maximum effluent COD value, mg/dm^3	92	98	+6,5%
Reagent consumption for pH neutralization, L	45	34	-24,4%

The results demonstrate that the use of KPIs as dynamic setpoints enables the automation system to operate in a "smart compromise" mode between environmental safety and energy efficiency. The dynamic model allows equipment power to be adapted to the actual oxidation demand, which is particularly critical for food industry enterprises characterized by uneven operating schedules. Despite the overall reduction in aeration intensity, the system responds predictively to shock loads, maintaining wastewater variables within regulatory environmental limits.

Thus, practical recommendations for automation engineers can be formulated as follows:

1) Implementation of soft sensors. For facilities where real-time direct measurement of COD/BOD is economically infeasible, the use of regression-based models relying on power consumption and water flow data is recommended;
2) Priority configuration. When programming PLCs, environmental KPIs should be implemented as hard constraints that override energy-saving algorithms. Specifically, when COD approaches 90% of the maximum permissible concentration (MPC), energy optimization functions should be disabled in favor of intensified treatment;
3) Phased digitalization. The implementation of dynamic KPIs should begin with process units characterized by the highest energy consumption (aeration tanks, secondary pumping stations), followed by gradual expansion of the architecture to the entire facility.

5 CONCLUSIONS

This study demonstrates that the implementation of a sustainable development strategy at industrial facilities is feasible only if KPIs are transformed from static reporting indicators into dynamic state variables of automation systems. The proposed mathematical model, based on the calculation of an integral sustainability index, enables real-time balancing between environmental standards and energy efficiency through adaptive recalculation of local controller setpoints, without requiring fundamental changes to the existing technological structure.

The object-oriented approach, combined with the use of soft sensors for predicting complex variables such as COD, ensures a high level of control reliability even under limited digital maturity of instrumentation and control systems. Simulation results obtained for a dairy wastewater treatment case study confirm the possibility of reducing specific

energy consumption by 28% while fully complying with environmental limits. This demonstrates the practical applicability of the proposed approach for retrofitting existing facilities.

Future research will focus on the integration of digital twin technologies, adaptive learning mechanisms for soft sensors, and large-scale validation across different industrial sectors. Overall, the developed architecture can be regarded as a universal and scalable tool for modernizing contemporary automated process control systems within the framework of the global "green" transition.

REFERENCES

[1] R. Salles, J. Mendes, A. J. Baptista, and P. Moura, "A review of indicators-based computational methodologies to improve the sustainability of wastewater treatment plants," Comput. Chem. Eng., 2025, Art. no. 109279, doi: 10.1016/j.compchemeng.2025.109279.

[2] D. Renfrew, V. Vasilaki, and E. Katsou, "Indicator based multi-criteria decision support systems for wastewater treatment plants," Sci. Total Environ., vol. 915, Art. no. 169903, 2024, doi: 10.1016/j.scitotenv.2024.169903.

[3] S. Rahimzade, M. R. A. Moghaddam, M. Shafiei, and M. Namavar, "Sustainability assessment of urban wastewater management (collection, treatment, and reuse): Developing a multi-dimensional indicator-based framework," Sustain. Cities Soc., 2025, Art. no. 106646, doi: 10.1016/j.scs.2025.106646.

[4] G. Sabia, L. Petta, F. Avolio, and E. Caporossi, "Energy saving in wastewater treatment plants: A methodology based on common key performance indicators for the evaluation of plant energy performance, classification and benchmarking," Energy Convers. Manage., vol. 220, Art. no. 113067, 2020, doi: 10.1016/j.enconman.2020.113067.

[5] M. O. van Schaik, S. Sucu, H. J. Cappon, W. S. Chen, D. B. Martinson, D. Ouelhadj, and H. H. Rijnaarts, "Mathematically formulated key performance indicators for design and evaluation of treatment trains for resource recovery from urban wastewater," J. Environ. Manage., vol. 282, Art. no. 111916, 2021, doi: 10.1016/j.jenvman.2020.111916.

[6] K. Chen, H. Wang, B. Valverde-Pérez, S. Zhai, L. Vezzaro, and A. Wang, "Optimal control towards sustainable wastewater treatment plants based on multi-agent reinforcement learning," Chemosphere, vol. 279, Art. no. 130498, 2021, doi: 10.1016/j.chemosphere.2021.130498.

[7] Y. Chen, H. Ma, and Y. K. Wang, "Key performance indicator-related process monitoring for irregular scenarios with incomplete data," Expert Syst. Appl., vol. 298, pt. A, Art. no. 129440, 2026, doi: 10.1016/j.eswa.2025.129440.

[8] Q. Maqsood, R. Fatima, F. Rafaqat, T. Mehmood, S. W. Ali, and M. Hussain, "Revolutionizing water and wastewater treatment: Data-driven approaches for advanced solutions," Desalin. Water Treat., vol. 324, Art. no. 101432, 2025, doi: 10.1016/j.dwt.2025.101432.

[9] L. Allen and J. Cordiner, "Knowledge-enhanced data-driven modeling of wastewater treatment processes for energy consumption prediction," Comput. Chem. Eng., vol. 194, Art. no. 108982, 2025, doi: 10.1016/j.compchemeng.2024.108982.

[10] S. Longo, A. Hospido, and M. Mauricio-Iglesias, "Energy efficiency in wastewater treatment plants: A framework for benchmarking method selection and application," J. Environ. Manage., vol. 344, Art. no. 118624, 2023, doi: 10.1016/j.jenvman.2023.118624.

[11] K. G. Aparna and R. Swarnalatha, "Dynamic optimization of a wastewater treatment process for sustainable operation using multi-objective genetic algorithm and non-dominated sorting cuckoo search algorithm," J. Water Process Eng., vol. 53, Art. no. 103775, 2023, doi: 10.1016/j.jwpe.2023.103775.

[12] Q. Liu and X. Jiang, "Dynamic multi-objective optimization control for wastewater treatment process based on modal decomposition and hybrid neural network," J. Water Process Eng., vol. 61, Art. no. 105274, 2024, doi: 10.1016/j.jwpe.2024.105274.

[13] I. Kalogeropoulos, A. Alexandridis, and H. Sarimveis, "Economic oriented dynamic matrix control of wastewater treatment plants," J. Process Control, vol. 118, pp. 202–217, 2022, doi: 10.1016/j.jprocont.2022.08.006.

[14] C. Yang, Y. Zhang, M. Huang, and H. Liu, "Adaptive dynamic prediction of effluent quality in wastewater treatment processes using partial least squares embedded with relevance vector machine," J. Cleaner Prod., vol. 314, Art. no. 128076, 2021, doi: 10.1016/j.jclepro.2021.128076.

[15] H. Xiao, B. Bai, X. Li, J. Liu, Y. Liu, and D. Huang, "Interval multiple-output soft sensors development with capacity control for wastewater treatment applications: A comparative study," Chemom. Intell. Lab. Syst., vol. 184, pp. 82–93, 2019, doi: 10.1016/j.chemolab.2018.11.007.

[16] P. M. Ching, R. H. So, and T. Morck, "Advances in soft sensors for wastewater treatment plants: A systematic review," J. Water Process Eng., vol. 44, Art. no. 102367, 2021, doi: 10.1016/j.jwpe.2021.102367.

[17] W. Liu, S. He, J. Mou, T. Xue, H. Chen, and W. Xiong, "Digital twins-based process monitoring for wastewater treatment processes," Reliab. Eng. Syst. Saf., vol. 238, Art. no. 109416, 2023, doi: 10.1016/j.ress.2023.109416.

[18] A. N. Matheri, B. Mohamed, F. Ntuli, E. Nabadda, and J. C. Ngila, "Sustainable circularity and intelligent data-driven operations and control of the wastewater treatment plant," Phys. Chem. Earth, vol. 126, Art. no. 103152, 2022, doi: 10.1016/j.pce.2022.103152.

[19] N. Zaiets, N. Lutska, L. Vlasenko and S. Junge, "Efficient Production Management System with a Packaging Quality Control Module,"Proceedings of International Conference on Applied Innovation in IT, vol. 13, issue 5, pp. 411–416, 2025. doi: 10.25673/123060.

SSH Connection During Web Automation Testing

Oleksii Cherkashyn
Blynk Technologies Inc., Brickell Avenue 951, 33131 Miami, FL, USA
oleksii.sciencepapers@gmail.com

Keywords: SSH, Web Application Testing, Web Test Automation, WebdriverIO, JavaScript.

Abstract: Web automated testing has become increasingly relevant as web applications continue to grow in complexity and serve as critical components for businesses and organizations. Ensuring the reliability of server responses during test execution, as well as validating the presence or absence of uploaded files, is a key aspect of maintaining application integrity. This study demonstrates an approach using the WebdriverIO framework in combination with the ssh2 library to perform automated verification of server logs and file states in real time. The results show that web automated tests reliably detect both the presence and removal of files on the server, while the proposed solution proves stable across different operating systems, including Ubuntu, Windows 10, and macOS. Furthermore, it supports execution through CI/CD pipelines, such as Jenkins, enabling seamless integration into automated development workflows. Overall, this approach provides a practical and efficient method for enhancing the accuracy and reliability of web application testing, particularly in scenarios requiring server log monitoring and file validation. Across 790 test executions, the proposed solution demonstrated high stability, with 787 successful runs and 3 failures, resulting in an overall stability rate of approximately 99.62%.

1 INTRODUCTION

Web applications are now widespread, enabling essential operations across areas such as e-commerce, finance, healthcare, and education. With their growing complexity and scale, maintaining their reliability, performance, and security has become increasingly demanding. Web testing—which involves both verification and validation of web-based systems—remains a key factor in ensuring the overall quality of these applications [1].

In recent years, the number of websites has grown rapidly, making it essential to evaluate them against various quality criteria to ensure they meet expected standards [2]. Modern software development is marked by a rapid increase in the complexity and scale of web applications, which requires quality assurance (QA) specialists to adopt efficient testing methods to maintain reliability and security. Traditional manual testing approaches are no longer sufficient to satisfy industry demands, driving the adoption of automated testing frameworks [3]. The web applications is therefore usually more complex than testing other software products, and methods and tools must be found to make testing more efficient [4]. While automated web testing is well-developed, several unresolved challenges still exist. The present study contributes by addressing the verification of diverse server logs and the presence or absence of files on the server via SSH connection in the context of automated web tests. The WebdriverIO [5] framework was chosen as the basis for web automation due to its stability and reliability. WebdriverIO is a Node.js framework, available through npm [6], offering flexibility and compatibility with a wide range of testing tools and environments.

SSH (Secure Shell) offers users a secure channel to connect to computers over potentially insecure networks. It is widely used by system administrators to perform remote logins and execute commands through a command-line environment [7]. This study utilizes the SSH2 client module to establish SSH connections, enabling required access during automated web testing [8]. In this context, SSH connections target the Ubuntu servers [9] directly. The deployment of a web application on this or any other server is irrelevant, as the connection is needed only on the server where the required tests and verifications will take place.

2 RELEVANCE

Web applications now constitute a major part of software produced by organizations. Proper testing is vital to uphold their quality, as they are instrumental in enhancing brand visibility and improving communication with users [4]. Automated testing plays a vital role in modern software development, enabling faster releases and higher-quality software [10]. Web applications are especially difficult to test because of their dynamic nature, asynchronous behavior, and constantly changing interfaces - challenges widely recognized in the literature. As these systems rapidly evolve to meet user needs, testing approaches must also adapt to ensure reliability and efficiency [11].

Given the diversity of web applications, file upload functionality is often present in multiple parts of a system. In the context of web automation, automating file uploads is not always a straightforward task. A clear understanding of the fundamental steps involved in the file upload process is essential for identifying UFU vulnerabilities. The client-side handling of file uploads—such as allowing users to select a file and initiating its transfer—is defined by HTML and HTTP specifications implemented within the browser. In contrast, server-side processing may vary significantly depending on the web application's implementation and the configuration of the web server [12]. While it is certainly possible to validate the server response in automated tests (e.g., checking for a 200 status code) and perform visual verification, not all relevant information can be obtained from the browser console. For example, it is not always possible to confirm whether the file is stored in the expected directory on the server, whether its size matches the expected value, or whether duplicate files have been created. Similarly, in the case of file deletion, only direct verification on the server side can ensure that the file has actually been removed and that no residual copies remain.

A similar situation arises after executing automated tests: not all relevant information can be verified through browser logs alone. Although this largely depends on the server architecture, many systems maintain server-side log files. This study contributes by leveraging SSH connections during the execution of web tests to perform the kinds of server-side validations described above.

3 ANALYSIS OF CURRENT RESEARCH

In the context of server log verification in web automation, scientific studies are typically focused on validating server responses through browser network traffic or by analyzing logs obtained from the browser console after the execution of specific actions by automated web tests.

The authors Tunc and Kulcu, in their research, propose an approach for analyzing server logs to reconstruct user navigation paths in web applications. Their method includes preprocessing steps such as log cleaning, session identification, and path extraction to transform raw log data into structured sequences of user actions. By analyzing these sequences, the authors identify common navigation patterns and detect inefficiencies in user interaction with the system. The study demonstrates that server logs can serve as a reliable source of information about actual system usage without requiring direct user involvement. This approach highlights the potential of server logs for evaluating system behavior and supporting further analysis in web application environments [13].

The authors Tan, Yu, Leners, and Walfish propose an approach for verifying the correctness of web application execution using server logs in untrusted environments. In their work, server logs are treated as auxiliary and potentially untrusted data that help reconstruct the execution order and dependencies between requests. They introduce the Efficient Server Audit Problem, where logs are combined with recorded requests and responses to validate that the server behaved correctly. The proposed solution relies on re-executing requests and comparing the results with original outputs, using logs to guide and optimize this verification process. Additionally, they improve efficiency through reduplicated re-execution, allowing similar execution paths to be verified together rather than individually [14].

The authors Dharmaadia, Athanasopoulos, and Turkmen in their work survey fuzzing frameworks for server-side web applications and emphasize how server-side feedback, including logs and responses, can be used to assess application behavior. They discuss techniques for generating diverse request inputs that exercise different server paths and produce varied server log traces. The study highlights the importance of extracting meaningful feedback from the server, such as HTTP status codes and log information, to guide further fuzzing iterations. By analyzing how different fuzzing tools interpret and respond to server output, the authors illustrate how

logs can serve as a valuable oracle for detecting unexpected or erroneous server behavior. Overall, the paper proposes that effective fuzz testing must leverage server-generated information, including logs, to improve coverage and identify security and reliability issues in web applications [15].

Suryaningrat, Ramayanti, Taberima, and Kurniawan propose a multi-layered approach to secure file uploads that can be effectively applied in testing [16]. They emphasize validating file extensions and MIME types, which suggests testing that invalid or disguised files are properly rejected. They also highlight file size restrictions, making it important to verify boundary values and ensure oversized files are not accepted. The authors recommend sanitizing file names, so test cases should include special characters, long names, and potentially malicious inputs. In addition, they propose validating upload paths and storage locations, which can be tested by attempting unauthorized access or path manipulation. Finally, they suggest using unique identifiers (e.g., hashing), which implies testing that files are stored uniquely and not overwritten.

The authors Anwar, Fadlil, and Riadi, in their research, propose a multi-layered validation approach for SVG file uploads that can be applied in testing scenarios. They emphasize checking the file's magic number, ensuring that uploaded files are truly SVGs and not disguised malicious files. They also recommend analyzing the XML structure using the DOM, which allows testers to verify that files are well-formed and comply with SVG standards. Their approach suggests creating combined test cases that validate extension, MIME type, magic number, and DOM structure simultaneously to detect partially invalid or malicious files. They further highlight the importance of negative testing with corrupted or manipulated SVG files to evaluate the system's robustness. Finally, they propose black-box testing techniques to assess how the application handles potentially dangerous files from an external user perspective [17].

The authors Srimathi, Smilin, and Gowri [18], in their research, propose a security framework for direct user-to-cloud file upload systems, which can be applied in testing file upload functionality. They emphasize validating file types and sizes, ensuring that unauthorized or oversized files are correctly rejected. They also recommend testing that existing files cannot be overwritten without proper permissions. Their approach suggests creating negative test cases for invalid or forged temporary tokens, verifying that the system properly blocks unauthorized uploads. Additionally, they highlight the importance of checking server responses to incorrect or manipulated callback notifications after file upload. Finally, they propose testing all stages of the upload process - token generation, file transfer, and callback handling - to identify potential failures and security vulnerabilities.

It can be concluded that existing research is generally focused on validating uploaded files and server logs, primarily through responses observed in the browser console and network traffic. This study contributes to the field and addresses a research gap by introducing validation of server-side logs and uploaded files directly on the server via an SSH connection during the execution of web automated tests.

4 SOFTWARE CONFIGURATION

4.1 Test Infrastructure Configuration

The study utilized Node.js version 20.19.4, along with WebdriverIO version 8 for automated testing. The Mocha test runner was employed, and the test scripts were implemented in JavaScript [19]. Mocha provides a flexible and adaptable structure that can be tailored to various testing approaches, including support for asynchronous testing, which is essential for modern web applications that rely on dynamic content [20]. The full list of dependencies and libraries is specified in the package.json file, which is included in Appendix (Listing A1).

4.2 Preparing an SSH Connection

First, it is necessary to generate SSH keys. This task is handled by the built-in program ssh-keygen [21], which is available on almost all operating systems. The key generation can be performed using the command below. The example is given for an Ubuntu server, and it is important to run the command under the same user account that will execute the tests.

Bash command:

```
ssh-keygen -t rsa -b 4096
```

After executing this command, two files are created: id_rsa and id_rsa.pub in the ~/.ssh/ directory. The SSH public key from id_rsa.pub must be placed on the Ubuntu server where the web application is deployed. It should be added to the authorized_keys file located in the ~/.ssh/ directory for the root user. t is also important to ensure the security of the connection. Modern providers that offer virtual Ubuntu servers typically allow configuring access

restrictions in their web console. You should restrict access by IP address - specifically, allow only the IP of the Ubuntu server where the tests are executed, which will be connecting to the Ubuntu server hosting the web application.

4.3 Test Preparation

In Appendix (Listing A2), an example of the exported checkQuery function used in the test is provided. It is also important to copy the previously created id_rsa key into the appropriate directory, the path to which is specified in the code above. Within the sshClient.exec method of the function described above, any Ubuntu bash command can be specified. In this particular case, an example using the tail command is provided. After establishing the SSH connection, it monitors the log file in real time for the presence of Command 1 and also verifies the number of occurrences of this command after certain test actions are performed. An example of using the exported function in a test, as well as an example of the test itself, is provided below.

JavaScript code:

```
await checkQuery();
await browser.pause(1500);
const clickBtn = await $('#btn');
await clickBtn.click();
```

In the context of the example above, the function is executed first, establishing an SSH connection and reading the log file while searching for Command 1. Then, a 1.5 seconds wait is introduced, followed by a button click, with the expectation that the required command will be written to the log file after the click. This approach is implemented with consideration for the asynchronous nature of web tests. Within the previously described sshClient.exec method, the tail command can be replaced with any other Bash command. For example, it can be used to search for and count uploaded files during tests. The one-and-a-half-second delay is an average and approximate value, and it depends on the implementation and the server's speed in printing the expected commands to the logs. The command below, for instance, will check the number of files in the specified directory.

JavaScript code:

```
ls -1 uploads/files | wc -l
```

5 RESULTS AND DISCUSSION

In the context of this study, React web applications with a Java server architecture and an Express.js server deployed on an Ubuntu server were tested. The tests themselves were similarly executed on the Ubuntu server and on local machines running Windows 10 and macOS.

The results demonstrate a generally high level of stability of the proposed approach, although not all test executions were fully successful. A key factor influencing reliability is the specific implementation of server-side logging and the level of access to files stored on the server. The tests were executed through Jenkins, which was responsible for running the automated test suites, and this setup similarly demonstrated stable overall behavior. An important requirement is that SSH keys must be generated for the Jenkins user to ensure proper authentication and controlled access.

In this study, 158 tests were conducted on each server architecture described above, with each test executed five times, resulting in a total of 790 test executions per configuration. The study also showed that, for efficient real-time log monitoring, the tail command should include sufficient contextual information, such as IP addresses. However, the effectiveness of this approach strongly depends on the logging implementation to ensure that only relevant log entries are captured, without interference from other users' commands.

Although the overall system demonstrated high stability, three failures were observed during execution. One test failed due to the detection of a duplicate command in the log output, which affected the expected validation logic. Another failure was caused by an unstable internet connection, which interrupted the test execution flow. The third failure occurred because the test did not complete within the defined 1.5-second waiting threshold, which was likely also influenced by temporary network instability on the browser side. Overall, out of 790 test executions, 787 were successful and 3 failed, resulting in a stability rate of approximately 99.62%.

In addition to the approach based on generating SSH keys under the root account, scenarios involving users with restricted permissions were also analyzed. In general, the required permission set depends on the specific task being performed. For example, in the case of reading log files, granting read access to the corresponding files is sufficient. From a security perspective, however, the recommended approach is to generate and use SSH keys for a user account with limited privileges rather than for the root account.

The secure storage of private SSH keys remains a subject of ongoing discussion from the perspective of selecting the most effective security strategy. Since private keys provide direct authentication capabilities, their protection is critically important and requires strict access control mechanisms, secure storage policies, and minimization of unauthorized exposure risks. For access to the Jenkins web interface, it is strongly recommended to enforce IP-based access restrictions in combination with strong user authentication credentials, thereby providing an additional layer of protection against unauthorized access. Furthermore, periodic SSH key rotation should be incorporated into the overall security policy, as regular key regeneration reduces the potential impact of key compromise and limits the exposure window of leaked credentials. The optimal key rotation interval depends on the security requirements, infrastructure scale, and organizational risk model; however, security best practices generally recommend performing key renewal on a regular basis and immediately after any suspected security incident or personnel change.

6 CONCLUSIONS

Although some studies focus on server log verification and validation of uploaded files, this study contributes to the topic of SSH connections during the execution of web automated tests, as well as the verification of logs and files on the server. Based on the results of this study, the following conclusions can be drawn:

- Web automated tests using the WebdriverIO framework with the ssh2 library reliably perform verification of server logs and the presence or absence of files on the server after they are uploaded or deleted;
- The solution is stable, regardless of the operating system on which the tests are executed;
- The solution supports test execution through CI/CD pipelines, for example, using Jenkins.

REFERENCES

[1] K. Kapula, “A Survey on Web Testing: on the Rise of AI and Applications in Industry,” Journal of Information Systems Engineering and Management, vol. 10, no. 2, 2025, , doi: 10.52783/jisem.v10i2.9469.

[2] M. Kundra, “Selenium – A Trending Automation Testing Tool,” International Journal of Trend in Scientific Research and Development (IJTSRD), vol. 4, no. 4, pp. 1321-1324, Jun. 2020, [Online]. Available: https://www.ijtsrd.com/engineering/software-engineering/31202/selenium-–-a-trending-automation-testing-tool/manav-kundra

[3] A. Tymoshchuk, “The Evolution of Test Automation: From Selenium to Playwright. A Comparison of Automation Tools: Selenium vs. Playwright vs. Cypress,” International Journal of Computer (IJC), vol. 54, no. 1, pp. 37-45, Apr. 2025, [Online]. Available: https://ijcjournal.org/InternationalJournalOfComputer/article/view/2355

[4] S. Balsam and D. Mishra, “Web application testing - Challenges and opportunities,” Journal of Systems and Software, vol. 219, Jan. 2025, Art. no. 112186, , doi: 10.1016/j.jss.2024.112186.

[5] O. Cherkashyn, “Application Test Automation in Headless Android Emulator,” Proceedings of International Conference on Applied Innovation in IT, vol. 13, no. 5, pp. 445-452, 2025, , doi: 10.25673/123064.

[6] H. Sun, A. Rosà, D. Bonetta and W. Binder, “Automatically Assessing and Extending Code Coverage for NPM Packages,” in IEEE/ACM International Conference on Automation of Software Test (AST 2021), pp. 40-49, 2021, doi: 10.1109/AST52587.2021.00013.

[7] D. D. Tran, K. Ogata, S. Escobar, S. Akleylek and A. Otmani, “Formal Analysis of Post-Quantum Hybrid Key Exchange SSH Transport Layer Protocol,” IEEE Access, vol. 12, pp. 1672-1687, Dec. 2023, , doi: 10.1109/ACCESS.2023.3347914.

[8] A. Lekova, P. Tsvetkova, A. Andreeva, G. Dimitrov, T. Tanev, M. Simonska, T. Stefanov, V. Stancheva Popkostadinova, G. Padareva, K. Rasheva and D. Vitanova, “A Design Based Research Approach to Streamline the Integration of High Tech Assistive Technologies in Speech and Language Therapy,” Technologies, vol. 13, no. 7, p. 306, Jul. 2025, , doi: 10.3390/technologies13070306.

[9] S. Sadeq, W. B. Abdulaziz, R. N. A. Alsaadi and M. M. A. Algherini, “Enhancing Server Security: Implementation and Evaluation of the Port Knocking Method on Ubuntu Virtual Servers,” Journal of Engineering and Sustainable Development, vol. 29, no. 5, Nov. 2025, , doi: 10.31272/jeasd.2809.

[10] R. Barua, S. K. Ghosh and M. Andalibur Rahman, “The Future of Test Automation: A Comparative Analysis of Selenium vs. AI Driven Tools,” International Journal of Data Science, Bioinformatics and Cyber Security, vol. 1, no. 1, pp. 28-45, May 2025, , doi: 10.46610/IJDSBCS.2025.v01i01.003.

[11] M. Rahman, S. K. Ghosh and R. Barua, “AutoQALLMs: Automating Web Application Testing Using Large Language Models (LLMs) and Selenium,” Computers, vol. 14, no. 11, p. 501, Nov. 2025, , doi: 10.3390/computers14110501.

[12] S. Neef and M. Oudeh, “Bringing UFUs Back into the Air with FUEL: A Framework for Evaluating the Effectiveness of Unrestricted File Upload Vulnerability Scanners,” in Proceedings of the 21st International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA 2024), Lausanne, Switzerland, Jul. 2024, pp. 207-226, , doi: 10.1007/978-3-031-64171-8_11.

[13] S. K. Tunc and O. Kulcu, “A Web Application Path Analysis through Server Logs,” in Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2018) - Volume 1: KDIR, pp. 427-430, 2018, , doi: 10.5220/0007231804270430.

[14] C. Tan, L. Yu, J. B. Leners and M. Walfish, “The Efficient Server Audit Problem, Deduplicated Re-execution, and the Web,” in Proceedings of the 26th Symposium on Operating Systems Principles (SOSP ’17), pp. 546-564, Oct. 2017, doi: 10.1145/3132747.3132760.

[15] P. A. Dharmaadi, E. Athanasopoulos and F. Turkmen, “Fuzzing frameworks for server side web applications: a survey,” International Journal of Information Security, vol. 24, article no. 73, Feb. 2025, doi: 10.1007/s10207-024-00979-w.

[16] A. Suryaningrat, D. Ramayanti, G. M. Taberima and P. P. Kurniawan, “File Upload Security: Essential Practices for Programmers,” CCIT Journal, vol. 17, no. 2, 2024, , doi: 10.33050/ccit.v17i2.3172.

[17] F. Anwar, A. Fadlil and I. Riadi, “Validation Analysis of Scalable Vector Graphics (SVG) File Upload using Magic Number and Document Object Model (DOM),” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 11, no. 11, 2020, , doi: 10.14569/IJACSA.2020.0111133.

[18] S. Srimathi, K. Shanthini Smilin and G. Gowri, “A Security Framework for Direct User-to-Cloud Upload Systems with Automation for Vulnerability Detection,” International Journal for Research Trends and Innovation, vol. 10, no. 12, pp. a192-a198, Dec. 2025, [Online]. Available: https://ijrti.org/papers/IJRTI2512025.

[19] A. W. Putra and N. Legowo, “Greedy Algorithm Implementation for Test Case Prioritization in the Regression Testing Phase,” Journal of Computer Science, vol. 21, no. 2, pp. 290-303, 2025, , doi: 10.3844/jcssp.2025.290.303.

[20] E. Kesavan, “Comparative Study of JavaScript-Based Testing Frameworks Integrated with Selenium WebDriver,” International Scientific Journal of Engineering and Management, vol. 1, no. 5, Dec. 2022, , doi: 10.55041/ISJEM00087.

[21] F. Rascoussier, “Predicting SSH keys in Open SSH Memory dumps,” in University of Passau, Germany; INSA Lyon, 2024, p. 148, [Online]. Available: https://hal.science/hal-04562462v1.

APPENDIX

The Appendix contains the scripts and sections of programming languages code used in the study.

```
{
  "name": "app_name",
  "type": "module",
  "devDependencies": {
    "@wdio/cli": "^8.32.3",
    "@wdio/local-runner": "^8.32.3",
    "@wdio/mocha-framework": "^8.16.3",
    "@wdio/spec-reporter": "^8.16.3",
    "chai": "^5.1.0",
    "chromedriver": "^144.0.0",
    "geckodriver": "^4.3.0",
 "wdio-chromedriver-service": "^8.1.1",
    "ssh2": "^1.15.0"
  },
  "scripts": {
    "test": "wdio run ./wdio.conf.ts"
  },
}
```

Listing A1: Package.json.

```
import { Client } from 'ssh2';
import fs from 'fs';

const sshConfig = {
host: 'server_ip',
    username: 'root',
    privateKey:
fs.readFileSync('./sshkey/id_rsa')
};

export async function checkQuery ()
{
    const sshClient = new Client();
    let lastOutput = '';
    let outputCount = 0;
    sshClient.on('ready', () => {
    console.log('SSH connection
established');

    sshClient.exec('tail -f logs.log
| grep --line-buffered "Command 1"',
(err, stream) => {
    if (err) throw err;

    stream.on('data', data => {
    const output = data.toString();
    if (output !== lastOutput) {
    lastOutput = output;
    outputCount++;
    console.log('Output:', output);
    if (outputCount > 1) {
    sshClient.end();
    throw new Error (`Wrong count
Command 1. Count is ${outputCount}`)
    }
}

    }).stderr.on('data', data => {
    });
});
     setTimeout(() => {
     sshClient.end();
     console.log('SSH connection
closed after 3 seconds');
        }, 3000);
    })
    sshClient.connect(sshConfig);
}
```

Listing A2: JavaScript code.

Experimental Evaluation of Automated Intermittent Pneumatic Mixing in Moisture-Controlled Anaerobic Digestion

Sergey Pochinok[1], Alexey Lavrentyev[1], Alaibek Obozov[2], Raya Sultanalieva[1], Sairagul Nasirdinova[1] and Veronika Kirova[3]

[1] *Kyrgyz State Technical University named after I. Razzakov, Ch. Aitmatov Avenue 66, 720044 Bishkek, Kyrgyz Republic*

[2] *Laboratory of Renewable Energy Sources, National Academy of Sciences of the Kyrgyz Republic, Skribina Str. 23, 720055 Bishkek, Kyrgyz Republic*

[3] *Anhalt University of Applied Sciences, Bernburger Str. 55, Koethen, Germany*

info@incomcompany.kz, teolidz@mail.ru, alaibek.obozov@gmail.com, raia-ktu@mail.ru, snasirdinova1967@gmail.com

Keywords: Anaerobic Digestion, Pneumatic Mixing, Automated Control, Intermittent Mixing, Biogas Yield, Moisture Control, Substrate Preparation.

Abstract: Stable anaerobic digestion depends on both substrate preparation and reproducible mixing conditions inside the bioreactor. This paper presents an experimental evaluation of automated intermittent pneumatic mixing combined with moisture-controlled substrate preparation for anaerobic digestion systems. Moisture content and dry matter fraction of a prepared corn silage substrate were determined by thermogravimetric analysis and used to calculate the required water addition for target moisture levels before reactor loading using a wet-basis mass-balance approach. The main experimental comparison was performed between manual pneumatic mixing and automated intermittent pneumatic mixing controlled by a time relay and solenoid valve. The automated mode provides repeatable gas injection into the distribution chamber, reduces operator-dependent variability, and improves the reproducibility of the mixing regime. Experimental results showed that automated intermittent mixing increased total biogas production from 99.35 m³ to 125.63 m³, corresponding to an absolute increase of 26.28 m³ and a relative improvement of 26.45% compared with manual operation. In addition, intermittent operation reduced compressor duty cycle compared with hypothetical continuous mixing, indicating potential energy savings during long-term operation. The study introduces operational performance indicators for comparing mixing modes and applies a practical moisture-standardization procedure for substrate preparation. Unlike previous studies focused separately on substrate preparation or mixing automation, the proposed approach integrates moisture-controlled substrate standardization, automated pneumatic mixing, and operational performance evaluation within a single experimental workflow. The proposed method can be applied to small and medium-scale biogas plants where simple, low-cost automation is preferable to complex control systems while improving process stability and biogas productivity.

1 INTRODUCTION

Anaerobic digestion is widely used for the conversion of organic waste into biogas and stabilized digestate and is considered one of the key technologies for renewable energy production and sustainable waste management [1], [7]. The efficiency and stability of anaerobic digestion depend on a combination of biological, thermal, hydraulic, and mechanical factors inside the bioreactor. In addition to temperature control and microbial activity, substrate preparation and mixing conditions strongly influence mass transfer, gas release, and the uniformity of the digestion process [2], [6]. Insufficient or irregular mixing may lead to the formation of floating crusts, sediment accumulation, local temperature gradients, and stagnant zones with limited substrate circulation, which negatively affects biogas productivity and process repeatability [2], [7].

For small and medium-scale biogas plants, pneumatic bubbling is an attractive alternative to conventional mechanical mixing because it reduces the number of moving mechanical components inside the reactor and simplifies maintenance requirements. Pneumatic agitation can be implemented using relatively simple compressors, gas receivers, and bubbling pipelines, which makes this approach

suitable for decentralized agricultural and rural biogas installations. However, in many practical systems the pneumatic mixing process is activated manually by the operator. As a result, the duration and frequency of agitation depend on operator attention and routine practice, introducing variability into reactor operation. Under such conditions, the same biogas plant may operate under different mixing regimes from one production cycle to another.

Automated intermittent pneumatic mixing offers a practical solution to this limitation. By integrating a solenoid valve and a time relay into the gas-distribution system, pneumatic bubbling can be activated according to a predefined schedule without continuous operator intervention. Such automation improves the repeatability of the mixing regime and stabilizes substrate agitation conditions inside the reactor. In addition, intermittent operation reduces compressor duty cycle compared with continuous mixing and therefore may decrease operational energy demand during long-term plant operation.

Substrate preparation is also an important factor influencing anaerobic digestion performance. The moisture content of biomass affects substrate rheology, pumpability, microbial accessibility, and the tendency of heterogeneous feedstock to form dense layers or local accumulations [5], [6]. In practical operation, the same nominal feedstock composition may exhibit different actual moisture levels because of seasonal variation, storage conditions, or raw-material heterogeneity. Consequently, uncontrolled moisture variation may influence biogas yield independently of the selected mixing mode. For this reason, moisture determination and water-addition calculation should be considered as part of a substrate-standardization procedure prior to evaluating mixing performance.

This paper presents an experimental and calculation-based evaluation of automated intermittent pneumatic mixing combined with moisture-controlled substrate preparation for anaerobic digestion systems. The study integrates:

1) thermogravimetric determination of substrate moisture content and dry matter fraction;
2) wet-basis mass-balance calculation of required water addition before loading;
3) experimental comparison of manual and automated intermittent pneumatic mixing modes;
4) operational performance indicators for evaluating biogas-yield improvement, repeatability, and potential energy reduction.

Experimental results demonstrated that automated intermittent pneumatic mixing increased total biogas production from 99.35 m³ to 125.63 m³, corresponding to an absolute increase of 26.28 m³ and a relative improvement of 26.45% compared with manual operation. The automated mode also provided more reproducible mixing conditions and reduced operator-dependent variability.

Unlike previous studies focused either on substrate preparation or on mixing-system automation separately, the proposed approach integrates moisture-controlled substrate standardization, automated intermittent pneumatic mixing, and operational performance evaluation within a single practical workflow for anaerobic digestion systems. The proposed method is intended primarily for small and medium-scale biogas plants where low-cost and operationally simple automation solutions are preferable to complex industrial control systems.

The remainder of the paper is organized as follows. Section 2 describes the experimental system, substrate preparation procedure, moisture calculations, and the compared mixing modes. Section 3 presents the experimental results and discusses the influence of automated intermittent pneumatic mixing on biogas production, process repeatability, and operational performance. Section 4 summarizes the main findings and outlines directions for future research.

2 MATERIALS AND METHODS

This section describes the experimental system and calculation procedures used to evaluate automated intermittent pneumatic mixing in moisture-controlled anaerobic digestion. The methodology includes: (i) a description of the biogas plant and pneumatic bubbling system; (ii) the comparison of manual and automated mixing modes; (iii) substrate moisture determination and water-addition calculation prior to loading; and (iv) performance indicators used to assess biogas-yield improvement and operational repeatability. Moisture content was determined by thermogravimetric analysis with time–mass recording, while the required water addition was calculated using a wet-basis mass balance [5], [6]. The automated mixing mode was implemented using a solenoid valve and a time relay, enabling repeatable gas injection into the distribution chamber according to a predefined schedule [4].

2.1 Experimental Biogas Plant and Pneumatic Mixing System

The experimental system was based on a biogas plant equipped with a pneumatic bubbling system for

substrate mixing. The system included a biochemical reactor, gas holder, water seal, compressor, receiver, distribution chamber, solenoid valve, and a set of perforated bubbling pipes installed in the lower part of the reactor. During operation, biogas accumulated in the upper part of the reactor was compressed and supplied through the distribution chamber into the bubbling pipes. The gas jets passed through the substrate layer and provided mixing by pneumatic agitation. This configuration reduces the need for mechanical moving elements inside the reactor and is suitable for small and medium-scale biogas installations.

The automated control unit consisted of a solenoid valve connected to a time relay. In the automated mode, the relay opened and closed the solenoid valve according to a predefined schedule, allowing gas injection into the distribution chamber without manual operator intervention. In the experimental configuration, intermittent mixing was applied every 4 h 30 min for 30 s, which was selected as the operating schedule for automated pneumatic agitation.

2.2 Mixing Modes Compared

Two pneumatic mixing modes were considered in this study: manual pneumatic mixing and automated intermittent pneumatic mixing. In the manual mode, gas supply to the distribution chamber was activated by the operator through a valve. Therefore, the timing and duration of mixing depended on operator attention and routine practice. This mode represents the conventional operation of simple pneumatic mixing systems used in small and medium-scale biogas plants.

In the automated intermittent mode, gas supply was controlled by a solenoid valve connected to a time relay. The valve was opened and closed according to a predefined schedule, which provided repeatable activation of pneumatic bubbling and reduced operator-dependent variability. In the experimental configuration, the mixing system was activated every 4 h 30 min for 30 s. This mode was selected to provide regular substrate agitation while avoiding continuous energy demand.

For methodological comparison, continuous mixing was also considered as a reference concept. Unlike intermittent mixing, continuous operation provides permanent agitation but may increase energy consumption and may not be necessary for small-scale systems where periodic substrate agitation is sufficient. Therefore, the main experimental comparison in this paper focuses on manual and automated intermittent pneumatic mixing, while continuous operation is used as an energy-related reference scenario. The considered mixing modes are summarized in Table 1 for clarity of comparison.

The following subsection describes the substrate preparation and moisture-control procedure used to standardize the feedstock before comparing mixing performance.

2.3 Substrate Preparation and Moisture Determination

A prepared corn silage substrate was used as the test material for moisture determination. Because heterogeneous biomass can exhibit local moisture variation, the bulk substrate was manually mixed before sampling to improve representativeness and reduce sampling bias [5], [6]. A test portion was then taken from the mixed substrate and immediately placed on the analyzer pan to minimize unintended moisture change during handling.

Moisture measurements were performed using a laboratory moisture analyzer Kern MLB 50-3N. The instrument operates by thermogravimetric moisture determination: the sample is weighed, heated under a controlled program, and the mass loss is recorded until the measurement endpoint is reached. This procedure is consistent with standard gravimetric approaches for moisture and solids determination widely used in biomass and wastewater analysis [3]. The analyzer provides moisture readings with a resolution of 0.01% and mass readings with a resolution of 0.001 g, with a maximum load of 50 g [8].

Table 1: Mixing modes considered in the study.

Mixing mode	Control principle	Mixing activation	Main limitation / feature
Manual pneumatic mixing	Operator-controlled	Manual opening of gas supply valve	Human factor; irregular timing
Automated intermittent pneumatic mixing	Time relay + solenoid valve	Every 4 h 30 min for 30 s	Repeatable operation; reduced operator dependence
Continuous mixing	Constant operation	Continuous or near-continuous agitation	Higher energy demand

The determination procedure included the following steps. First, the analyzer was prepared and tared with the empty pan. Second, the test portion was placed on the pan and evenly distributed as a thin layer to promote uniform heating and drying. Third, the drying program was started, and the analyzer recorded sample mass as a function of time. The run was completed when the endpoint criterion of the analyzer was reached, corresponding to stabilization of the measured mass. For the demonstrated run, the total measurement duration was 18 min 30 s.

The recorded time–mass data were subsequently used to calculate wet-basis moisture content, dry matter fraction, and the initial moisture value required for the water-addition calculation.

The next subsection presents the mass-balance equations used to calculate moisture content, dry matter fraction, and the amount of water required to reach target substrate moisture before loading.

2.4 Moisture and Water-Addition Calculations

To standardize the initial substrate condition before comparing mixing modes, moisture content and dry matter fraction were determined for the prepared substrate portion. Moisture content was determined by a gravimetric approach based on the initial (wet) sample mass and the final (dry) mass after completion of the drying program [5], [6]. The following quantities were used:

Initial mass: m_0;

Final (dry) mass: m_d;

Using m_0 and m_d, the moisture content on a wet basis, W (%), was calculated according to (1), where ($m_0 - m_d$) represents the mass of water removed during drying:

$$W = \frac{m_0 - m_d}{m_0} * 100\%. \quad (1)$$

The dry matter fraction, DM (%), was then determined as the complement of moisture on a wet basis according to (2):

$$DM = 100\% - W. \quad (2)$$

For reporting purposes, the water mass contained in the analyzed portion was calculated according to (3):

$$m_w = m_0 - m_d. \quad (3)$$

To estimate the amount of mixing water required to reach a target wet-basis moisture, W_t (%), for a given substrate mass, M, a wet-basis mass balance was applied [2], [6]. Let W_0 (%) denote the measured initial moisture content of the substrate (wet basis). Assuming that added water affects only the water fraction while the dry matter mass remains constant, the required added water mass, M_{add}, was calculated according to (4):

$$M_{add} = M * \frac{W_t - W_0}{100 - W_t'}, \quad (4)$$

where mw is the water mass in the analyzed portion; M and M_{add} are expressed in the same mass units (e.g., kg); and W_0 and W_t are expressed in percent.

In this study, the moisture-control calculation was used as a substrate-standardization step, ensuring that differences observed between manual and automated pneumatic mixing modes were not primarily caused by uncontrolled variation in initial substrate moisture. For operational use, the water-addition calculation can be implemented as a simple spreadsheet or summarized as guideline values per 100 kg of substrate for typical target moisture levels.

In Section 3, the recorded time–mass data are first used to compute W, DM, and practical guideline values of M_{add} per 100 kg of substrate. These standardized substrate-preparation parameters are then linked to the comparison of manual and automated pneumatic mixing modes**.**

2.5 Performance Indicators for Comparing Mixing Modes

To evaluate the influence of pneumatic mixing mode on anaerobic digestion performance, a set of operational and energy-related indicators was introduced. The indicators were selected to characterize biogas productivity, process repeatability, and potential reduction in energy consumption under automated intermittent operation.

The primary performance metric was the total biogas production obtained under each operating mode. The absolute increase in biogas production was calculated according to (5):

$$\Delta V = V_{auto} - V_{manual}, \quad (5)$$

where V_{auto} is the total biogas volume obtained under automated intermittent pneumatic mixing, and V_{manual} is the total biogas volume obtained under manual pneumatic mixing.

To compare the efficiency of the two operating modes, the relative biogas-yield improvement was determined according to (6):

$$I_{BG} = ((V_{auto} - V_{manual}) / V_{manual}) \times 100\%, \quad (6)$$

where I_{BG} represents the percentage increase in biogas production relative to the manual mode.

In addition to gas yield, process stability and operational repeatability were considered. In manual operation, the activation of pneumatic bubbling depends on the operator and may vary in duration and timing between cycles. In contrast, the automated intermittent mode provides repeatable activation intervals through the use of a time relay and solenoid valve. Therefore, repeatability of mixing activation was used as a qualitative indicator of process stability.

To evaluate operational efficiency, the energy demand associated with pneumatic mixing was estimated. The energy consumption of the mixing system during operation was calculated according to (7):

$$E_{mix} = P \times t_{on}, \quad (7)$$

where P is the power of the compressor or gas-supply equipment, and ton is the total operating time during the considered period.

For comparative analysis, continuous mixing was considered as a reference operating scenario. The potential energy saving achieved by intermittent mixing relative to continuous operation was estimated according to (8):

$$SE = (1 - E_{int} / E_{cont}) \times 100\%, \quad (8)$$

where E_{int} is the energy consumption under intermittent operation and E_{cont} is the estimated energy consumption for continuous mixing over the same period..

Under the experimental schedule used in this study, pneumatic mixing was activated every 4 h 30 min for 30 s. Such intermittent operation substantially reduces the cumulative compressor operating time compared to continuous agitation while maintaining regular substrate mixing. Consequently, the automated intermittent mode combines improved operational repeatability with lower expected energy demand.

The proposed indicators provide a practical framework for comparing pneumatic mixing modes in terms of biogas productivity, process stability, and operational efficiency. The following section presents the experimental results obtained from substrate moisture measurements and from the comparison of manual and automated intermittent mixing modes.

3 RESULTS AND DISCUSSION

3.1 Substrate Moisture and Water-Addition Results

Following the procedure described in Section 2.3, a time–mass record was obtained during thermogravimetric drying of the prepared corn silage substrate. Table 2 reports the measured sample mass as a function of time. The sample mass decreased monotonically from 2.000 g to 0.475 g due to moisture removal and approached a quasi-stable value toward the end of the run, which is consistent with the analyzer endpoint criterion based on mass stabilization [5]. The total duration of the demonstrated measurement was 18 min 30 s.

Table 2: Time–mass data recorded during moisture determination of the substrate.

Time, t (min)	Measured mass,m (t) (g)
0	2.000
1	1.800
2	1.710
3	1.560
4	1.410
5	1.280
6	1.167
7	1.060
8	1.000
9	0.930
10	0.833
11	0.750
12	0.690
13	0.674
14	0.630
15	0.600
16	0.570
17	0.520
18	0.480
18.5	0.475

Using the final stabilized mass as m_d, the next subsection calculates the wet-basis moisture W, dry matter fraction DM, and the corresponding mixing-water requirement for typical target moisture levels. These values are used to standardize substrate preparation before comparing manual and automated pneumatic mixing modes.

3.2 Moisture Content and Dry Matter Fraction

Using the initial wet mass m_0 and the final dry mass md obtained at the analyzer endpoint, the substrate moisture content on a wet basis and the dry matter fraction were calculated according to (1)–(3). The initial wet mass was 2.000 g, while the final dry mass after thermogravimetric drying was 0.475 g. Therefore, the water mass removed during drying was 1.525 g.

The resulting values of W, DM, and mw for the tested substrate portion are summarized in Table 3. These parameters quantify the initial condition of the substrate batch and can be used as input data for standardized substrate preparation. In the context of this study, they help reduce the influence of uncontrolled feedstock-moisture variation when comparing manual and automated pneumatic mixing modes.

Table 3: Moisture content and dry matter fraction for the tested substrate portion.

Parameter	Symbol	Value
Initial wet mass (g)	m_0	2.000
Final dry mass at endpoint (g)	m_d	0.475
Water mass in portion (g)	m_w	1.525
Moisture content, wet basis (%)	W	76.25
Dry matter fraction (%)	DM	23.75

The measured initial moisture W_0 = 76.25% was then used to calculate the amount of mixing water required for typical target moisture levels before evaluating the influence of mixing mode on biogas production.

3.3 Visual Observation of the Sample Before and After Drying

A visual comparison of the tested substrate portion before and after the thermogravimetric moisture-determination run is presented in Fig. 1. After drying, the sample exhibited the expected reduction in apparent wetness and a more compact structure, which is consistent with the measured mass decrease from 2.000 g to 0.475 g [5]. Although visual assessment was not used for quantitative evaluation, the image provides additional documentation of the tested material and supports the traceability of the laboratory procedure [6].

The measured initial moisture W_0 = 76.25% is used in the next subsection to estimate the required amount of added water for target moisture levels before comparing the effect of mixing mode on biogas production.

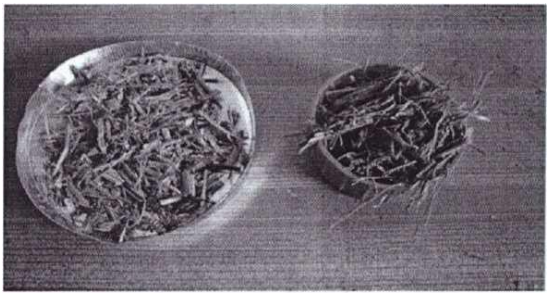


Figure 1: Tested substrate portion before and after thermogravimetric drying (moisture determination run).

3.4 Practical Estimation of Mixing Water or Target Moisture Levels

For operational use, the measured initial moisture W_0 = 76.25% can be applied to estimate the amount of added water required to reach a target moisture level before loading. For a substrate batch of mass M with initial wet-basis moisture W_0 and target wet-basis moisture W_t, the required added water mass Madd was estimated using the wet-basis mass balance defined in (4). The calculation assumes that the added water contains no dry matter and that the dry matter mass of the substrate remains constant during dilution [2], [6].

For routine operation, (4) was evaluated for a reference batch mass of M = 100 kg to obtain practical guideline values. Table 4 summarizes the corresponding added water mass M_{add} required to reach typical target moisture levels.

Table 4: Example guideline values of added water per 100 kg of substrate for typical target moisture levels.

Target moisture, W_t (%)	Added water per 100 kg, M_{add} (kg)
85	58.33
92	196.88

The values in Table 4 were calculated directly from the measured initial moisture W_0 = 76.25% using the wet-basis mass balance in (4). Therefore, they represent exact calculation values for the tested substrate portion and may differ from rounded operational guideline values based on tabulated initial-moisture categories.

The estimate applies when $W_t > W_0$, that is, when substrate moisture is increased by adding water.

If $W_t \leq W_0$, water addition is not required; achieving a lower moisture level would require dewatering or mixing with a drier co-substrate. These values provide standardized substrate-preparation parameters and reduce the risk that differences in biogas production are caused by uncontrolled variation in initial moisture.

The next subsection presents the main experimental comparison of manual and automated pneumatic mixing modes using biogas-production data.

3.5 Biogas Production Comparison Under Manual and Automated Mixing Modes

The main experimental comparison was performed between manual pneumatic mixing and automated intermittent pneumatic mixing [9]. In the manual mode, gas supply to the pneumatic bubbling system was activated by the operator. In the automated mode, gas injection was controlled by a time relay and solenoid valve, providing repeatable intermittent mixing according to a predefined schedule.

The experimental results demonstrated that automated intermittent pneumatic mixing increased the total biogas production from 99.35 m³ in the manual mode to 125.63 m³ in the automated mode. The absolute increase in biogas production was 26.28 m³, corresponding to a relative improvement of 26.45% compared to manual operation.

Table 5 summarizes the comparative biogas-production results for the investigated mixing modes.

Table 5: Comparison of biogas production under manual and automated pneumatic mixing modes.

Mixing mode	Total biogas production, V (m³)	Absolute increase, ΔV (m³)	Relative increase, IBG (%)
Manual pneumatic mixing	99.35	—	—
Automated intermittent pneumatic mixing	125.63	26.28	26.45

The obtained results indicate that automated intermittent pneumatic mixing improves both biogas productivity and operational repeatability. The predefined activation schedule ensured more stable and reproducible substrate agitation compared with manual operation, where the mixing regime depends on operator intervention. More regular pneumatic bubbling likely contributed to improved substrate homogenization and reduction of stagnant zones inside the reactor, which positively influenced the anaerobic digestion process.

The results additionally suggest that intermittent automated mixing may provide operational advantages over continuous agitation by reducing unnecessary operating time of the pneumatic system while maintaining stable reactor performance. Further analysis of operational stability and energy-related aspects is presented in Section 3.6.

3.6 Operational Stability and Energy Considerations

In addition to the increase in total biogas production, the automated intermittent pneumatic mixing mode demonstrated improved operational stability and reduced dependence on manual operator intervention. In the manual mode, the activation time and duration of pneumatic bubbling depended on routine operational practice and therefore could vary between operating cycles [10]. Such variability may lead to irregular substrate agitation, uneven gas distribution, and the formation of stagnant zones inside the reactor.

In contrast, the automated intermittent mode ensured repeatable activation of pneumatic bubbling through the use of a time relay and solenoid valve. The predefined operating schedule (30 s activation every 4 h 30 min) maintained regular substrate agitation throughout the digestion process while minimizing unnecessary compressor operation. As a result, the automated mode provided a more reproducible mixing regime and reduced the influence of human-factor-related variability.

From an operational perspective, intermittent pneumatic mixing also decreases the cumulative operating time of the gas-supply equipment compared with continuous agitation. Assuming a continuous operating mode as a reference scenario, the intermittent schedule used in this study corresponds to a substantially lower duty cycle and therefore lower expected electrical energy consumption of the compressor system. Although direct electrical measurements were not performed in the present work, the reduction in operating time indicates the potential for energy savings during long-term reactor operation.

The combination of increased biogas yield, improved repeatability of mixing activation, and reduced expected operating time of the pneumatic system demonstrates the practical suitability of automated intermittent mixing for small and medium-scale anaerobic digestion systems. These results

support the use of simple automation elements as an effective approach for improving both process stability and operational efficiency in biogas plants.

4 CONCLUSIONS

This paper presented an experimental and calculation-based evaluation of automated intermittent pneumatic mixing combined with moisture-controlled substrate preparation for anaerobic digestion. The substrate preparation stage included thermogravimetric determination of moisture content and dry matter fraction in a prepared corn silage substrate. The measured initial wet mass was 2.000 g, and the final dry mass at the endpoint was 0.475 g, corresponding to a wet-basis moisture content of 76.25% and a dry matter fraction of 23.75%. These values were used to calculate the required water addition for target substrate moisture levels before loading.

The main experimental comparison demonstrated that automated intermittent pneumatic mixing increased total biogas production from 99.35 m³ in the manual mode to 125.63 m³ in the automated mode. The absolute increase was 26.28 m³, while the relative improvement compared to the manual mode reached 26.45%. The automated mode provided repeatable gas injection through a time relay and solenoid valve, reducing operator-dependent variability and improving the stability and reproducibility of the mixing regime.

The study additionally introduced operational indicators for evaluating mixing performance, including biogas-yield improvement, process repeatability, and estimated energy efficiency. Although direct electrical measurements were not performed, the intermittent operating schedule substantially reduced compressor operating time compared with continuous mixing, indicating the potential for lower energy consumption during long-term operation.

The novelty of the proposed approach lies in combining substrate moisture standardization, automated intermittent pneumatic mixing, and operational performance assessment within a single practical workflow for anaerobic digestion systems. The proposed method is particularly suitable for small and medium-scale biogas plants where low-cost automation and stable operating conditions are important.

Further research should include direct monitoring of compressor energy consumption, long-term evaluation of daily biogas productivity, investigation of different intermittent mixing schedules, and validation of the proposed approach under varying substrate compositions and reactor loading conditions.

ACKNOWLEDGMENTS

We acknowledge support by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) and the Open Access Publishing Fund of Anhalt University of Applied Sciences.

REFERENCES

[1] A. G. Vedenev and A. V. Lavrentyev, “Biogas plant,” Kyrgyz Republic Patent 1624, May 30, 2014.

[2] A. Dzh. Obozov and A. V. Lavrentyev, “Algorithm of software control of the feedstock mixing mode in the methane tank of a biogas plant,” Problems of Automation and Control, no. 2(41), pp. 56-60, 2021.

[3] APHA, AWWA, and WEF, Standard Methods for the Examination of Water and Wastewater, 23rd ed. Washington, DC, USA: American Public Health Association, 2017, Method 2540 “Solids.”

[4] ISO 18134-1:2015, Solid Biofuels - Determination of Moisture Content - Oven Dry Method - Part 1: Total Moisture - Reference Method. Geneva, Switzerland: International Organization for Standardization, 2015.

[5] ISO 18134-2:2017, Solid Biofuels - Determination of Moisture Content - Oven Dry Method - Part 2: Total Moisture - Simplified Method. Geneva, Switzerland: International Organization for Standardization, 2017.

[6] D. J. Batstone, J. Keller, I. Angelidaki, S. V. Kalyuzhnyi, S. G. Pavlostathis, A. Rozzi, W. T. M. Sanders, H. Siegrist, and V. A. Vavilin, “The IWA anaerobic digestion model no. 1 (ADM1),” Water Science and Technology, vol. 45, no. 10, pp. 65-73, 2002, doi: 10.2166/wst.2002.0292.

[7] P. Weiland, “Biogas production: Current state and perspectives,” Applied Microbiology and Biotechnology, vol. 85, no. 4, pp. 849-860, 2010, doi: 10.1007/s00253-009-2246-7.

[8] KERN & Sohn GmbH, Operating Manual: Electronic Moisture Analyzer MLB 50-3N (Thermogravimetric Moisture Determination), Doc. no. MLB_N-BA-e-1021, 2010.

[9] B. Singh, Z. Szamosi, and Z. Siménfalvi, “Impact of mixing intensity and duration on biogas production in an anaerobic digester: A review,” Critical Reviews in Biotechnology, vol. 40, no. 4, pp. 508-521, 2020, doi: 10.1080/07388551.2020.1731413.

[10] A. Abdellah, M. N. U. Rehman et al., “Influence of mixing on anaerobic digestion efficiency in stirred tank digesters: A review,” Water Research, vol. 143, pp. 503-517, 2018, doi: 10.1016/j.watres.2018.06.065.

Posture Detection and Assistance for Mobility Aid: A Sensor Fusion and Machine Learning Approach

Ivan Omeko, Stefan Twieg and Martin Obert
Department of Electrical, Mechanical and Industrial Engineering, Anhalt University of Applied Sciences, Bernburger Str. 55, 06366 Köthen, Germany
I.Omeko@outlook.com, Stefan.Twieg@hs-anhalt.de, Martin.Obert@hs-anhalt.de

Keywords: K-Means, CNN, Neural Networks, Posture Detection, Elderly Care, Machine Learning, Jetson Orin, MQTT.

Abstract: The increasing proportion of elderly individuals poses significant challenges for mobility assistance and independent living. Rollators are widely used mobility aids, yet conventional devices remain largely passive and provide no feedback on improper usage that may lead to discomfort or musculoskeletal strain. This paper presents the design and evaluation of a smart posture detection system integrated into a commercially available rollator as part of the AktiMuW project. The proposed system combines multiple sensors - including inertial measurement units, ultrasonic distance sensors, and strain gauges - with machine learning techniques to assess rollator usage and user posture in real time. A two-stage classification approach is employed. First, the operational state of the rollator is identified using supervised learning methods. Feed-forward neural networks, convolutional neural networks, and random forest classifiers are evaluated, with the random forest model demonstrating the best balance of accuracy and computational efficiency, achieving over 97% validation accuracy across all device states while significantly reducing inference time and resource usage. Second, user posture is analyzed using unsupervised k-means clustering. Different posture granularities are investigated, ranging from five detailed posture classes to simplified configurations. A three-class posture model ("Comfortable," "Too Close," and "Too Far") is selected as an optimal compromise between classification accuracy and actionable feedback, achieving validation accuracies of up to 99%. The complete system is deployed on an embedded NVIDIA Jetson Orin Nano platform and integrated via MQTT-based communication. Real-time benchmarking confirms that the combined models operate within acceptable computational limits while maintaining reliable posture detection. The presented approach demonstrates the feasibility of lightweight, sensor-based posture monitoring for rollator users and provides a foundation for future assistive feedback systems aimed at improving safety and comfort for elderly individuals.

1 INTRODUCTION

1.1 Motivation

People around the world are living longer due to improvements in quality of life and healthcare. In Germany, the share of people aged 65+ reached 23% in 2024 - 2 percentage points higher than ten years earlier - and by 2050, this figure is expected to rise to 28% [1].

With age, physical abilities decline, and walking often becomes more challenging. Among individuals over the age of 65, 66.5% of those with a disability report significant walking difficulties [2].

To address these public health and technological challenges, many older adults rely on mobility aids such as rollators. Rollators (wheeled walkers) are a primary mobility aid for the elderly, used by a lot of older adults in Germany. However, traditional rollators are passive and limited in functionality.

The AktiMuW project [3] was initiated to bridge this gap by integrating smart technologies into rollators. The project aims to implement assistive features such as obstacle and hazard detection, navigation support, and posture detection to enhance user safety and independence.

There were previous approaches to track posture with walkers, which involved complex setups [4] or relied on additional sensors like cameras [5]. To the authors' knowledge, the approach of combining motion, weight-bearing force, and the person's distance using sensor fusion and machine learning to detect posture in a rollator, which will be discussed later, represents a novelty.

1.2 Problem Statement

When elderly individuals use rollators incorrectly, it can place unnecessary stress on parts of the spine or certain muscle groups, particularly in the back, leading to discomfort or even injury. Ensuring proper posture while using a rollator is essential for maintaining comfort and preventing fatigue.

This paper describes the development of a posture detection system for rollator users. The system is based on data collected from various sensors, including distance measuring sensors, inertial measurement units (IMUs), and strain gauges. These inputs are used to train and deploy machine learning models capable of assessing user posture in real-time and providing corrective feedback.

2 SYSTEM DESCRIPTION

The AktiMuW project adopts a strategy of incorporating smart technologies like voice integration [6] or detection of daily objects in urban traffic into a commercially available rollator without altering its structural integrity. A combination of sensors was selected specifically for posture analysis, including inertial measurement unit (IMU), distance measurement sensors, and strain gauges. Recognizing the complexity and variability of human movement, a machine learning-based approach was selected for interpreting the gathered sensor data. This methodology represents the foundational step toward enabling real-time posture detection. The process of the posture detection is shown in Figure 1.

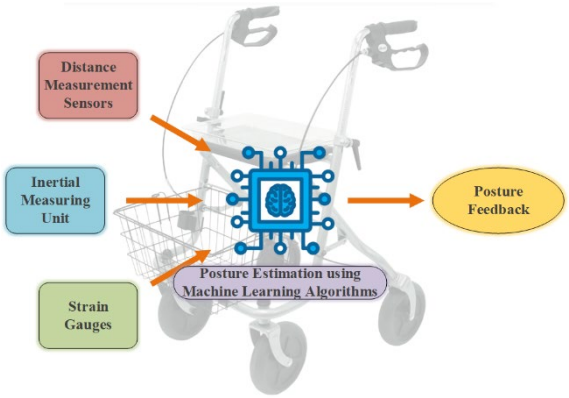


Figure 1: Visualisation of the posture detection in the AktiMuW project.

2.1 Hardware

The project was based upon the Drive Migo 2G rollator. The smart system integration required additional hardware components to enable sensor data collection and processing capabilities. The selected hardware architecture consists of:

- Jetson Orin Nano Developer Kit for machine learning processing;
- Raspberry Pi Zero W for sensor data acquisition;
- MPU6050 IMU that measures rollator acceleration and angular velocity for movement detection;
- HC-SR04-P ultrasonic sensors to monitor proximity between rollator and user's legs for positioning analysis;
- strain gauges for measuring the load applied to rollator handles indicating user grip and weight distribution;
- HX711 amplifiers for amplification and digitization of the strain gauge signals for accurate load measurements.

Raspberry Pi with IMU, amplifiers and other complimentary electronics have been placed on a custom in-house manufactured printed circuit board (PCB). Strain gauges have been mounted to each of the rollator's handles using clear silicone adhesive sealant.

To keep the electronic hardware safe, a custom housing was developed using SolidWorks 2023 and 3D-printed to accommodate the Jetson and the PCB. The housing was placed under the seat of the rollator. Two HC-SR04-P sensors have been attached to the housing via mounting part. Hardware placement is shown in Figure 2.

Additionally, some of the sensors were calibrated prior to use. The MPU6050 IMU's output was adjusted for its static tilt offset. The selection of the primary distance sensor was based on a comparative test of the HC-SR04, HC-SR04-P, and TFmini-S models; the full evaluation is detailed in a separate study [7].

2.2 Software

The Raspberry Pi Zero W serves as the primary data acquisition unit, responsible for sensor data collection and transmission to the Jetson Nano for processing and analysis. The system implements a structured data flow from sensor collection through network transmission to final processing.



Figure 2: Housing with components fitted onto the rollator.

Data transmission utilizes the MQTT (Message Queuing Telemetry Transport) protocol, a lightweight messaging standard optimized for IoT applications with limited bandwidth and processing requirements. The Jetson Nano plays the role of the MQTT broker while simultaneously hosting a dedicated wireless access point. The Raspberry Pi connects to this network as an MQTT client, establishing a reliable communication channel between the data acquisition and processing units.

Sensor data from utilized sensors is merged into a unified data string format to ensure consistent parsing and processing. Individual sensor readings are delimited by semicolons, creating a structured message that maintains data integrity during transmission. All sensor data is published under the standardized MQTT topic "Rollator" to enable organized message routing and subscription management. Upon reception at the MQTT broker, incoming messages are processed according to the predefined data structure. Each message contains sensor readings in the following sequence:

- ultrasonic sensor distance measurements (left and right sensors);
- IMU linear acceleration data (x, y, z axes);
- IMU angular velocity data (x, y, z axes);
- strain gauge load measurements (left and right sensors);
- environmental temperature reading (acquired via IMU sensor).

2.3 Posture Detection Methodology

Accurate posture classification depends on the reliable identification of the rollator's operational state. Therefore, a two-stage classification approach was implemented, where the system first determines the current state of the device before proceeding to analyse the user's posture.

Three distinct device states were defined to represent the primary usage conditions of the rollator system:

- State 0: Device is not used and not moved;
- State 1: Device is used but not moved (user applies pressure on the handles but without movement);
- State 2: Device is used and moved.

This classification provides essential context for subsequent posture detection since posture assessment is only meaningful when the device is actively used and moved (the system is in State 2).

In the context of the posture detection, additional classification classes were defined:

- State 3: Comfortable;
- State 4: Too low, middle;
- State 5: Too low, close;
- State 6: Too high, middle;
- State 7: Too high, far.

The only adjustable parameter on the rollator related to the user's comfort is the height of the handles. Improper handle height can cause the user to adopt compensatory postures: handles positioned too low may lead to excessive forward leaning, while handles set too high may result in backward leaning. However, users may also modify their posture by changing their distance from the rollator. For instance, leaning may be reduced by stepping closer to the rollator when handles are too low or stepping farther away when handles are too high, but the device usage will remain incorrect.

Table 1: Defined postures and their descriptions.

Posture	Description
Comfortable	Handles at 13 cm (hands at pelvic height), user's legs approx. 45 cm from ultrasonic sensors
Too low, middle	Handles at 18 cm on the bar, user's legs approx. 45 cm from ultrasonic sensors
Too low, close	Handles at 18 cm on the bar, user's legs approx. 15-20 cm from ultrasonic sensors
Too high, middle	Handles at 8 cm on the bar, user's legs approx. 45 cm from ultrasonic sensors
Too high, far	Handles at 8 cm on the bar, user's legs approx. 60 cm from ultrasonic sensors

Detailed descriptions of the postures are provided in Table 1. Here, "middle" refers to the distance between the user and the rollator associated with the comfortable posture, while "close" and "far" describe compensatory adjustments in user distance in response to low or high handle positions, respectively. Since the testing was performed by the

author, every configuration was determined according to personal ergonomic preferences.

3 CHOSEN MACHINE LEARNING APPLICATION

For the purpose of the device state classification (States 0-2) the following approaches have been chosen for comparison: feed-forward neural networks (FNN), convolutional neural networks (CNN) and random forests (RF).

FNNs are one of the most fundamental and widely used types of neural networks. In FNNs, information flows in one direction: from input to output, without cycles or loops. During the learning process, FNNs adjust their weights using algorithms such as gradient descent and Widrow-Hoff's learning rule, which minimize the prediction error. A classic example is the perceptron: a simple FNN used for binary classification tasks like face recognition. FNNs are particularly effective for pattern recognition and function approximation tasks due to their ability to generalize. Once trained on a dataset, the network can produce accurate predictions for previously unseen inputs that share structural patterns with the training data [8].

CNNs were specifically designed to address the limitations of FNNs in dealing with high-dimensional data such as images. While they share the same conceptual foundation - layers of neurons with learnable weights - The defining operation in a CNN is convolution. In this process, small filters (also known as kernels) slide over the input image, extracting local patterns such as edges, textures, or shapes [9].

A random forest is an ensemble-based classification algorithm that constructs multiple decision trees and aggregates their predictions through majority voting. Each tree in the forest is trained on a bootstrap sample - a randomly drawn subset of the training data with replacement - and, during training, each split in a tree considers only a random subset of features rather than the full feature set. This dual randomness through both data and feature selection, encourages diversity among trees and reduces overfitting, a common problem with individual decision trees [10].

For the task of the posture detection, the k-means clustering algorithm was selected. In contrast to the previous approaches, k-means is an unsupervised learning method. While supervised models are trained using labeled data (the model knows the target class during training), unsupervised methods like k-means operate on unlabeled data, grouping data points based solely on inherent patterns and similarities.

K-means is a partitional clustering algorithm that organizes a dataset of n samples into k clusters by minimizing the within-cluster variance. Each cluster is represented by a centroid, and the goal of the algorithm is to assign data points to the cluster whose centroid is nearest, using a distance metric such as Euclidean distance [11]. The algorithm proceeds iteratively through two main steps:

1) assignment step: each data point is assigned to the cluster with the closest centroid;
2) update step: the centroids are recalculated as the mean of all data points assigned to each cluster.

These steps repeat until the cluster assignments no longer change significantly, typically indicating convergence [12].

4 MODELS TRAINING

4.1 Device State Classification

To perform the model training, data was collected for each class individually and saved to separate CSV files using a dedicated Python script. Later they were combined into a single dataset.

For the device state classification such features were selected:

- Linear acceleration along the X-axis (Acc_x);
- Linear acceleration along the Y-axis (Acc_y);
- Angular velocity around the Z-axis (Gyro_x);
- Strain gauge readings from both handles (DMS_R, DMS_L).

FNN, CNN and RF models were trained on the same dataset with the same dataset parameters. Results of the conducted trainings are presented in the cross-validation confusion matrixes below. The validation results in Figure 3 are: State 0: 99.03%, State 1: 91.50%, State 2: 94.52%. This indicates that the FNN trained model can detect the correct state of the rollator in the vast majority of cases.

The CNN in Figure 4 produced validation results similar to the FNN, with a modest increase in performance: approximately 5 and 0.3 percentage points higher accuracy at detecting States 1 and 2 respectively.

Among the three classification methods, all achieved validation accuracies exceeding 90%. The RFC in Figure 5 demonstrated the highest overall performance, with validation accuracies of 99.76% for State 0, 98.61% for State 1 and 97.81% for State 2.

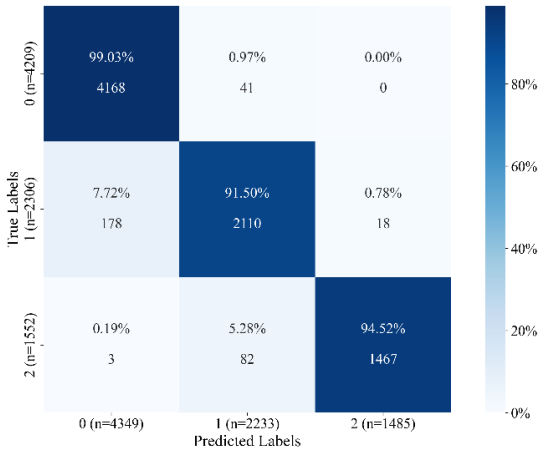


Figure 3: Cross-validation confusion matrix of the trained FNN model.

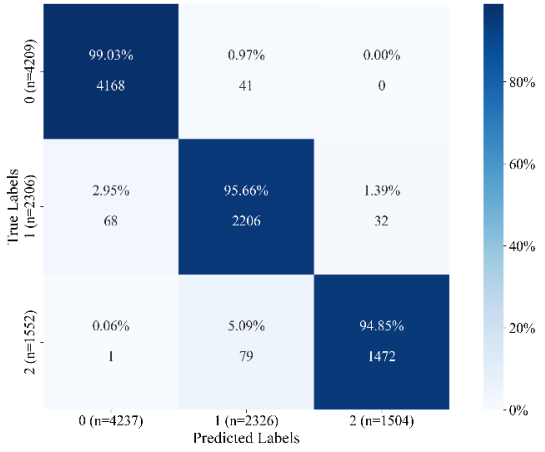


Figure 4: Cross-validation confusion matrix of the trained CNN model.

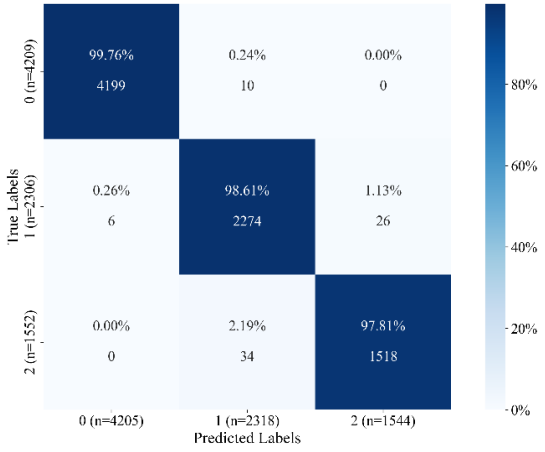


Figure 5: Cross-validation confusion matrix of the trained random forests model.

To evaluate the real-time performance of the trained models and choose the best performing one, dedicated Python scripts were written for each classifier. These scripts executed the models on the Jetson Orin Nano for 300 seconds (5 minutes) under controlled conditions - no additional processes were active beyond essential system services. During the test run, the following performance metrics were recorded:

- Inference time for each prediction;
- RAM usage by the process during inference;
- System-wide CPU usage at 1-second intervals.

As can be observed from Table 2, both FNN and CNN models show similar performances with CNN being slightly more resource efficient, using on average less RAM and CPU. The RFC model, on the other hand, showed a significant advantage in speed, with an average inference time of 38.16 ms, compared to 114.75 ms for FNN and 115.20 ms for CNN. This translates to roughly 67% faster inference compared to both FNN and CNN. Random forests model performs utilizing system resources in much more efficient way: on average it uses around 150 MB of RAM, which is 73% less than both FNN and CNN.

On average, RFC consumed only 5.28% of total CPU load, while FNN and CNN both hovered around 14%: RFC used around 63% less CPU than the neural networks. This lower CPU usage is especially important for maintaining overall system responsiveness on the Jetson Orin Nano.

Even under peak conditions (top 1% and top 0.1% of samples), RFC consistently maintained lower CPU usage: 8.90% with RFC vs. ~18% with FNN/CNN for top 1% CPU usage, 9.30% with RFC vs. ~19% with FNN/CNN for top 0.1% CPU usage.

Overall, the RFC model is the most resource-efficient option across all metrics: it has fastest inference time, lowest RAM consumption and lowest average and peak CPU usage. For this reason, the RFC-trained model was chosen to be used for real-time deployment on the Jetson Orin Nano due to its lightweight and efficient operation.

4.2 K-Means Clustering for Posture Detection

Data collection for posture classification was conducted using a procedure similar to that employed for the initial three-class device state classification. Individual datasets corresponding to each defined posture were recorded separately and save to CSV files. These datasets were subsequently merged into a single training set during the clustering process.

Table 2: Real-time performance metrics of device state classification models.

Metric	FNN	CNN	RFC
Total number of inferences	2100	2042	2030
Average inference time	114.75 ms	115.20 ms	38.16 ms
Average RAM usage	559.04 MB	555.19 MB	153.30 MB
Average CPU usage	14.40%	14.12%	5.28%
Average Top 1% CPU usage	18.36%	18.20%	8.90%
Average Top 0.1% CPU usage	19.20%	18.70%	9.30%

The k-means clustering approach was applied to the dataset in an unsupervised manner using the trained sensor data. Since clustering does not inherently assign semantic meaning to its output, the resulting cluster labels (0 through 4) do not correspond directly to the actual posture classes.

To overcome this, majority voting was employed: each cluster was analyzed to determine which actual posture label appeared most frequently within it. These dominant labels were then mapped to the corresponding cluster indices using a cluster_to_label dictionary, which was saved to disk for future use.

Once training was completed, the trained model was used to predict cluster assignments for the validation set. The predictions were then translated into meaningful posture classes via the mapping dictionary.

Clustering with earlier defined five postures was performed. Cluster formation is shown in Figure 6, using features from ultrasonic sensors and strain gauges. The cross-validation was conducted as well.

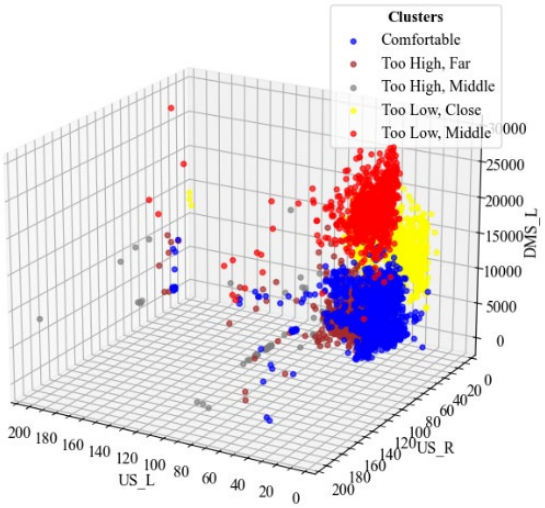


Figure 6: Graphical representation of 5 posture cluster using US_L, US_R and DMS_L.

The confusion matrix (Fig. 7) shows mixed results. Clustering struggled to identify certain postures correctly. Although “Comfortable”, “Too High, Far”, “Too Low, Close” and “Too Low, Middle” were predicted correctly 81% or more, "Too High, Middle" posture was consistently misclassified, most often as "Comfortable" (75%). This suggests overlap in feature space between the two postures, particularly in distance measurements. Although "Too High, Middle" is defined by increased handle load, the difference in strain gauge data appears insufficiently distinct to separate it from the "Comfortable" posture during unsupervised learning.

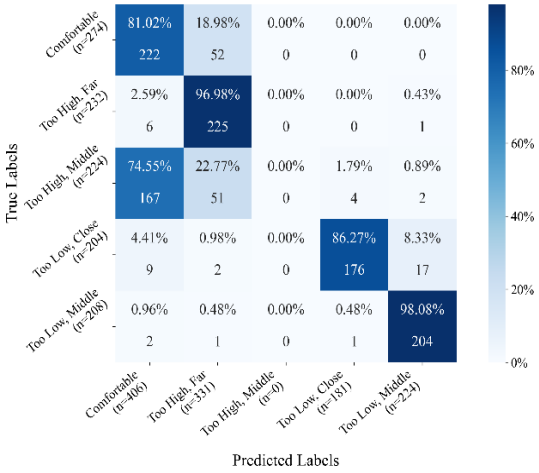


Figure 6: Cross-validation confusion matrix of the performed 5-posture clustering.

To assess whether reducing the number of clusters could improve clarity and accuracy, a second trial was conducted using just two clusters:

- Comfortable;
- Not Comfortable.

“Comfortable” posture remains the same as it was in the first trial. The dataset for “Not Comfortable” includes the data for every specific uncomfortable posture out of four that was established, combined into one. Cluster visualization and the resulting confusion matrix are presented in Figure 8 and Figure 9 respectively. This reduced-class setup showed very high classification accuracy:

"Comfortable" - 86%, "Not Comfortable" - 98%. While these results demonstrate improved performance, a binary distinction between postures is too coarse for the needs of the project. More informative feedback is necessary.

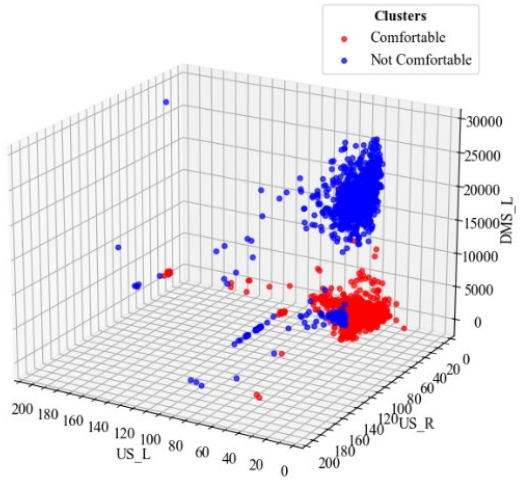


Figure 7: Graphical representation of 2 posture clusters using US_L, US_R and DMS_L.

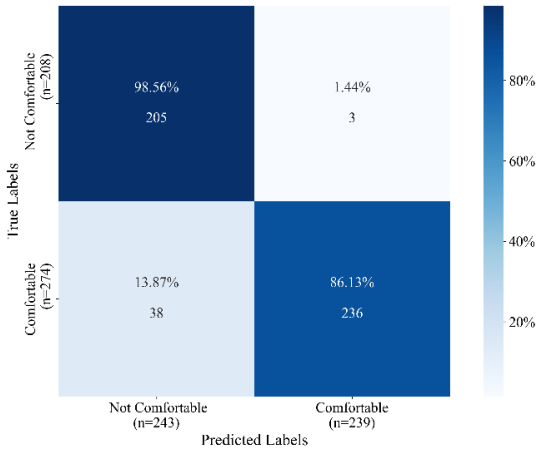


Figure 8: Cross-validation confusion matrix of the performed 2-posture clustering.

For that reason, a third trial was conducted with three postures:

- Comfortable;
- Too Close - Comfortable handle height (13 cm) and short distance (15-20 cm);
- Too Far - Comfortable handle height (13 cm) and long distance (60 cm).

This configuration provides clear, interpretable posture feedback to the user while remaining manageable for the clustering algorithm. Clusters are shown in Figure 10 and validation results in Figure 11.

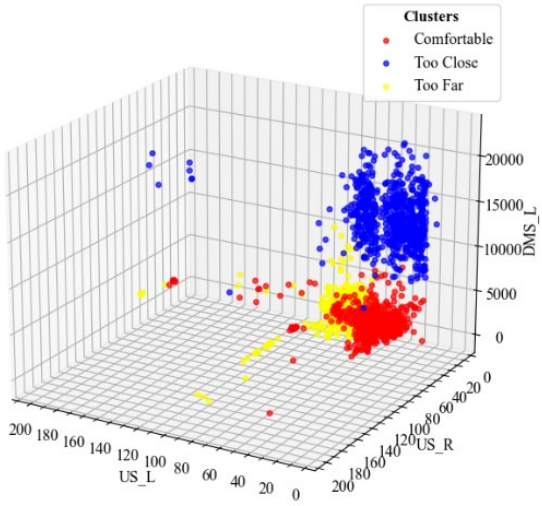


Figure 9: Graphical representation of 3 posture clusters using US_L, US_R and DMS_L.

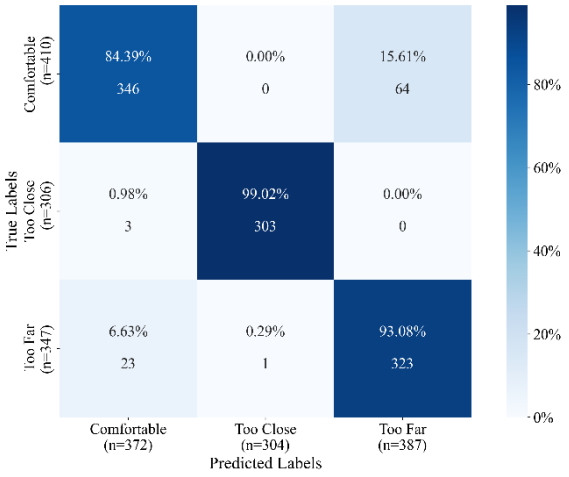


Figure 10: Cross-validation confusion matrix of the performed 3-posture clustering.

Cross-validation confusion matrix shows high accuracy at predicting 3 postures. "Comfortable" posture was correctly predicted 84% of the time, "Too Close" - 99% and "Too Far" - 93%.

The results indicate that while five-class clustering provides detailed classification it struggles with accuracy due to overlapping features among similar postures. Reducing to two clusters improves accuracy but sacrifices granularity. The three-cluster setup stands as a strong compromise, maintaining high classification accuracy while offering meaningful posture differentiation.

This balanced approach enables the system to deliver meaningful, actionable posture feedback in real-time without adding unnecessary computational complexity. For these reasons, the three-posture clustering model was chosen for deployment on the rollator, running on the Jetson Orin Nano.

5 FINAL SETUP

With the models for device state classification and posture detection trained, they were integrated into a unified deployment script for the rollator system. A Python script was developed that uses the random forest classifier to classify the device state (labels 0, 1, or 2). When State 2 is detected, indicating walking, the script triggers the k-means clustering model to identify the user's posture. All necessary components, including the machine learning models and scalers, were installed on the Jetson Orin Nano within the ~/posture_detection/ directory. A dedicated virtual environment was created to ensure safe and isolated execution.

To evaluate the system's resource usage when both models run concurrently, a benchmarking test was conducted, similar to that described in Section 4.1. The results are summarized in Table 3.

Table 3: Real-time performance metrics: RFC only vs. RFC + k-means.

Metric	RFC	RFC + k-means
Total number of inferences	2030	2017
Average inference time	38.16 ms	41.08 ms
Average RAM usage	153.30 MB	155.12 MB
Average CPU usage	5.28%	10.72%
Average Top 1% CPU usage	8.90%	15.75%
Average Top 0.1% CPU usage	9.30%	16.30%

The results show that while the system experiences a moderate increase in load when both models are active, resource usage remains within acceptable limits. The slight increases in inference time (from 38.16 ms to 41.08 ms) and RAM usage (from 153 MB to 155 MB) are minimal. Although average CPU usage doubled, peak usage remained below 17%, indicating that the script operates efficiently and leaves system headroom.

During live posture estimation, predicted posture labels are published to the MQTT topic Rollator/Labeled (Fig. 12), enabling external systems or interfaces to access real-time feedback. The integrated posture detection setup was tested extensively and consistently performed well, accurately identifying postures in the majority of cases. Minor deviations were occasionally observed, likely due to sensor interference or transient signal noise.

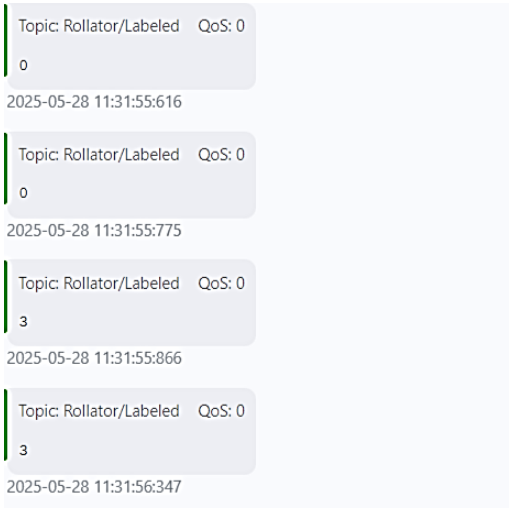


Figure 11: Posture estimated labels published to the MQTT "Rollator/Labeled" topic.

6 CONCLUSIONS

For device state classification (States 0, 1 and 2) feed-forward network, convolutional neural network and random forest approaches have been compared. The RFC model outperformed others in both accuracy and resource efficiency.

For posture detection, k-means clustering approach was used to differentiate between five postures, but for the accuracy it was decided to use the three-posture configuration with "Comfortable", "Too Close" and "Too Far" postures. Finally, both the RFC and k-means trained models were successfully integrated and deployed on the Jetson Orin Nano for the real-time posture detection.

The underlying approach demonstrates that real-time tracking of the body posture while using a rollator is successfully achievable through the use of multiple sensors and their fusion via machine learning.

To further advance the goal of preventing falls and postural damage, future iterations will incorporate a broader dataset to achieve the highest possible level of user compatibility.

ACKNOWLEDGEMENTS

The research project underlying this contribution was funded by the Federal Ministry of Education and Research under the grant number FKZ 03WIR3122A. The responsibility for the content of this publication lies with the author.

REFERENCES

[1] Destatis, "Bevölkerungspyramide: Altersstruktur Deutschlands von 1950 - 2070," [Online]. Available: https://service.destatis.de/bevoelkerungspyramide/index.html#!y=2024&a=20,65&l=en&g.

[2] W. He and L. J. Larsen, Older Americans With a Disability: 2008-2012, Washington, DC: U.S. Government Printing Office, 2014.

[3] Hochschule Anhalt, "AktiMuW," [Online]. Available: https://www.hs-anhalt.de/aktimuw/uebersicht.html.

[4] R. Pérez-Rodríguez, P. Moreno-Sánchez, M. Valdés-Aragonés, M. Oviedo Briones, S. Divan, N. García-Grossocordón, and L. Rodríguez-Mañas, "FriWalk robotic walker: usability, acceptance and UX evaluation after a pilot study in a real environment," Disability and Rehabilitation: Assistive Technology, vol. 15, pp. 1-10, 2019, [Online]. Available: https://doi.org/10.1080/17483107.2019.1617795.

[5] M. Palermo, J. M. Lopes, J. André, A. C. Matias, J. Cerqueira, and C. P. Santos, "A multi-camera and multimodal dataset for posture and gait analysis," Scientific Data, vol. 9, no. 1, 2022, [Online]. Available: https://doi.org/10.1038/s41597-022-01722-7.

[6] S. Rajanayagam, M. A. Ingrisch, P. Müller, P. Jahn, and S. Twieg, "Enhancing Voice Activity Detection for an Elderly-Centric Self-Learning Conversational Robot Partner in Noisy Environments," Proceedings of the International Conference on Applied Innovations in IT, vol. 13, 2025, [Online]. Available: https://doi.org/10.25673/119209.

[7] I. Omeko, S. Twieg, and M. Obert, "Sensor-Based Gait Analysis: A Comparative Study of Ultrasonic and Laser Sensors for Gait Monitoring in Rollator-Assisted Walking," Proceedings of the International Conference on Applied Innovations in IT, vol. 13, pp. 213-222, 2025, [Online]. Available: https://doi.org/10.25673/119236.

[8] I. Aleksander and H. Morton, An Introduction to Neural Computing, London: Chapman and Hall, 1990.

[9] C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[10] L. Breiman, "Random Forests," Machine Learning, vol. 45, pp. 5-32, 2001, [Online]. Available: https://doi.org/10.1023/A:1010933404324.

[11] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification, 2nd ed., Wiley, 2006.

[12] R. Nascimento, A. Neto, Y. Shalom, H. Nascimento, L. Lucena, and J. Freitas, "A new hybrid optimization approach using PSO, Nelder-Mead Simplex and Kmeans clustering algorithms for 1D Full Waveform Inversion," PLOS ONE, vol. 17, 2022, [Online]. Available: https://doi.org/10.1371/journal.pone.0277900.

Impact of Structural Uncertainty on the Time Parameters of the Project Schedule in a Coal Mining Application

Danylo Nazymko
Université Côte d'Azur, Valrose Avenue 28б 06103 Nice, France
danylo.nazymko@etu.univ-cotedazur.fr

Keywords: Project Schedule, Structure, Uncertainty, Stochastic Simulation, Causal Analytics, Expediting.

Abstract: Stochastic Monte Carlo simulation has been used to simulate a mining program schedule in uncertain parametric and structural conditions. Task duration was entered as distributions, reflecting parametric uncertainty. To simulate structural uncertainty, the probability of the task occurrence and the correlation between tasks were considered. The combined effect of these factors was a novel aspect of this research. The probabilities and distributions of the critical path were analyzed as a function of structural and parametric uncertainties. Nonlinear dependences were registered as an effect of the uncertainties. A case study was conducted on an example of a mining program schedule having 107 tasks. This experience confirmed that parametric and structural uncertainties, when combined, can dramatically alter the distribution of the critical path, thereby altering decisions in project management and expediting. Enhancing the sustainability of the coal mining program can be essentially improved by eliminating task correlations and minimizing program structural uncertainties.

1 INTRODUCTION

Despite advances in the development of the theory of project management, more than half of them end up with significant delays in project timelines, budget overruns, and cuts in project scope and objectives [1]. The main reason is the turbulent project environment in which parametric and structural uncertainty arises. Naderpour et al. [2] made a systematic literature review and developed a new model for uncertainty management in the time estimation of construction projects.

As early as 1975, Ang, Chaker, and Abdelnour [3] developed a probabilistic network evaluation technique algorithm. They accounted for mean durations and the correlations among the network paths to evaluate the probability distribution of the project completion time. However, they emphasized parametric uncertainty, leaving in a shadow the structural ambiguity as a separate important factor, which may affect the project.

Mahmoudi and Javed [4] upgraded the review of fast-tracking and crashing techniques that project managers usually deploy to speed up the processes and ensure the delivery of projects on time. These two main techniques are considered separate and relatively independent approaches. However, Nazimko and Zakharova [5] demonstrated that the crashing unavoidably induces and accelerates the fast-tracking effect, increasing the probability of multi-chain critical paths occurrence. This new positive effect exploits the structural uncertainty that was not previously noted.

De Meyer, Loch, and Pich [6] identified four project uncertainty types: variation, foreseen uncertainty, unforeseen uncertainty, and chaos or turbulence. In this paper, the foreseen uncertainty was considered more strictly as the uncertainty in the structural characteristics of a project schedule.

The sensitivity of the project to structural uncertainty may be significantly higher than to parametric uncertainty [5]. Under these circumstances, the management of project structure takes on special importance.

Mining is a typical industry that organizes its activities into projects and even large programs, carried out in uncertain geologic environments.

Task uncertainty may be represented by its probability, which can alter the project schedule structure. Task's interdependence can be expressed through their correlation.

Geological uncertainty can directly affect structural uncertainty, such as task probability.

The elastic and strength parameters of the rock mass significantly affect advance rates and can be treated as stochastic parameters, characterized by a specific distribution within a given range. When determining the longwall face advance rate using the deterministic method, these uncertainties within the rock mass are ignored, making the final result neither accurate nor reliable. Therefore, statistical and probabilistic risk analysis can provide more reliable results.

Increased depths, adverse geological uncertain conditions, elevated variable stress regimes, and weaker ground conditions, coupled with an industry-wide need for higher development rates, continue to make ground control management challenging.

Stochastic modeling is the most powerful tool for studying the structural uncertainty of projects [7].

A stochastic project scheduling model was employed, which takes into account the parametric uncertainty of the project by using any distributions of the tasks' durations as inputs, and the structural uncertainty considering the occurrence probability of scheduling tasks or activities and their correlation [5]. This model was selected because it accounted for multi-chain critical paths and provided an optimal expediting process for a project schedule. All critical paths are equitable items of a project schedule [5]. Furthermore, if the project is deployed in an uncertain environment, the most critical paths may swap places.

Table 1: Input data for the stochastic model.

Task number	Task probability	From node	To node	Duration mode	Minimum	Maximum	Correlated activity	Correlation coefficient
1	2	3	4	5	6	7	8	9
1	1	1	2	20	10	30	0	0
2	1	2	3	20	10	30	0	0
3	1	2	4	20	10	30	0	0
4	1	2	5	20	10	30	0	0
5	1	3	6	20	10	30	0	0
6	1	4	6	20	10	30	0	0
7	1	5	6	20	10	30	0	0
8	1	6	7	20	10	30	0	0

This paper investigates the influence of structural uncertainty on project parameters, addressing a highly relevant problem for understanding the patterns that can shape project management actions in an uncertain environment.

2 DESCRIPTIONS OF THE MODEL (DEMONSTRATIVE EXAMPLE)

Figure 1 shows a calendar graph for a simple, demonstrative project with eight activities. The project graph follows the “activity on the arches” convention, where “boxed numbers” or nodes represent events, while arrows or arcs represent activities or tasks [8]. The “dashed” arrows (Fig.1) represent activities that can have structural uncertainty, as discussed in the following text. Graph nodes were used to express the logical relationships between tasks, which is very effective in handling strongly correlated tasks.

The graph has a sequential-parallel structure to study the influence of uncertainty factors in both the sequential execution of activities and their simultaneous deployment. According to the Monte Carlo-based algorithm, the same schedule is simulated many times (thousands of times) to obtain a representative set of data points, which are subsequently used for analysis. The required number of simulations is determined using the relevant criterion at a predetermined confidence level.

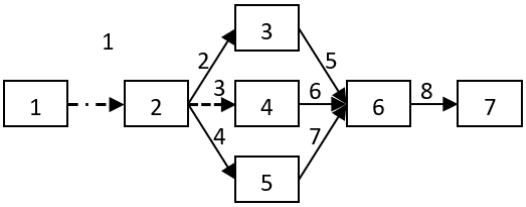


Figure 1: Project schedule graph.

Table 1 presents the input data entered during the stochastic modeling process. Figure 2 shows the histogram of the project completion dates. With a deterministic formulation of the problem and using the modal durations as inputs, the total project duration would be 80 days.

However, the most probable project completion date for the stochastic variant is 85±7.28 days, whereas the most optimistic project completion time varies between 58±8 days. The distributions of task completion times exhibit an approximately symmetric form and align well with the normal distribution, as was proved with the Shapiro–Wilk criterion [9].

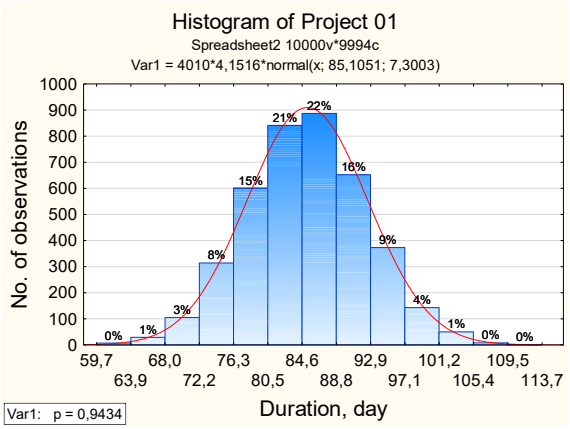


Figure 2: Project duration distribution.

Special experiments have shown that regardless of the shape of the distribution of input parameters, at the output, the experimental data points are approaching a normal distribution, which agrees with the central limit theorem [10] and demonstrates the validity of stochastic modeling results.

A whole set of critical paths has been obtained because there are a large number of possible variants and combinations of initial factors in the process of stochastic modeling. Furthermore, the larger the graph of the project schedule, the greater the number of possible critical paths [5]. Therefore, even for a small representation, seven critical paths were obtained, as shown in Table 2.

Table 2: Critical path frequencies.

Critical path (tasks)	Frequency
1-2-5-8	0.317
1-3-6-8	0.306
1-4-7-8	0.302
1-3-6-8 || 1-4-7-8	0.027
1-2-5-8 || 1-4-7-8	0.022
1-2-5-8 || 1-3-6-8	0.022
1-2-5-8 || 1-3-6-8 || 1-4-7-8	0.003

Since all initial deadlines were set equally, the main critical paths ran through the chains of tasks 1-2-5-8, 1-3-6-8, and 1-4-7-8, and their probabilities were approximately 0.3. Due to the probabilistic nature of usual processes, reflected in the stochastic model as distributions of input factors, parallel critical paths emerged from random arrangements. These paths are rarely encountered in the deterministic formulation of the problem. Thus, with probability 0.022-0.027, the critical path can pass simultaneously through all possible combinations of pairs of nodes 3 and 4, 3 and 5, or 4 and 5. The critical path passes through all three parallel nodes 3, 4, and 5 with a probability of 0.003. This regularity was documented by Nazimko and Zakharova [5], who proved the importance of taking into account multi-chain critical paths.

3 INVESTIGATIONS OF THE STRUCTURAL UNCERTAINTY

3.1 Task Probability Effect

The influence of the schedule structure on the uncertainty in the project parameters of the base model was studied. For example, the probability of task 3 passing through nodes 2 and 4 was set to 0.5. This means that in about 2000 cases out of 4000 iterations, task 3 will be absent (as indicated in Fig. 1 by a dashed arrow), and the graph structure will change. It turned out that the overall project completion dates have not changed much despite the aleatoric presence of task 3 in the calendar. Thus, the mode of the project end date is 84±7.4 days. According to the Kolmogorov-Smirnov criterion, the difference in terms compared to the full graph of the project was insignificant.

Table 3: Frequencies of critical paths.

Critical path (tasks)	Frequency
1-4-7-8	0.397
1-2-5-8	0.384
1-3-6-8	0.155
1-2-5-8 || 1-4-7-8	0.038
1-2-5-8 || 1-3-6-8	0.013
1-3-6-8 || 1-4-7-8	0.010
1-2-5-8 || 1-3-6-8 || 1-4-7-8	0.002

However, a significant difference was observed in the probabilities of critical paths (Table 3). Thus, Table 3 shows that the frequency of the critical path 1-3-6-8 has decreased by about half, while the frequencies of the two remaining parallel paths, 1-2-5-8 and 1-4-7-8, have noticeably increased. Notably, all other less probable paths that include node 4, through which task 3 passed, significantly decreased their probability to about 1%, while the path 1-2-5-8 || 1-4-7-8, which does not have this node, significantly increased the frequency of its occurrence to 0.038. The reason for this effect is that the function of the temporarily absent task 3 was taken over by tasks 2 and 4. Consequently, it was possible to conclude that reducing the probability of tasks on parallel chains in the project schedule graph does not affect the overall value of the project execution time. This only reduces

the probability of passing through the critical path when there is a task with a decreasing probability along the chain.

It should be assumed that removing a task from the sequential chain will have a different effect. To test this assumption, the probability of 0.5 was now given to task 1 (marked by the dashed-dotted vector in Fig. 1). The other tasks had a probability equal to one. The distribution of critical paths and their probabilities was approximately the same as in the first variant since the symmetry of the graph was not broken in this variant. Therefore, the span of the whole project has significantly decreased, namely, the mode of the project execution time became 75.3±11.5 days, or reduced by 10 days. At the same time, the project span has increased by 1.6 times. Thus, the decrease in the probability of the schedule's tasks on its consecutive chains diminishes the project's average span but increases its variation.

3.2 Effect of Tasks Correlation

One of the most important characteristics of the project schedule structure is the mutual dependence between activities, which is another structure parameter. Table 4 shows the results of the stochastic simulation for a symmetric correlation between the durations of activities 2 and 4. Both positive and negative correlations were investigated with a correlation coefficient of ±0.8.

It turned out that the effect of task interdependence does not depend on the sign of the correlation. In addition, the project completion time and its variation do not change, regardless of whether there is a correlation between the activities. The effect of mutual dependence between activities affects primarily the probabilities of the critical paths to which the correlated activities belong. This is clearly demonstrated in Table 4, where the critical path 1-2-5-8, containing task 2, and the path 1-4-7-8, containing task 4, significantly increased their probability to 0.33-0.36, while the path 1-3-6-8, not containing correlated tasks, significantly decreased its frequency of occurrence. The same can be said of paths with low probability.

Table 4: Critical path frequencies.

Critical path (tasks)	Frequency
1-4-7-8	0.355
1-2-5-8	0.338
1-3-6-8 || 1-4-7-8	0.038
1-3-6-8	0.230
1-2-5-8 || 1-3-6-8	0.023
1-2-5-8 || 1-4-7-8	0.018

4 CASE STUDY

4.1 Description of a Mining Development Program

The variants of the year program schedule were analyzed in the case of mining development in a large Ukrainian coal mine. This mine produces coking coal, which is an important source for Ukrainian steelmaking and defense. The productive seam is mined at depths of 950-1100 meters under difficult mining and geological conditions. The coal seam is extracted using longwall technology, with full caving of the roof as the ground control. Methane content in the coal seam ranges from 8 to 15 m^3/ton, which impedes longwall ventilation and necessitates maintaining sustainable underground stability of the roadways.

The initial data file, which digitally outlines the program of 107 real underground mining activities, is shown in [11]. The duration of tasks in this case was expressed in weeks. The longwall faces were planned to provide 6,840 thousand tons of raw coal during the planning period. Parametric uncertainty was accounted for by using a normal distribution for task durations. Structural uncertainty was described by the probability of tasks and their correlations.

The basic mining tasks were considered, namely longwall extraction (14 tasks) and development of the reserves or roadways and entries driving (93 tasks). Other auxiliary tasks, such as construction, transportation, ventilation, and power supply, were outside the scope of this project. In fact, it was a megaproject or program, because even extracting a panel or driving a 3-kilometer-long roadway is a significant project.

The process of modeling the annual mining development program has been carried out iteratively. To determine the program schedule's critical path, 5,000 iterations were used. In each iteration, the computer selected random values for the task rates based on predetermined distribution laws derived from previous mining activity experience.

At each iteration, the critical path was determined, and all characteristics of each task were stored in a special file: its duration, earliest and latest start and end dates, and free time reserve. During the iterations, critical paths were tracked, and their occurrences were counted. At the end of the simulation, the frequencies of all critical paths were calculated, reflecting their probabilities with a predetermined level of reliability. The more iterations, the closer the frequency aligns with the actual probability.

4.2 Investigation of the Tasks' Probability and Their Correlation Effect

Three variants of the mining development program were simulated. In the first variant, the probability of all activities was equal to one, and there was no dependence between activities [11]. In the second variant, the probability of task 88 (1st south entry, block 10, duration mode=40 weeks) was set at 0.5.

The technologies used to prepare the panel for coal extraction can vary significantly depending on whether a low-amplitude transient disturbance occurs. This is precisely the situation with the development of a panel that consists of tasks 88-89 (1stand 2nd south entry block 10, duration mode=40+8=48 weeks). Therefore, the structural uncertainty of the schedule was accounted for with a probability of 0.5 in the second version of the model. This means that if a low-amplitude fault is present and cannot be passed by the longwall, a pillar should be left around the fault, and two separate longwalls will be required. Consequently, two startup rooms and the first south entry (task 88) will be needed. However, if the low-amplitude fault is not problematic and can be bypassed, the panel will be extracted using a single longwall.

Often, the activities depend on each other. This is especially true in underground coal mining. A typical example is when two underground roadways share a single source of fresh air for ventilation during their development. This is exactly the situation that occurred during the development of two roadways, tasks 88-89 and 90 (north ventilation roadway, horizon 708 m, duration mode=48 weeks). The extraction of coal seams at a depth of 1000 meters is accompanied by an intensive release of explosive methane. Therefore, its concentration should be maintained at a safe level, not exceeding 1%. Hence, both ventilation roadways and the 1st south entry draw fresh air from the same source: the roadway that supplies it. Thus, there was an inverse relationship between the consumption of fresh air: the more air consumed by one production unit, the less air is left for the other, and vice versa. Accordingly, the rate of development of the roadways changed in proportion to the consumption of fresh air. This inverse-proportionality dependence is taken into account by the correlation between the rates of 88-89 and 90 task realizations. The correlation coefficient was -0.95. This difference characterized the third variant of the schedule model.

The finish of the last program variant ranged from 45 to 135 weeks, with the most likely being 77 weeks. Table 5 lists the most likely critical paths for all program schedule variants. The program contains both sequential and parallel paths.

A total of 34 critical paths were detected, both sequential and parallel. Only activities with a frequency or probability greater than 0.05 were selected. In the first variant of the program schedule, the total frequency of six such significant critical paths was 0.596, with the maximum frequency of path 66-67 (duration mode = 41 + 7 = 48 weeks) being 0.163.

In the second option, the number of critical paths increased to eight, with the most likely critical path remaining 66-67. The probability of path 88-89 decreased by about half, which corresponds to a structural probability coefficient of 0.5 [11]. Critical path 88-89 gave way to paths 41-42-43 (duration mode = 22 + 14 + 12 = 48 weeks) and 90, which made it possible to enter the number of significant critical paths. It is noteworthy that the total probability of significant critical paths increased to 0.662 or 11% compared to the first version of the program schedule.

However, the most significant changes occurred in the third variant. In this case, the correlation of pathways 88-89 and 90 led to a change in leaders, as critical pathway 88-89 became the most likely. At the same time, the dependent critical path 90 became a component of the significant ones, and as a result, it took second place in the list. It should be noted that in the first version, it was completely absent, and in the second version, it barely exceeded the level of 0.05.

Now its frequency has increased four times and amounts to 0.248. These two pathways removed almost all critical pathways with probability less than 0.1 from the list, and the former leader, 66-67, barely managed to stay on the list.

Even though the number of critical paths in the third variant is minimal, their total probability has increased to the highest level, up 33% compared to the first variant. Thus, a nonlinear dependence of the program schedule structure on parameters such as path probability and their correlation was observed. Not only did the probability of critical paths change, but also their positions in the list of more significant critical paths. Both the decrease in task probabilities and the increase in their correlation negatively affect the project schedule, reducing its balance and increasing differences among the critical paths' probabilities.

Table 5: The most likely critical paths.

1st variant		2nd variant (Probability of 88=0.5)		3rd variant (Correlation 88-89/90)	
CP	Probability	CP	Probability	CP	Probability
66-67	0.163	66-67	0.169	88-89	0.423
88-89	0.159	68	0.085	90	0.248
27-28-29	0.078	76-77	0.083	85-86-87	0.071
68	0.076	88-89	0.083	66-67	0.053
76-77	0.067	27-28-29	0.08		
46-47-48-49	0.053	46-47-48-49	0.057		
		41-42-43	0.053		
		90	0.052		
Total	0.596		0.662		0.795

According to [5], the greater the number of multiple critical paths and the closer their probabilities, the more balanced the project schedule becomes. This point was further elaborated by a new finding: the likelihood that tasks take less than one unit of time and the correlations between tasks are the most influential factors contributing to project delay risks and complicating the expediting process. Besides, it was shown that decreasing the probability of tasks in parallel chains does not affect overall project duration, but it lowers the chance that those chains become critical paths. In addition, reducing the probability of activities in sequential chains shortens project duration but increases variability.

Another new finding showed that the effect of task interdependence on project parameters is independent of the sign of the correlation. Project duration and variability also remain unchanged regardless of whether tasks are correlated. However, correlations between tasks increase the probability of critical paths in the affected chains. Changes were observed in both the probabilities of critical paths and their ranking among the significant critical paths.

From a practical standpoint, it is important to avoid combining unlikely tasks with correlated tasks.

4.3 Effect of Structural Uncertainty on Expediting Process

The deployment of expediting is a common measure to ensure a program meets its due date or at least to reduce delays. The mining program was conducted for two scenarios: one without structural uncertainty and one with structural uncertainty. According to [8], [5], the process of crushing from the most critical task 66 located on the most critical path 66-67 (refer to Table 5, 1st variant) was started. The reduction of task duration was carried out in small increments, with each cut being no more than one week of the task duration. Then a new set of critical paths was found, and expediting iteration was repeated. Such an iterative approach ensured an optimal path of the expediting evolution because this irreversible process was path-dependent.

Table 6 demonstrates essential differences in the list of critical paths during the program's expediting. Four iterations have been needed to finish the expediting process for 1st variant and minimize the critical path frequency to 0.1. The further expediting made no sense because any of the critical paths 77-76, 121-28-29, or 46-47-48-49 showed approximately the same probability. On the contrary, twice as many iterations were needed to accomplish expediting the program with correlated tasks. The process of compressing the critical paths began with path 88-89, where tasks 88 and 90 were correlated. Task 88 was modified in the first iteration, changing the most probable critical path to 90.

This path prevailed from the second iteration until the seventh. The frequency of the critical task 90 was reducing slowly. At last, it approximately matched the frequency of tasks' sequence 88-89 (0.16≈0.10).

Table 6: Critical paths evolution during expediting.

	No correlation							
Iteration	1	2	3	4	5	6		
Critical path	66-67	88-89	88-89	77-76	121-28-29	46-47-48-49		
Frequency	0.163	0.208	0.156	0.1	0.098	0.096		
		Correlated tasks						
Iteration	1	2	3	4	5	6	7	8
Critical path	88-89	90	90	90	90	90	90	66-67
Frequency	0.423	0.454	0.458	0.386	0.312	0.206	0.176	0.16(0.11)

Further expediting was not expedient because several critical paths leveled their frequencies [5] (Fig. 3). The curves in Figure 3 represent the frequency of critical paths, which consisted of different tasks and replaced each other during the expediting processes (refer to Table 6).

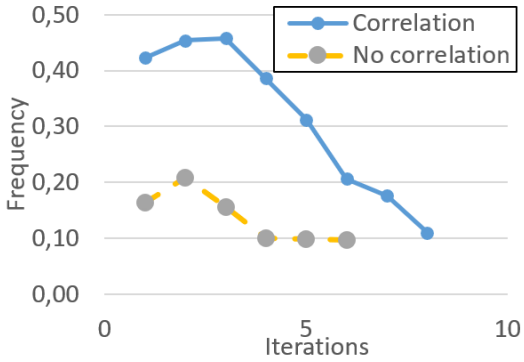


Figure 3: Frequency of critical paths variation.

This experiment has shown that structural uncertainty may essentially delay the process of expediting the mining development program.

To reduce the risk of program delays, the following two measures have been proposed to control the program schedule's structural uncertainty.

Firstly, operators should apply microseismic tomography to clarify low-amplitude faults in the coal seam [12], [13]. This will reduce the geological uncertainty and increase the probability of an underground roadway driving in a disturbed coal seam. The risk of economic losses from the loss of a mechanized complex is three orders of magnitude higher than the cost of seismic tomography.

Secondly, operators should provide an independent ventilation source for tasks 88 and 90 to eliminate the correlation between the two roadways.

As the project progresses, discrepancies between the plan and reality emerge, often requiring the originally single critical path to be crashed [5]. Expediting is frequently done in haste, leading to impulsive, ill-considered decisions. This common issue degrades project quality, causes engineering stagnation, and may even lead to safety violations, particularly in the mining industry.

Our model enables agile assessment of the project schedule. If the schedule has a single critical path, it should be redesigned by restructuring it to split some chains into parallel sequences. When multiple critical paths exist, their frequency should be analyzed. If the variance in critical path frequencies exceeds a reasonable threshold (e.g., 20–30%), iterative crashing should be applied as described earlier [5].

Throughout project execution, periodic schedule assessments should be performed using a combination of earned value management protocols [14] and the results of this investigation. The earned value management detects early signs of schedule delays, while the revealed negative effect of correlation facilitates the search for the optimal strategy for project acceleration.

5 CONCLUSIONS

Stochastic modeling on project networks is a rather effective tool for investigating the structural uncertainty of a project that is implemented in a turbulent environment. The following conclusions can be drawn:

1) The probability of tasks being less than one and the correlation between tasks are the most influential factors that cause the risk of project delays and complicate the expediting process.
2) Decreasing the probability of tasks being in parallel chains of the project schedule structure does not affect the overall project delivery time, but only decreases the probability of a critical path through a chain where there is a task that has a decreasing probability.
3) Reducing the probability of activities on sequential chains diminishes the project execution time but increases its variation.
4) The effect of the mutual dependence of tasks on project parameters does not depend on the sign of correlation. The project span and its variation also do not change, regardless of whether there is a correlation between the tasks. However, the probability of critical paths increases where correlated tasks are located.
5) Changes were observed both in the probabilities of the critical paths and in their positions within the list of significant critical paths.
6) When developing the program schedule structure, one should be cautious and possibly avoid combining unlikely tasks and their correlation.

ACKNOWLEDGMENTS

The author would like to thank the anonymous reviewer for helpful suggestions on revising the section concerning the fundamentals of probability theory and for highlighting the novelty of the results.

FUNDING

The National Academy of Sciences of Ukraine, grant 0120U100081, and the Ministry of Education and Science, Ukraine, grants 0121U110527 and 0123U100356 supported this work.

REFERENCES

[1] P. L. Guida and G. Sacco, "A method for project schedule delay analysis," Computers and Industrial Engineering, vol. 128, pp. 346-357, Feb. 2019, [Online]. Available: https://doi.org/10.1016/j.cie.2018.12.046.

[2] A. Naderpour, J. Majrouhi Sardroud, M. Mofid, Y. Xenidis and T. Pour Rostam, "Uncertainty Management in Time Estimation of Construction Projects: A Systematic Literature Review and New Model Development," Scientia Iranica, vol. 0, no. 0, pp. 0-0, Dec. 2017, [Online]. Available: https://doi.org/10.24200/sci.2017.4605.

[3] H.-S. Ang, A. A. Chaker and J. Abdelnour, "Analysis of Activity Networks under Uncertainty," Journal of the Engineering Mechanics Division, vol. 101, no. 4, pp. 373-387, Aug. 1975, [Online]. Available: https://doi.org/10.1061/JMCEA3.0002028.

[4] Mahmoudi and S. A. Javed, "Project scheduling by incorporating potential quality loss cost in time-cost tradeoff problems," Journal of Modelling in Management, vol. 15, no. 3, pp. 1187-1204, Feb. 2020, [Online]. Available: https://doi.org/10.1108/JM2-12-2018-0208.

[5] V. V. Nazimko and L. M. Zakharova, "Project Schedule Expediting under Structural and Parametric Uncertainty," Engineering Management Journal, vol. 35, no. 1, pp. 29-49, Jan. 2023, [Online]. Available: https://doi.org/10.1080/10429247.2022.2030179.

[6] A. De Meyer, C. H. Loch and M. T. Pich, "Managing Project Uncertainty: From Variation To Chaos," IEEE Engineering Management Review, vol. 30, no. 3, pp. 91-91, 2002, [Online]. Available: https://doi.org/10.1109/EMR.2002.1032403.

[7] B. Black, R. Ainslie, T. Dokka and C. Kirkbride, "Distributionally robust resource planning under binomial demand intakes," European Journal of Operational Research, vol. 306, no. 1, pp. 227-242, Apr. 2023, [Online]. Available: https://doi.org/10.1016/j.ejor.2022.08.019.

[8] A Guide to the Project Management Body of Knowledge (PMBOK® Guide), 6th ed. Project Management Institute, 2017.

[9] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," Biometrika, vol. 52, nos. 3-4, pp. 591-611, 1965, [Online]. Available: https://doi.org/10.1093/biomet/52.3-4.591.

[10] D. C. Montgomery and G. C. Runger, Applied Statistics and Probability for Engineers. Hoboken, NJ, USA: John Wiley & Sons, 2010, [Online]. Available: https://books.google.fr/books?id=_f4KrEcNAfEC.

[11] D. Nazymko, "Results of stochastic simulation of projects' schedules," Mendeley Data, vol. 1, 2023, [Online]. Available: https://doi.org/10.17632/77n5v9tgzp.1.

[12] H. Barthwal, F. J. Calixto and M. van der Baan, "Attenuation tomography using recorded microseismicity in a mine," in SEG Technical Program Expanded Abstracts 2019, Society of Exploration Geophysicists, Aug. 2019, pp. 3011-3015, [Online]. Available: https://doi.org/10.1190/segam2019-3213938.1.

[13] C.-P. Lu, G.-J. Liu, N. Zhang, T.-B. Zhao and Y. Liu, "Inversion of stress field evolution consisting of static and dynamic stresses by microseismic velocity tomography," International Journal of Rock Mechanics and Mining Sciences, vol. 87, pp. 8-22, Sep. 2016, [Online]. Available: https://doi.org/10.1016/j.ijrmms.2016.05.008.

[14] H. L. Chen, "Early Prediction of Project Duration: A Longitudinal Study," Engineering Management Journal, vol. 30, no. 3, pp. 191-202, Jul. 2018, [Online]. Available: https://doi.org/10.1080/10429247.2018.1483169.

Enhancing Web Test Automation Through Screenshot and Video-Based Execution Reporting

Oleksii Cherkashyn
Blynk Technologies Inc., Brickell Avenue 951, 33131 Miami, FL, USA
oleksii.sciencepapers@gmail.com

Keywords: Video Reporter, Screenshot Reporter, Web Application Testing, Web Test Automation, WebdriverIO, JavaScript.

Abstract: Automated web testing increasingly relies on comprehensive reporting mechanisms to facilitate bug reproduction and analysis. This study focuses on enhancing test automation through screenshot and video-based execution reporting using the WebdriverIO framework. The built-in method saveScreenshot was employed for capturing screenshots at each step of the test execution, while the Video Reporter library was used to record videos for failed tests. The study demonstrates that both GUI and headless modes of the Chrome browser do not affect the quality of screenshots and videos. Captured screenshots have an average size of approximately 61.5 KB to 136.9 KB, and videos range from 18.4 KB to 71.7 KB, making them suitable for execution on low-budget or resource-limited remote servers. Furthermore, with Jenkins supporting PNG screenshots and WEBM video playback, visual analysis can be performed directly within the CI/CD interface. The results confirm that the visual quality of both screenshots and video recordings is sufficient to identify the exact point of test failure and to reproduce the corresponding bugs, providing practical guidance for improving test reporting and debugging in automated web testing environments.

1 INTRODUCTION

Web applications play a crucial role in modern companies and organizations. They enhance user engagement by providing simple and accessible ways for interaction, allowing businesses to communicate effectively with their users. Given their importance, thorough and continuous testing is essential to ensure high quality. However, this process introduces several challenges. As a result, testing web applications is generally more complex than testing other types of software, requiring efficient tools and well-designed methodologies [1].

Manual testing is generally considered less reliable compared to automated testing carried out with specialized tools and scripts. In addition, it tends to require more time and effort, whereas automated approaches enable faster execution of test cases. As a result, automation significantly improves testing speed and efficiency. Its implementation allows developers to detect defects more quickly and streamline the overall testing process. Moreover, automated testing supports effective regression testing by identifying issues that may arise after changes are introduced into the codebase [2].

A critically important aspect of automated web testing is the reporting of test results, particularly in real time. Real-time insights enable teams to identify patterns, detect anomalies, and highlight areas that require improvement [3]. For additional analysis of failed tests, screenshot and video reporting tools can be used, as they provide a faster and more intuitive way to identify the location and root cause of errors.

In the context of this study, the WebdriverIO framework [4] is utilized as a robust and reliable solution for automated testing. WebdriverIO, a Node.js-based framework available via npm [5], provides flexibility and seamless compatibility with a wide variety of testing tools and environments. The official version of the WebdriverIO video reporter (wdio-video-reporter) [6] is used in the test framework to record videos of test executions. In addition, standard WebdriverIO methods are applied to capture screenshots on test failures, ensuring that both video and image artifacts are available for debugging and reporting purposes. This approach allows detailed visualization of test results and facilitates faster identification of issues.

This paper contributes to the practices of test analysis and debugging by utilizing screenshots and

video reporting for failed tests, implemented using the WebdriverIO library.

The research question addressed in this study is whether screenshot and video-based execution reporting can improve the efficiency of failure analysis and bug reproduction in automated web testing while remaining suitable for execution in resource-constrained CI/CD environments. In particular, the study investigates the feasibility of integrating lightweight visual reporting mechanisms into automated testing pipelines without significantly increasing storage or computational requirements.

The novelty of this work lies not in the introduction of a new reporting tool, but in the experimental evaluation and practical integration of screenshot and video-based execution reporting within the WebdriverIO framework under different browser execution modes. Unlike existing studies that primarily describe reporting mechanisms conceptually, this paper provides a quantitative analysis of artifact sizes, evaluates compatibility with Jenkins CI/CD environments, and demonstrates that lightweight visual artifacts generated in both GUI and headless browser modes remain sufficient for effective debugging and failure reproduction. Therefore, the study contributes practical evidence regarding the applicability of visual execution reporting for low-budget and resource-limited automated testing infrastructures.

2 RELEVANCE

Reproducing bugs depends on the information contained in bug reports to replicate the issues described. Functionally, a bug report generally consists of key elements such as Steps to Reproduce (S2R), Expected Behavior (EB), and Observed Behavior (OB). Regarding media types, reports may include textual descriptions, images, and videos or GIFs. Images refer specifically to UI screenshots or any static visual content provided by users that can aid the bug reproduction process [7].

In their study, Mallipeddi, Yaqoob, Khan, Mehmood, Mylonas, and Pitropakis describe how testers manually follow test plans, execute test cases, and capture evidence, such as screenshots, comparing actual results with expected outcomes to identify discrepancies and unexpected behavior [8].

Based on the analysis, the following conclusions can be drawn:

- Screenshots and videos of failed tests assist in manually reproducing bugs;
- They provide a visual understanding of differences between previous and current builds;
- Unlike stack traces or textual descriptions of potential errors, visual evidence offers a faster and more intuitive way to locate bugs, thereby saving time.

Thus, the topic of reproducing failed automated tests is relevant, highlighting the need to study and improve existing practices related to screenshot and video reporters.

3 ANALYSIS OF CURRENT RESEARCH

Overall, there is a lack of scientific research focused on the use of screenshot or video reporters for failed automated tests specifically with the WebdriverIO library. Therefore, studies related to other frameworks and existing practices were examined.

In their study, Mon and Panczyk note that Cypress stands out for its simplicity and speed, as well as its built-in screenshot and video capture capabilities, which simplify the testing process. Playwright, on the other hand, is a modern tool offering advanced features, stability, and multi-browser support. Selenium supports the largest number of programming languages but lacks native screenshot and video capture functionality. The authors also provide a table indicating that both Cypress and Playwright support screenshot and video reporters. However, their study does not include an analysis of how these reporters are actually used in practice [9].

In their research, Zhao, Talebipour, Baral, Park, Yee, Khan, Brun, Medvidovic, and Moran introduce *Avgust*, an automated approach that uses video recordings of app executions to generate usage-based UI test cases. They propose processing these videos with neural image-understanding models to automatically infer user interaction sequences from visual state changes captured in screenshots and frames. This enables the synthesis of an app-agnostic model of user behavior, which can then be used to generate new test cases for unseen applications. Unlike traditional test automation that focuses on code coverage or crash detection, their method leverages visual artifacts from videos as a primary source for understanding real usage patterns. The authors show that a significant portion of the tests Avgust generates successfully exercise desired behaviors, demonstrating the effectiveness of using visual evidence (screenshots/video) to construct meaningful tests. Overall, their work highlights how

visual reporters of UI interactions can be used not just for evidence, but as input data to drive automated test creation [10].

Wang, Li, Wang, Menzies, and Wang in their study propose leveraging screenshots in bug reports to improve duplicate bug report detection. They suggest extracting and analyzing visual features from screenshots and combining them with textual information to better identify similar issues. Their approach demonstrates that visual data provides complementary context that is often missing in text-only descriptions, leading to improved detection accuracy. From a web automation perspective, this implies that screenshots captured during test execution can be utilized not only for reporting but also for automated failure analysis. Overall, their work highlights the potential of integrating screenshot reporters with intelligent techniques to enhance bug triaging and reduce redundancy [11].

Ashrafi, Bouktif, and Mediani in their study propose enhancing bug reports by leveraging screenshots as a primary source of information. They suggest automatically extracting meaningful visual features from screenshots to better represent the state of the user interface at the moment of failure. Their approach aims to complement or clarify textual descriptions, reducing ambiguity and improving bug understanding. From a web automation perspective, this implies that screenshots generated during test execution can be used not only for reporting but also for automated analysis and interpretation of failures. Overall, their work highlights the potential of transforming screenshot reporters into intelligent tools for improving bug reproducibility and diagnosis [12].

Existing research and practices primarily focus on using screenshots and video reporters from failed tests as visual tools for reproducing potential bugs, debugging, and analyzing issues or changes in the platform after development. This study extends and complements these practices by applying them specifically within the WebdriverIO framework.

Unlike prior research that mainly investigates screenshots and videos as auxiliary debugging evidence or as datasets for intelligent processing techniques, this study introduces a resource-oriented evaluation of visual execution reporting within WebdriverIO-based automated testing. The scientific contribution of the work lies in demonstrating that lightweight screenshot and video artifacts generated during automated test execution maintain sufficient diagnostic quality for failure analysis and bug reproduction, even in headless execution environments and resource-constrained CI/CD systems. Additionally, the study provides quantitative evidence regarding artifact sizes and CI/CD compatibility, which is largely absent in existing literature. This shifts the focus from conceptual usage of visual reporting toward experimentally validated applicability in practical industrial automation infrastructures.

4 SOFTWARE CONFIGURATION

4.1 Test Infrastructure Configuration

The study was conducted using Node.js version 20.19.4 in combination with WebdriverIO version 8 for automated testing. Mocha was employed as the test runner, with test scripts written in JavaScript [13]. Mocha offers a flexible and adaptable framework suitable for a variety of testing strategies, including support for asynchronous testing, which is crucial for modern web applications that rely on dynamic content [13]. Chromedriver is used to control the Chrome browser during automated testing [14]. It serves as the WebDriver implementation that translates test commands into browser actions, enabling the execution of test scripts and interaction with web elements. Using chromedriver ensures that the tests run reliably on the Chrome browser and allows integration with the WebdriverIO framework for end-to-end automation. A complete list of dependencies and libraries is provided in the package.json file, included in Appendix (Listing A1).

4.2 Screenshot and Video Reporter Configuration

In general, the key settings for screenshot and video reporters are defined in the framework's configuration file, wdio.conf.js. For screenshots, the built-in framework method saveScreenshot is sufficient. To facilitate storage and later analysis, each screenshot can be named using the exact date and time. This process can be assisted by libraries such as Moment.js [15]. The main configuration of the reporters in the framework is provided in Appendix (Listing A2). In the video reporter, the videoSlowdownMultiplier option was set to 10, which analysis showed to be a sufficient value. This setting is used to adjust the playback speed of the recorded videos - higher values produce slower videos, while lower values produce faster videos. Acceptable values range from 1 to 100.

5 RESULTS AND DISCUSSION

In this study, the latest stable version of the Chrome browser, Version 145.0.7632.117 (Official Build) (64-bit), was used, along with the latest chromedriver 146.0.0. Automated tests were executed on a local Windows 10 machine as well as on a remote Ubuntu server. The tests were conducted on three different React web applications and organized into three separate test packs containing 200, 100, and 50 tests, respectively. The browser was configured with a resolution of 1920 × 1080 and launched with the start maximized option.

The tests were executed in both GUI and headless modes of the browser. The study showed that the screenshot and video reporters do not affect the mode, producing outputs of the same quality and resolution in both cases.

Regarding screenshots taken during the tests, in the case of a failed test, the screenshot accurately captured the exact point at which the test failed. The screenshots are of high quality, and when comparing 100 screenshots, their file sizes ranged from 61.5 KB - 136.9 KB.

Regarding the video reporter, as indicated in Appendix 2, video recording is performed only for failed tests. However, screenshots are captured at every stage of each test regardless of its outcome. For each individual test, a folder is created with the name of the test, and all corresponding screenshots are saved. The number of screenshots depends on the size of the test, but on average, 5 to 30 screenshots are generated. The file sizes of these screenshots range from approximately 31.0 KB to 154.3 KB.

In the case of a failed test, these screenshots are compiled into a single WEBM video file, with sizes ranging from approximately 18.4 KB to 71.7 KB. The video quality is moderate but sufficient to understand the test execution flow and the conditions under which the test fails. Tests were also executed via Jenkins on an Ubuntu server, and analysis showed that both screenshots and videos can be viewed directly through the Jenkins UI.

The test packs were executed 10 times as part of this study. The results also indicate that a key limitation is that the framework does not automatically delete all screenshots captured by the video reporter during test execution. Thousands of screenshots can be generated in a single day, so it is necessary to include scripts to remove these files.

Although the study primarily focuses on the practical applicability of screenshot and video-based reporting, the experimental evaluation demonstrates several measurable characteristics of the proposed approach. In particular, the generated artifacts maintained stable quality across different execution environments, including GUI and headless browser modes, as well as Windows and Ubuntu operating systems. Furthermore, the relatively small file sizes of screenshots and WEBM videos indicate that the approach can be applied in CI/CD infrastructures with limited storage and computational resources without introducing significant overhead.

From a debugging perspective, the collected screenshots and videos enabled rapid identification of the exact failure step and the sequence of actions preceding the error. While the present work does not include a formal user study or quantitative measurements of debugging speed, repeated execution of the test packs demonstrated that visual reporting artifacts consistently improved the interpretability of failed test executions compared to log-only reporting approaches. Therefore, the proposed method can reduce the effort required for reproducing defects and analyzing unstable automated tests in practical testing environments.

A potential threat to validity is that the experiments were conducted using a specific configuration involving the Chrome browser, WebdriverIO framework, React-based applications, and Jenkins CI/CD infrastructure. Different browser engines, operating systems, or reporting plugins may produce different storage characteristics or execution overhead. However, because the proposed approach relies on standard WebdriverIO APIs and widely supported PNG and WEBM formats, the findings are expected to be transferable to similar modern web automation environments.

6 CONCLUSIONS

Although there are studies on screenshot and video reporters in automated web testing, this research complements existing practices in the context of using the WebdriverIO library. Based on the results of this study, several conclusions can be drawn:

- The GUI and headless modes of the Chrome browser do not affect the quality of the screenshots and videos captured for failed tests;
- The saved screenshots have an average size of approximately 61.5 KB to 136.9 KB, while the videos range from 18.4 KB to 71.7 KB, indicating relatively small file sizes. This makes them suitable for running tests on low-budget or resource-limited remote servers;
- Since Jenkins supports viewing PNG screenshots and WEBM video files, visual analysis can be performed directly within the Jenkins interface;

- The visual quality of the screenshots and video recordings is sufficient to clearly identify the point at which a test failed and to reproduce the bug.

REFERENCES

[1] S. Balsam and D. Mishra, "Web application testing - Challenges and opportunities," Journal of Systems and Software, vol. 219, Jan. 2025, art. 112186, [Online]. Available: https://doi.org/10.1016/j.jss.2024.112186.

[2] H. At Thooriqoh, T. N. Annisa and U. L. Yuhana, "Selenium Framework for Web Automation Testing: A Systematic Literature Review," JUTI: Jurnal Ilmiah Teknologi Informasi, vol. 19, no. 2, Jul. 2021, [Online]. Available: https://doi.org/10.12962/j24068535.v19i2.a1021.

[3] R. M. Salagundi and S. R. Savitha, "Automated reporting and statistical analysis of test case results in continuous integration: A custom dashboard approach," World Journal of Advanced Engineering Technology and Sciences, vol. 13, no. 1, pp. 027-033, 2024, [Online]. Available: https://doi.org/10.30574/wjaets.2024.13.1.0374.

[4] O. Cherkashyn, "Application Test Automation in Headless Android Emulator," Proceedings of International Conference on Applied Innovation in IT, vol. 13, no. 5, pp. 445-452, 2025, [Online]. Available: https://doi.org/10.25673/123064.

[5] H. Sun, A. Rosà, D. Bonetta and W. Binder, "Automatically Assessing and Extending Code Coverage for NPM Packages," in 2021 IEEE/ACM International Conference on Automation of Software Test (AST 2021), pp. 40-49, 2021, [Online]. Available: https://doi.org/10.1109/AST52587.2021.00013.

[6] "Video Reporter," [Online]. Available: https://webdriver.io/docs/wdio-video-reporter/.

[7] D. Wang, Z. Zhang, S. Feng, W. G. J. Halfond and T. Yu, "An Empirical Study on Leveraging Images in Automated Bug Report Reproduction," in Proceedings of the 2025 IEEE/ACM 22nd International Conference on Mining Software Repositories (MSR), pp. 123-133, Apr. 2025, [Online]. Available: https://doi.org/10.1109/MSR66628.2025.00019.

[8] S. Mallipeddi, M. Yaqoob, J. A. Khan, T. Mehmood, A. Mylonas and N. Pitropakis, "AutoQALLMs: Automating Web Application Testing Using Large Language Models (LLMs) and Selenium," Computers, vol. 14, no. 11, p. 501, Nov. 2025, [Online]. Available: https://doi.org/10.3390/computers14110501.

[9] M. Mon and B. Panczyk, "A Comparative Analysis of Web Application Test Automation Tools," Journal of Computer Sciences Institute, vol. 35, pp. 159-165, Jun. 2025, [Online]. Available: https://doi.org/10.35784/jcsi.7119.

[10] Y. Zhao, S. Talebipour, K. Baral, H. Park, L. Yee, S. A. Khan, Y. Brun, N. Medvidovic and K. Moran, "Avgust: Automating Usage Based Test Generation from Videos of App Executions," in Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2022), pp. 421-433, Nov. 2022, [Online]. Available: https://doi.org/10.1145/3540250.3549134.

[11] J. Wang, M. Li, S. Wang, T. Menzies and Q. Wang, "Images don't lie: Duplicate crowdtesting reports detection with screenshot information," Information and Software Technology, vol. 110, pp. 139-155, Jun. 2019, [Online]. Available: https://doi.org/10.1016/j.infsof.2019.03.003.

[12] A. Ashrafi, M. Bouktif and R. Mediani, "ImageR: Enhancing Bug Report Clarity by Screenshots," arXiv preprint arXiv:2505.01925, 2025, [Online]. Available: https://arxiv.org/abs/2505.01925.

[13] E. Kesavan, "Comparative Study of JavaScript-Based Testing Frameworks Integrated with Selenium WebDriver," International Scientific Journal of Engineering and Management, vol. 1, no. 5, Dec. 2022, [Online]. Available: https://doi.org/10.55041/ISJEM00087.

[14] B. García, F. Ricca, M. Leotta and M. Munoz-Organero, "Test automation with Selenium: A survey," Information and Software Technology, vol. 194, art. 108077, 2026, [Online]. Available: https://doi.org/10.1016/j.infsof.2026.108077.

[15] K. Lim, Y. Kwon and D. Kim, "A Longitudinal Study of Vulnerable Client-side Resources and Web Developers' Updating Behaviors," in Proceedings of the 2023 ACM on Internet Measurement Conference (IMC '23), pp. 162-180, Oct. 2023, [Online]. Available: https://doi.org/10.1145/3618257.3624804.

APPENDIX

The Appendix contains the scripts and sections of programming languages code used in the study.

```
{
  "name": "app_name",
  "type": "module",
  "devDependencies": {
    "@wdio/cli": "^8.32.3",
    "@wdio/local-runner": "^8.32.3",
    "@wdio/mocha-framework": "^8.16.3",
    "@wdio/spec-reporter": "^8.16.3",
    "chai": "^5.1.0",
    "moment": "^2.29.4",
    "chromedriver": "^144.0.0",
 "wdio-chromedriver-service": "^8.1.1",
 "wdio-video-reporter": "^4.0.5",
 "@moroo/wdio-slack-reporter":"^8.0.1",
  },
  "scripts": {
    "test": "wdio run ./wdio.conf.ts"
  },
}
```

Listing A1: Package.json.

```
import video from 'wdio-video-
reporter';
import moment from 'moment';

export const config = {
reporters: ['spec',
```

```
        [ video, {
          saveAllVideos:
false,
          videoSlowdownMultiplier:
10,
        }],
    ],

afterTest: async function (test,
context, { error, result, duration,
passed }) {
        if (error &&
browser.capabilities.browserName ===
'chrome') {
        const timestamp =
moment().format('YYYYMMDD_HHmmss')
              await
browser.saveScreenshot(`./errorscree
n/error_${timestamp}.png`);
        }
    }
}
```

Listing A2: JavaScript code.

Hochshule Anhalt
EDITION
ICAIIT

STATISTICAL MODELLING AND ANALYSIS AND THEIR APPLICATIONS IN LIFE SCIENCES

Section 5

APRIL 2026



Proceedings of the International Conference on Applied Innovations in IT

Patterns of Behavioural Risk Factors and their Associations with Cardiovascular Outcomes in Ukraine

Oleh Kovalyshyn[1], Yaroslav Dorogyi[2], Vladyslav Ostash[1,3] and Nadiia Kasianchuk[1]
[1]*Department of Computer Science, Kyiv School of Economics, Shpaka Str. 3. 03113 Kyiv, Ukraine*
[2]*Department of Artificial Intelligence and Cybersecurity, Donetsk National Technical University, 82100 Drohobych, Ukraine*
[3]*Faculty of Biotechnology and Biotechnics, Igor Sikorsky Kyiv Polytechnic Institute, 03056 Kyiv, Ukraine*
oleh.kovalyshyn.25@kse.org.ua, yaroslav.dorohyi@donntu.edu.ua, vostash@kse.org.ua, nkasianchuk@kse.org.ua

Keywords: Ukraine, WHO STEPS, Cardiovascular Disease, Blood Pressure, Hypertension, Behavioural Risk Factors, Tobacco, Alcohol, Substance Use, Body Mass Index, Obesity, Logistic Regression.

Abstract: Evidence on behavioural and metabolic cardiovascular risk patterns in Ukraine remains limited beyond descriptive reporting from the WHO STEPS 2019 survey. We reconstructed the STEPS 2019 analytic framework and examined generational patterns of tobacco and alcohol use, body-mass index (BMI), and their associations with elevated blood pressure and self-reported cardiovascular history. We analysed de-identified microdata from the survey (adults aged 18-69; n = 4,409). The complex survey design was reconstructed in R and calibrated to Ukraine's 2019 age-sex population distribution. Prevalence estimates were produced for tobacco and alcohol indicators, including product-type patterns by age and sex. Elevated blood pressure (SBP ≥140 and/or DBP ≥90 mmHg) was compared across BMI categories. Predictors of self-reported cardiovascular history (stroke and/or heart disease event) were assessed using standard logistic regression, survey-weighted logistic regression, and Firth penalized logistic regression, with backward AIC-based selection; performance was evaluated using ROC AUC and F-score on complete cases (n = 2,279). Current tobacco use was 32.9% overall (49.9% men; 17.6% women), with younger adults more likely to use non-cigarette products. Current alcohol use was widespread (≈69%). Blood pressure differed significantly by BMI, while self-reported cardiovascular history did not. In multivariable modelling, the best-performing specification was Firth logistic regression (AUC = 0.721; F-score = 0.447), identifying age, BMI category, history of high blood sugar, and salt consumption frequency as significant predictors ($p < 0.05$); tobacco and alcohol use were not significant, consistent with potential reverse causation. Findings support prioritising sodium-reduction and weight-management strategies, while interpreting behavioural associations cautiously due to self-report and the cross-sectional design.

1 INTRODUCTION

According to the World Health Organization, non-communicable diseases (NCD) are responsible for 43 million deaths in 2021, of which 73% were in low- and middle-income countries. Cardiovascular diseases are the most common among NCD and cause 90% more deaths than cancers [1]. Moreover, recent findings suggest that younger adults are diagnosed with cardiovascular diseases more often than in previous years, potentially breaking the established pattern of them being associated with older people [2]. Finding socioeconomic and habitual patterns can increase understanding of the NCDs and help to form policies and treatments to reduce the risk for the population.

For that, it's important to understand the current risk factors, prevalence, generational trends and impact of local cultures on the level of NCDs in the population. Such analysis is lacking for Ukraine which is a nation with a transitioning economy and rapidly aging population.

STEPS is a survey, conducted by the World Health Organization that focuses on NCD risk factors among 18-69 years old adults. Conducted for the first, and, until today, only, time in Ukraine it provides insights into Ukrainian population's habits regarding alcohol and tobacco use, diet, sport activity, income and objective health measurements such as body-

mass index (BMI), heart rate, blood pressure, cholesterol level.

The original report provides general analysis of the current state of Ukrainians' health and prevalence of different habits and risk factors. It doesn't try to look into dependencies between the data and connection between risk factors and measured parameters/habits. Analyzing the age and sex differences in habits and risk factors can lead to new insights regarding Ukrainian health and cultural norms.

2 METHODS

2.1 Data Source and Weighting

This study uses de-identified microdata from the WHO STEPS Ukraine 2019 survey, a nationwide multistage cluster household survey of adults aged 18-69 years conducted under the WHO STEPS protocol. The dataset includes questionnaire-based behavioral risk factors and sociodemographics (Step 1) and standardized physical measurements, including blood pressure and anthropometrics (Step 2). The analytic sample comprised 4,409 respondents drawn through the original sampling framework documented in the STEPS report [3].

We reconstructed the complex sample design in R using the *Survey* package. Individual-level analyses applied the provided sampling weights (*wstep1*). To align estimates with the national population structure, we additionally calibrated/post-stratified the weighted sample to Ukraine's 2019 age-sex distribution. Because the STEPS report does not specify the exact population source used to construct these demographic corrections, we used the 2019 population counts from the State Statistics Service of Ukraine as the reference distribution for calibration [4].

To validate the reconstruction, we compared reconstructed estimates against the STEPS report using pre-specified anchor indicators: active smoking prevalence, active alcohol use prevalence, and mean body mass index (BMI), evaluated overall and across age-sex cohorts. During validation, we assessed missing-data patterns in these anchor variables and identified substantial missingness in the alcohol-use indicator. To examine whether discrepancies with the published report could be driven by missing-data handling, we performed a sensitivity check that recoded missing alcohol-use responses as non-use (NA = false). Because this recoding imposes a strong and generally unjustified assumption, it was used only for sensitivity assessment; all subsequent substantive analyses treat missing alcohol-use responses as missing and use standard survey-based handling.

2.2 Substance Use Analysis: Tobacco and Alcohol

Tobacco and alcohol indicators were analysed by sex and WHO STEPS age groups (18-29, 30-44, 45-59, 60-69); tobacco analyses additionally considered place of residence (urban vs rural). All estimates accounted for the complex survey design and applied sampling weights (*wstep1*) using the R *survey* package, producing survey-weighted prevalence estimates with 95% confidence intervals.

Associations with sociodemographic factors were assessed using survey-weighted logistic regression (svyglm, binomial). For tobacco, current smoking status was modelled as a function of sex and age group, with urban/rural residence evaluated via stratification or as a covariate as appropriate. Among current tobacco users, product-type patterns were described using survey-weighted prevalence (95% CI) for combustible cigarettes, heated tobacco products, electronic cigarettes/vapes, hookah/waterpipe, and non-smoked tobacco, overall and by sex and age group.

For alcohol, we examined current use and lifetime abstinence (never consumed alcohol), each modelled separately by sex and age group using survey-weighted logistic regression. Alcohol analyses used outcome-specific complete-case handling (excluding respondents with missing values for the relevant alcohol outcome).

2.3 Body-Mass Index Analysis

Body-mass index (BMI; kg/m²) was summarised descriptively across respondent characteristics without survey weights to compare groups rather than produce population estimates. Mean BMI (95% CI) was reported by sex and WHO STEPS age groups (18-29, 30-44, 45-59, 60-69). Respondents were classified using WHO cut-offs: underweight (<18.5), normal (18.5-24.9), overweight (25.0-29.9), and obesity (≥30.0); these categories were used in all subsequent BMI-stratified analyses.

Physiological correlates of BMI were assessed by comparing heart rate and blood pressure across BMI categories. For each respondent, mean heart rate and mean systolic/diastolic blood pressure were calculated as the average of three repeated

measurements; analyses used indicator-specific complete cases (excluding refusals/missing readings). Category-specific means (95% CI) were estimated overall and by sex.

We estimated the prevalence of elevated blood pressure (SBP ≥140 and/or DBP ≥90 mmHg) within each BMI category and tested differences in systolic and diastolic blood pressure across categories using the Kruskal-Wallis test; when significant, Dunn's post-hoc pairwise comparisons with Bonferroni adjustment were applied (two-sided $\alpha = 0.05$). Cardiovascular morbidity was captured via self-reported history of heart attack, stroke, or heart-related chest pain; for each BMI category we reported prevalence and odds ratios relative to the normal-weight group.

2.4 Models for Heart Disease Prediction

Building on the descriptive results, we fit multivariable regression models to quantify associations between respondent characteristics and prior cardiovascular events. Finding significant associations gives better understanding of the social and behavioral context of the disease. We asked if any of the previous findings could translate into cardiovascular diseases in patients. The primary outcome was self-reported disease history, defined as reporting a previous stroke and/or heart disease event in the STEPS questionnaire. We compared three specifications: (i) standard logistic regression (GLM), (ii) survey-weighted logistic regression accounting for the complex design (R *survey* package), and (iii) Firth penalized logistic regression to reduce small-sample bias and separation (R *logistf* package).

Models were estimated on a complete-case subset excluding missing or ambiguous responses ("Refused", "Don't know", NA) for the outcome and all candidate predictors, yielding n = 2,279 (from 4,409), with 18.69% reporting the outcome. Predictors were selected based on prior analyses and included physical and lifestyle factors (e.g., BMI category, blood pressure indicators, smoking and alcohol measures, diet and physical activity variables) and core sociodemographics (age group, sex). Continuous variables were checked for plausibility and, where appropriate, transformed into clinically interpretable forms (e.g., BMI categories, hypertension indicator).

Variable selection was performed using backward stepwise elimination based on the Akaike Information Criterion (AIC), a model selection metric that balances goodness-of-fit against model complexity. Lower AIC values indicate a better trade-off between explanatory power and parsimony, reducing the risk of overfitting while retaining informative predictors.

Model discrimination was assessed via ROC analysis, reporting AUC and F-score, computed consistently across model families to compare the impact of survey weighting and penalized estimation on predictive performance.

3 RESULTS

3.1 Survey Reconstruction and Benchmarking Against the STEPS Report

To evaluate reconstruction accuracy and the effect of calibration to Ukraine's 2019 age-sex population structure, we benchmarked reconstructed estimates against the STEPS 2019 Ukraine report using three pre-specified anchor indicators: current (active) smoking prevalence, current (active) alcohol use, and mean BMI. Differences (reconstructed minus reported) are reported overall and by sex-age strata (Tables 1-3).

Table 1: Difference in reconstructed vs STEPS report estimates of current smoking prevalence and mean BMI by sex-age cohort (reconstructed − reported, percentage points), after calibration to the 2019 age-sex population distribution.

Cohort	Δ smoking, %	Δ BMI, kg/m^2
All ages and genders	-0.96	0.2
Male 18-29	0.87	0
Male 30-44	-0.58	0.2
Male 45-59	-0.3	0
Male 60-69	-0.07	0
Male 18-69	-0.4	0.1
Female 18-29	6.11	0
Female 30-44	0.75	0
Female 45-59	-0.03	0
Female 60-69	0.01	0
Female 18-69	0.94	0.2

Reconstructed smoking prevalence and mean BMI were broadly consistent with the report across most strata (overall smoking difference −0.96 percentage points; overall BMI difference +0.2 kg/m²), with the main deviation observed among females aged 18-29 (Table 1). In contrast, reconstructed estimates of current alcohol use exceeded report values by ~13-15 percentage points,

while lifetime abstinence ("sober") closely matched (difference +0.09 percentage points) (Table 2). Inspection of the microdata indicated substantial missingness in the alcohol-use indicator (1,122 observations; 25% coded as NA), compared with no missingness for smoking and 5% missingness for BMI (233 observations), suggesting that differences in missing-data handling likely explain the alcohol discrepancy.

Table 2: Difference in reconstructed vs STEPS report estimates of current alcohol use prevalence (and lifetime abstinence, "sober") (reconstructed − reported, percentage points), after calibration to the 2019 age-sex population distribution.

Cohort	Difference, %
All ages and genders	13.45
Male 18-69	14.81
Female 18-69	13.1
Sober, all ages and genders	0.09

Table 3: Sensitivity analysis: difference in reconstructed vs STEPS report estimates of current alcohol use prevalence after recoding missing alcohol-use responses as non-use (NA → false); (reconstructed − reported, percentage points), after calibration to the 2019 age-sex population distribution.

Cohort	Difference, %
All ages and genders	2
Male 18-69	-1.3
Female 18-69	-1.1

To assess this possibility, we conducted a diagnostic sensitivity analysis recoding missing alcohol-use responses as non-use (NA → false). Under this assumption, reconstructed alcohol-use estimates aligned much more closely with the report (overall difference +2 percentage points; −1.3 for men and −1.1 for women; Table 3). Because this recoding imposes a strong and generally unjustified missingness assumption, it was used for sensitivity assessment only and was not applied in subsequent substantive analyses.

3.2 Substance Use Analysis

Survey-weighted estimates indicate that current tobacco use in Ukraine in 2019 was 32.94% (design-based SE = 1.61%), with a large sex gap (49.91% in men vs 17.64% in women). Among women, tobacco use declined sharply with age, from 26.62% (18-29) and 25.05% (30-44) to 13.57% (45-59) and 3.61% (60-69). Among men, prevalence remained high across age groups, rising from 46.67% (18-29) to 50.52% (30-44) and peaking at 55.00% (45-59), before decreasing to 43.23% at ages 60-69 (Fig. 1).

We tested two hypotheses: (i) younger adults have lower overall tobacco use, and (ii) younger adults are more likely to use non-cigarette products [5], [6]. The first hypothesis was only weakly supported, as overall prevalence was not consistently lower in younger groups-particularly among men (Fig. 1). The second hypothesis was supported: although combustible cigarettes were the most common product in every age group, non-cigarette products (hookah/waterpipe, heated tobacco, and electronic cigarettes/vapes) were concentrated among younger adults and declined with age (Fig. 2). Because respondents could report multiple product types, product-specific prevalences are non-exclusive and should not be summed.

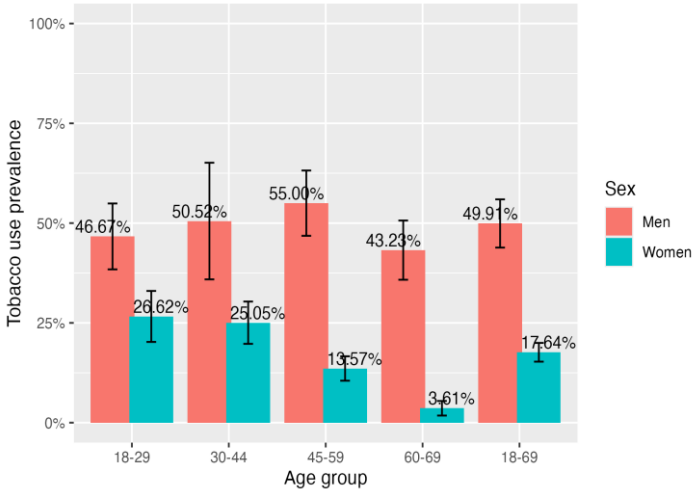


Figure 1: General tobacco use prevalence.

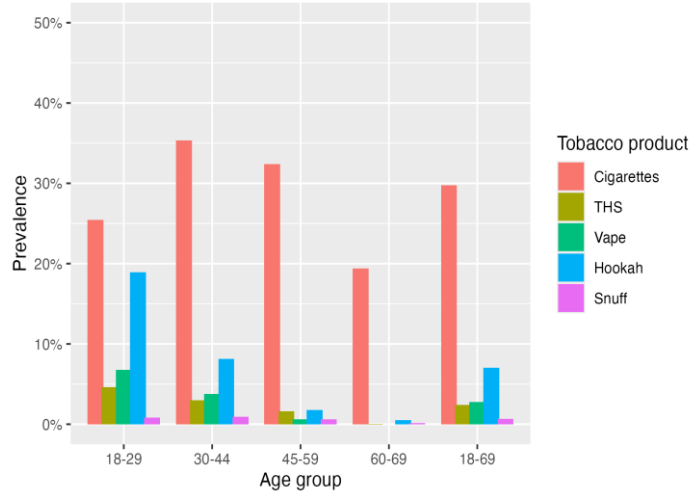


Figure 2: Survey-weighted prevalence of tobacco product types by age group (population prevalence). Product categories are non-exclusive; error bars indicate uncertainty (95% CI or SE).

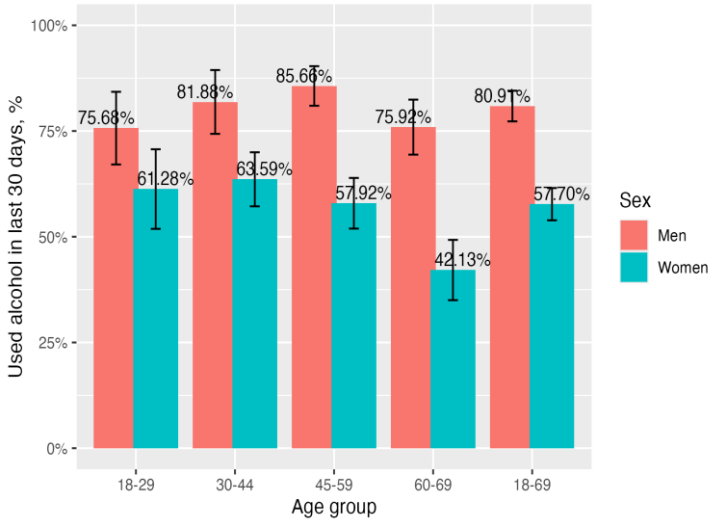


Figure 3: Active alcohol use prevalence.

Active use of alcohol was defined by the fact of consumption during the last 30 days at the time of survey. The metric is high among all age groups but is significantly higher in men vs women (80.91% vs 57.70%). The same pattern of the youngest and the oldest consuming less than middle-aged respondents is visible with active alcohol use (Fig. 3).

9.5% (SE = 0.85%) of Ukrainians have never tasted alcohol (self-reported), the pattern is inverse of that that we see for active use. Younger and older people are sober more than those whose youth coincided in time with the collapse of the Soviet Union and following economic shock. Also, women are sober more often than men (Fig. 4).

3.3 BMI Analysis

Consistent with prior evidence that BMI tends to increase with age, mean BMI in the STEPS 2019 sample rose across successive age groups for both sexes (Fig. 5) [7]. Among men, this upward trend plateaued with a small decline in the 60-69 group relative to ages 45-59, whereas among women the increase was more pronounced at older ages. The overall mean BMI in the sample was 25.0 kg/m² (design-based SE = 0.165), placing the average respondent at the threshold of the overweight category.

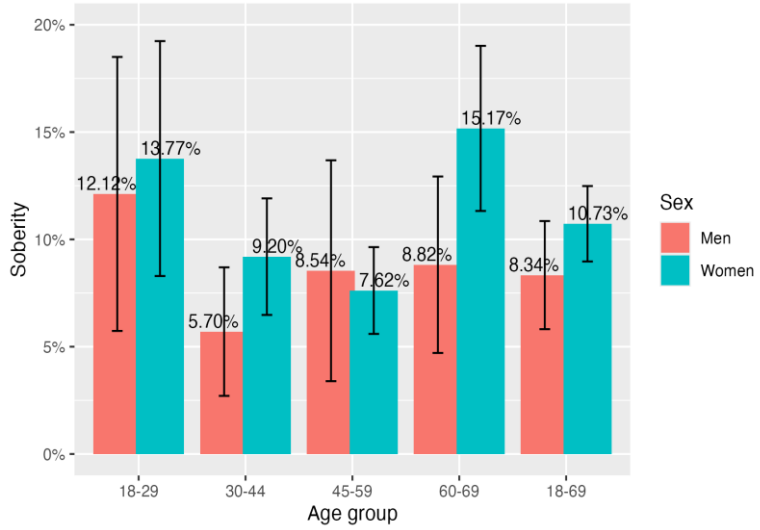


Figure 4: Sobriety prevalence.

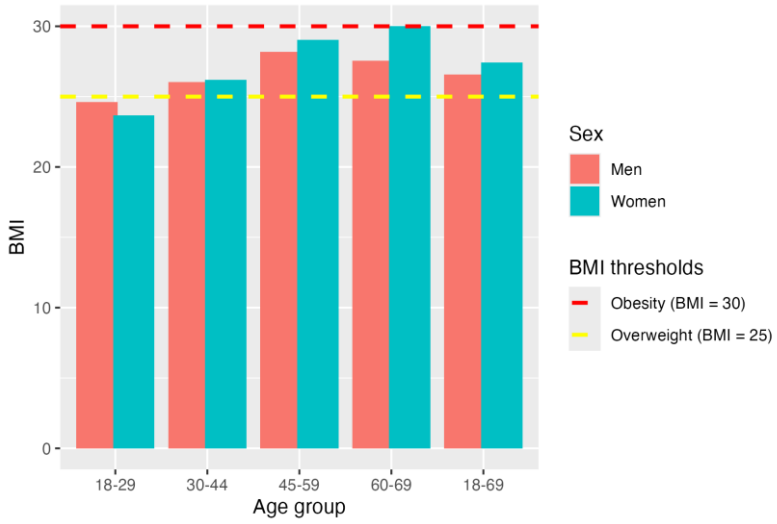


Figure 5: Mean BMI (kg/m²) by age group and sex.

Most age-sex strata exhibited mean BMI values above the overweight threshold (BMI ≥ 25), and women aged 60-69 had a mean BMI at approximately 30.0 kg/m² (SE = 0.33), corresponding to the obesity threshold. In survey-weighted GLM analysis of BMI, age group was strongly associated with BMI relative to the 18-29 reference group, whereas sex was not statistically significant at the 0.05 level (Table 4).

Given the well-established positive relationship between BMI and blood pressure [8], [9], we examined the prevalence of elevated blood pressure across BMI categories. The proportion of respondents meeting the elevated blood pressure criterion was substantially higher in the overweight and obese groups, and lowest in the underweight group (Fig. 6). This pattern was consistent with odds ratios comparing each BMI category to the normal-weight reference group: the odds of elevated blood pressure increased monotonically with the BMI category (Table 5).

Table 4: Survey-weighted GLM p-values for predictors of BMI.

Parameter	p-value
sex	0.098378
Age range 30-44	0.000182
Age range 45-59	< 2e-16
Age range 60-69	< 2e-16

We next assessed self-reported history of cardiovascular events/symptoms across BMI categories. The proportion of respondents reporting a cardiovascular history was higher in the overweight and obese groups, while normal and underweight groups appeared similar (Fig. 6). However, inference for the underweight category is constrained by limited sample size: after restricting to respondents with an unambiguous "yes/no" response, only 60 individuals were classified as underweight (including 8 reporting cardiovascular history), which reduces precision and statistical power. Consistent with this limitation, the Kruskal-Wallis test did not detect statistically significant differences across BMI categories for self-reported cardiovascular history ($\chi^2 = 3$, df = 3, p = 0.392). Odds ratios by BMI category (Table 6) should therefore be interpreted cautiously, particularly for the underweight group.

Table 5: Odds ratios for elevated blood pressure by BMI category (reference: normal weight).

Cohort	OR
underweight	0.410
normal	1
overweight	1.82
obesity	4.84

Table 6: Odds ratios for self-reported cardiovascular history by BMI category (reference: normal weight).

Cohort	OR
underweight	1.06
normal	1
overweight	1.54
obesity	2.70

3.4 Key Predictor Selection for Cardiovascular History Using Linear Models

The final multivariable model identified four significant independent predictors: age, BMI category, history of high blood sugar, and frequency of salt consumption ($p<0.05$). Other potential risk factors, including physical activity, household income, marital status, and education level, current tobacco and alcohol use, did not reach statistical significance in the adjusted model ($p>0.05$). This lack of association likely reflects reverse causality, where individuals diagnosed with cardiovascular conditions may have ceased substance use following their diagnosis.

Firth's penalized logistic regression yielded the highest discriminative performance with an AUC of 0.721 and an F-Score of 0.447, slightly outperforming the standard survey-weighted GLM (AUC = 0.709) (Fig. 7).

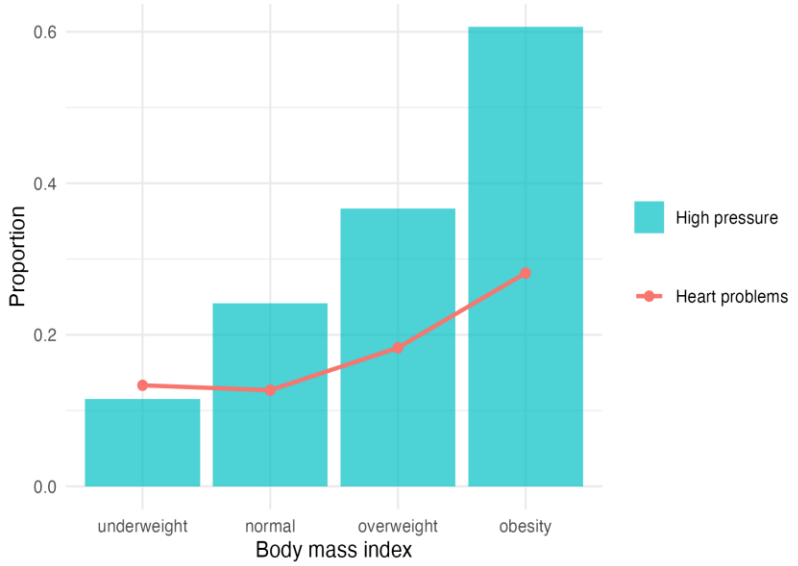


Figure 6: Prevalence of elevated blood pressure and self-reported heart problems by BMI category.

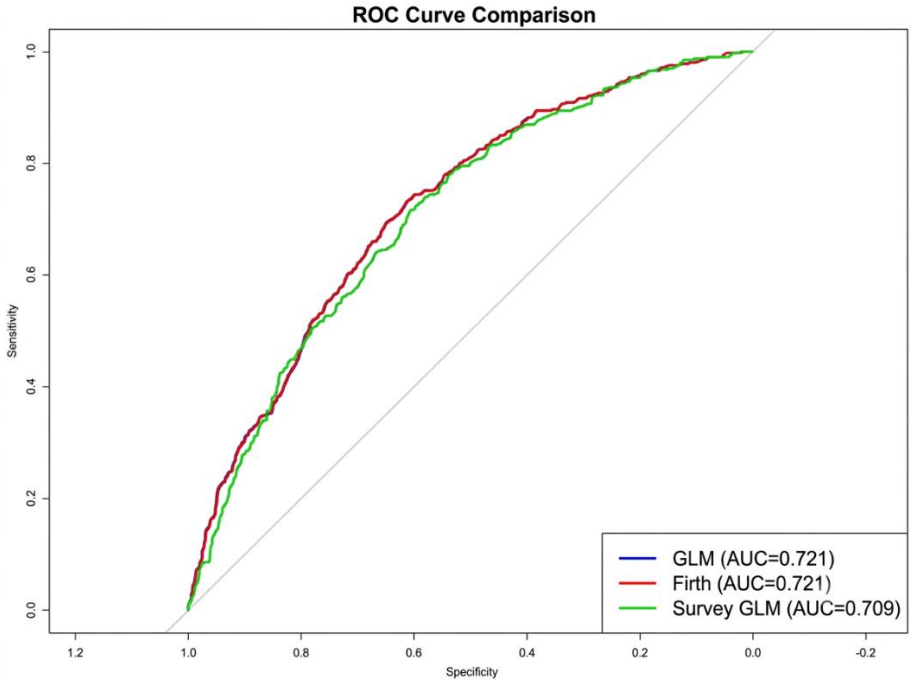


Figure 7: ROC curves comparison.

4 CONCLUSIONS

Increased BMI is an established factor of risk of cardiovascular diseases [10]. Habits such as alcohol and tobacco consumption are also related to increased mortality and lifetime disease risk [11], [12].

These findings have specific implications for public health policy in Ukraine. While alcohol and tobacco control remains essential, our multivariable model highlights that BMI and salt consumption frequency are the most robust predictors of existing cardiovascular conditions in this dataset. Unlike behavioral habits (smoking/drinking) which might be modified after a diagnosis (reverse causality), high BMI and dietary salt intake appear to be persistent risk drivers. Consequently, interventions targeting sodium reduction (e.g., labeling regulations, public awareness campaigns) and weight management strategies may yield the most immediate benefits for secondary prevention and reducing cardiovascular burden.

This study reveals that Ukrainians continue to be active smokers and drinkers despite educational efforts and generational differences. Trends in tobacco consumption seem to shift rather than decrease, so more studies might be needed regarding the influence of novel consumption methods on one's health. The relation between high blood pressure (which is also one of risk factors of cardiovascular diseases) and body mass index is confirmed. Confirmed/established probable predictors of heart problems: age, BMI, high blood sugar, increased salt consumption.

A significant limitation is the reliance on self-reported data for behavioral factors and medical history. This introduces the risk of social desirability bias, where respondents may underreport stigmatized behaviors such as alcohol consumption or tobacco use, and recall bias regarding medical diagnoses. Furthermore, the cross-sectional nature of the STEPS survey precludes causal inference, limiting our ability to distinguish between risk factors that caused the disease and lifestyle changes adopted post-diagnosis. We were also not able to assess the influence of socioeconomic factors (income, relocation due to war) because these data were too vague or missing. Another limitation is that the data were actual as for 2019 and no subsequent STEPS surveys were conducted to check the dynamics. Since the russian invasion and beginning of the full-scale war, many factors might have changed due to increased background and active stress level, uncertainty and massive population movements.

REFERENCES

[1] World Health Organization, "Noncommunicable diseases," World Health Organization, Dec. 23, 2024, [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases.

[2] H. C. Gooding et al., "Challenges and Opportunities for the Prevention and Treatment of Cardiovascular Disease Among Young Adults: Report from a National Heart, Lung, and Blood Institute Working Group," Journal of the American Heart Association, vol. 9, no. 19, Oct. 2020, [Online]. Available: https://doi.org/10.1161/jaha.120.016115.

[3] "STEPS: prevalence of noncommunicable disease risk factors in Ukraine 2019," WHO Regional Office for Europe, Copenhagen, 2020, [Online]. Available: https://phc.org.ua/sites/default/files/users/user90/2019_STEPS_report_eng.pdf, [Accessed: Feb. 03, 2026].

[4] "Data Bank | State Statistics Service of Ukraine," Stat.gov.ua, 2026, [Online]. Available: https://stat.gov.ua/uk/explorer?md5=22fb4077f27cf51088adc40eb6e2caec, [Accessed: Feb. 11, 2026].

[5] K. E. Warner, "Kids No Longer Smoke Cigarettes. Why Aren't We Celebrating?" American Journal of Public Health, vol. 114, no. 11, pp. 1191-1194, Nov. 2024, [Online]. Available: https://doi.org/10.2105/AJPH.2024.307841.

[6] R. Dahal, A. Bhattarai, and K. Adhikari, "Age and sex-related patterns of electronic cigarette use in the general population: Supporting a de novo substance use pattern," Population Medicine, vol. 4, no. December, pp. 1-10, Dec. 2022, [Online]. Available: https://doi.org/10.18332/popmed/157270.

[7] T. Zhang et al., "Rate of change in body mass index at different ages during childhood and adult obesity risk," Pediatric Obesity, vol. 14, no. 7, p. e12513, Jan. 2019, [Online]. Available: https://doi.org/10.1111/ijpo.12513.

[8] L. Song et al., "Body Mass Index is Associated with blood pressure and vital capacity in medical students," Lipids in Health and Disease, vol. 22, no. 1, Oct. 2023, [Online]. Available: https://doi.org/10.1186/s12944-023-01920-1.

[9] L. Yuan et al., "Dose-response relationship between body mass index and hypertension: A cross-sectional study from Eastern China," Preventive Medicine Reports, vol. 46, p. 102852, Aug. 2024, [Online]. Available: https://doi.org/10.1016/j.pmedr.2024.102852.

[10] S. S. Khan et al., "Association of Body Mass Index with Lifetime Risk of Cardiovascular Disease and Compression of Morbidity," JAMA Cardiology, vol. 3, no. 4, pp. 280-287, Apr. 2018, [Online]. Available: https://doi.org/10.1001/jamacardio.2018.0022.

[11] M. Piano, "Alcohol's Effects on the Cardiovascular System," Alcohol Research: Current Reviews, vol. 38, no. 2, p. 219, 2017, [Online]. Available: https://doi.org/10.35946/arcr.v38.2.06.

[12] M. Rahman et al., "Cardiovascular Effects of Smoking and Smoking Cessation: A 2024 Update," Global Heart, vol. 20, no. 1, Jan. 2025, [Online]. Available: https://doi.org/10.5334/gh.1399.

Organo-Mineral Complexes for Water Purification: Experimental and Statistical Analysis

Sergii Guzii[1], Vitalina Lukianova[1], Yevheniia Anpilova[2,3], Oleksii Soloviov[4] and Andriy Tovmachenko[5]

[1]*Department of Nuclear-Physical Technologies, State Institution "The Institute of Environmental Geochemistry of National Academy of Sciences of Ukraine", Academician Palladin Avenue 34a, 03142 Kyiv, Ukraine*

[2]*Department of Computation Hydrosystems, Helmholtz Centre for Environmental Research GmbH, Permoserstraße 15, 04318 Leipzig, Germany*

[3]*Department of Natural Resources, Institute of Telecommunications and Global Information Space of the National Academy of Sciences of Ukraine, Chokolivskiy boulevard 13, 03186 Kyiv, Ukraine*

[4]*Department of Mineral Resources and Nuclear Energy, State Institution "The Institute of Environmental Geochemistry of the National Academy of Sciences of Ukraine", Academician Palladin Avenue 34a, 03142 Kyiv, Ukraine*

[5]*Department of Radioactive Waste Management, State Institution "The Institute of Environmental Geochemistry of National Academy of Sciences of Ukraine", Academician Palladin Avenue 34a, 03142 Kyiv, Ukraine*

sguziy2@gmail.com, igns@ukr.net, yevheniia.anpilovat@ufz.de, asolovjov0102@gmail.com, andrewtovmachenko@gmail.com

Keywords: Industrial Water, Suspended Solids, Turbidity/Transparency, Sediment, Plastic Strength, Mathematical Models, Optimisation, Organo-Mineral Complex.

Abstract: The article presents an assessment of the sanitary and hygienic condition of water in storage ponds at solid waste landfill facilities, with the aim of enabling their reconstruction, expansion, and reclamation of bottom sediments. Based on experimental and statistical modelling, the optimal composition of an organo-mineral complex for the removal of suspended solids from technically contaminated water was determined. Water transparency with turbidity below 1 Nephelometric Turbidity Unit is achieved using 2.2 to 4.2 % quicklime, 7.5 to 11.25 % bentonite, and 6 to 10 % enzyme. The optimal composition of the organo-mineral complex for sediment formation was also established. The resulting precipitate exhibits plastic strength up to 160 kPa after 28 days of exposure, achieved with 30 to 35 % quicklime, 0 to 2.5 % bentonite, and 0 to 1.5 % enzyme. The formed sediment is self-sealing and water-resistant, making it suitable for constructing embankments around storage ponds at solid waste landfills to increase both industrial water capacity and evaporation area.

1 INTRODUCTION

Storage ponds at municipal solid waste landfill facilities accumulate technically polluted water whose turbidity and chemical load complicate both reclamation and any subsequent expansion of pond capacity [1]. The water column typically holds an emulsified mixture of petroleum hydrocarbons, fine mineral particles, colloidal organics, and microbial biomass that resists gravity settling alone. Producing a clarified effluent while also generating a consolidated, load-bearing precipitate from a single treatment step would substantially streamline landfill pond management - a goal that motivates the present study [12]. If the treated precipitate can be engineered to meet geotechnical performance criteria, it may serve directly as fill material for pond embankments, turning a waste stream into a construction resource.

2 LITERARY REVIEW AND PROBLEM STATEMENT

Suspended sediment monitoring reported in [2] and [3] show that peak single-event turbidity in drainage systems can exceed 460 nephelometric turbidity units (NTU). Passive retention structures
reduce the downstream transport of suspended solids but cannot alone deliver effluent clarity suitable for reuse or discharge without additional treatment.

Plant-derived coagulants - cactus-pad mucilage and watermelon-seed extract - were assessed for pond-water clarification in [4], achieving turbidity removal between 78.58% and 94.18%; however, the required settling times were considerably longer than those typical of mineral-based coagulants, limiting their practical throughput.

A hybrid clarifier integrating rotating biological discs into a horizontal settling chamber was engineered and piloted for lightly polluted source water in [5]; the device removed turbidity, dissolved organics, and NH^{+4}-N simultaneously over a 2-hour hydraulic retention time, with effluent quality held stable for over two months; removal reached 73.65 ± 5.15% for turbidity, 53.98 ± 5.17% for TOC, and 77.01 ± 10.02% for NH^{+4}-N leaving residuals of 1.96 NTU, 1.98 mg/L TOC, and 0.46 mg/L NH^{+4}-N in the treated effluent.

Injection of polyacrylamide into turbid pumped water was trialled in [6] using two charge types. Cationic PAM consistently outperformed the non-ionic variant: turbidity in the receiving basin fell by 98% and 90%, respectively, with charge-driven destabilisation identified as the key mechanism responsible for the superior performance of the cationic grade.

A three-stage coagulation–sedimentation–filtration sequence was field-tested on a seasonally turbid urban river in [7]. Applying 50 ppm polyaluminium chloride together with 1.5 ppm PAM, the process removed 99.53% of turbidity and 94.69% of total phosphorus, with the clarified effluent consistently reaching turbidity of < 1 NTU and a TP concentration of up to 0.017 mg/L, which fully complies with Class II for EQSSW.

Electrocoagulation with paired iron–aluminium electrodes was evaluated for aquaculture pond effluent in [8], with a Box–Behnken response-surface design used to map the joint effects of electrolysis time, settling duration, and current density. The optimised parameter set (11.97 min electrolysis, 29.99 min settling, 2.389 A) yielded 91.67% turbidity removal in measured trials, deviating by less than 0.2% from the quadratic model prediction of 91.84%.

Coagulation and flocculation of travertine quarry wastewater were systematically compared in [9] across a pH range of 6–9. Single-stage coagulation proved insufficient regardless of the reagent used, whereas the two-stage coagulation–flocculation sequence restored transparency effectively. Among three coagulants tested, aluminium chloride showed the best pH-6 and pH-9 performance, while differences between reagents largely disappeared at neutral pH. The dominant removal pathway shifted from charge-neutralisation in acidic conditions to sweep coagulation near neutrality, with bridging flocculation governing the polymer-aided stage.

Dubyniak et al. [10] developed predictive models for clarifier operation that couple tank geometry, flow rate, and sediment rheology to measurable output indicators: effluent turbidity, suspended-solids concentration, and particle settling velocity. The models allow reagent dosing and temperature regimes to be adjusted in advance of performance drift, supporting consistent long-term clarification without reactive intervention.

Laboratory compaction and shear-strength tests on silt treated with hydraulic binders (0–8% lime, cement, or both) are presented in [11]. After curing under conditions that included freeze–thaw cycling, all treated specimens exhibited higher internal friction angles and cohesion values relative to untreated controls, with the degree of gain depending on the binder combination and the soil's grain-size distribution. These results underline the potential of lime-rich mineral precipitates - such as those produced by the organo-mineral complex studied here - as a structural fill for embankment construction.

The reviewed literature confirms that no single treatment mechanism is universally adequate: chemical coagulation, biological treatment, and electrochemical methods each address a subset of the pollutants present in landfill pond water, and their performance is highly reagent- and condition-specific. Integrated approaches combining mineral coagulants with sorbents and enzyme preparations offer a route to simultaneous turbidity removal and sediment consolidation, consistent with the resource-efficiency principles articulated in recent environmental-engineering studies [12], [13]. The present work extends our earlier investigation of the lime–zeolite–ECO-NOVA system [14] by substituting bentonite for zeolite and Zeta-Zim for ECO-NOVA, keeping all other experimental conditions identical so that the two sorbent–enzyme combinations can be compared directly. The goal is to identify the compositions that minimise treated-water turbidity and maximise sediment plastic strength, with the end use of building and reinforcing storage-pond embankments at solid waste landfill sites.

3 MATERIALS AND METHODS

3.1 Water Sampling

The studies were conducted using water contaminated by Storage Pond No. 1 on the land plot of the Solid Waste Landfill (SWL) in Chernihiv, Ukraine (the same site as reported in [14]).

The complete physicochemical and radiological characterisation of this water (Measurement Protocols Nos. 1528 and 1528/P dated 15.07.2025, LLC "EKODIYA", Kyiv), including all 24 chemical parameters and radiation indicators, is reported in full in the authors' prior work [14]. The same water source is used in both studies to enable direct comparison of the two organo-mineral systems under identical conditions.

3.2 Selection of Precipitating Reagents

Three components were selected for the organo-mineral complex. Quicklime (CaO) served as the reactive mineral phase; its alkalising action promotes the precipitation of dissolved anions as water-insoluble hydroxide and carbonate salts. Bentonite, a smectite-group clay with a high specific surface area, was incorporated as the adsorbent phase to capture oil-derived compounds, surfactants, and radionuclides via surface sorption. Zeta-Zim (an industrial multi-enzyme preparation) was included to catalyse the hydrolysis and transesterification of fats and petroleum residues, and to cross-link the aluminosilicate layers of bentonite into a cohesive, water-repellent matrix [15]. This combination differs from our earlier formulation [14], which used zeolite and ECO-NOVA enzyme; the present reagent set was selected to evaluate whether a smectite-based system achieves comparable or superior treatment outcomes.

3.3 Experimental and Statistical Modelling

Composition optimisation employed a three-factor simplex-centroid mixture design analysed within the STATISTICA 12 environment. A special cubic model was fitted to account for non-linear binary and ternary interactions among the components. This methodology follows the approach validated in our previous study [14] but is applied here to the novel bentonite–Zeta-Zim system.

The coefficients of variation and the matrix for planning the experiment are given in Tables 1 and 2, and the implementation of the experiment is given in Tables 3 and 4.

Table 1: Variation intervals and values of variable factors.

Factors, type	Natural	Coded	Levels of variation		Variation interval
			0	1	
CaO	%	X1	0	35	17.5
Bentonite	%	X2	0	15	7.5
Zeta-Zim	%	X3	0	10	5

Table 2: Experiment planning matrix.

Plan points	Matrix plan in codes			Full size matrix plan		
	X1	*X2*	*X3*	CaO	B	Z
1	0.00	1.00	0.00	0	15	0
2	0.33	0.33	0.33	11.7	5	3.3
3	1.00	0.00	0.00	35	0	0
4	0.50	0.50	0.00	17.5	7.5	0
5	0.00	0.00	1.00	0	0	10
6	0.50	0.00	0.50	17.5	0	5
7	0.00	0.00	0.50	0	7.5	5

Table 3: Correspondence of coded values to natural values.

Coded	Natural		
	CaO (*X1*)	Bentonite (*X2*)	Zeta-Zim (*X3*)
0	0	0	0
0.25	8.75	3.75	2.5
0.5	17.5	7.5	5
0.75	26.25	11.25	7.5
1.0	35	15	10

Table 4: Summary of experimental results.

Plan points	Turbidity, NTU, after day			Tensile strength, kPa, after day		
	14	21	28	14	21	28
1	519	170	163	0.12	0.8	1.41
2	971	229	169	0.5	0.96	1.94
3	1000	1000	1000	3.77	57.86	160
4	1000	1000	1000	1.18	12.24	57.86
5	8	3.49	1.02	0	0	0
6	772	376	212	0.1	0.7	1.49
7	98	103	74	0.1	0.12	1.94

The initial parameters selected were water turbidity $T \rightarrow min$ and sediment plastic strength $T_s \rightarrow max$, which were determined on days 14, 21 and 28 of the holding periods.

3.4 Experimental Methods

To determine water turbidity, we used a portable TB 1 turbidimeter manufactured by VELP SCIENTIFICA, Italy.

The TB1 portable turbidimeter quickly and accurately measures the turbidity of water samples,

with results displayed directly in nephelometric turbidity units (NTU).

To determine the plastic strength of the sediment in kPa, a Rebinder cone plastometer weighing 224 g and with an angle of $\alpha=45°$was used. The dependence of plastic strength on the depth of penetration of the Rebinder cone was approximated by the function $y=14.77x^{-1.97}$.

The sample weight was 200 g, depending on the number of components of the organo-mineral complex presented in Table 2. Water was drained from the samples to determine turbidity, and the sediment was used to determine plastic strength.

4 RESULTS

4.1 Determination of Water Turbidity

Regression analysis of the data in Table 4 yielded the following empirical models describing water turbidity evolution as functions of the organo-mineral complex composition at three ageing intervals:

$$T^{14d}=1000X1+519X2+8X3+962X1X2+1072X1X3-662X2X3+8358X1X2X3 \quad (1)$$

$$T^{21d}=1000X1+170X2+3,49X3+1660X1X2-502,98X1X3+65,02X2X3-8044,53X1X2X3 \quad (2)$$

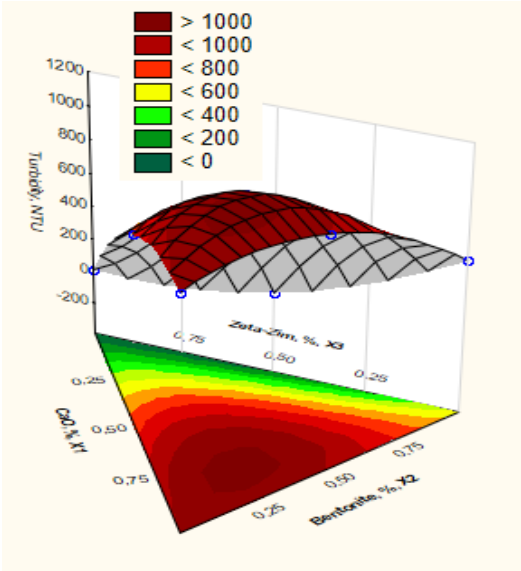
$$T^{28d}=1000X1+163X2+1,02X3+1674X1X2-1154,04X1X3-32,04X2X3-7376,94X1X2X3 \quad (3)$$

Ternary response surfaces (Fig. 1-3) were constructed from the regression equations using the factor designations in Table 2. Compositions at plan points 3 and 4 (Table 4) completely absorbed the water phase; these were assigned a ceiling value of 1000 NTU.

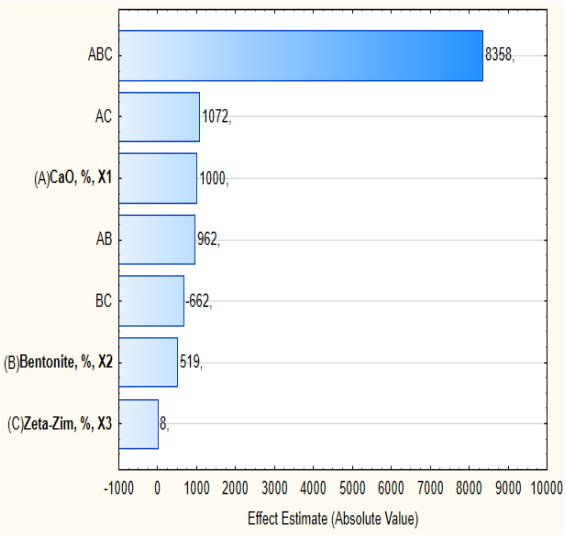
Analysis of regressions (1)–(3) showed that water turbidity after 14, 21, and 28 days of exposure is primarily influenced by the three main factors X1, X2, and X3, which together account for up to 80% of the total effect (see Figures 1b–3b).

As shown in Figure 1a, water turbidity decreases from 519 to 8 NTU after 14 days of exposure when the quicklime content (X1) decreases from 8.75% to 0%, the bentonite content (X2) increases from 14% to 15%, and the enzyme content (X3) increases from 0% to 10%.

As shown in Figure 2a, turbidity decreases from approximately 400 to 3.5 NTU after 21 days of exposure when the quicklime content (X1) decreases from 17.5% to 0%, the bentonite content (X2) increases from 12.5% to 15%, and the enzyme content (X3) increases from 0% to 10%.

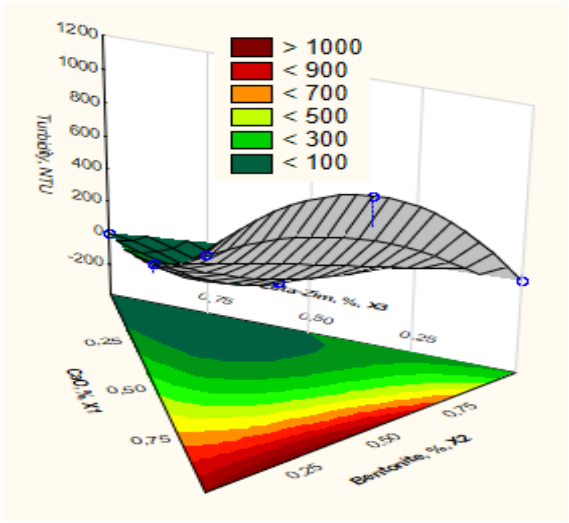


(a)

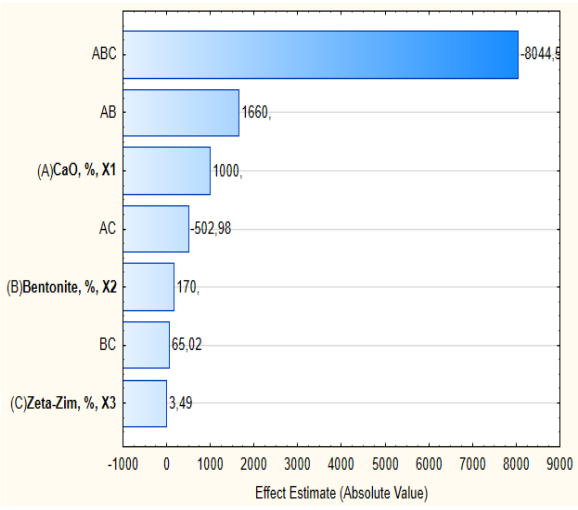


(b)

Figure 1: Terminal surface: a) and Pareto diagram, b) of water turbidity on the 14th day of exposure to factors of the organo-mineral complex.



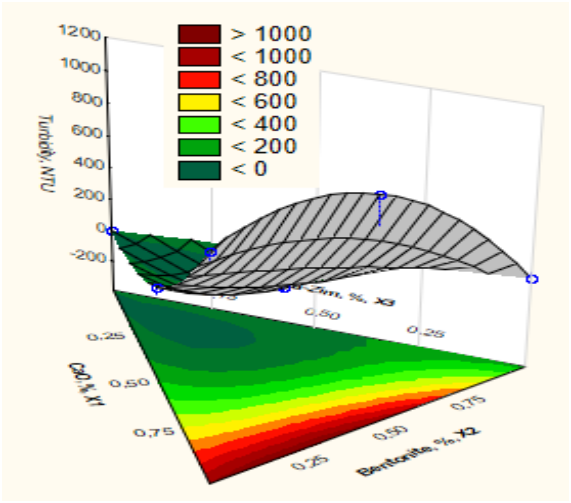
(a)



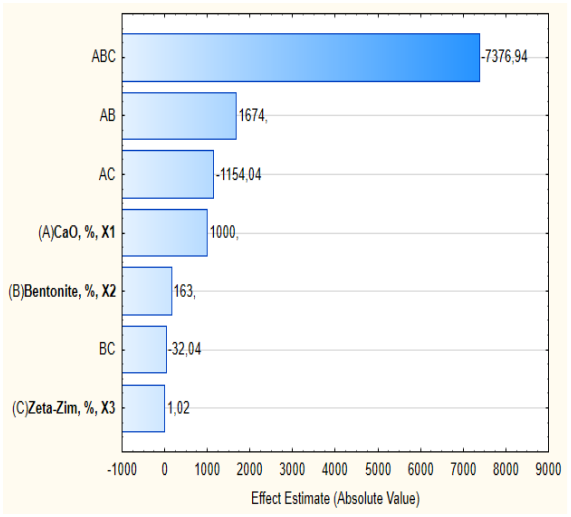
(b)

Figure 2: Terminal surface: a) and Pareto diagram, b) of water turbidity on the 21th day of exposure to factors of the organo-mineral complex.

Complete optical transparency (turbidity ≤ 1 NTU) is attained after 28 days of exposure, as shown in Figure 3a. This condition is achieved at quicklime contents ranging from 0 to 17.5% (X1), bentonite contents from 0 to 15% (X2), and enzyme contents from 0 to 10% (X3).



(a)



(b)

Figure 3: Terminal surface: a) and Pareto diagram, b) of water turbidity on the 28th day of exposure to factors of the organo-mineral complex.

Superimposition of the three ternary response surfaces at the transparency threshold ($T \rightarrow 0$), the optimal range of the organo-mineral complex composition was determined. It is defined by quicklime content from 2.2 to 4.2% ($X1$), bentonite from 7.5 to 11.25% ($X2$) and enzyme from 6 to 10% ($X3$). This composition ensures the production of transparent water with turbidity not exceeding 1 NTU.

4.2 Determination of the Plastic Strength of Sediment

Analogous regression modelling was applied to the sediment plastic strength data (Table 4), yielding (4)-(6) for the three ageing periods:

$$Ts^{14d}=3,77X1+0,12X2+6.7E^{-16}X3-3,06X1X2-7,14X1X3+0,16X2X3+8,61X1X2X3 \quad (4)$$

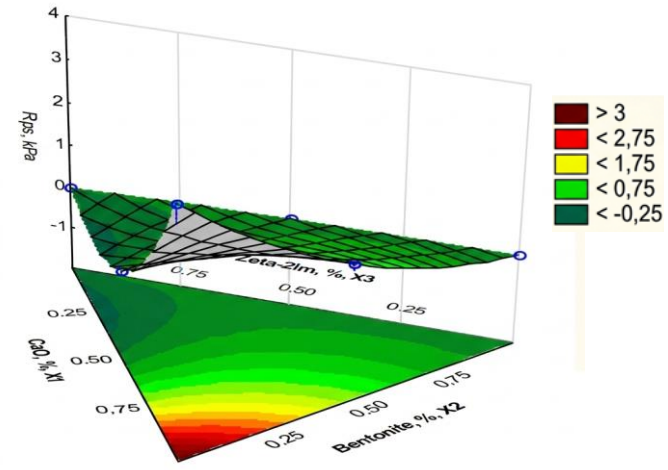
$$Ts^{21d}=57,86X1+0,8X2+1.01E^{-14}X3-68,36X1X2-112,92X1X3-1,12X2X3+45,18X1X2X3 \quad (5)$$

$$Ts^{28d}=160X1+1,41X2+1.05E^{-14}X3-91,38X1X2-314,04X1X3+4,94X2X3-198,87X1X2X3 \quad (6)$$

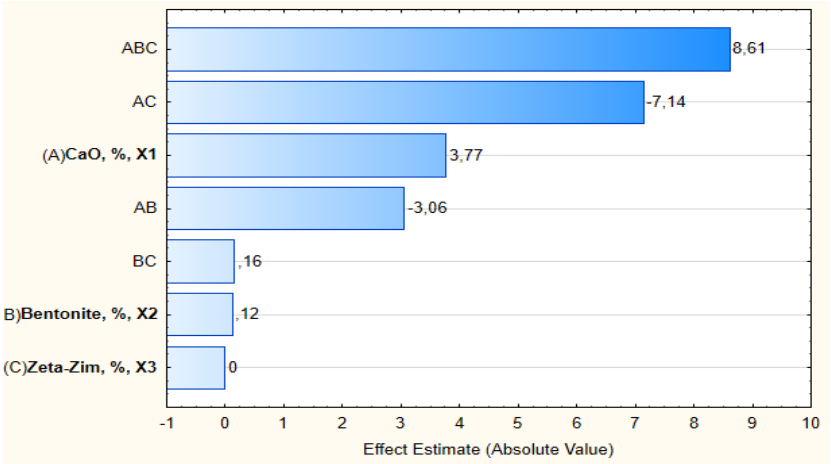
Ternary strength response surfaces (Fig. 4–6) were derived from the regression equations using the factor designations in Table 2. Point 5 as presented in Table 4 did not form sediment and was therefore assigned a value of 0. Analysis of (4)-(6) shows that plastic strength after 14, 21, and 28 days is primarily governed by factors X1, X2, and X3, which account for up to 80 % of the variation as shown in Figures 4b-6b.

Figure 4a shows that by day 14, sediment plastic strength increases from 0.1 to 3.77 kPa as a function of quicklime content (*X1*) ranging from 26.5 to 35%, the bentonite (*X2*) from 0 to 3.75% and the enzyme (*X3*) from 0 to 2%.

At day 21 (Fig. 5a), plastic strength rises from 0.12 to 57.86 kPa, governed by quicklime content (*X1*), ranging from 27 to 35%; the bentonite (*X2*) from 0 to 4.25%; and the enzyme (*X3*) from 0 to 2%.

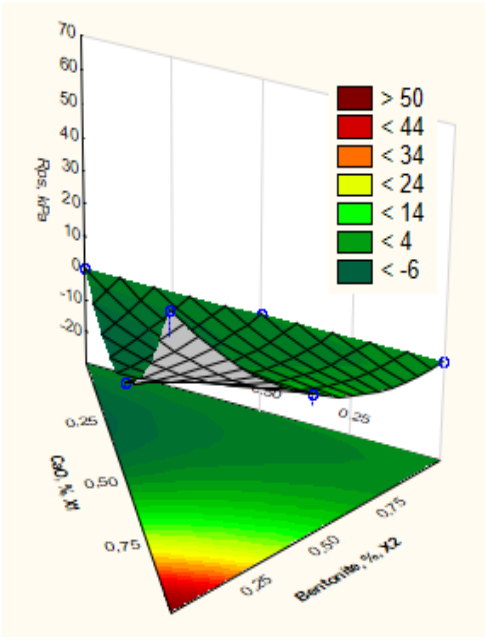


(a)

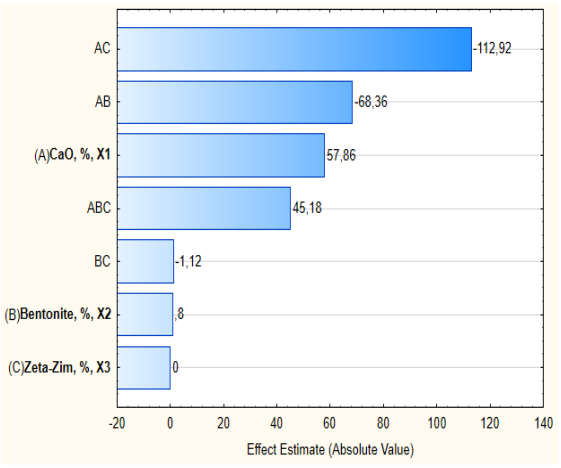


(b)

Figure 4: Ternary surface: a) and Pareto diagram, b) of the influence of organo-mineral complex factors on the plastic strength of sediment after 14 days of agin.



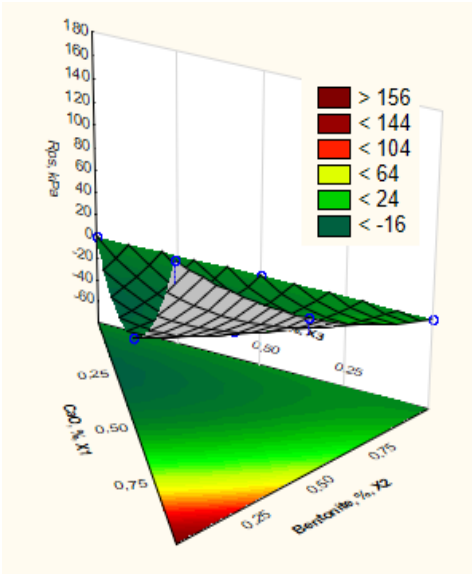
(a)



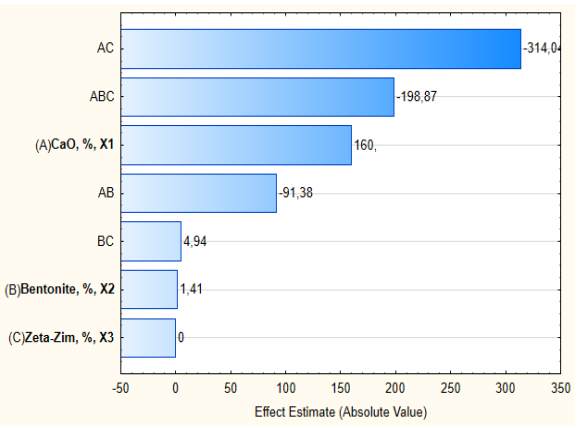
(b)

Figure 5: Ternary surface: a) and Pareto diagram, b) of the influence of organo-mineral complex factors on the plastic strength of sediment after 21 days of aging.

The maximum recorded strength of 160 kPa is attained at day 28 (Fig. 6a), with quicklime spanning 27.5–35% (*X1*), the bentonite from 0 to 5.5% (*X2*) and the enzyme from 0 to 2% (*X3*). The isolines of maximum plastic strength values are located in the lower left corner of the factor space as seen in Figures 4a-6a.



(a)



(b)

Figure 6: Ternary surface: a) and Pareto diagram, b) of the influence of organo-mineral complex factors on the plastic strength of sediment after 28 days of aging.

5 CONCLUSIONS

This study applied experimental-statistical mixture modelling to determine the optimal compositions of a bentonite–Zeta-Zim organo-mineral complex for technically polluted storage pond water, extending the reagent comparison initiated in [13].

Overlay of the three strength response surfaces at the maximum strength criterion $Ts \rightarrow max$, the feasible range of the organo-mineral complex composition was determined. It is defined by quicklime content from 30 to 35% (X1), bentonite from 0 to 2.5% (*X2*) and enzyme from 0 to 1.5% (*X3*). Compositions within the range ensure the formation of artificial stone with a plastic strength of 160 kPa.

For removing turbidity (down to 1 NTU) and for forming mechanically stable sediments. Transparent water was achieved using a mixture of 2.2–4.2% quicklime, 7.5–11.25% bentonite and 6–10% enzyme. Sediments with a plastic strength of up to 160 kPa after 28 days were obtained using a mixture of 30–35% quicklime, ≤2.5% bentonite and ≤1.5% enzyme. The resulting sediment is self-compacting and waterproof, and can be used to construct embankments for storage ponds at solid waste landfills.

ACKNOWLEDGMENTS

This work was supported by the Horizon Europe project (Grant No. 101131382, CLEANWATER).

REFERENCES

[1] J. A. Hargreaves, “Control of clay turbidity in ponds,” SRAC Publication No. 460, 1999.

[2] R. Hoess and J. Geist, “Effect of fish pond drainage on turbidity, suspended solids, fine sediment deposition and nutrient concentration in receiving pearl mussel streams,” in Environmental Pollution, vol. 274, p. 116520, 2021, [Online]. Available: https://doi.org/10.1016/j.envpol.2021.116520.

[3] C. S. Thaxton and R. A. McLaughlin, “Sediment capture effectiveness of various baffle types in a sediment retention pond,” in Transactions of the ASAE, vol. 48, no. 5, pp. 1795-1802, 2005, [Online]. Available: https://doi.org/10.13031/2013.20013.

[4] M. M. Said and N. O. Msuya, “Effects of coagulant dosage, particle size, and settling time on pond water treatment with cactus pads and watermelon seeds,” in Tanzanian Journal of Science, vol. 50, no. 2, pp. 253-268, 2024, [Online]. Available: https://doi.org/10.4314/tjs.v50i2.7.

[5] W. Wang et al., “Removal performances of turbidity, organics, and NH4+-N in a modified settling tank with rotating biological discs used for enhancing drinking water purification,” in Water, vol. 14, p. 4066, 2022, [Online]. Available: https://doi.org/10.3390/w14244066.

[6] J. J. Kang, J. W. Vetter, and R. A. McLaughlin, “Chemical treatment to reduce turbidity in pumped construction site water,” in Journal of Environmental Engineering, vol. 144, no. 12, pp. 04018120-1-04018120-7, 2018, [Online]. Available: https://doi.org/10.1061/(ASCE)EE.1943-7870.0001498.

[7] Y. Yuan et al., “Innovative adaptation of coagulation-sedimentation-filtration process in lightly polluted urban rivers with seasonal high turbidity,” in Scientific Reports, vol. 15, p. 20430, 2025, [Online]. Available: https://doi.org/10.1038/s41598-025-09223-4.

[8] Ch. Igwegbe, O. D. Onukwuli, and P. Ch. Onyechi, “Optimal route for turbidity removal from aquaculture wastewater by electrocoagulation-flocculation process,” in Journal of Engineering and Applied Sciences, vol. 15, no. 1, pp. 99-108, 2019.

[9] B. Ersoy, I. Tosun, A. Günay, and S. Dikmen, “Turbidity removal from wastewaters of natural stone processing by coagulation/flocculation methods,” in Clean - Soil, Air, Water, vol. 37, no. 3, pp. 225-232, 2009, [Online]. Available: https://doi.org/10.1002/CLEN.200800209.

[10] T. Dubyniak, P. Mykulyk, V. Nevozhai, V. Bukhovets, and T. Lepkyi, “Mathematical modeling of the clarifier performance for water coagulation,” in Scientific Journal of Ternopil National Technical University, vol. 117, no. 1, pp. 28-41, 2025, [Online]. Available: https://doi.org/10.33108/visnyk_tntu2025.01.

[11] A. Gruchot, K. Kamińska, and A. Woś, “The effects of lime and cement addition on the compaction and shear strength parameters of silty soils,” in Materials, vol. 18, p. 974, 2025, [Online]. Available: https://doi.org/10.3390/ma18050974.

[12] A. Akmataliev, A. Jalilov, S. Abdumomun Uulu, A. Orozalieva, S. R. Muradova, T. Tashkenbayev, Z. Irmatova, and I. Valixanov, “Innovations in hydropower development and their impact on the environmental safety of aquatic systems,” in Proceedings of International Conference on Applied Innovation in IT, vol. 13, issue 5, pp. 493-505, 2025, doi:10.25673/123075.

[13] S. Kadirova, S. Kharatova, A. Sherov, Z. R. Asadov, Z. Mametov, Y. Li, K. Mirzayev, A. A. Uulu, and S. Nasriddinov, “Water-saving technologies and the development of the green economy: an integrated approach,” in Proceedings of International Conference on Applied Innovation in IT, vol. 13, issue 5, pp. 1097-1108, 2025, doi:10.25673/123152.

[14] S. Guzii, A. Tovmachenko, and V. Viter, “Modelling of organo-mineral complexes in the system lime-zeolite-enzyme for purifying technically polluted water from suspended solids and compacting sediment,” Academic Journal. Industrial Machine Building, Civil Engineering, vol. 65, no. 2, pp. 118-125, 2025, doi:10.26906/znp.2025.65.4209.

[15] S. Guzii and N. Klimenko, “Investigation of the influence of enzymes on the physical and mechanical properties of building materials,” in Proc. IV International Water Forum Aqua-Ukraine, International Forum Environmental Technologies, Kyiv, Ukraine, Sep. 19-21, 2006, pp. 425-430.

Hochshule Anhalt
EDITION
ICAIIT

INFORMATION TECHNOLOGY IN EDUCATION

Section 6

APRIL 2026



Proceedings of the International Conference on Applied Innovations in IT

Transformation of Higher Education in the Age of Artificial Intelligence: Analysis of Digital Learning Practices

Daria Lys[1] , Vladyslav Suvorov[2] , Tetiana Borysenko[3], Maryna Tsynova[4] and Svitlana Hryhorenko[5]
[1]*Department of Cultural Studies and Philosophy of Culture, Odesa Polytechnic National University, Shevchenko Avenue 1, 65044 Odesa, Ukraine*
[2]*Department of Power Supply and Power Consumption Management, Odesa Polytechnic National University, Shevchenko Avenue 1, 65044 Odesa, Ukraine*
[3]*Foreign Languages Department, Odesa Polytechnic National University, Shevchenko Avenue 1, 65044 Odesa, Ukraine*
[4]*Department of English Language in Marine Engineering, National University "Odessa Maritime Academy", Didrikhson Str. 8, 65052 Odessa, Ukraine*
[5]*Department of Computerized Systems and Software Technologies, Odesa Polytechnic National University, Shevchenko Avenue 1, 65044 Odesa, Ukraine*
lys.daria@op.edu.ua, suvorov@op.edu.ua, t.i.boryseko@op.edu.ua, marinatsinovaya@gmail.com, s.m.hryhorenko@op.edu.ua

Keywords: Digital Technologies, Artificial Intelligence (AI), Digitalization of Higher Education, Digital Learning Practices, Digital Culture, Transformation of Universities, Distance Learning System (DLS), Moodle, Google Classroom.

Abstract: The article analyzes the transformation of higher education in the era of Artificial Intelligence (AI) in the context of digital learning practices and cultural changes within the university environment. The aim of the study is to understand the impact of AI technologies on institutional models of universities and everyday learning practices. The methodological basis of the work is an interdisciplinary approach that combines cultural and pedagogical analysis with a comparative study of digital educational platforms. Special attention is paid to the role of AI as a tool for personalized learning, learning analytics, and the automation of educational and managerial processes, as well as to issues of ethics, autonomy, and cultural responsibility. The practical part of the article includes a comparative analysis of the distance learning systems Moodle and Google Classroom, using the case of the Odesa Polytechnic National University. It is shown that these platforms represent different models of organizing digital learning: a functionally oriented LMS with a high level of customization, and a cloud-based distance learning system focused on interface convenience and user experience. The results of the study confirm that the integration of AI into higher education is not only a technological but also a cultural process, which requires rethinking academic roles, educational strategies, and the value foundations of universities.

1 INTRODUCTION

1.1 Knowledge Transmission as a Cultural-Historical Phenomenon

Today our cultural heritage and access to information are being reinterpreted through digital archives and Artificial Intelligence (AI)-based analytics. Thanks to Natural Language Processing (NLP) technologies, machines can interact with human language, decode its meaning, and even generate language [1]. This means that today AI-based translation tools help us overcome language barriers, while voice assistants and chatbots make natural language interaction between humans and machines easier [2]. These technological achievements highlight the importance of language not only as a means of cultural communication but also as a way to preserve and ensure the accessibility of linguistic and cultural heritage. For example, AI-based tools can revive and preserve endangered languages [3]. In addition, digital archives and AI-driven analysis can make historical texts and cultural artifacts accessible and understandable to a wider audience of students. However, the inability of automatic speech recognition systems to adequately respond to

linguistic diversity may limit the quality of language input for multilingual students. This shows that AI is not only a tool created by human hands but also a natural extension of human intelligence, language, and culture [4]. From an analytical point of view, the given examples show the transformation of the function of language from a means of communication to a tool of algorithmic mediation. This makes it possible to consider AI as a mediator of cultural experience, influencing the structure of knowledge rather than only its transmission.

1.2 The Relationship Between Technologies and Education

Today screen technologies, mobile applications, social robots, and algorithmic interactions have become integral elements of the cognitive, sociocultural, and emotional development of modern students [5]. The users of such technologies are «digital natives» and «digital immigrants». These concepts were first introduced by Marc Prensky in 2001. According to his definition, digital natives are the generation that grew up in close contact with digital technologies from birth, for example, schoolchildren or students. Digital immigrants, on the other hand, are people who encountered digital technologies only at a certain stage of life and had to master them later, such as teachers. Digital natives tend to perceive information quickly, to take in to consideration polysemy, and learn through audiovisual materials, while digital immigrants prefer more linear and text-centered way of processing information.

It follows that even if we consider such elements of educational technology as theoretical foundations, human resources, methods and techniques, environment, goals, learners, learning situations, and assessment separately, it becomes clear how significant educational technology is for the practice of learning [6]. Liu and co-authors describe modern interaction between generative AI and education through four paradoxes. The first paradox is that AI is both a friend of the educational process and a source of potential danger. The second paradox is that AI is very talented but at the same time extremely dependent: its answers, recommendations, and ideas are limited by the sources available to it. The third paradox is that AI is both accessible to everyone and restricted for most. The fourth paradox is that even if this technology is banned, it will still remain attractive [7]. According to Coghlan, in order to deprive people of autonomy, maturity, and historical memory, no authoritarian leader (Big Brother) is needed. People themselves will worship technologies that take away their ability to think and put pressure on them [8]. Thus, it can be assumed that the effectiveness of digital learning directly depends on the consistency between students' cognitive styles and the digital tools used in the learning process.

2 LITERATURE REVIEW

The connection between AI and education goes much deeper than is usually assumed. The first work devoted to the application of AI in education dates back to 1954. In the 2000s, the number of studies in this field was only 173. In the 2010s, it reached thousands. The 2020s became a period of explosive growth: if in 2020 the number of studies exceeded 10,000, then by 2024 it surpassed 20,000. Thus, AI overcame very quickly its «formation period» and became one of the key actors both in the research and in the practical dimension of educational processes.

In 2023 according to UNESCO, AI is among the key technologies and applications that will shape the future of learning and teaching in higher education [9]. The literature highlights several advantages of using technological learning materials and computer support to improve learning efficiency, which can be grouped into the following directions:

- work with computer-mediated learning (web design, presentation, animation, sound, colour, simulation, use of visual and audio elements) [10];
- use of interactive boards, which provide teachers and students with a wide range of opportunities: they allow teachers to easily manage material, support different teaching styles, facilitate group classes online, and help students better understand abstract topics;
- creation of diverse learning environments, meeting needs related to individual differences, and increasing interest [11];
- provision of reliable observations, facilitation and enrichment of laboratory classes in institutions where materials for experiments are lacking [12];
- support for the immediate application of such learning principles as feedback, error correction, and reinforcement [13];
- giving learners the opportunity to determine their own learning time;
- multiple use of materials and simplification of information presentation [14];
- saving time.

Holmes concludes that student learning requires a transition from an environment where the basis was the book and learner autonomy to a complex and interactive environment that offers vivid sounds, video clips, symbols, and signs [15]. It should be noted that most of these studies were conducted with students, assistants, and teachers, and most importantly - with parents. Insufficient attention to parents, who play an important role in learning processes, may lead to the formation of negative attitudes among students towards the humanities, technical sciences, scientific thinking, and digital technologies [16].

The results of studies of this topic show that computer-mediated learning is most often considered as a more effective method compared to traditional approaches, which contributes to improving students' academic success [17]. It is noted in literature notes that the use of technologies in teaching the humanities, especially for students of technical specialties, increases learners' performance and motivation, and also supports long-term knowledge acquisition [18].

At the same time, in the review by Zhao et al., it is noted that AI has begun to integrate into teaching and learning processes, and three types of pedagogical applications are distinguished: feedback (16 studies), reasoning (10 studies), and adaptive learning (9 studies) [19]. A systematic literature review by Sadiku and co-authors identified four main functions of AI in online higher education: performance prediction, resource recommendations, automatic assessment, and improvement of the learning experience. Thus, information technologies are closely connected with the educational process.

3 METHODOLOGY

The research methodology is based on a combination of qualitative and quantitative approaches. Methods used include comparative analysis to contrast the functional and interface features of Moodle and Google Classroom; UX analysis to evaluate the usability of interfaces according to user behavior patterns; analysis of user sessions with the HotJar tool to record interactions with the system; and generalization and interpretation of empirical data to draw conclusions about the effectiveness of digital platforms. The comparison criteria for LMS were: navigation complexity (number of page transitions); time required to perform typical learning actions; number of clicks needed to achieve a target action; frequency of user difficulties; and the overall level of user experience (UX). The study analyzed a sample of users and their interaction sessions with the system, which made it possible to identify typical scenarios of platform use. In addition, two possible directions of educational transformation were highlighted: conservative and liberal, whose absolute guidelines in pedagogical practice are quite evident. During the analysis, the advantages and limitations of each were outlined.

3.1 Paternalism or Conservative Education and Libertarianism or Free Education

An educational concept can be formed through the expression of beliefs, principles, and broad strategies. For example, the main goal of education is to prepare for the economic success of the nation. Therefore, all students must acquire the skills necessary to work with information technologies. At the same time, society should not allow division. Social integration can be ensured by the fact that all students follow the same curriculum. Thus, a standard program must be adopted. Educational institutions should be responsible, and teachers should be controlling.

However, with a paternalistic understanding of education, society may face at least the problem of the «ideology of computer literacy», which Ahmad and Rahmat considered relevant [20]. Its main points can be summarized as follows: computers are everywhere; the need for computer skills in work is growing; therefore, everyone must study computer disciplines in the first years of university. Yet most employment forecasts indicate that only a small number of workers need to be specialists in the field of information technologies, and this minority should have higher education. The rest of the workers who use computers do not need skills beyond following instructions on the screen.

Libertarianism seeks to soften some consequences of postmodernism. Achieving this goal is possible only through the formation of an alternative view, for example:

1) The main goal of education is to prepare individuals who will become «smart clients». Information technologies will constitute the main market, managed by individuals. Therefore, they should be at the center of youth education;
2) In the future, individualism will dominate. Thus, imposing the same curriculum on everyone has lost relevance. Each student will follow a personal learning program, for example, elective courses based on computer

educational environments and virtual communities;

3) As a result, most learning will take place at home, that is, online. Nevertheless, some universities will continue to exist. They will be known as centers of lifelong education and will perform the function of safe provision of online learning.

Perelman, the author of «School's Out», believes that this means the approaching end of the educational empire. That is, if we want to know what will destroy the American academic empire in the next ten years, here are the key words: videotelephony, individual communication systems, virtuality, and so on. An example can be Hartley's observation, who argues that it is increasingly difficult for students to sit for 90 minutes, concentrating on a single source of information without the possibility of choice and control. However, the flow of media messages, which makes consumption and focus on oneself and the present moment attractive, may lead not so much to a «desire for autonomy» as to deep intellectual, sociocultural, and moral weakening.

At the same time, Neil Postman in his book «Amusing Ourselves to Death» notes that from Confucius to Plato, Cicero, Locke, and Dewey, philosophers of education emphasized that no one ever claimed or even suggested that education turned into entertainment ensured meaningful, effective, and sustainable success. Cultural integration is complex, as it requires overcoming obstacles.

3.2 Integration of AI-Supported Systems into Educational Activities

Especially after the COVID-19 pandemic and now war conditions in Ukraine, the acceleration of digitalization has encouraged universities to create more flexible, data-driven, and student-oriented structures. In this context, the integration of AI-supported systems into educational activities has become an inevitable necessity. For all participants - students, teachers, administrative staff, and researchers - AI offers functional solutions in such key areas as learning, assessment, research, and decision-making.

Adaptive learning systems analyze students' performance levels, learning speed, styles, and interests, turning educational processes into more effective, flexible, and student-centered ones. Unlike the classical «one program for all» approach, such systems take into account the cognitive and affective profile of each student and provide differentiated learning materials. This increases students' motivation and strengthens learning outcomes. Data obtained through AI-based systems, when analyzed within learning analytics, contribute to the development of learning materials, the identification of students' strengths and weaknesses, and the restructuring of general educational strategies. When processing student data, it is necessary to consider the principles of privacy, transparency, and student autonomy. Learning analytics should be structured not only to measure performance but also to support the holistic development of students.

Such approaches reduce the workload of teachers related to routine assessment and allow them to focus on more meaningful tasks, such as pedagogical planning, group interaction, and strategy development. Timeliness of feedback is also an important factor. Hasibuan and Azizah note that immediate feedback is effective in preventing incorrect learning [21], while Wang emphasizes its role in supporting self-regulation and internal motivation [22]. According to the self-determination theory of Veytia Bucheli and Gomez-Galan, such practices deepen the learning process by supporting students' sense of autonomy and competence [23]. For example, chatbots based on natural language processing provide round-the-clock support in administrative matters such as academic advising, registration, scheduling, and informing about exam dates. These systems reduce administrative workload while enriching the student experience through user-friendly approaches. They improve the quality of digital services and play an important role in combating digital inequality.

It should be noted that modern LMS integrate artificial intelligence tools, in particular: automated grading, student support chatbots, and/or recommendation systems for learning materials. For example, in Google Classroom the integration with Google AI allows tasks to be checked automatically, while Moodle supports AI plugins for learning activity analytics. Thus, the UX design of LMS should be considered in connection with the possibilities of AI integration. AI-based applications can analyze class attendance, assignment deadlines, exam results, and student behavior on online platforms, forming early warning systems. Such systems identify at-risk students, inform teachers, and provide personalized support. In addition, these data offer valuable feedback for updating learning design and improving assessment methods. In this context, the role of the teacher is transformed: he or she ceases to be only a transmitter of knowledge and becomes a mentor who stimulates social interaction and critical

thinking. AI takes over analytical and administrative processes, while educators focus on creating creative and cooperative learning environments in the digital space.

3.3 Analysis of the Interface of the Moodle Distance Learning System in Ukraine by Odesa Polytechnic National University Example

In Ukraine, the development of the distance learning system (DLS) started much later than in Western European countries and became popular during the Covid-19 pandemic and now in war conditions. The introduction of distance learning in Ukrainian educational institutions marked the beginning of raising the level of informatization of Ukrainian society and increasing the number of institutions equipped with computer technology. The Odesa Polytechnic National University, like most universities in the world, uses the most popular system Moodle as its DLS, which directly influences the quality of student learning. If a student does not have the opportunity to interact with the system, he or she loses motivation to use it as a tool for organizing self-education, which leads to lower academic performance and, accordingly, to a decrease in the relevance of studied disciplines for them.

It is appropriate to consider the interface of the Moodle DLS from the perspective of UX design. UX design is a broad field of knowledge that describes the feelings arising in users` interaction with a system. When designing digital products, three factors are considered: form, behavior, and content. Moodle DLS is a website that can be accessed from different places with Internet connection. The audience of this website is specific - students and teachers of the university. From the point of view of form and content, the useful component is considered first, that is, how many learning materials from each discipline will be placed, how the learning process will be organized (whether it will be a distance course consisting only of lecture notes and methodological support, or whether interactive and assessed elements such as practical tasks, testing, surveys, etc. will also be included). It can be concluded that the appearance of the website (form) is practically not considered as a factor influencing students' interaction with the system.

Moodle DLS has the ability to adapt and build on already prepared graphic solutions (themes). The Adaptable theme has been applied to the considered DLS. Although the appearance of the website is not a fundamental factor of analysis, Adaptable is a flexibly customizable adaptive two-column theme designed for use both in large university institutions and in small training companies. The DLS interface analysis to improve its quality for students should be conducted using data directly from the students themselves. There are enough tools that help track user behavior on a website. For this study, the HotJar service was used. HotJar is a behavior analytics service that helps Internet companies understand user behavior on a website. One of the popular tools of HotJar is the recording of user sessions on the site, which demonstrates their behavior (number of clicks, mouse movements, etc.). The analysis of student behavior on the site is usually conducted with different pages of the site: the main page, the course page, the test page. The data are video files, which are later analyzed manually. The study was conducted within one course that involved 128 users (students) and 742 sessions of interaction with the Moodle system. The data were collected using the HotJar tool, which records user behavior indicators such as the number of clicks, cursor movement, and navigation paths. The average session duration was 12.4 minutes (SD = 5.8), the average number of clicks was 18.7, and the average time to search for learning material was 34 seconds. In 41.6% of cases, students made repeated transitions between pages, which indicates navigation complexity. The obtained results allow Moodle to be interpreted as a platform with increased navigation difficulty and serve as an empirical basis for its comparison with alternative digital environments, particularly Google Classroom. Thus, students use the Moodle distance learning system differently, offered as a tool for organizing online learning. Nevertheless, not all students can use the system effectively and spend more time searching for information on the site. Therefore, Moodle DLS requires interface optimization, reduction of the number of links leading directly to course pages.

3.4 Analysis of the Google Classroom Interface on the Example of the Licensed Platform of the Odesa Polytechnic National University

Alongside the Moodle system, the Odesa Polytechnic National University also uses Google Classroom in its educational process, which performs the functions of a distance learning system (DLS). Google Classroom is a cloud-based educational platform integrated into the Google services ecosystem and operates on the basis of corporate accounts for teachers (@op.edu.ua), students (@stud.op.edu.ua), and postgraduates (@as.op.edu.ua). The use of corporate

email ensures centralized access, control, and protection of educational content, as well as the creation of a unified digital educational space of the university. From a regulatory perspective, Google Classroom can be classified as a DLS, since it provides the organization of the learning process, communication between participants, distribution of learning materials, and monitoring of task completion in an online environment. Unlike Moodle, which is a classical Learning Management System (LMS) with wide customization and functional expansion options, Google Classroom is oriented toward a simplified model of managing the learning process, focusing on quick interaction, transparent course structure, and intuitive navigation.

The Google Classroom interface was initially designed with UX principles and mass users in mind. The course structure is built according to the logic of the "Stream", as well as the tabs "Assignments", "People" and "Grades", which reduces the cognitive load on the user and minimizes the number of navigation actions needed to find learning materials. Unlike Moodle, where the depth of page nesting and the number of transitions can complicate orientation, Google Classroom offers a more linear and visually clear interaction model with fewer transitions between interface elements.

A particularly important factor for the effective use of Google Classroom in the university environment is the previous user experience of students. Most learners have already encountered this system at the level of general secondary education, which forms basic digital skills in working with the interface, assignments, and comments. Thus, the adaptation period for students is minimal, and the platform itself is perceived as a familiar tool for organizing learning. This positively influences motivation, engagement, and regularity in completing academic tasks.

Another important functional element of Google Classroom is the grading system, which is oriented toward flexible management of learning outcomes and step-by-step informing of students. The platform allows teachers to preliminarily assign grades for unsubmitted or unassessed tasks. These grades are stored in the internal system and are not displayed to students until their official publication, which makes it possible to adjust results, consider additional attempts or retakes, and build more individualized model of knowledge control.

Google Classroom also includes a built-in grade calculator that automatically calculates the final result for a distance course based on completed tasks. The teacher can choose a grading system, for example, points, percentages, letters, etc., and also decide whether the final grade will be displayed to students during the learning process or only after the course is completed. This approach allows the grading system to be adapted to the requirements of a specific discipline and educational program.

Among other significant features of Google Classroom, it is worth noting its deep integration with Google Workspace services (Google Docs, Slides, Sheets, Drive, Meet). This makes it possible to organize collective and individual work in real time, automate the process of submitting assignments, and minimize technical barriers during the review of student work. All changes in documents are recorded by the system, which facilitates the control of academic integrity and tracking of each student's contribution. At the same time, the functionality of comments and private messages ensures asynchronous communication between teachers and students, including the possibility of leaving private remarks on specific tasks. This strengthens feedback and supports the personalization of the educational process. In addition, the system supports automatic notifications that inform students about new assignments, deadlines, and published grades. Compared to Moodle, which recorded a higher average number of clicks and navigation transitions, the interface of Google Classroom requires fewer navigation actions and offers a more linear structure of interaction.

4 RESULTS AND DISCUSSION

4.1 Comparative Analysis of the Use of Moodle and Google Classroom in the University Educational System

The practical analysis of LMS is considered as an example of implementing digital educational environments in which artificial intelligence technologies are potentially integrated. It should be noted that detailed quantitative behavioral data were available only for the Moodle platform. The analysis of Google Classroom was carried out on the basis of structural and functional comparison of interfaces, as well as indirect UX indicators (in particular, the number of navigation actions, the logic of the interface, and the previous user experience of students). When comparing the use of Moodle and Google Classroom, it is appropriate to speak about two types of distance learning systems (DLS) applied in the university:

1) A functionally rich LMS (Moodle), oriented toward comprehensive management of educational courses, detailed analytics, and multi-level knowledge control;
2) A simplified cloud-based DLS (Google Classroom), focused on quick communication, fast organization of the learning process, and ease of use.

From a practical point of view, Google Classroom is perceived as a system that is easier to master, mainly due to its familiar interface and students' previous school experience. This is reflected in the reduction of navigation actions and the shorter time needed to access learning materials compared to Moodle. At the same time, Moodle retains its importance as a universal platform for implementing complex distance courses. The results of the comparative analysis of Moodle and Google Classroom in the university educational system are presented in Table 1.

According to the observations, 65% of students prefer Google Classroom, while 35% use Moodle only when necessary. At the same time, the parallel use of two distance learning systems allows the Odesa Polytechnic National University to flexibly adapt to different educational tasks, combining the functional depth of Moodle with the user accessibility of Google Classroom. It should be noted that Google Classroom, together with its intuitive interface and students' previous user experience, is an effective and methodologically appropriate tool of distance learning in university practice, especially during the period of war.

To summarize the results of the comparative analysis, a radial diagram (Fig. 1) was used, which reflects qualitative evaluations of the platforms. This is a visual analytical tool that shows the relative advantages and limitations of Moodle and Google Classroom according to key UX and didactic criteria. The values of the indicators serve to synthesize qualitative observations obtained during the analysis of the interface and user experience, reflecting relative differences between the platforms. They are intended to generalize qualitative conclusions and are not aimed at providing quantitative measurement.

Table 1: Analytical and criteria-based comparison of the Moodle and Google Classroom distance learning systems.

Analytical Criterion	Moodle	Google Classroom
Type of Platform	Functionally rich open-source LMS	Cloud-based DLS
Educational Paradigm	Management of complex, structured courses	Operational organization of the learning process
Interface Architecture	Hierarchical, multi-level structure	Linear, stream-like structure
UX Design	A higher average number of page transitions (several levels of nesting)	A lower average number of transitions (limited number of transitions)
AI integration	Limited (through plugins)	Embedded (Google AI ecosystem)
Cost of Use and Integration with Other Services	Free of charge, but requires expenses for server and administration	Free of charge as part of Google Workspace for Education
Scalability	Suitable for large universities	Better suited for individual courses
Adaptation to Students' User Experience	Many features, but a higher level of navigation complexity; requires additional training and adaptation	Lower level of navigation complexity, based on previous school experience
Convenience for Teachers	Needs training for effective use	Intuitively easy to use
Organization of Learning Materials	Flexible, but difficult to perceive	Unified and transparent
Assessment System	Multi-level, with extensive customization options	Automated, with preliminary calculation of results
Learning Activity Analytics	Implemented through external tools	Built-in mechanisms for activity and deadlines
Functionality	Very broad: courses, tests, forums, analytics, plugins	Limited, but sufficient for basic learning
Typical Use Scenarios	Complex distance courses with multi-level control	Current organization of learning and communication
Preliminary Assessment	Limited or implemented through plugins	Built-in function of preliminary assessment
Main Limitations	Overloaded interface, complicated navigation	Limited advanced LMS functions

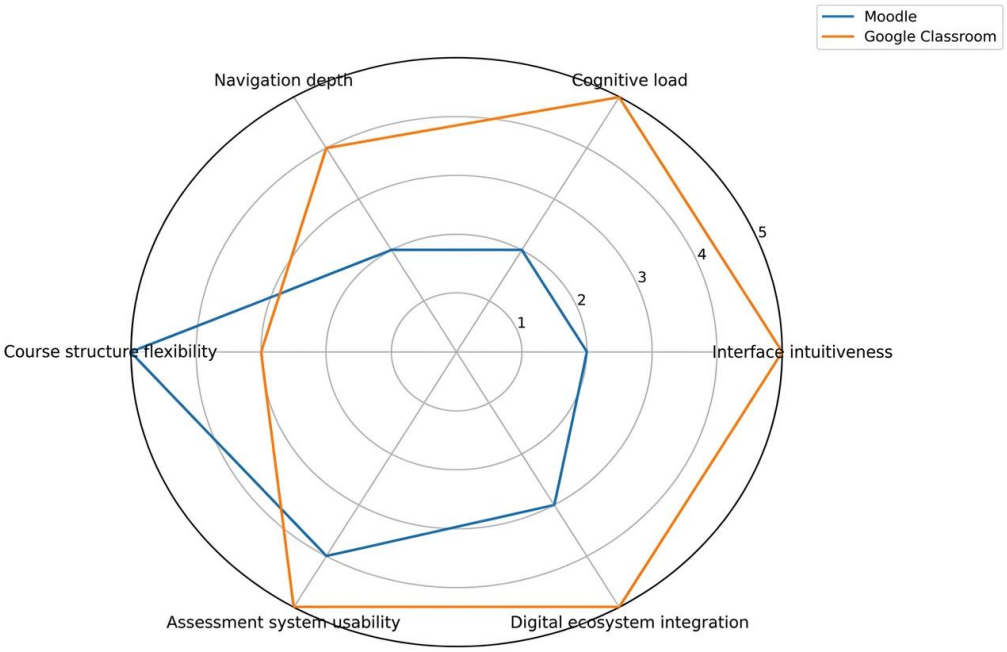


Figure 1: Radial diagram of comparative UX and functional characteristics of Moodle and Google Classroom.

4.2 Recommendations for the Implementation of AI in Higher Education

For the effective integration of AI into the higher education system, it is extremely important to introduce comprehensive training programs and internships for teachers aimed at developing their AI literacy. For example, the International Internship «Digital Future: Blended Learning» DigIn.Net2. Such programs should provide knowledge and skills related to the use of AI tools in a pedagogical context. Including courses on the topic «Artificial Intelligence and Education» in the curricula of teacher training institutes may help increase the awareness of future educators about the application of AI in the educational sphere. For example, this may include the creation of AI workshops as well as a university center for digital competences.

5 CONCLUSIONS

The conducted study was aimed at understanding the transformation of higher education in the era of AI in the context of digital learning practices, cultural changes, and institutional transformations of the university environment. Based on interdisciplinary analysis, it was shown that the introduction of AI into the educational sphere cannot be considered only as a technological update but requires a conceptual and value-based rethinking of the goals and functions of education.

The paternalistic model presented in the article emphasizes the role of standardization, control, and institutional leadership in preparing specialists for the information economy but carries the risk of reducing the educational process to the instrumental acquisition of technologies. In contrast, the libertarian approach highlights the individualization and autonomy of learning but underestimates the social-integrative function of education and may contribute to the fragmentation of cultural and intellectual space. Therefore, there is a need to search for a balanced model in which technological innovations are combined with the cultural and ethical foundations of education.

It has been shown that AI performs the functions of a tool for personalized learning, automated assessment, analytical support for decision-making, and the development of student services. At the same time, it is emphasized that the effectiveness of these processes directly depends on the level of digital and AI competence of teachers, as well as on the institutional readiness of universities for cultural change. Therefore, attention should be paid to preparing teachers for the correct and appropriate use of technological learning materials in the educational environment.

The comparative analysis of Moodle and Google Classroom revealed that these platforms represent two different approaches to organizing distance learning and showed that the complexity of Moodle's interface and navigation may reduce students' motivation and the effectiveness of self-learning, while Google Classroom, thanks to its intuitive UX design and students' previous user experience, promotes greater engagement and regularity of learning activity. The results show the need to consider not only the functional characteristics of the platforms but also the cultural, cognitive, and behavioral features of users in the process of optimizing digital educational environments. The UX analysis data open up opportunities for improving the integration of AI into LMS, particularly in the context of learning personalization and interaction automation. In the war conditions in Ukraine, the hybrid use of different DLS has become a stable educational strategy. The study has certain limitations. First, quantitative UX data were collected only for Moodle, while the analysis of Google Classroom is partly based on qualitative comparison. Second, the sample is limited to one course and one higher education institution, which may affect the generalizability of the results.

In conclusion, it has been emphasized that AI in education is not an end, but a tool whose effectiveness is determined by the presence of a clear philosophical position, ethical guidelines, and a well-thought-out pedagogical strategy. The prospects for further research are connected with a deeper analysis of the interaction between AI, UX design of educational platforms, and the digital culture of universities, as well as with empirical studies of the impact of AI-supported systems on students' academic performance, motivation, and cultural socialization. It should be remembered that AI is not a goal, but a tool for improving learning and management processes. The correct use of this tool is our shared responsibility.

REFERENCES

[1] D. A. Lys, "Evolution of communication technologies: The socio-cultural transformation of humans into digital users," in Culturological Almanac, vol. 1, pp. 352-359, 2025, [Online]. Available: https://doi.org/10.31392/cult.alm.2025.1.40.

[2] C. Boscardin, B. Gin, P. Golde, and K. Hauer, "ChatGPT and generative artificial intelligence for medical education: Potential impact and opportunity," in Academic Medicine, vol. 99, no. 1, pp. 22-27, 2023, [Online]. Available: https://doi.org/10.1097/acm.0000000000005439.

[3] A. D. Ismoilovna, "Harnessing AI tools to enhance foreign language acquisition: Innovations and impacts," in American Journal of Social Sciences and Humanity Research, vol. 5, 2025, [Online]. Available: https://doi.org/10.37547/ajsshr/volume05issue01-07.

[4] V. Heilala, R. Araya, and R. Hamalainen, "Beyond text-to-text: An overview of multimodal and generative artificial intelligence for education using topic modeling," in Computers & Education, vol. 203, 2024, [Online]. Available: https://doi.org/10.1145/3672608.3707764.

[5] D. A. Lys, "The contemporary human body as a social category: The concept of the 'digital body' in virtual reality," in Global World: Contemporary Challenges: Collective Monograph, pp. 20-27, 2025, [Online]. Available: https://doi.org/10.61718/mon202501.

[6] M. C. Laupichler, A. Aster, J. Schirch, and T. Raupach, "Artificial intelligence literacy in higher and adult education: A scoping literature review," in Computers and Education: Artificial Intelligence, vol. 3, 2023, [Online]. Available: https://doi.org/10.1016/j.caeai.2022.100101.

[7] M. Liu, Y. Ren, L. Nyagoga, F. Stonier, Z. Wu, and L. Yu, "Future of education in the era of generative artificial intelligence: Consensus among Chinese scholars on applications of ChatGPT in schools," in Future in Educational Research, vol. 1, no. 1, pp. 72-101, 2023, [Online]. Available: https://doi.org/10.1002/fer3.10.

[8] S. Coghlan, T. Miller, and J. Paterson, "Good proctor or 'Big Brother'? AI ethics and online exam supervision technologies," arXiv preprint, 2020, [Online]. Available: https://arxiv.org/abs/2011.07647.

[9] UNESCO, "Harnessing the era of artificial intelligence in higher education: A primer for higher education stakeholders," 2023, [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000386670.

[10] B. Qiu, Y. Zhu, L. Du, and C. Feng, "Analysis and reflection on the teaching application of artificial intelligence technology in the context of big data," in Curriculum and Teaching Methodology, vol. 7, no. 4, 2024, [Online]. Available: https://doi.org/10.23977/curtm.2024.070404.

[11] D. Macaday-Quioco, "The prevalent skills and competencies of emotional intelligence for effective educational leadership: A systematic review of literature," in International Journal of Multidisciplinary Research and Analysis, vol. 7, no. 10, 2024, [Online]. Available: https://doi.org/10.47191/ijmra/v7-i10-22.

[12] M. Nasir, M. Hasan, A. Adlim, and M. Syukri, "Utilizing artificial intelligence in education to enhance teaching effectiveness," in Proc. ICE, vol. 2, no. 1, pp. 280-285, 2024, [Online]. Available: https://doi.org/10.32672/pice.v2i1.1367.

[13] L. K. Nhan, "Exploring students' perceptions of AI-powered assistants in enhancing English speaking proficiency," in International Journal of Innovative Science and Research Technology, vol. 9, no. 9, 2024, [Online]. Available: https://doi.org/10.38124/ijisrt/ijisrt24sep792.

[14] A. Paida, "The role of AI technology in enhancing students' writing creativity: A case study on Indonesian language learning in secondary schools," in International Journal for Multidisciplinary Research, vol. 6, no. 6, pp. 1-7, 2024, [Online]. Available: https://doi.org/10.36948/ijfmr.2024.v06i06.32998.

[15] W. Holmes, M. Bialik, and C. Fadel, Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Boston: Center for Curriculum Redesign, 2021, [Online]. Available: https://curriculumredesign.org/wp-content/uploads/AIED-Book-Excerpt-CCR.pdf.

[16] W. Holmes, M. Bialik, and C. Fadel, "Ethical principles for artificial intelligence in education," in Education and Information Technologies, vol. 27, pp. 2941-2960, 2022, [Online]. Available: https://doi.org/10.1007/s10639-022-11316-w.

[17] M. Muthukrishna, J. Dai, D. P. Madrid, R. Sabherwal, K. Vanoppen, and H. Yao, "AI can revolutionise education but technology is not enough: Human development meets cultural evolution," in Journal of Human Development and Capabilities, 2025, [Online]. Available: https://doi.org/10.1080/19452829.2025.2517740.

[18] I. U. Begum, "Role of artificial intelligence in higher education - An empirical investigation," in International Research Journal on Advanced Engineering and Management, vol. 2, no. 3, pp. 49-53, 2024, [Online]. Available: https://doi.org/10.47392/irjaem.2024.0009.

[19] H. Zhao, H. Zhang, J. Li, and H. Liu, "Performance motivation and emotion regulation as drivers of academic competence and problem-solving skills in AI-enhanced preschool education: A SEM study," in British Educational Research Journal, pp. 1-22, 2025, [Online]. Available: https://doi.org/10.1002/berj.4196.

[20] S. Ahmad, M. Rahmat, M. Mubarik, M. Alam, and S. Hyder, "Artificial intelligence and its role in education," in Sustainability, vol. 13, no. 22, 2021, [Online]. Available: https://doi.org/10.3390/su132212902.

[21] R. Hasibuan and A. Azizah, "Analyzing the potential of artificial intelligence (AI) in personalizing learning to foster creativity in students," 2023, [Online]. Available: https://doi.org/10.61996/edu.v1i1.2.

[22] M. Wang, A. Tlili, M. K. Khribi, C. K. Lo, and R. Huang, "Generative artificial intelligence in special education: A systematic review through the lens of the mediated-action model," in Information Development, 2025, [Online]. Available: https://doi.org/10.1177/02666669251335655.

[23] M. G. Veytia Bucheli, J. Gomez-Galan, M. L. Caceres Mesa, and L. Lopez Catalan, "Digital technologies as enablers of universal design for learning: Higher education students' perceptions in the context of SDG4," in Discover Sustainability, vol. 5, p. 473, 2024, [Online]. Available: https://doi.org/10.1007/s43621-024-00699-0.

Information Technologies in the Digital Ecosystem of Higher Education Institutions

Iryna Polishchuk[1], Yuriy Dovhan[1], Oksana Yakushevska[1], Larysa Dovhan[1], Oleksandr Nepochatenko[2] and Viktoriia Petrova[1]
[1]*Vinnytsia Institute of Trade and Economics, State University of Trade and Economics, Soborna Str. 87, 21050 Vinnytsia, Ukraine*
[2]*Uman National University, Instytutska Str. 1, 20301 Uman, Ukraine*
i.polishchuk@vtei.edu.ua, y.dovhan@vtei.edu.ua, o.yakushevska@vtei.edu.ua, l.dovhan@vtei.edu.ua, nepochatenko2@gmail.com, v.petrova@vtei.edu.ua

Keywords: Information Technologies, Digital Ecosystem, Digital Transformation, Higher Education, Learning Management System, SEL System.

Abstract: The article examines the issues of modern information technologies (IT) and their application in higher education institutions. The purpose of the article is to analyze the main areas of application of IT in the digital ecosystem of a higher education institution, covering educational process management, teaching and learning, and scientific research. Comparative, systemic, and structural-functional methods are used to conduct the study. The evolution of IT in education is briefly highlighted from the invention of the printing press to the formation of modern digital ecosystems. The main areas of IT application in higher education institutions are as follows: automation of management and administrative processes; computerization of the educational process; creation of an interactive learning environment; and integration of information and communication technologies into scientific and research activities. Modern IT solutions for each area are described. The practical experience of the Vinnytsia Institute of Trade and Economics of the State University of Trade and Economics is presented as a case study of an integrated digital ecosystem built on its automated management system "Smart Electronic Learning," and the Moodle educational platform. It is emphasized that the maximum effect of digital transformation is achieved not through the mechanical implementation of individual technologies, but their organic integration into a single digital ecosystem. An integrated approach to implementing IT ensures data integration, eliminates information duplication, and creates a synergistic effect for all participants in the educational process.

1 INTRODUCTION

Digital transformation of higher education is one of the key development directions changing traditional educational paradigms, creating new opportunities for managing educational institutions, organizing the educational process, and conducting scientific research. The formation of digital ecosystems in higher education is becoming an essential condition for ensuring the competitiveness of higher education institutions (HEIs), improving the quality of educational services, and enhancing the efficiency of all subsystems' operations. In the context of globalization and rapid technological change, the importance of the comprehensive implementation of IT across all HEI activities is continually growing.

2 LITERATURE REVIEW

The international scientific community has actively studied the issue of integrating IT into the educational process over the past decades.

The fundamental aspects of the use of IT in education are considered [1]-[3], aiming to examine the theoretical foundations of the digitalization of the educational space.

The current stage of research is characterized by a transition from general issues to specific aspects of digital transformation. The legal and managerial challenges of the digital era in education have been clarified [4] and the mechanisms of managing educational processes in the conditions of digitalization has been examined [5]. A systematic review of the impact of information and

communication technologies on educational management has been conducted [6] and educational information management systems has been investigated [7].

The processes of implementing educational technologies have been summarized in a systematic review [8], which analyzes the factors of successful adaptation of IT in academic institutions. A comprehensive analysis of the impact of digital technologies on education and aspects of digital transformation has been carried out [9]. The issue of information architecture transformation in the context of university digitalization has been considered [10].

The integration of IT into education and their impact on teaching and learning processes have been investigated [11], [12]. The use of IT in a collaborative approach to education has been considered [13]. The practical challenges of using technologies in classrooms and ways to address them have been analyzed [14], [15].

The dichotomy of hopes and fears regarding IT in education has been investigated [16], and trends in the development of IT in education have been analyzed [17].

Studies on the use of Big Data and AI in education during the COVID-19 pandemic have become particularly relevant [18].

Despite numerous studies on specific aspects of IT applications, the issue of forming an integrated digital ecosystem in HEIs requires further systematic analysis that considers real cases and the practical experience of individual higher education institutions.

3 IT EVOLUTION IN EDUCATION

The introduction of IT roots, stemming from humanity's desire to optimize cognitive processes, systematize knowledge, and increase the efficiency of learning through the creation of specialized tools and methods. The evolution of IT in education began long before the emergence of digital computers.

The invention of the printing press by Johannes Gutenberg around 1440 marked the first technological revolution in education. The possibility of mass reproduction of books transformed the educational space when knowledge ceased to be the privilege of a narrow circle and became available to broad segments of the population. This laid the foundation for the development of mass education in subsequent centuries.

At the beginning of the 20th century, the first mechanical educational devices appeared. In particular, in the 1920s, S. Pressey developed the testing machine, which allowed students to test their knowledge independently and receive instant feedback. This invention laid the foundation for standardized knowledge assessment and automation of the learning process, foreshadowing the ideas of programmed learning.

In the 1950s-1970s, the era of computer technology in education began. The concept of programmed instruction, developed by B. F. Skinner and J. G. Holland in 1954, demonstrated the possibility of an individualized educational process and the gradual control of the acquisition of knowledge through technical tools.

A revolutionary breakthrough was the emergence of the Programmed Logic for Automatic Teaching Operations system (PLATO) in 1960. It was the first full-fledged interactive computer learning environment that supported multi-user mode, a graphical interface, and elements of social interaction. In parallel, the direction of computer-based learning (Computer-Assisted Instruction, CAI) developed, laying the methodological and technological foundations of modern digital educational systems.

The spread of personal computers in the 1980s enabled the mass implementation of educational software in educational institutions. The creation of computer classrooms and laboratories enabled the integration of traditional teaching methods with interactive simulations, training programs, and automated testing systems, significantly expanding teachers' didactic capabilities.

The emergence of the Internet and the World Wide Web in the early 1990s marked a new stage - the transformation of education into a global, networked space. The first university websites, the use of e-mail in academic communication, and the emergence of learning management systems (LMS). i.e., WebCT (1995) and Blackboard (1997), launched the era of electronic learning (e-learning). These platforms were the first to provide standardized organization of educational content, automated assessment, and systematic interaction between students and teachers in the digital environment.

In the 2000s, a massive introduction of LMS, videoconferencing technologies, electronic libraries, and institutional repositories was observed. These tools provided flexible access to educational resources and interactive participation in the educational process, regardless of participants' geographical location. In addition, integrated

management systems for educational institutions were developed, combining administrative and academic processes in a single digital environment, thereby increasing the efficiency of educational organization and management.

The 2010s were marked by the Massive Open Online Courses movement on platforms such as Coursera, edX, and Udacity, which made high-quality education accessible to millions of people worldwide. Mobile learning, cloud technologies, and the use of social networks as educational tools opened up new forms of communication and collaboration. Education has acquired features of flexibility, personalization, and continuity, supporting various models of distance and self-study.

The current stage of the digital transformation of education, which began in the 2020s, is characterized by the integration of advanced technologies: AI for adaptive learning and personalization of educational trajectories, virtual and augmented reality for creating immersive learning environments, blockchain technologies for verifying and protecting academic documents, and Big Data analytics for predicting success and optimizing educational processes.

The COVID-19 pandemic has become a powerful catalyst for the accelerated digitalization of education, underscoring the crucial importance of educational institutions' technological readiness. This has led to a massive shift toward integrated digital ecosystems that combine LMS, interactive platforms for synchronous and asynchronous interaction, mobile applications, and cloud services. Such convergence of technologies ensures the continuity, flexibility, personalization, and global accessibility of the educational process, forming a qualitatively new model of education in the 21st century.

4 KEY DIRECTIONS OF IT APPLICATION IN HIGHER EDUCATION

The higher education system includes various institutions, from governance and quality assurance bodies to research institutions, coordination bodies, library consortia, as well as information and analytical centers. At the same time, higher education institutions are the primary drivers of educational, scientific, and innovative activities, and their digital transformation determines the level of technological development across the entire system. HEI's digital ecosystem is a complex multi-level system of interconnected technological solutions that provide information support for all aspects of the university's operation. The functional structure of IT applications in HEIs is universal, determined by their typical tasks, regardless of organizational and legal form, profile, or scale. The concept of a digital ecosystem involves not only the implementation of individual IT solutions, but also systemic integration, which creates a synergistic effect and qualitatively new opportunities for the development of educational, scientific, and managerial activities.

Analysis of general patterns of IT application in higher education enables us to identify typical areas of digitalization.

HEIs' activities are traditionally structured into three main areas: the management of the academic process, teaching and learning, and scientific research. Accordingly, digital transformation encompasses implementing IT across the following areas: automating management and administrative processes, computerizing the educational process and creating an interactive academic environment, and integrating IT into scientific and research activities.

4.1 IT in the Educational Process Management

The main goal of implementing IT in management activities is to increase the efficiency of administrative processes, ensure transparent decision-making based on objective data, optimize resource use, and create a single information space for the institution.

The use of IT in management is defined by the need to integrate heterogeneous organizational processes, ensure differentiated access to information for different user categories, guarantee the security of personal data in accordance with the law, and provide comprehensive analytical reporting to support management decisions.

Currently, comprehensive corporate management information systems tailored for the educational sector are being introduced. They include Campus Management, Ellucian Banner, Oracle PeopleSoft Campus Solutions, etc. These systems integrate the functions of managing the student contingent, human resources, finance, the material and technical base, as well as scientific activities on a single platform. Electronic document management systems (M.E.Doc, ASKOD EDMS, DocuSign, etc.) automate administrative procedures, enable electronic approval, and support document digital signatures. Data analytics systems enable multidimensional analysis of educational data, the creation of interactive management dashboards, the

prediction of performance indicators (e.g., the risk of student expulsion and the demand for educational programs), and the identification of problem areas.

Electronic Dean's offices automate the accounting of academic performance, the creation of individual curricula, and the issuance of academic certificates and diploma documents. Student relationship management systems optimize the processes of the admission campaign and communication with potential students. Promising areas of development of managerial IT include the implementation of AI for predictive analytics and automation of routine administrative processes, the use of blockchain technologies to verify academic documents and ensure their international recognition, the development of mobile applications for quick access to management information, integration with national educational information systems, and global data exchange platforms, as well as the implementation of systems for automatic monitoring of the quality of educational services based on Big Data analysis.

4.2 IT in Teaching and Learning

The purpose of implementing IT in the educational process is to improve the quality of educational services by ensuring interactive learning, personalizing educational trajectories, expanding access to high-quality academic resources, and developing students' digital competencies.

The use of educational technologies requires their pedagogical adaptation to the specifics of academic courses, user-friendly interfaces for different user groups, multimodal forms of learning (synchronous, asynchronous, blended), and integration with systems for assessing academic performance. The modern technological infrastructure of the academic process is based on the integration of various digital tools. LMS (Moodle, Canvas, Blackboard, Google Classroom, etc.) provide structured placement of educational content, interactive elements (forums, wikis, glossaries), automated testing, tracking student progress and reporting, and communication among participants in the educational process.

Video conferencing platforms (Zoom, Microsoft Teams, Google Meet, BigBlueButton, etc.) provide synchronous remote interaction, including document collaboration, screen sharing, and breakout rooms.

Adaptive AI-based learning systems (Knewton, Smart Sparrow, Century Tech, etc.) manage personal development planning according to individual needs, learning pace, and students' information perception.

Virtual and augmented reality (VR/AR) technologies create immersive learning environments for practical training in medicine, engineering, and architecture.

Academic integrity control systems (such as Turnitin, Unicheck, StrikePlagiarism, etc.) ensure the originality of student work.

Electronic library systems provide access to digital educational resources, scientific databases, and electronic textbooks. Interactive content creation tools (H5P, Articulate, Adobe Captivate, etc.) allow teachers to develop multimedia learning materials.

Electronic assessment systems (Gradescope, Crowdmark, ZipGrade, ExamSoft, etc.) and LMS assessment modules automate testing and grading.

The prospects for the development of educational technologies are associated with the deepening of personalization of learning through the use of AI and machine learning to analyze individual academic needs, the expansion of using VR/AR to create realistic simulations of complex processes and phenomena, the introduction of gamification technologies and serious games to increase student motivation, the development of microlearning and adaptive mobile applications for learning at any time and place, the use of AI-based chatbots for 24/7 student support, the integration of Learning Analytics technologies for in-depth analysis of understanding behavior and prediction of success, as well as the development of social learning through the integration of educational platforms with social networks.

4.3 IT in Scientific Research

The use of IT in scientific activities aims to intensify research processes, ensure the openness and accessibility of scientific results, increase the visibility of the institution's publication activity in the international scientific space, and build an infrastructure for international scientific cooperation.

It is specified by the need to ensure long-term storage of research data, support international academic metadata standards, integrate with global scientific databases, and adhere to the principles of Open Science.

The modern infrastructure of digital scientific activities includes various technological components.

Institutional repositories (based on DSpace, EPrints, Fedora, etc.) provide long-term storage and open access to scientific publications, dissertations,research reports, and conference proceedings in accordance with the principles of Open Access.

Current research information systems (PURE, Converis, VIVO, etc.), integrate data on research projects, publications, patents, scientific events, and grants, and allow researchers to be profiled.

Bibliometric platforms (Scopus, Web of Science, Google Scholar, etc.) provide analysis of scientific impact through citation indices (h-index, impact factor), the detection of scientific connections, and the analysis of cooperation networks.

Data visualization and analysis tools (Tableau, Power BI, RStudio, Python libraries, etc.) enable the processing of large amounts of research data and the presentation of results in visual form.

Plagiarism-detection systems for scientific texts (iThenticate, CrossCheck, etc.) ensure the originality of research and academic integrity.

Collaborative research platforms (ResearchGate, Academia.edu, Mendeley, etc.) facilitate academic communication, the exchange of scientific results, and the formation of research networks.

Bibliography management systems (Zotero, EndNote, Mendeley, etc.) automate work with scientific sources and the formation of references.

Cloud platforms for scientific computing (Google Colab, Amazon Web Services, Microsoft Azure, etc.) provide power for resource-intensive computing.

Promising directions for the development of scientific IT include the introduction of AI for automated analysis of scientific literature, trend identification and hypothesis generation, expanding the use of semantic technologies and linked data for the integration of scientific knowledge, development of FAIR data infrastructure to ensure the repeatability of research, implementation of blockchain technologies for intellectual property protection and authorship verification, creation of Virtual Research Environments for international scientific cooperation, integration with global scientific infrastructures (ORCID, Crossref, DataCite, etc.), development of open peer review systems, as well as the use of digital twins technologies for modeling complex scientific systems.

Analysis of the current state of IT applications in HEIs reveals a wide range of solutions tailored to each functional area of the university's activities. At the same time, a maximum effect of digital transformation is achieved not through the mechanical accumulation of individual technologies, but through their organic integration into a single digital ecosystem. Such integration ensures data connectivity, eliminates information duplication, optimizes organizational processes, and creates qualitatively new opportunities for the institution's development. The practical implementation of the concept of an integrated digital ecosystem requires a systematic approach and strategic planning, as evidenced by the experience of particular educational institutions.

5 INTEGRATED DIGITAL ECOSYSTEM: VITE SUTE EXPERIENCE

The formation of the digital ecosystem is represented based on the practical experience of the Vinnytsia Institute of Trade and Economics of the State University of Trade and Economics (VITE SUTE). The Institute's integrated digital ecosystem is built on two key platforms: its own automated educational process management system ("Smart Electronic Learning", SEL), and the open-source LMS Moodle. Being interconnected, these systems provide comprehensive information support for the Institute's management, educational, and scientific activities.

The SEL development began in 2020 to create a tool aligned with the Institute's existing operational processes and the specifics of higher education in Ukraine. The system was developed in accordance with state standards governing the creation of automated systems. This approach is substantiated by the fact that most automated management systems on the Ukrainian IT market do not provide a complete cycle of automation for educational processes [19].

A modern web technology stack was used in the development of SEL, ensuring scalability, performance, and ease of maintenance. The backend is implemented using the PHP programming language (versions 7.4/8.0) with the Yii2 framework (version 2.0.13+), which supports MVC architecture and a component-based approach to application development. MySQL 8.0 is used for data storage, guaranteeing reliability through transaction support and ACID properties. The client-side of the system is built using the AdminLTE 3.0 template, which is based on the Bootstrap 4 framework. The jQuery library is used to implement dynamic user interaction. The system's deployment infrastructure is built using Docker containerization and Docker Compose for service orchestration, along with an Nginx (Alpine) web server acting as a reverse proxy and PHP-FPM for handling server-side requests.

The system architecture is built on a modular principle and comprises five main components. The "Teaching" module integrates E-Register, the Lecturer's Individual Work Plan, and the Individual Repository. The "Student" module provides an electronic gradebook, a grade register, and an elective course selection system. The "Services" module combines the functions of the Dean's office, the Human Resources Department, workload distribution, and timetable generation. The "Repository" module provides storage and access to electronic publications of the teaching staff. The

"Internal Resources" module integrates document flow and connects to the Moodle platform.

The "Lecturer's Individual Work Plan" module is an tool for planning, monitoring, and evaluating the professional activities of staff. The module integrates all types of work of the academic staff. The system automatically imports data on teaching workload distribution, generates reports on plan implementation for a selected period, and provides real-time monitoring of task completion. Integration of this module with other system components (timetable, workload, and research publications) ensures control over the lecturer's activities.

The "Timetable" module operates on the basis of a personalized teaching workload database and provides the creation and maintenance of an electronic register of the lecturer's workload. The system automatically registers additional classes, tracks class substitutions, and generates schedules for lecturers and academic groups.

A unique SEL module is the automated system for elective course selection, which enables students to plan their personal development. The system allows learners to access elective course information via the attached syllabi, select the course they are interested in, and automatically subscribe to it. All processes, including registration, data processing, and analysis, are automated by the system, significantly reducing the time and labor costs for administrative staff.

To maintain the academic process, VITE SUTE uses the Moodle platform, a learning environment widely used as one of the most commonly used open-source LMS platforms. Moodle hosts teaching materials for each particular course and provides interactive tools, student activity tracking, and an electronic gradebook.

Integration between Moodle and the SEL is carried out through the "Internal Resources" module. It connects the Institute's management infrastructure and the learning space. It ensures an ongoing information flow: data on students, their enrollment, and curricula are generated in SEL and made available on the Moodle platform without duplication. Moodle acts as a platform for hosting educational materials, while SEL functions as an performance monitoring system, with data exchange via API and single sign-on support.

The SEL also supports the representation of the teaching staff's scientific activity through several key modules. The Institutional repository is integrated into the system through the "Repository" module and provides long-term storage and open access to electronic publications of the teaching staff. After checking student work for originality, the system automatically places it in the teacher's repository, optimizing the scientific archiving process.

The "Plagiarism" module verifies the originality of students' Master's research projects by comparing them with a database of research papers submitted in previous years.

The construction of the digital ecosystem of the VITE SUTE, based on the integration of the SEL and Moodle, demonstrates the practical implementation of the concept of an integrated digital ecosystem for HEIs. The module architecture of the SEL enables integration with other Institute tools and systems, as well as scalability to meet the needs of different user categories.

Integration of systems ensures several key outcomes. First, a single information space is formed in which all participants of the educational process, including students, teachers, and administrative staff, work with up-to-date data in real time. Second, duplication of information is eliminated, so that data on students, curricula, and assessment results are entered once and automatically distributed between related modules. Third, a comprehensive analysis of the Institute's activities, from planning the educational process to monitoring teachers' scientific activities, is provided.

Thus, the experience of VITE SUTE confirms that a comprehensive, integrated approach to implementing IT is an effective tool for the digital transformation of a higher education institution, ensuring increased efficiency in management, educational, and scientific processes.

6 CONCLUSIONS

The analysis of the current state of IT applications in higher education shows that digital transformation covers all the main areas of HEI activity: educational process management, teaching and learning, and scientific research. The evolution of educational technologies from simple learning machines to integrated digital ecosystems indicates a constant growth of their role in ensuring the efficiency and high quality of education. The analysis of a wide range of modern IT solutions - from educational institution management systems and LMS to bibliometric systems and data analysis tools - reveals a robust technological infrastructure capable of meeting the needs of modern universities.

The experience of VITE SUTE confirms that the maximum effect of digital transformation is achieved not through the mechanical implementation of individual technologies, but their organic integration into a single digital ecosystem. The developed integrated system, combining its SEL with the Moodle, ensures data integration, eliminates

information duplication, and creates a synergistic effect for all participants in the educational process. The results of the study can serve as a basis for further development in the field of forming integrated digital ecosystems of HEIs, in particular for developing methodological principles for planning and implementing IT strategies.

REFERENCES

[1] H. Amca, "Information Technology in Education," in School of Computing and Technology, Eastern Mediterranean University, pp. 1-4, Mar. 2001.

[2] W. R. Muir, "Information Technology and Education," in Quality of Human Resources: Education, Encyclopedia of Life Support Systems, vol. 3, UK: EOLSS Publishers, 2009.

[3] F. Hamidi, M. Meshkat, M. Rezaee, and M. Jafari, "Information Technology in Education," in Procedia Computer Science, vol. 3, pp. 369-373, 2011, [Online]. Available: https://doi.org/10.1016/j.procs.2010.12.062.

[4] Z. Li and W. Zhang, "Technology in education: Addressing legal and governance challenges in the digital era," in Education and Information Technologies, vol. 30, pp. 8413-8443, 2025, [Online]. Available: https://doi.org/10.1007/s10639-024-13036-9.

[5] L. Rahm, "Educational imaginaries: Governance at the intersection of technology and education," in Journal of Education Policy, vol. 38, no. 1, pp. 46-68, 2023, [Online]. Available: https://doi.org/10.1080/02680939.2021.1970233.

[6] T. Liu, Y. T. Luo, P. C.-I. Pang, and H. Y. Kan, "Exploring the impact of information and communication technology on educational administration: A systematic scoping review," in Education Sciences, vol. 15, no. 9, p. 1114, 2025, [Online]. Available: https://doi.org/10.3390/educsci15091114.

[7] S. D. Kurniasih, "Information and communication technology in educational management information system," in Indonesian Journal of Islamic Communication, vol. 6, no. 1, pp. 125-138, 2023, [Online]. Available: https://doi.org/10.35719/ijic.v6i1.2212.

[8] A. Granić, "Educational technology adoption: A systematic review," in Education and Information Technologies, vol. 27, no. 7, pp. 9725-9744, 2022, [Online]. Available: https://doi.org/10.1007/s10639-022-10951-7.

[9] S. Timotheou, O. Miliou, Y. Dimitriadis, S. Villagrá Sobrino, N. Giannoutsou, R. Cachia, A. Martínez Monés, and A. Ioannou, "Impacts of digital technologies on education and factors influencing schools' digital capacity and transformation: A literature review," in Education and Information Technologies, vol. 28, pp. 6695-6726, 2023, [Online]. Available: https://doi.org/10.1007/s10639-022-11431-8.

[10] N. Bobro, "Transforming information architecture in the context of university digitalization," in Information Technologies in Education, no. 1 (57), pp. 7-13, 2025, [Online]. Available: https://doi.org/10.14308/ite000788.

[11] D. Mdhlalose and G. Mlambo, "Integration of technology in education and its impact on learning and teaching," in Asian Journal of Education and Social Studies, vol. 47, no. 2, pp. 54-63, 2023, [Online]. Available: https://doi.org/10.9734/ajess/2023/v47i21021.

[12] F. Duraku, "Information technology in teaching and learning in education," in E-Journal VFU, vol. 23, pp. 244-258, 2025, [Online]. Available: https://doi.org/10.53606/evfu.23.244-258.

[13] N. Zamkova, I. Polishchuk, Yu. Dovhan, L. Dovhan, and R. Novytskyi, "Training of future logistics and supply chain managers: A competency approach," in Financial and Credit Activity: Problems of Theory and Practice, vol. 1, no. 48, pp. 427-440, 2023, [Online]. Available: https://doi.org/10.55643/fcaptp.1.48.2023.3946.

[14] A. M. Johnson, M. E. Jacovina, D. E. Russell, and C. M. Soto, "Challenges and solutions when using technologies in the classroom," in Adaptive Educational Technologies for Literacy Instruction, pp. 13-29, 2016, [Online]. Available: https://doi.org/10.4324/9781315647500-2.

[15] M. Liu, R. Zhou, J. Dai, and X. Feng, "Analysis and practice of using modern information technology for classroom teaching mode reform," in Mobile Information Systems, vol. 2022, 2022, [Online]. Available: https://onlinelibrary.wiley.com/doi/epdf/10.1155/2022/2565735, [Accessed: 25 Jan. 2026].

[16] J. Łaszczyk, "Information technology in education - hopes and fears," in International Journal of Electronics and Telecommunications, vol. 61, no. 3, pp. 261-266, 2015, [Online]. Available: https://doi.org/10.1515/eletel-2015-0034.

[17] D. Hawkridge, New Information Technology in Education, 1st ed., London, UK: Routledge, 2022, [Online]. Available: https://doi.org/10.4324/9781003312826.

[18] J. Zhao and Q. Li, "Big data-artificial intelligence fusion technology in education in the context of the new crown epidemic," in Big Data, vol. 10, no. 3, pp. 262-276, 2022, [Online]. Available: https://doi.org/10.1089/big.2021.0245.

[19] L. M. Radzikhovska, O. V. Ivashchuk, R. M. Novytskyi, V. V. Romanuke, and S. A. Yaremko, "The automated system 'SEL' as a tool for higher education institution management," in Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu, vol. 2025, no. 3, pp. 212-220, 2025, [Online]. Available: https://doi.org/10.33271/nvngu/2025-3/212.

Online and Face-to-Face Learning Effectiveness in Ukrainian Higher Education: A Multidimensional Comparison

Iryna Zhalinska[1], Serhii Nikolaienko[1], Kateryna Parnus[2], Marina Kosheleva[2] and Olena Mykhaylovska[2]

[1]*Department of Psychology and Social Welfare, Zhytomyr Polytechnic State University, Chudnivska Str. 103, 10005 Zhytomyr, Ukraine*

[2]*Department of Pedagogical Technologies and Language Training, Zhytomyr Polytechnic State University, Chudnivska Str. 103, 10005 Zhytomyr, Ukraine*

zhalinska@gmail.com, nikolayenko_sm@ukr.net, katemarchuk38@gmail.com, maryshelle68@gmail.com, olena.feliz4@gmail.com

Keywords: Online Learning, Face-to-Face Learning, Higher Education, Learning Effectiveness, Student Engagement, Psychological Perception, Digital Learning Environment.

Abstract: The rapid digital transformation of higher education has intensified scholarly interest in the effectiveness of online learning. This study provides a multidimensional comparison of online and face-to-face learning effectiveness in Ukrainian higher education. Drawing on a mixed-methods research design, the study integrates a systematic review of recent empirical literature (2019-2025) with quantitative survey data collected from 140 university students. Learning effectiveness is conceptualized as a multidimensional construct encompassing academic outcomes, student engagement and motivation, interaction quality, technological and organizational conditions, and psychological perceptions of the educational process. The findings indicate that online learning demonstrates clear advantages in technological accessibility, organizational efficiency, and perceived psychological safety, while face-to-face learning shows stronger outcomes in fostering social interaction and the development of transferable skills. Differences in academic outcomes and motivation are comparatively moderate, suggesting that effectiveness is shaped less by learning modality per se and more by instructional design and institutional support. The results support a complementary interpretation of online and face-to-face formats rather than a dichotomous opposition between them. The study contributes empirical evidence from an underrepresented Eastern European context and offers practical implications for the balanced integration of digital learning formats in contemporary higher education systems.

1 INTRODUCTION

In the twenty-first century, digitalization has permeated all spheres of social life, triggering profound transformations in educational practices, particularly within higher education systems. One of the most significant consequences of this process has been the expansion of online education as an independent mode of organizing the educational process and an integral component of contemporary university curricula. Online education has become a key element of modern higher education, offering flexibility, accessibility, and cost efficiency. Across Europe, Ukraine, and globally, universities increasingly integrate online learning formats into their academic programs, reflecting broader trends toward digitalization and lifelong learning initiatives. In academic literature, online education is understood as a set of distance learning formats that involve the use of digital platforms, multimedia resources, and networked communication to ensure access to educational content regardless of learners' location or time constraints. The considerable scholarly interest in the effectiveness of this mode of learning is driven not only by its large-scale implementation during the COVID-19 pandemic but also by ongoing structural transformations in higher education systems in the post-pandemic period [1].

The effectiveness of online education in higher education institutions is conceptualized as a multidimensional construct encompassing students' academic outcomes (achievement and academic

performance), levels of engagement and learning motivation, as well as the capacity of digital environments to support communication among participants in the educational process. Findings from systematic reviews and meta-analyses indicate that the success of online learning largely depends on a complex set of factors, including course design quality, instructional structure, students' digital literacy, instructors' professional competence, and the availability of appropriate technological and organizational support at the institutional level. Moreover, some studies emphasize that the positive impact of online education on academic outcomes can be realized only when high-quality interactive elements, active participant interaction, and adaptive instructional design are incorporated [2].

Alongside its advantages related to flexibility, accessibility, and individualized learning pathways, online education faces significant challenges. These include sustaining student motivation, ensuring meaningful social interaction, mitigating feelings of isolation among participants, as well as overcoming technical and infrastructural barriers that limit equitable access to digital resources. Without adequate institutional support and proper instructor training, these factors can substantially undermine the effectiveness of online courses.

Despite the growing body of research on online learning, several limitations remain. Comparative analyses that integrate academic, motivational, interactional, technological, and psychological dimensions within a single analytical framework are still relatively limited. Furthermore, empirical evidence from Eastern European contexts, particularly Ukraine, remains underrepresented in the international scholarly discourse. Post-pandemic evaluations also tend to insufficiently address psychological and regulatory aspects of the educational experience, including perceptions of safety, stress, and well-being.

These limitations point to the need for a more integrative approach to assessing learning effectiveness, one that captures not only academic outcomes but also the behavioural, relational, technological, and psychological conditions under which learning occurs. In this regard, investigating online and face-to-face learning effectiveness in higher education constitutes not only a relevant academic issue but also an important societal task.

Accordingly, this study adopts a multidimensional analytical framework that incorporates academic, motivational, interactional, technological, and psychological perspectives as interrelated domains of educational effectiveness. The study aims to provide a systematic comparison of online and face-to-face learning in Ukrainian higher education. By doing so, it contributes empirical evidence to the ongoing discussion on the role of digital learning formats in contemporary university education.

2 LITERATURE REVIEW

In contemporary international research, the effectiveness of online education in higher education institutions examined as a multidimensional phenomenon encompassing academic, socio-psychological, technological, and organizational dimensions. Scholars emphasize that online learning provides learners with flexibility, opportunities for self-paced study, and access to interactive digital resources, thereby fostering higher engagement and deeper comprehension of learning materials. According to the literature, these characteristics can positively affect students' academic performance and overall achievement [3].

Research by Mohammadi et al. (2025) and Younas (2025) demonstrates that online learning environments can significantly enhance student motivation, increase opportunities for interaction, and create favorable conditions for developing digital competence. The authors highlight the necessity of standardized platforms, pedagogically supported student services, and continuous professional development for instructors as key elements for the successful implementation of such educational models [3], [4].

At the same time, the effectiveness of online learning depends not only on technological infrastructure but also on meaningful course design, participants' digital readiness, and students' psychological motivation. In particular, the development of interactive content, adaptive learning pathways, and tools supporting student self-regulation plays a decisive role in determining engagement levels and the ability to achieve intended learning outcomes. These findings align with the results of Yuan et al. (2024), who emphasize the importance of comprehensive analyses of learners' behavioral patterns for improving the predictability of online learning outcomes [5].

Studies on the effectiveness of online education indicate that digital platforms have a positive impact on student motivation to learn and self-regulation, as well as on student engagement and autonomy in the learning process. In turn, these factors correlate with improved academic performance, provided that

instructional content and pedagogical strategies are carefully designed and effectively implemented [6].

Instructors benefit from centralized content delivery systems, streamlined administrative processes, and improved capacity to monitor and assess student progress. For higher education institutions, online learning enhances cost efficiency, supports scalability, and expands access to a broader and more diverse student population, including international learners [7].

However, alongside these advantages, most reviews draw attention to substantial limitations and challenges associated with the implementation of online education. Some relate to limited social interaction, which may lead to feelings of isolation and decreased engagement, particularly in fully distance learning formats. Other studies highlight technical inequalities in access to high-quality digital resources, insufficient digital competence among some students, and difficulties in delivering laboratory-based and practice-oriented courses in virtual environments. These challenges underscore the importance of pedagogically grounded instructional design and comprehensive support for participants in online learning contexts [8].

Analytical reviews also emphasize the role of psychological factors in online education, such as autonomy, self-esteem, and self-regulation capacity, which function as critical determinants of learning success in digital environments. This perspective is reinforced by recent research asserting that motivation is not only a key predictor of success but also a compensatory factor capable of mitigating the absence of direct instructor-student interaction [9].

The study by Pashchenko et al. (2025) demonstrates that online learning environments incorporating interactive digital elements and structured content can contribute to higher student satisfaction and improved academic performance. Evaluations of such approaches also reveal positive dynamics in the development of critical thinking, analytical, and communicative skills, which are essential competencies for the contemporary labor market [10].

Overall, literature published over the past five years converges on a consistent conclusion: online education represents a significant component of contemporary educational practice capable of enhancing academic performance, motivation, and digital competence among students; however, its effectiveness largely depends on the integration of technology, pedagogically sound course design, and comprehensive support for participants in the educational process.

3 RESEARCH METHODOLOGY

To assess the effectiveness of online education in higher education, this study employed a mixed-methods research design combining quantitative and qualitative analytical approaches. This study adopts this design based on the premise that educational processes - particularly those mediated by digital technologies - are inherently multidimensional and cannot be adequately captured through a single methodological paradigm.

3.1 Research Design

The study adopts a descriptive-analytical and explanatory design based on the integration of secondary data analysis (scholarly literature) and primary empirical data collected through surveys. The application of methodological triangulation enhances the validity and reliability of the findings by minimizing the limitations inherent in individual methods.

3.2 Literature Review

At the first stage, a systematic review of scholarly literature was conducted, encompassing peer-reviewed journal articles, meta-analyses, and analytical reports published between 2019 and 2025. The selected timeframe reflects the period of accelerated digital transformation in higher education, particularly following the global expansion of online learning. The literature review enabled the identification of key indicators of online education effectiveness, as well as its principal advantages, challenges, and research gaps.

3.3 Sample and Data Collection

The empirical phase involved a quantitative survey administered to 140 students from the Faculty of Information and Computer Technologies at a higher education institution in the Zhytomyr region of Ukraine. Respondents represented different academic years and levels of experience with online and blended learning formats.

Data collected using a structured online questionnaire consisting of closed-ended and semi-open questions designed to measure:

- perceptions of online learning effectiveness;
- levels of student engagement and motivation;
- quality of student-instructor interaction;

- technological and organizational challenges associated with the digital learning environment;
- psychological perceptions of the educational process.

The online administration mode ensured methodological consistency with the digital context under investigation.

3.4 Research Instruments, Validity and Reliability

Survey items were developed in alignment with regulatory frameworks defined by strategic European Union policy documents in digital education, particularly the Digital Education Action Plan 2021-2027, as well as the core principles of the European Higher Education Area. Within these frameworks, online learning effectiveness is conceptualized as a multidimensional phenomenon encompassing not only academic outcomes but also student engagement and motivation, learning autonomy, and the quality of interaction in digital educational environments [11], [12]. This ensures comparability with European empirical practices and enables interpretation of the results as indicators of alignment between national educational contexts and European standards for digital higher education.

Likert-type scales were used for quantitative measures, facilitating subsequent statistical analysis. Content validity was ensured by aligning questionnaire items with theoretical models of online learning presented in contemporary academic literature.

3.5 Data Analysis Methods

Quantitative data were analyzed using descriptive statistics to identify general trends in perceptions of online education among students and instructors. Qualitative responses to open-ended questions were processed through thematic analysis involving coding and categorization of recurring themes, including advantages, challenges, and recommendations for improving online learning.

3.6 Ethical Considerations

The study was conducted in accordance with established ethical principles of academic research. Participation was voluntary, and all collected data were anonymized. Participants were informed about the purpose of the study and the exclusive use of the results for scientific research purposes.

3.7 Methodological Limitations

Key limitations of the study include reliance on self-reported data, which may affect response subjectivity, as well as the limited geographical representativeness of the sample. Nevertheless, the combination of quantitative and qualitative methods partially compensates for these limitations and provides a more holistic understanding of the phenomenon under investigation.

4 RESULTS AND DISCUSSION

This dimension reflects students' generalized perceptions of the effectiveness of the learning process in online and face-to-face learning in terms of achieving educational outcomes and the overall quality of the learning experience (Fig. 1).

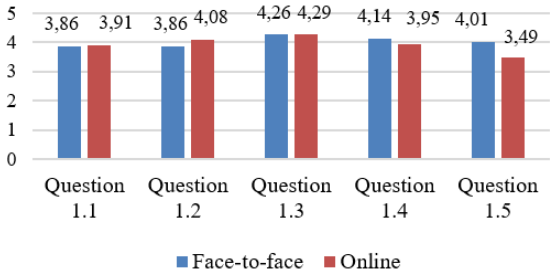


Figure 1: Perceived learning effectiveness across online and face-to-face formats.

In Figure 1:

- Question 1.1 - The learning process supports the achievement of programme learning outcomes and the development of required competencies.
- Question 1.2 - Learning materials and information are presented clearly and are easy to understand during classes
- Question 1.3 - I receive timely, meaningful, and constructive feedback from academic staff.
- Question 1.4 - Interaction with academic staff and fellow students helps me to better understand the learning material.
- Question 1.5 - The learning process effectively supports the development of my transferable skills (e.g. communication, critical thinking, teamwork).

The highest values in both formats are associated with receiving timely and substantive feedback from instructors (4.26 and 4.29 for face-to-face and online formats, respectively), confirming the consistently high importance of instructor support regardless of the mode of delivery.

The most pronounced gap in favor of face-to-face learning is observed in the development of soft skills (-0.52), which may indicate the limited capacity of the online environment to foster communicative, collaborative, and social competencies. A similar, though less pronounced, tendency is evident in the domain of interaction with instructors and peers (- 0.19). This underscores the need for targeted pedagogical and organizational reinforcement of social and communicative components within the digital learning environment.

Overall, the findings suggest that online learning is perceived as effective in terms of academic outcomes, accessibility of materials, and the quality of feedback. However, it lags behind the face-to-face format in fostering soft skills and social interaction. These results allow online and face-to-face learning formats to be interpreted as relatively holistic educational models in students' perceptions, each demonstrating distinct strengths and limitations.

Engagement and motivation (Fig. 2) represent the lowest-rated dimension across both formats, suggesting that the motivational-behavioral component constitutes the most vulnerable layer of the educational process. While mean differences between formats are small, the consistently moderate scores indicate that engagement is not format-specific but structurally constrained within both delivery modes.

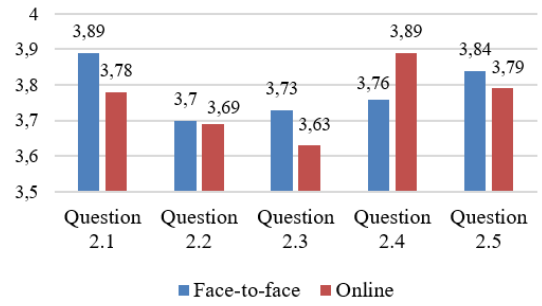


Figure 2: Student engagement and motivation across online and face-to-face formats.

In Figure 2:

- Question 2.1 - The learning format encouraged my active engagement in the educational process (e.g. completing tasks, participating in discussions, independent study).
- Question 2.2 - The learning format had a positive impact on my motivation to engage in regular study.
- Question 2.3 - The learning format contributed to the development of my self-discipline.
- Question 2.4 - I am interested in continuing my studies in this format in the future.
- Question 2.5 - Interaction with academic staff within this format supported my motivation to learn.

Online learning demonstrates relatively higher future-oriented acceptance (3.89), reflecting its perceived sustainability. However, lower ratings in self-discipline (3.63) and active participation (3.78) point to increased demands for autonomous regulation in digital environments.

Importantly, the absence of large format effects suggests that motivational dynamics are shaped less by modality and more by instructional design and regulatory demands. Theoretically, these findings support models of self-determination and academic engagement that emphasize structural and pedagogical scaffolding over delivery format per se. Thus, engagement appears to function as a cross-format systemic challenge rather than a modality-dependent deficit.

Interaction quality is evaluated positively across both formats, reinforcing the instructor's central role as a stabilizing agent within the educational system (Fig. 3). However, effect magnitudes reveal differentiated interaction logics.

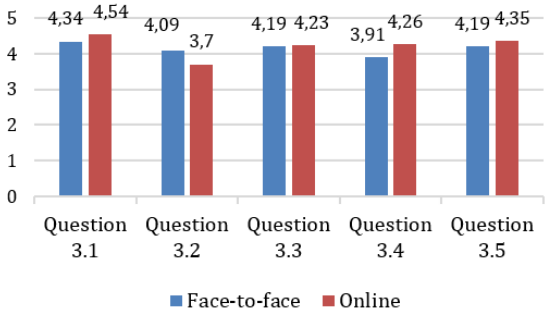


Figure 3: Student-instructor interaction quality across formats.

In Figure 3:

- Question 3.1 - During the educational process, academic staff adhere to the principles of academic integrity and demonstrate respect towards students.
- Question 3.2 - During classes, I experience a sense of presence and engagement from academic staff.

- Question 3.3 - I feel that I am an equal and valued participant in the educational process.
- Question 3.4- The system for organising and submitting assessed work is convenient and efficient.
- Question 3.5 - The assessment of my academic performance is fair and objective.

The most substantial face-to-face advantage is observed in perceived instructor presence (-0.39), indicating that embodied co-presence continues to shape relational and affective dimensions of teaching. This gap likely reflects the reduced transmission of nonverbal immediacy cues and spontaneous interaction in digital settings.

Conversely, online learning demonstrates moderate advantages in assignment organization and timeliness (+0.35), objectivity of assessment (+0.16), and academic integrity adherence (+0.20). These effects suggest that digital environments enhance procedural transparency and institutional clarity.

Taken together, these findings support a dual-process interpretation of interaction quality: face-to-face learning privilege relational immediacy, whereas online formats reinforce procedural structure. This distinction contributes to the growing literature on digital pedagogy by differentiating affective presence from organizational efficiency as parallel, format-contingent dimensions of instructional interaction.

Technological and organizational conditions constitute the strongest domain of online learning advantage (Fig. 4). Effect sizes are substantively larger here than in previous dimensions, indicating a structurally embedded digital benefit.

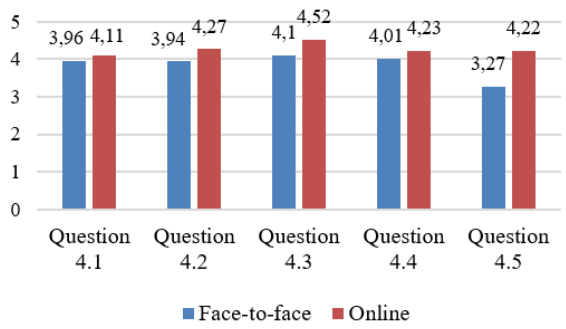


Figure 4: Technological and organizational conditions across formats.

In Figure 4:

- Question 4.1 - The institution provides timely technical support when difficulties arise.
- Question 4.2 - Learning materials are well structured and readily accessible.
- Question 4.3 - Academic staff make effective use of digital tools in the teaching and learning process.
- Question 4.4 - Overall, the technological and organisational conditions support the achievement of learning outcomes.
- Question 4.5 - During my studies, I am able to maintain a satisfactory balance between academic responsibilities and personal life.

The most pronounced effect concerns study-life balance (+0.95), representing the largest cross-format difference observed across all dimensions. This substantial magnitude underscores flexibility and temporal autonomy as defining affordances of digital learning. Additional moderate advantages are found in effective use of digital tools (+0.42) and accessibility of learning materials (+0.33), confirming structural optimization within the online environment.

In contrast, differences in technical support are minimal (+0.15), suggesting institutional parity across formats.

Theoretically, these results align with affordance-based models of digital learning, wherein technology-mediated environments reconfigure spatial-temporal constraints and resource accessibility. However, variability within online ratings indicates that digital effectiveness is contingent upon institutional implementation quality rather than inherent technological superiority.

Psychological perceptions display the most context-sensitive pattern (Fig. 5).

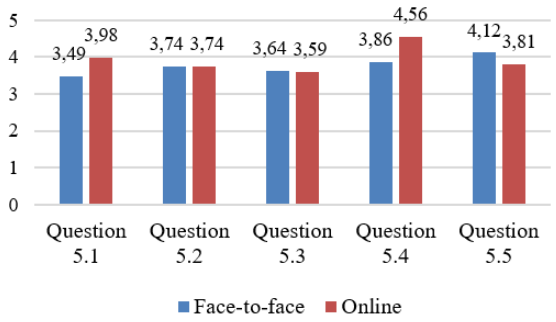


Figure 5: Psychological perceptions across online and face-to-face formats.

In Figure 5:

- Question 5.1 - The level of stress associated with my studies is manageable.
- Question 5.2 - During classes, I am able to concentrate and sustain my attention.

- Question 5.3 - I feel that I receive adequate psychological and emotional support from academic staff and fellow students.
- Question 5.4 - I feel safe within the educational environment.
- Question 5.5 - The level of social interaction available to me during my studies is sufficient.

Online learning shows clear advantages in perceived safety (+0.70) and acceptable stress levels (+0.49), both representing moderate-to-large effects relative to other dimensions. These magnitudes suggest that psychological regulation may be more strongly influenced by learning format than academic outcomes.

Conversely, face-to-face learning retains an advantage in social interaction (-0.31), indicating the enduring importance of direct interpersonal contact for socio-emotional integration.

Given the Ukrainian national context of the sample, elevated sensitivity to safety and stress indicators may reflect broader environmental instability. Thus, format effects should be interpreted within macro-contextual conditions influencing psychological security.

Conceptually, these findings extend the discourse on digital learning by introducing a regulatory perspective: online environments may function as psychological buffers by reducing situational stressors, while face-to-face learning sustain social belonging and relational depth. The complementarity of these effects supports hybrid models that integrate digital psychological safety with face-to-face social capital.

The analysis of students' open-ended responses regarding the effectiveness of online learning and its potential improvement identified a set of interrelated organizational, pedagogical, and technological factors shaping perceptions of educational quality.

Primarily, students associate the effectiveness of the online format with a clearly structured course design, including a logical division of content into topics, transparent requirements, explicit assessment criteria, and well-defined deadlines. Such structural clarity reduces uncertainty and facilitates more effective planning of learning activities.

A major advantage of online learning is perceived as temporal and spatial flexibility. The elimination of commuting, the opportunity to study in a comfortable environment, and the possibility of combining studies with employment are considered factors enhancing productivity. Comfortable conditions are seen as supporting concentration and reducing psychological strain. At the same time, some respondents report decreased motivation and limited social interaction, reflecting an ambivalent perception of distance education.

Instructional design quality emerges as another critical component of effectiveness. Students highlight the importance of clear explanations, real-time demonstrations during practical sessions, access to recorded lectures, and supplementary video materials. There is also a strong demand for greater interactivity, including quizzes, polls, short in-class tasks, and brief feedback activities. Interactivity is interpreted as a mechanism for sustaining attention and cognitive engagement.

Many responses emphasize the need to strengthen the practical and applied orientation of courses by increasing the number of practice-based activities, incorporating real-life cases, and reconsidering the balance between theory and practice. The effectiveness of online learning is further linked to the development of learner autonomy, self-discipline, and intrinsic motivation, indicating the growing importance of self-regulation in digital learning environments.

The technological dimension is regarded as a foundational condition for quality education, including stable internet access, user-friendly platforms, digital automation tools, and the integration of innovative solutions. Additionally, some responses extend beyond organizational improvements and raise broader concerns about the impact of artificial intelligence on academic integrity, motivation, and future professional prospects.

Overall, enhancing the effectiveness of online learning, from the students' perspective, requires a comprehensive approach integrating structured course organization, interactive teaching methods, accessible learning materials, practical orientation, technological optimization, and consideration of long-term digital transformations in education.

5 CONCLUSIONS

The present study provides a multidimensional comparison of online and face-to-face learning effectiveness in Ukrainian higher education, integrating academic, motivational, interactional, technological, and psychological dimensions. The findings demonstrate that learning effectiveness cannot be reduced to a single modality-based hierarchy; rather, both formats exhibit distinct strengths and limitations across different dimensions of the educational process.

In terms of academic outcomes and feedback quality, differences between formats are relatively moderate. Both online and face-to-face learning demonstrate comparable levels of perceived support for learning outcomes and constructive instructor feedback. However, face-to-face learning shows a

stronger contribution to the development of transferable skills and social interaction, underscoring the importance of embodied co-presence for communicative and collaborative competencies.

Online learning demonstrates its most pronounced advantages in technological and organizational conditions. Students report higher levels of material accessibility, effective use of digital tools, structured assessment procedures, and improved study-life balance, with flexibility and temporal autonomy emerging as defining affordances of digital environments.

Psychological perceptions reveal a differentiated pattern: online learning is associated with higher perceived safety and lower stress, whereas face-to-face learning maintains an advantage in social integration and interpersonal interaction. These findings suggest that digital environments may enhance psychological stability, while face-to-face formats strengthen relational depth and social belonging.

Engagement and motivation appear only moderately influenced by modality. The relatively small differences across formats indicate that student engagement is shaped more by instructional design and institutional support than by delivery mode itself.

Overall, the results support a complementary rather than oppositional understanding of online and face-to-face learning formats. Each modality demonstrates specific affordances that contribute to different aspects of educational effectiveness. Therefore, strategic integration of digital and face-to-face elements may enhance institutional resilience and educational quality.

The study is limited by its reliance on self-reported data and a sample drawn from a single faculty, which may influence generalizability. Future research should incorporate cross-disciplinary samples, multi-institutional comparisons, and longitudinal designs to further examine the evolving relationship between digital and face-to-face learning environments.

REFERENCES

[1] C. van Dorresteijn, D. Fajardo-Tovar, N. Pareja Roblin, F. Cornelissen, M. Meij, J. Voogt, and M. Volman, "What Factors Contribute to Effective Online Higher Education? A Meta-Review," Tech Know Learn, vol. 30, pp. 171-202, 2025, [Online]. Available: https://doi.org/10.1007/s10758-024-09750-5.

[2] C. N. Akpen, S. Asaolu, S. Atobatele, H. Okagbue, and S. Sampson, "Impact of online learning on student's performance and engagement: a systematic review," Discov Educ, vol. 3, art. 205, 2024, [Online]. Available: https://doi.org/10.1007/s44217-024-00253-0.

[3] M. Mohammadi, M. Paasivara, and J. Kasurinen, "Blended learning in higher education: good practices in platforms and teachers support, enhancing students motivation," Educ Inf Technol, vol. 30, pp. 26001-26024, 2025, [Online]. Available: https://doi.org/10.1007/s10639-025-13770-8.

[4] M. Younas, D. A. S. El-Dakhs, and Y. Jiang, "Knowledge construction in blended learning and its impact on students' academic motivation and learning outcomes," Front. Educ., vol. 10, 2025, [Online]. Available: https://doi.org/10.3389/feduc.2025.1626609.

[5] J. Yuan, X. Qiu, J. Wu, J. Guo, W. Li, and Y.-G. Wang, "Integrating behavior analysis with machine learning to predict online learning performance: A scientometric review and empirical study," pp. 1-23, 2024, [Online]. Available: https://doi.org/10.48550/arXiv.2406.11847.

[6] E. Clark, O. Davis, and S. Al Jabri, "The Impact of Using Online Learning Platforms on Student Learning Motivation," International Journal of Educational Narratives, vol. 3, no. 2, pp. 123-132, 2025, [Online]. Available: https://doi.org/10.70177/ijen.v3i2.2148.

[7] V. Arkorful and N. Abaidoo, "The Role of e-Learning, the Advantages and Disadvantages of Its Adoption in Higher Education," International Journal of Education and Research, vol. 2, pp. 397-410, 2014, [Online]. Available: https://www.ijern.com/E-learning.php.

[8] N. Latyk, "Онлайн-освіта в університетах: трансформація та ефективність навчання," Youth and market, vol. 9, no. 241, pp. 172-176, 2025, [Online]. Available: https://doi.org/10.24919/2308-4634.2025.342319.

[9] L. Sudarnoto, M. Handoko, A. Riyanto, and D. Putri Arini, "The impact of online learning, learning motivation, and interpersonal relationships on students' wellbeing," Social Sciences & Humanities Open, vol. 11, pp. 1-9, 2025, [Online]. Available: https://doi.org/10.1016/j.ssaho.2025.101485.

[10] O. Pashchenko, Y. Koroviak, O. Mamaikin, T. Kozhushkina, and V. Rastsvietaiev, "Effectiveness of blended learning in the informatics course," Youth and market, vol. 7-8, no. 239-240, pp. 50-56, 2025, [Online]. Available: https://doi.org/10.24919/2308-4634.2025.334056.

[11] European Commission, "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Digital Education Action Plan 2021-2027: Resetting education and training for the digital age (COM/2020/624 final)," 2020, [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0624.

[12] European Higher Education Area, "European Higher Education Area (EHEA)," [Online]. Available: https://ehea.info.

Predictive Modeling of Information Systems Success in Higher Education: A WarpPLS-SEM Analysis of User-Centered Portals

Coravil Joy C. Avila
College of Information Technology Education, North Eastern Mindanao State University, 8300 Tandag City, Philippines
cjcavila@nemsu.edu.ph

Keywords: UTAUT, ISSM, User Satisfaction, Continuance Intention, Self-Efficacy, Higher Education, Enrollment Portal.

Abstract: This study evaluates the success of a university enrollment web portal through a predictive modeling approach that integrates the Unified Theory of Acceptance and Use of Technology (UTAUT) and the DeLone and McLean Information Systems Success Model (ISSM). Although higher education institutions increasingly deploy digital portals to support academic services, technical availability alone does not ensure user satisfaction, sustained use, or perceived institutional benefits. This study therefore examined the factors influencing user satisfaction, continuance intention, and net benefits in the context of the NEMSU enrollment portal. A mixed-methods design was employed, with quantitative data gathered from 176 students, faculty, and staff and analyzed using Structural Equation Modeling through WarpPLS 8.0. Qualitative inputs were used to contextualize users' experiences and support the interpretation of the quantitative findings. The results showed that system quality did not significantly influence user satisfaction, whereas information quality and service quality were strong predictors. User satisfaction significantly influenced continuance intention, while facilitating conditions and self-efficacy emerged as important drivers of sustained portal use. Mediation analysis further confirmed that continuance intention mediated the relationship between user satisfaction and net benefits. The findings suggest that university ICT administrators should prioritize accurate and timely enrollment information, responsive support services, reliable facilitating conditions, and user training to strengthen self-efficacy. The study contributes to the predictive validation of a hybrid UTAUT-ISSM model and provides practical guidance for improving sustainability, continuance adoption, and success of user-centered portals in higher education.

1 INTRODUCTION

University information systems now carry much of the routine work that students, faculty, and administrative staff previously handled through counter-based transactions. At NEMSU, the enrollment portal forms part of this shift because it supports enrollment-related access, coordination, and information retrieval. The rapid digital turn during and after COVID-19 made such systems necessary, but the more important question for universities is whether users will still rely on them once emergency adoption has passed. For this reason, system effectiveness in this study is treated as more than system availability; it includes continued use, user satisfaction, and perceived performance during actual portal transactions [1] - [3].

Enrollment portals in higher education are expected to simplify institutional services, but they do not become successful simply because they are online. Their value depends on how users experience the system while completing tasks such as checking information, following enrollment procedures, or requesting assistance. UTAUT-based studies show that continued educational technology use is affected by performance expectancy, effort expectancy, and facilitating conditions [1], [4], [5]. Evidence from online examination portal research also indicates that students' intention and actual use are shaped by perceived usefulness, ease of use, and institutional support, showing that portal success involves both system design and the resources surrounding its use [6].

The ISSM perspective provides another basis for evaluating the NEMSU enrollment portal because it connects user satisfaction with the quality of the

system, the quality of information, and the quality of service support [7] - [11]. Previous studies do not assign equal weight to these dimensions. In student information systems and educational technology settings, some users respond more strongly to system/interface quality, while others judge the system through the accuracy of information or the responsiveness of support services. This variation supports the need for a predictive model that is tested within the specific institutional setting rather than assumed from another university context [8], [9], [12].

For this paper, information system success refers to the practical value produced by the NEMSU enrollment portal after users interact with it, not merely the fact that the portal has been deployed [13]. The model therefore includes system quality, information quality, service quality, performance expectancy, effort expectancy, facilitating conditions, self-efficacy, user satisfaction, continuance intention, and net benefits. The ISSM side of the model links the three quality dimensions to user satisfaction, while the UTAUT side accounts for behavioral and support-related factors that may influence continued use [4], [13]. This distinction is important in the NEMSU context because students normally use the portal as part of required enrollment procedures, whereas faculty and administrative staff use it in relation to advising, coordination, verification, and support functions. Thus, satisfaction and continuance intention are interpreted as context-sensitive indicators of whether the portal remains useful, dependable, and worth sustaining over time [14] - [16].

Satisfaction is especially important in this study because it connects the user's immediate experience with the likelihood of future portal use. When users find the portal helpful during enrollment periods, their intention to continue using it becomes a stronger indicator of successful design and support. Prior higher education studies consistently position satisfaction as a predictor of continuance intention in digital learning and portal environments [7], [17]. In addition, studies using integrated acceptance and success models show that UTAUT and ISSM can be combined to explain how users move from initial acceptance to continued engagement [4], [18], [19].

The study therefore tests a user-centered predictive model for the NEMSU enrollment portal. It asks which quality factors predict user satisfaction, which user and support factors predict continuance intention, and how satisfaction and continuance intention explain net benefits. The measures used in the analysis were aligned with higher education research on satisfaction and continuance intention [6], [7], [12], [20] while the interpretation was anchored on broader work on user experience, sustained use, and system outcomes [21], [22]. The expected contribution is practical: the results can guide portal redesign, user assistance, training, and support planning during enrollment periods [1], [17].

2 RELATED LITERATURE

The DeLone and McLean Information Systems Success Model is widely used to evaluate higher education portals because it links system quality, information quality, and service quality with satisfaction and perceived benefits [13] - [15]. In practice, however, these relationships differ across portal settings. Some studies report stronger effects for system or interface design, while others emphasize service responsiveness or information accuracy [9], [14], [15]. For the NEMSU enrollment portal, this means that the analysis must identify the factors that matter most to its own users rather than rely on a general assumption that all quality dimensions contribute equally [23].

Continuance intention extends portal evaluation beyond first-time acceptance. Evidence from student portal studies suggests that system quality can affect satisfaction and future use when the system supports actual tasks through clear information presentation, functionality, ease of use, and perceived performance [17]. Related e-learning acceptance research also shows that system quality may influence acceptance through perceived usefulness and intention to use [24]. These findings support a user-centered interpretation of portal success: users are more likely to remain engaged when the portal structure helps them finish enrollment-related tasks with less confusion and more confidence [14], [17], [24].

PLS-SEM is frequently applied in portal and educational technology studies because it allows researchers to estimate relationships among quality variables, satisfaction, use, success, and perceived benefits. Recent studies further show that success determinants and facilitating conditions may shift depending on system context, user needs, and institutional readiness [14] - [16]. This is relevant to NEMSU because student users, academic personnel, and administrative personnel encounter the portal from different roles. A local predictive model is therefore more appropriate than directly transferring results from other platforms or populations [15], [16], [23].

Overall, the literature supports the view that portal success is shaped by user experience, perceived quality, satisfaction, continuance intention, and perceived benefits [13] - [15], [23]. This paper builds on that body of work by applying a combined ISSM-UTAUT model to one institutional enrollment portal and by examining how the model behaves among users who interact with the system during enrollment support and processing.

3 CONCEPTUAL FRAMEWORK

The conceptual framework integrates UTAUT and ISSM because the NEMSU enrollment portal must be evaluated from both adoption and system-success perspectives. UTAUT explains continued use through constructs such as performance expectancy, effort expectancy, and facilitating conditions [4]. ISSM, on the other hand, focuses on the quality of the system, the quality of the information it provides, the quality-of-service support, and the satisfaction and benefits that follow [13]. Combining the two models allows the study to examine both what users think the portal can help them do and how well the portal performs as an institutional information system.

In the ISSM portion of the framework, system quality covers the portal's functional and technical characteristics, including navigation, structure, availability, and user-friendliness. Information quality refers to whether enrollment-related content is complete, clear, organized, useful, and timely. Service quality refers to the assistance provided by support personnel, including responsiveness, competence, and availability. These three quality dimensions are expected to shape user satisfaction, or the user's overall judgment of the portal experience [7], [9], [13].

The UTAUT portion explains why users may continue or discontinue portal use after initial exposure. Performance expectancy refers to the perceived usefulness of the portal in completing enrollment-related tasks. Effort expectancy concerns how easy the portal is to learn and navigate. Facilitating conditions describe the resources, connectivity, device compatibility, and support structures available to users. Self-efficacy represents the user's confidence in completing portal tasks independently. Within the proposed model, user satisfaction also contributes to continuance intention, and both satisfaction and continuance intention are linked to net benefits [1], [4], [20].

The framework also accounts for differences among NEMSU user groups. Student users are treated as mandatory users because portal access is normally tied to enrollment requirements. Faculty members, program coordinators, college secretaries, and administrative staff may use the system in a more task-specific or semi-voluntary manner, depending on advising, coordination, or processing responsibilities. Their continued use may therefore depend more on support availability, task efficiency, and perceived benefits than on compliance alone. This distinction helps explain why the same success factors may not operate in the same way for all portal users [14] - [16]. The integrated relationships among these constructs are presented in Figure 1.

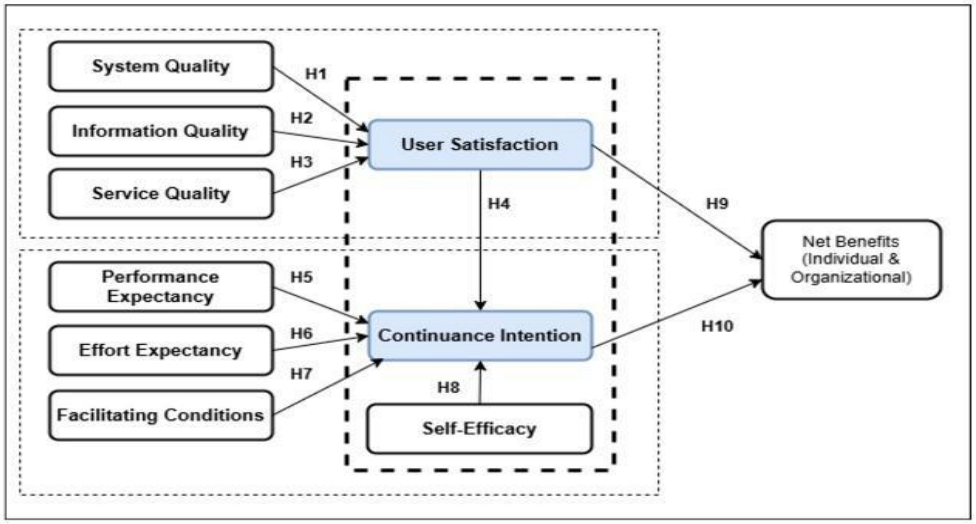


Figure 1: Proposed hybrid UTAUT-ISSM model.

4 METHODOLOGY

4.1 Research Design

The study used a mixed-methods design to examine the NEMSU enrollment portal from both measurement-based and user-experience perspectives. The quantitative component gathered survey responses on portal quality, satisfaction, continuance intention, self-efficacy, and perceived benefits. The qualitative component used exploratory interviews and focus group discussions to document user-reported difficulties, enrollment-related concerns, and suggested improvements. These qualitative inputs were not analyzed as a separate thematic-results section; they were used to support the interpretation of the SEM findings and to keep the discussion grounded in actual portal use. In this way, the numerical model was read alongside practical concerns raised by students and support personnel.

4.2 Flow of the Study

The study began with a review of literature on UTAUT, ISSM, portal success, satisfaction, and continued use. After the framework was established, survey questionnaires and interview guides were prepared for users involved in the NEMSU enrollment process. Data collection was conducted among students, faculty, and staff who interacted with the university information system. The survey was distributed through Google Forms within the main campus of Northeastern Mindanao State University.

After data collection, the responses were organized for quantitative analysis and checked against the intended research objectives. The analysis focused on identifying which portal quality and user-related factors should be prioritized for improvement. The qualitative inputs were used to clarify how users experienced the portal during enrollment-related activities. The process ended with recommendations for improving the NEMSU enrollment portal. Figure 2 summarizes the input-process-output flow followed in the study.

4.3 Research Variables and Evaluation Criteria

To make the evaluation criteria explicit, the study measured ten latent constructs from the hybrid UTAUT-ISSM model. The constructs were measured using Likert-scale indicators and interpreted through descriptive statistics and SEM path testing. Table 1 identifies each variable, its role in the model, and the basis used to evaluate it in relation to the NEMSU enrollment portal.

4.4 Participants

The respondents represented the main user groups of the NEMSU enrollment portal. Students were included as primary users, while administrative staff, college secretaries, program coordinators, and registrar personnel were included because they support student-facing enrollment processes [9], [11], [12]. This study employed a convenience sampling in that the target respondents were already within the university and had experiences directly with using the portal. These approaches are common in the studies of higher education portal [12] and e-learning portals [11], [14]; that collected data from the available population (typical end-users) via a campus-based or online method. This resulted in 176 valid responses which were used for the SEM estimation after screening of the dataset.

4.5 Analysis of Data

A total of 233 responses were collected from the original dataset. Data were also screened prior to model estimation for incomplete entries, inconsistent responding, outlying cases, and straight-lining. An approach of using a standard deviation cut-off, 0.00 to 0.25 was used to identify straight-lining [25], [26]. This left us with 176 valid cases after cleaning, removing fifty-seven responses.

The sample adequacy was examined by running a priori power analysis for F-test of multiple regression. The computation was based on nine predictors, an effect size of $f^2 = 0.15$, $\alpha = 0.05$, and power $(1 - \beta) = 0.95$. The sample size of 166 was required, yet the final dataset entry comprised a total of 176 valid respondents. So the retained sample was above the minimum requirement and sufficient for SEM testing [1], [20]. Table 2 presents the power-analysis parameters used to justify the sample size.

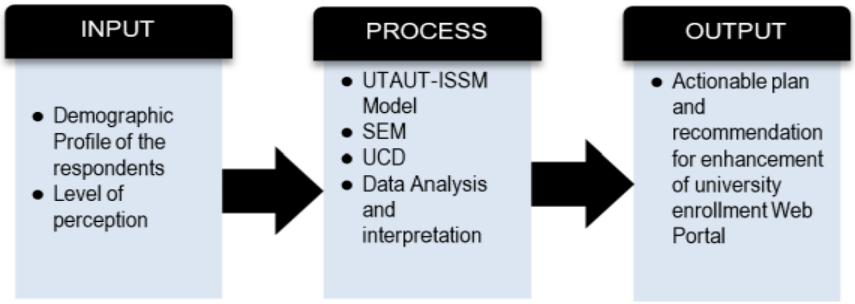


Figure 2: Flow of the study.

Table 1: Research variables and evaluation criteria.

Variable	Model Role	Evaluation Basis
System Quality	ISSM antecedent of user satisfaction	Navigation, user-friendliness, structure, reliability, and availability.
Information Quality	ISSM antecedent of user satisfaction	Completeness, clarity, organization, usefulness, timeliness, and decision support.
Service Quality	ISSM antecedent of user satisfaction	Responsiveness, competence, support availability, instructions, and training.
User Satisfaction	Central success construct	Overall experience, expectation fulfilment, efficiency, effectiveness, and contentment.
Performance Expectancy	UTAUT predictor of continuance intention	Perceived usefulness, productivity, task support, and time savings.
Effort Expectancy	UTAUT predictor of continuance intention	Ease of learning, clarity of interaction, navigation, and mobile convenience.
Facilitating Conditions	UTAUT predictor of continuance intention	Resources, knowledge, internet access, support team availability, and device compatibility.
Self-Efficacy	Extended user capability construct	Confidence in independent use and completion of portal tasks.
Continuance Intention	Post-adoption behavioral construct	Intention to continue, regular future use, preference, and sustained use.
Net Benefits	ISSM outcome construct	Time savings, decision support, resource reduction, data access, and administrative control.

Table 2: Distribution of respondents using G*power sample size estimation.

Parameter	Value
Test Family	F tests
Statistical Test	Linear multiple regression: Fixed model, R^2 deviation from zero
Type of Power Analysis	A priori: Compute required sample size
Effect size (f^2)	0.15
α error probability	0.05
Power (1 - β error probability)	0.95
Number of predictors	9
Noncentrality parameter (λ)	24.9000000
Critical F	1.9403478
Numerator degrees of freedom (df)	9
Denominator degrees of freedom (df)	156
Total sample size required	166
Actual Respondents	176
Actual Power	0.9500973

5 RESULTS

5.1 Hypothesis Testing

The WarpPLS-SEM analysis showed which paths in the proposed model were statistically supported. Information quality and service quality had significant effects on user satisfaction, while system quality did not. Among the satisfaction predictors, service quality produced the strongest coefficient, indicating that users responded strongly to the availability and responsiveness of assistance during enrollment-related transactions. Table 3 presents the path coefficients, probability values, effect sizes, and decisions for the hypothesized relationships.

User satisfaction had a significant positive effect on continuance intention, meaning that users who evaluated the portal experience more favorably were more likely to express future use. Continuance intention also had a significant effect on net benefits, indicating that the value of the NEMSU enrollment portal depends partly on sustained use rather than one-time access. These results show that satisfaction contributes to long-term use and to the realization of individual and institutional benefits. Figure 3 presents the corresponding structural model with beta coefficients.

Performance expectancy and effort expectancy were not significant predictors of continuance intention in this dataset. In the NEMSU setting, this may reflect the mandatory nature of enrollment-related portal use: students may continue using the portal because it is required, even when perceived usefulness or ease of use is not the strongest driver. In contrast, facilitating conditions and self-efficacy were significant predictors. Users were more likely to intend continued use when they had sufficient resources, support, and confidence to complete portal tasks independently.

Table 3 Results of hypothesis testing.

Hypothesis	Path	β	P-value	f^2	Decision
1	SysQ → US	-0.013	0.432	0.007	Not Supported
2	IQ → US	0.282	<0.001	0.180	Supported
3	ServQ → US	0.484	<0.001	0.333	Supported
4	US → CI	0.263	<0.001	0.170	Supported
5	PE → CI	0.056	0.228	0.016	Not Supported
6	EE → CI	0.064	0.196	0.024	Not Supported
7	FC → CI	0.184	0.006	0.097	Supported
8	SE → CI	0.422	<0.001	0.304	Supported
9	US → NB	0.271	<0.001	0.143	Supported
10	CI → NB	0.419	<0.001	0.245	Supported
11	US → CI → NB	0.110	0.018	0.058	Supported

Note: SysQ = System Quality; IQ = Information Quality; ServQ = Service Quality; US = User Satisfaction; PE = Performance Expectancy; EE = Effort Expectancy; FC = Facilitating Conditions; CI = Continuance Intention; SE = Self-Efficacy; NB = Net Benefits; $f^2 \geq 0.02$ = small effect; $f^2 \geq 0.15$ = medium effect; $f^2 \geq 0.35$ = large effect.

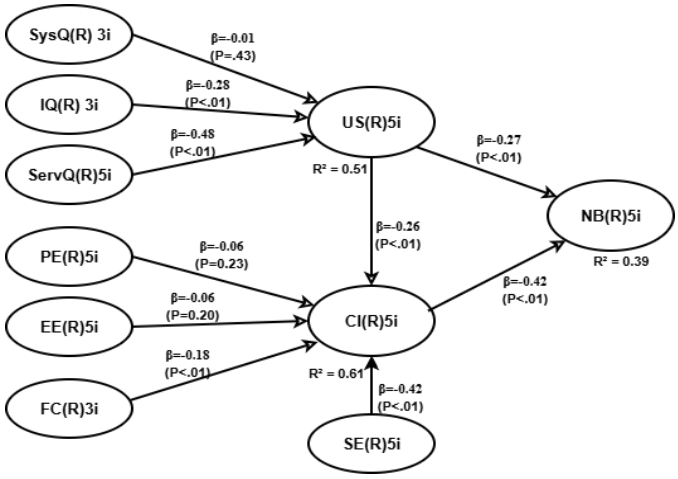


Figure 3: Structural model with beta coefficients.

Table 4: Coefficient of determination, full collinearity VIF, and Q^2.

Endogenous Construct	R^2	Full Collinearity VIF	Q^2
User Satisfaction	0.507	2.598	0.507
Continuance Intention	0.612	2.417	0.614
Net Benefits	0.388	2.112	0.395

Note: R^2 = coefficient of determination; Q^2 = Stone-Geisser's predictive relevance value.

5.2 Coefficient of Determination and Predictive Relevance

The model showed acceptable explanatory and predictive performance for the endogenous constructs. As shown in Table 4, the R^2 values for user satisfaction, continuance intention, and net benefits were all above zero, indicating that the model explained meaningful variation in these outcomes. The positive Q^2 values further indicate that the model had predictive relevance.

For NEMSU, the results point to practical improvements that should be prioritized during enrollment periods. The portal should provide accurate and timely information, while support personnel should be available to assist users who encounter access, navigation, or transaction-related problems. The hybrid UTAUT-ISSM model therefore offers guidance not only for predicting portal use but also for improving the support conditions that make continued use possible.

5.3 Integrated Interpretation of Success Factors

The strongest success factors in the NEMSU enrollment portal context were information quality, service quality, facilitating conditions, and self-efficacy. Information quality and service quality predicted satisfaction more strongly than system quality, suggesting that users placed greater value on accurate enrollment information and responsive assistance than on technical features alone. Facilitating conditions and self-efficacy also mattered for continuance intention, indicating that sustained use depends on resources, institutional support, and user confidence.

The findings also indicate that portal success should be interpreted differently for each user group. Students may continue using the portal because enrollment procedures require it, but their satisfaction still depends on whether information is clear and support is accessible. Faculty and administrative personnel use the portal in relation to advise, coordination, verification, and processing tasks; therefore, they may judge success through efficiency, support availability, and work-related benefits. A successful NEMSU enrollment portal must therefore combine reliable technical operation with human support that responds to the needs of different users.

6 CONCLUSIONS

This study confirms the suitability of hybrid UTAUT-ISSM model for explaining the success in using NEMSU enrollment portal. The results indicated that for the hypothesized relationships, information quality and service quality affected user satisfaction to a significant extent; Moreover, user satisfaction considerably influenced continuance intention and net benefits in the context of WarpPLS-SEM. It concludes that user satisfaction is not only a short-term reaction to the portal experience but also serves as an essential prerequisite for sustained usage of the information portals.

Supporting factors and self-efficacy were also important for continued portal use. In practical terms, people were more likely to remain engaged with the NEMSU enrollment portal when they had resources available to them and a belief that they would successfully be able to do what needs doing on the portal. The findings also indicate that students, faculty, and administrative staff may assess the portal from different roles and expectations. Enrollment needs drive students more so than usefulness and support quality, as well as benefits to faculty members' work processes for administrative staff.

The model's explanatory ability, the R^2 values for user satisfaction, continuance intention, and net benefits, shows that the integrated framework can explain both adoption-related and sustained-use outcomes. For NEMSU, the model can guide decisions on portal support, information management, training, and future system improvement. The integration of UTAUT and ISSM therefore provides a structured basis for evaluating educational portal success.

Several limitations should be considered. The first was that the data were obtained from users of the NEMSU enrollment portal and, as such, the findings may be less applicable to institutions with different

infrastructure, level of digital maturity or support arrangements. Second, the study involved students, faculty and administrative staff but was heavily weighted toward student users whose responses may have driven the overall pattern. Third, using a cross-sectional design in the present study means user perceptions were assessed at a single point and cannot reveal how satisfaction, self-efficacy or continuance intention may change as a function of repeated portal use. Future research could examine the trans-campus/institutional hierarchies of the hybrid UTAUT-ISSM model, employ longitudinal methodologies, and investigate further factors like system trust, perceived institutional support, digital literacy skills and user group classification. Such extensions may clarify how satisfaction develops into continuance use and meaningful organizational outcomes.

REFERENCES

[1] E. Marandu, I. Mathew, T. Svotwa, R. Machera, and O. Jaiyeoba, "Predicting students' intention to continue online learning post-COVID-19 pandemic: Extension of the unified theory of acceptance and usage technology," J. Appl. Res. High. Educ., vol. 15, no. 3, pp. 681-697, 2022, [Online]. Available: https://doi.org/10.1108/jarhe-02-2022-0061.

[2] K. Vasuthevan, S. Vaithilingam, and J. Ng, "Academics' continuance intention to use learning technologies during COVID-19 and beyond," PLoS One, vol. 19, no. 1, art. no. e0295746, 2024, [Online]. Available: https://doi.org/10.1371/journal.pone.0295746.

[3] S. Nikou, "Web-based videoconferencing in online teaching during the COVID-19 pandemic: University students' perspectives," in Proc. IEEE Int. Conf. Adv. Learn. Technol. (ICALT), 2021, pp. 431-435, [Online]. Available: https://doi.org/10.1109/ICALT52272.2021.00137.

[4] Y. Li and M. Zhao, "A study on the influencing factors of continued intention to use MOOCs: UTAUT model and CCC moderating effect," Front. Psychol., vol. 12, art. no. 528259, 2021, [Online]. Available: https://doi.org/10.3389/fpsyg.2021.528259.

[5] C. Gan, Y. Bong, G. Goh, and T. Liew, "Supporting student-centred learning with mobile technology: UTAUT and Hofstede's dimensions of culture," Cult. Educ., vol. 37, no. 1, pp. 268-300, 2025, [Online]. Available: https://doi.org/10.1177/11356405251316068.

[6] S. Shirolkar and R. Kadam, "Determinants of adoption and usage of the online examination portal (OEP) in Indian universities," Educ. + Training, vol. 65, no. 6/7, pp. 827-847, 2023, [Online]. Available: https://doi.org/10.1108/ET-09-2022-0360.

[7] M. Salim, L. Alfansi, A. Sularsih, F. Saputra, and C. Afandy, "The role of perceived usefulness in moderating the relationship between the DeLone and McLean model and user satisfaction," Uncertain Supply Chain Manage., vol. 9, no. 3, pp. 755-766, 2021, [Online]. Available: https://doi.org/10.5267/j.uscm.2021.4.002.

[8] K. Çelik and A. Ayaz, "Validation of the DeLone and McLean information systems success model: A study on student information system," Education and Information Technologies, 2022, [Online]. Available: https://doi.org/10.1007/s10639-021-10798-4.

[9] P. E. Ramírez-Correa, F. J. Rondán-Cataluña, and J. Arenas-Gaitán, "Student information system satisfaction in higher education: The role of visual aesthetics," Kybernetes, vol. 47, no. 8, pp. 1604-1622, 2018, [Online]. Available: https://doi.org/10.1108/K-08-2017-0297.

[10] D. Al-Fraihat, M. Joy, R. Masa'deh, and J. Sinclair, "Evaluating e-learning systems success: An empirical study," Computers in Human Behavior, vol. 102, pp. 67-86, 2020, [Online]. Available: https://doi.org/10.1016/j.chb.2019.08.004.

[11] M. Shams, M. Niazi, H. Gul, T. Mei, and K. Khan, "E-learning adoption in higher education institutions during the COVID-19 pandemic: A multigroup analysis," Front. Educ., vol. 6, 2022, [Online]. Available: https://doi.org/10.3389/feduc.2021.783087.

[12] A. Shahzad, R. Hassan, A. Aremu, A. Hussain, and R. Lodhi, "Effects of COVID-19 in e-learning on higher education institution students: The group comparison between male and female," Qual. Quant., vol. 55, no. 3, pp. 805-826, 2020, [Online]. Available: https://doi.org/10.1007/s11135-020-01028-z.

[13] W. H. DeLone and E. R. McLean, "The DeLone and McLean model of information systems success: A ten-year update," J. Manage. Inf. Syst., vol. 19, no. 4, pp. 9-30, 2003.

[14] W. A. Cidral, T. Oliveira, M. Di Felice, and M. Aparicio, "E-learning success determinants: Brazilian empirical study," Computers & Education, vol. 122, pp. 273-290, 2018, [Online]. Available: https://doi.org/10.1016/j.compedu.2017.12.001.

[15] B. Bervell, D. O. Mireku, and D. Agyapong, "Modelling the antecedents of students' satisfaction and continuous use intentions of an electronic appraisal portal system in higher education," Computers in Human Behavior Reports, vol. 15, art. no. 100431, 2024, [Online]. Available: https://doi.org/10.1016/j.chbr.2024.100431.

[16] M. A. Almaiah, A. Al-Khasawneh, and A. Althunibat, "Exploring the critical challenges and factors influencing the e-learning system usage during COVID-19 pandemic," Education and Information Technologies, vol. 25, no. 6, pp. 5261-5280, 2020, [Online]. Available: https://doi.org/10.1007/s10639-020-10219-y.

[17] B. Bervell et al., "Modelling distance education students' satisfaction and continuous use intention of students' portal: An importance-performance map analysis," Inf. Dev., vol. 42, no. 1, pp. 104-124, 2023, [Online]. Available: https://doi.org/10.1177/02666669231206774.

[18] A. AL-Hawamleh, "Exploring the satisfaction and continuance intention to use e-learning systems," Int. J. Electr. Comput. Eng. Syst., vol. 15, no. 2, pp. 201-214, 2024, [Online]. Available: https://doi.org/10.32985/ijeces.15.2.8.

[19] D. Kala and D. Chaubey, "Cryptocurrency adoption and continuance intention among Indians: Moderating role of perceived government control," Digit. Policy Regul. Gov., vol. 25, no. 3, pp. 288-304, 2023, [Online]. Available: https://doi.org/10.1108/DPRG-09-2022-0108.

[20] S. Masrani, M. Amin, V. Sivakumaran, and S. Piaralal, "Important factors in measuring learners' satisfaction and continuance intention in open and distance learning (ODL) institutions," High. Educ. Ski. Work-Based Learn., vol. 13, no. 3, pp. 587-608, 2023, [Online]. Available: https://doi.org/10.1108/HESWBL-12-2022-0274.

[21] A. Siyal, H. Chen, G. Chen, M. Memon, and Z. Binte, "Structural equation modeling and artificial neural networks approach to predict continued use of mobile taxi booking apps: The mediating role of hedonic motivation," Data Technol. Appl., vol. 55, no. 3, pp. 372-399, 2020, [Online]. Available: https://doi.org/10.1108/DTA-03-2020-0066.

[22] K. Tang et al., "Predicting continuance intention and use of mobile shopping apps with PLS-SEM and necessary condition analysis in tandem," J. Appl. Struct. Equ. Model., vol. 8, no. 1, pp. 1-27, 2024, [Online]. Available: https://doi.org/10.47263/jasem.8(1)02.

[23] E. Tetteh, Z. Qin, and B. Kwofie, "Computer-mediated communication portal implementation framework: A higher education institutional perspective," Int. J. Emerg. Technol. Learn., vol. 15, no. 3, p. 180, 2020, [Online]. Available: https://doi.org/10.3991/ijet.v15i03.11641.

[24] M. I. Alkhawaja, M. S. Abd Halim, M. S. S. Abumandil, and A. S. Al-Adwan, "System quality and student's acceptance of the e-learning system: The serial mediation of perceived usefulness and intention to use," Contemporary Educational Technology, vol. 14, no. 2, art. no. ep350, 2022, [Online]. Available: https://doi.org/10.30935/cedtech/11525.

[25] S. Gautam and V. Singh, "Adaptive discretization using golden section to aid outlier detection for software development effort estimation," IEEE Access, vol. 10, pp. 90369-90387, 2022, [Online]. Available: https://doi.org/10.1109/ACCESS.2022.3200149.

[26] J. Karch, "Outliers may not be automatically removed," J. Exp. Psychol. Gen., vol. 152, no. 6, pp. 1735-1753, 2023, [Online]. Available: https://doi.org/10.1037/xge0001357.

Integration of Digital Language Resources into Foreign Language Teaching for Students of Non-Linguistic Degree Programmes

Svitlana Sytniakivska, Snizhana Kubrak, Svitlana Khmelivska, Karyna Razumna and Svitlana Petrova
Polissia National University, Staryi Boulevard 7, 10008 Zhytomyr, Ukraine
{skarlet7760, snezha.zt, svitlana.khmelivska, mardza.kr, svitlana.petrova1992}@gmail.com

Keywords: Digital Resources, Foreign Language, Non-Linguistic University, Pedagogical Technologies, Communicative Competence.

Abstract: The rapid digitalization of higher education and the growing demands of the labor market have made foreign language proficiency essential for students of non-linguistic specialties. This study examines the effectiveness of integrating digital language resources into foreign language teaching at Polissia National University. The research employed a mixed-methods approach and was conducted over one academic semester with 78 students majoring in Economics, Law, Public Administration, and Accounting. Participants were divided into an experimental group and a control group. The experimental group used digital tools, including Quizlet, Google Classroom, interactive online dictionaries, and the AI-based assistant ChatGPT, while the control group followed traditional teaching methods. Data were collected through vocabulary and grammar tests, written assignments, oral presentations, questionnaires, and classroom observations. The findings indicate that students in the experimental group demonstrated significantly higher academic performance compared to the control group. Moreover, the use of digital resources enhanced students' motivation, learning autonomy, communicative competence, and critical thinking skills. The results confirm that integrating learning management systems, interactive vocabulary platforms, and AI tools improves both language proficiency and student engagement. The study contributes to research on digital pedagogy and offers practical recommendations for foreign language teachers in non-linguistic higher education.

1 INTRODUCTION

In today's globalized and multicultural world, proficiency in foreign languages (particularly English) has become an essential competence for students' professional advancement and academic development. At Polissia National University, this especially applies to non-linguistic degree programmes such as Business Economics, International Economic Relations, Law, Public Administration, Accounting, and Finance. In these fields, effective command of a foreign language facilitates international communication, provides access to scholarly sources, and enables participation in conferences as well as project-based activities.

In this context, contemporary students tend to demonstrate a high level of digital literacy and expect interactive and personalized learning approaches. While traditional teaching methods remain effective for the basic acquisition of grammar and vocabulary, they do not always adequately address students' needs in developing communicative competence, analytical skills, and learner autonomy.

The functional roles of the digital tools in achieving learning outcomes were differentiated according to their pedagogical purpose within the instructional process. Quizlet served as a vocabulary acquisition and retention tool through spaced repetition, flashcards, and gamified practice activities that reinforced lexical knowledge. Google Classroom functioned as the organizational and communicative hub, enabling structured distribution of materials, timely feedback, assignment management, and support for collaborative learning. ChatGPT acted as an AI-assisted language support tool that facilitated sentence construction, idea formulation, and grammar self-correction during writing tasks. Interactive online dictionaries provided multimodal lexical input, including definitions, pronunciation models, visual illustrations, and contextual examples, thereby supporting autonomous vocabulary development and pronunciation accuracy. The use of

digital language resources, such as interactive dictionaries, vocabulary-learning platforms, learning management systems, and AI-based assistants makes it possible to solve several relevant pedagogical tasks simultaneously.

First and foremost, this approach contributes to increased student motivation and engagement. Secondly, this approach supports the individualization of the learning process. Thirdly, this approach expands opportunities for the development of communicative and critical skills.

Finally, and no less importantly, this approach contributes to students' preparation for professional activity. For degree programmes in Economics, Law, Public Administration, Finance, and Accounting, a high level of English proficiency constitutes a key factor in career success and effective integration into international professional environments.

Thus, the relevance of the present study is determined not only by global trends in the digitalization of education, but also by the fact that, despite extensive international research on the effectiveness of digital resources in foreign language teaching, empirical studies in Ukraine that offer a comprehensive analysis of the use of interactive platforms, AI-based assistants, and online dictionaries in the education of students of non-linguistic specializations remain limited. Accordingly, the present study aims to address this research gap by demonstrating the practical effectiveness and pedagogical benefits of integrating digital technologies into the teaching and learning process.

2 LITERATURE REVIEW

Recent studies in the field of foreign language teaching methodology confirm the high effectiveness of digital resources in the learning process. International scholars emphasize that the integration of online platforms, gamification technologies, and AI-based tools contributes to increased student motivation as well as to more effective acquisition of vocabulary and grammatical structures.

In their comprehensive review, C. Chapelle and S. Sauro emphasize that language learning technologies enable the implementation of individualized learning trajectories, allow instructional materials to be adapted to learners' proficiency levels, and provide rapid feedback. The authors place particular emphasis on interactive platforms and artificial intelligence–based resources, highlighting their potential to foster the development of communicative competence and creativity [1].

R. Godwin-Jones highlights that digital tools such as Quizlet and online dictionaries significantly facilitate the acquisition of new vocabulary and contribute to the development of learner autonomy. In particular, the systematic repetition of lexical items through interactive flashcards enables more effective consolidation of knowledge than traditional instructional methods [2].

In his study on the use of mobile technologies in language teaching, G. Stockwell demonstrates that interactive platforms stimulate student engagement, foster the development of cognitive skills and communicative competence, and create favorable conditions for the gamification of learning, which in turn enhances learner motivation [3], [4].

The use of Google Classroom has also been recognized as effective. K. Johnson describes how learning management systems (LMS) facilitate the organization of instructional materials, enable monitoring of student progress, and support collaborative learning. These features are particularly relevant for non-linguistic degree programmes, where group work and knowledge sharing play a central role in the learning process [5].

In Ukraine, the relevance of digital resources for foreign language teaching is supported by the studies of I. Riabchenko and L. Kovalchuk. I. Riabchenko notes that interactive exercises, online dictionaries, and gamified elements significantly enhance student motivation and contribute to more effective acquisition of vocabulary and grammatical structures [6]. Drawing on research conducted with students pursuing non-linguistic degrees, L. Kovalchuk demonstrates that the use of digital platforms promotes learner autonomy, improves both written and oral skills, and fosters the development of critical thinking [7].

The literature review also indicates that the use of AI-based assistants (particularly ChatGPT) in language learning represents a relatively new yet promising research direction. Studies by international scholars demonstrate that students actively employ AI tools for generating examples, checking grammatical accuracy, and preparing written assignments, which substantially enhances the effectiveness of self-directed learning [8], [9], [10], [11], [12].

Thus, the analysis of existing literature confirms that the integration of digital resources into foreign language instruction for students in non-linguistic degree programmes is both theoretically grounded and pedagogically effective. However, within the

Ukrainian context, there remains a limited number of empirical studies that provide a comprehensive assessment of the impact of digital platforms, AI-based assistants, and interactive dictionaries on students' academic achievement and learning motivation. This gap underscores the relevance and innovative character of the present study.

3 RESEARCH METHODOLOGY

In contemporary education, the integration of digital technologies into foreign language teaching is gaining increasing significance. Empirical evidence indicates that the use of resources, such as Quizlet, significantly improves vocabulary learning outcomes, resulting in higher levels of lexical retention. Research findings also demonstrate that both students and teachers show a high degree of acceptance of and readiness to employ artificial intelligence–based tools, particularly ChatGPT. The Google Classroom platform, in turn, facilitates collaborative learning by enhancing student motivation and academic performance in English language learning. In addition, interactive online dictionaries provide access to multimodal resources (including textual definitions, pronunciation models, and visual illustrations), which expands opportunities for autonomous vocabulary acquisition.

Accordingly, the aim of the present study is to determine the extent to which the integrated use of digital language resources (Quizlet, Google Classroom, ChatGPT, and interactive online dictionaries) enhances the effectiveness of foreign language learning among students majoring in Economics, Law, and related degree programmes. The methodological framework of the study was based on an experimental–control design, which made it possible to compare traditional instruction with technology-enhanced learning.

Use of Quizlet. During English language classes, students in the experimental group regularly worked with interactive Quizlet flashcard sets containing field-specific vocabulary relevant to their fields of study, including economics, law, public administration, and accounting. Each student was able to study the flashcards individually and complete timed tests designed to assess the speed and accuracy of lexical retention. In practical classes, the educator also organized group-based competitive activities in the form of live quizzes, during which students shared their results and analyzed errors. This approach made it possible to combine individual revision with group interaction, while fostering healthy competition and enhancing student motivation.

Use of Google Classroom. Google Classroom functioned as the primary system for course organization. Students received access to instructional materials, homework assignments, supplementary exercises, and thematic video content. Educators used the platform to review student work and deliver feedback in real time. Particular attention was given to tasks involving the use of professional terminology, such as the preparation of short analytical reports in English on topics related to economics or law. The use of Google Classroom also enabled students to plan their learning activities in advance and revise course materials at any convenient time, which significantly enhanced levels of learner autonomy and responsibility.

Use of ChatGPT. The AI-based assistant ChatGPT was employed to provide individualized support to students during the preparation of written assignments and classroom discussions. Students in the experimental group were able to obtain model formulations, check the grammatical accuracy and stylistic relevance of their texts, and generate new examples of terminology use in the context of profession-oriented tasks. Educators' observations indicated that the use of ChatGPT contributed to the development of critical thinking skills. Rather than relying solely on ready-made outputs, students learned to analyze AI-generated examples, compare them with alternative formulations, and draw independent conclusions. In addition, the AI assistant supported students in preparing oral presentations and role-play activities, which increased their confidence in using the foreign language.

Use of interactive dictionaries. Interactive dictionaries were utilised both during classroom practice and as part of homework assignments. Students had access to resources with audio pronunciation, contextualized example sentences, visual representations, and thematic illustrations. During practical classes, educatord organized tasks focused on the rapid identification of word meanings, the use of their combinations in professional contexts, and interactive text-translation activities. This approach enabled effective consolidation of vocabulary, supported the development of associative thinking, and enhanced students' skills in self-control of lexical knowledge.

Integration of digital platforms into the learning process. Over the course of the term, all four digital resources were integrated into the modular structure of the course. For example, students were introduced to a new topic via Google Classroom, acquired

relevant terminology using Quizlet, and verified the accuracy and contextual use of lexical items with the support of ChatGPT and interactive dictionaries. Final practice activities were completed through online tests and pair-based tasks during practical classes. This comprehensive integration ensured a balanced combination of theoretical knowledge, practical skills, and the development of learner autonomy.

Methodology. The study was conducted over the course of one academic term and involved 78 students from Polissia National University, enrolled in the following degree programmes: Business Economics, Law, Public Administration and Management, and Accounting.

Participants were recruited through convenience sampling from intact academic groups enrolled in the selected degree programmes, with eligibility criteria including regular class attendance and comparable baseline English proficiency as determined by pre-experimental testing. To minimize selection bias, students who met these criteria were proportionally represented from each programme. Assignment to the experimental and control groups was conducted using stratified random allocation to ensure equivalence in terms of specialization, gender distribution, and initial language proficiency. This procedure enhanced the internal validity of the study by ensuring the comparability of groups prior to the intervention.

The participants were divided into two groups: an experimental group (EG) and a control group (CG), each comprising 39 students. Students from all listed degree programmes were represented proportionally in both groups. The instruction process in the experimental group (EG, n=39) was supported by an integrated set of digital language resources, whereas the control group (CG, n=39) was taught using traditional instructional methods without the use of specialized digital apps.

Within the instructional process, the EG engaged in a range of tasks supported by the following tools and resources: Quizlet (to create courses and flashcard sets for vocabulary acquisition, including gamified practice activities and thematic quizzes aligned with each instructional unit); Google Classroom (the central platform for organizing course materials, distributing, collecting, and assessing assignments); ChatGPT (an AI-based assistant to support the preparation of written tasks, generate example sentences, and assist with grammar correction); interactive online dictionaries with audio pronunciation and visual illustrations (to expand learners' vocabulary, support pronunciation self-monitoring, and provide contextualized examples of word usage).

The experiment was conducted throughout the entire term, with regular classroom instruction maintained in both groups. However, the EG systematically completed additional activities using digital platforms, whereas the CG performed comparable lexical and grammatical tasks in a traditional format, including printed materials, conventional dictionaries, and communicative exercises without the use of the specified online resources. This design ensured a clear basis for comparison, as the only variable distinguishing the two groups was the integration of digital technologies. The methodological framework also included ongoing monitoring of both groups' progress through monthly tests, as well as the collection of participant feedback via questionnaires. The alignment of test tasks and the evaluation of students' language proficiency were conducted according to identical criteria for both groups.

Data collection instruments and methods. To evaluate the effectiveness, both quantitative and qualitative research methods were employed. Foreign language proficiency was assessed through pre- and post-experimental testing, which included structured tests measuring lexical knowledge, grammatical accuracy, and reading comprehension. Identical tests were used for both the EG and the CG. Academic performance was further analyzed by comparing the average term grades (final exam or pass/fail assessment) of students in the EG and CG. In addition, questionnaires and surveys were used to collect students' perceptions of digital resource use, focusing on motivation, learning satisfaction, and overall learning experience. Additional data were obtained through instructor observation, which involved systematic teaching notes documenting student engagement, classroom activity, and the dynamics of group interaction. Finally, a qualitative analysis of individual student work was conducted: samples of written assignments were examined with respect to lexical diversity and grammatical accuracy.

The survey was held in two stages: at the beginning of the experiment, in order to document participants' initial attitudes and expectations, and at the end of the term, to assess any changes over time. The instrument consisted of standardized questions, partially adapted from established methodological frameworks, and employed a Likert-scale format for response measurement. Data obtained through instructor observation made it possible to capture relevant contextual factors, such as changes in the level of classroom discussion and student

engagement, and the ways in which learners utilized AI-based tools, including ChatGPT, as well as online dictionaries during the learning process.

4 RESULTS

Quantitative analysis of academic performance. A comparison of mean grades received during the final tests in both groups revealed a statistically significant improvement in the experimental group. Specifically, the average grade on the vocabulary knowledge test in the EG increased by 18% compared to the baseline level, whereas the CG demonstrated an increase of only 7%. The difference between the two groups was statistically significant ($p<0.05$). In addition, the EG showed better progress in reading comprehension and writing skills. The mean grade on the written examination increased substantially compared to the previous term, with gains notably higher than those observed in the CG. Overall, 85% of students in the EG advanced to higher proficiency levels according to the CEFR scale, compared to 60% in the CG (see Fig. 1).

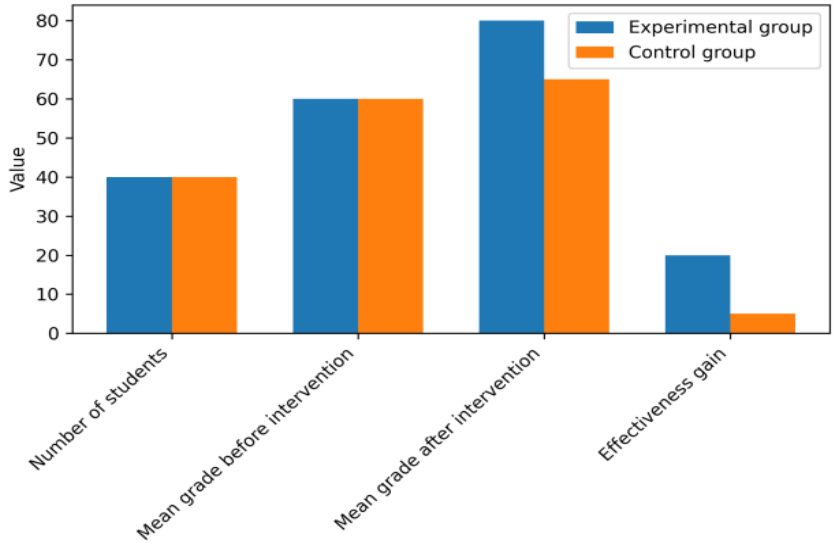


Figure 1: Learning outcomes of students in the EG and CG following the implementation of digital resources.

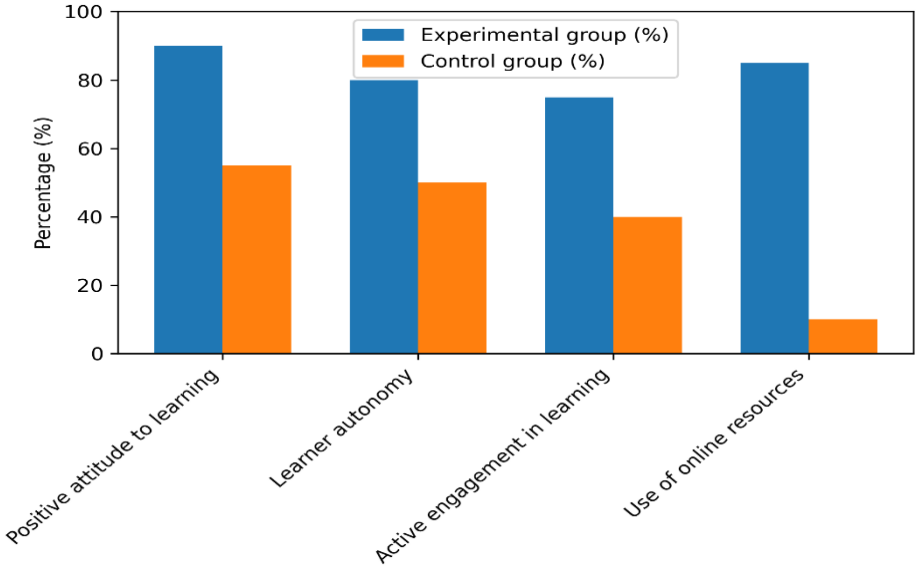


Figure 2: Increase in student motivation and engagement among learners using digital resources

Qualitative analysis and assessment of motivation. According to the questionnaire data, students in the experimental group reported an increase in motivation and engagement. Specifically, 90% of EG respondents indicated that the classes were more engaging due to the use of interactive activities, whereas only 55% of students in the control group reported a similar perception. The majority of EG participants positively evaluated the use of Quizlet as a gamified component, noting that it "makes learning easier and more enjoyable." Furthermore, 80% of students emphasized that the ability to complete tasks at a convenient time through online platforms
enhanced their learner autonomy, which they described as increased "independence from time and place of preparation."

Many participants also highlighted the convenience of Google Classroom, noting that the platform enables prompt feedback from the teacher and allows students to monitor their progress. In open-ended responses, students reported that ChatGPT assisted them in formulating their ideas more clearly in English by providing useful example sentences. The survey results indicated that 85% of students in the EG favor the use of ChatGPT in the learning process, perceiving it as an "interactive assistant".

Comparison of group interaction. Instructor observations indicated that students in the EG engaged more actively in pair and small-group work during class sessions. They willingly discussed responses to Quizlet-based quizzes and collaboratively analyzed the results. This level of collaborative engagement is consistent with findings of the research, which suggests that the use of Google Classroom fosters teamwork and contributes to more positive student attitudes toward the subject. In contrast, learning in the control group was more formalized, and classroom discussions occurred less frequently (see Fig. 2).

Thus, the quantitative analysis revealed significantly stronger academic outcomes in the EG, while the qualitative findings confirmed higher levels of student motivation and self-organisation. In particular, the majority of participants noted that "due to the use of digital tools, communication became more interactive," and that the process of foreign language learning was perceived as clearer and more accessible.

5 CONCLUSIONS

The pedagogical investigation confirmed the high effectiveness of integrating digital language resources into foreign language instruction for students in non-linguistic degree programmes. The application of Quizlet, Google Classroom, ChatGPT, and interactive online dictionaries resulted in a substantial increase in lexical proficiency, improved writing skills and reading comprehension, and greater activation of self-directed learning processes. Students' positive perception of ChatGPT as a virtual academic support tool indicates the potential of AI-based assistants to stimulate sustained interest in learning English.

The experimental findings demonstrate not only improved academic performance - with the experimental group achieving an 18% increase in vocabulary test scores compared to 4% in the control group - but also a significant enhancement of student motivation and engagement, with 90% of EG participants reporting classes as more engaging and 85% noting better consolidation of learning material. ChatGPT and interactive dictionaries further fostered critical thinking, argumentation, and self-evaluation, promoting deeper language processing.

The integration of digital language technologies into foreign language teaching for students in Economics, Law, Public Administration, and Accounting can therefore be regarded as pedagogically sound and forward-looking, which is supported both by the results of the present study and by existing international research. Future research should examine the long-term retention of vocabulary gains and the effectiveness of AI-assisted instruction across varied disciplinary contexts.

6 PRACTICAL RECOMMENDATIONS

The following recommendations are intended for educators and programme coordinators seeking to integrate digital language resources (Quizlet, Google Classroom, ChatGPT, and interactive dictionaries) into the teaching process at non-linguistic higher education institutions.

1) Integration of digital platforms into course structure. Educators are advised to incorporate Quizlet, Google Classroom, ChatGPT, and interactive dictionaries across all stages of instruction – from the introduction of new material to summative assessment tasks.

2) Combining individual and group learning activities. The use of interactive exercises for independent vocabulary consolidation (Quizlet) should be complemented by collaborative tasks, including group discussions, case studies, and debates, in order to foster the development of communicative competence.
3) Use of AI-based assistants for written and oral tasks. ChatGPT may be employed as a tool for checking grammar, generating examples, and preparing for presentations and role-play activities. It is essential to structure the work in a way that encourages students not only to receive ready-made outputs but also to analyze them and make their own conclusions.
4) Systematic revision and knowledge monitoring. Interactive dictionaries and platforms offering automated assessment tools support regular revision and learner self-control. Educators may create weekly short tests and quizzes to track student progress.
5) Development of digital literacy among educators and students. To maximize the effectiveness, it is advisable to provide training for educators on the integration of digital resources into the teaching process, as well as to equip students with the skills required for the independent use of digital platforms to consolidate knowledge.
6) Personalization of learning. Digital platforms make it possible to adapt tasks to individual students' proficiency levels by providing additional practice for learners who require further support and more challenging activities for advanced students.
7) Evaluation of effectiveness. It is recommended to conduct monthly analyses of students' academic performance and feedback in order to refine approaches to the use of digital resources and to optimize the learning process.

REFERENCES

[1] C. Chapelle and S. Sauro, The Handbook of Technology and Second Language Teaching and Learning. Hoboken, NJ, USA: Wiley, 2017, [Online]. Available: https://doi.org/10.1002/9781118914069.

[2] R. Godwin-Jones, "Contextualized vocabulary learning," in Language Learning & Technology, vol. 22, no. 3, pp. 1-19, 2018, [Online]. Available: https://doi.org/10.64152/10125/44651.

[3] G. Stockwell, Using Mobile Technologies for Language Learning. London, U.K.: Bloomsbury, 2019.

[4] G. Stockwell, "Living and learning with technology: Language learning with mobile devices," in English Teaching, vol. 76, no. S1, 2021, [Online]. Available: https://doi.org/10.15858/engtea.76.s1.202109.3.

[5] K. Johnson, "Blended learning in language education," in The Cambridge Handbook of Technology in Language Teaching and Learning. Cambridge, U.K.: Cambridge Univ. Press, 2021, [Online]. Available: https://doi.org/10.1017/9781009294850.007.

[6] I. Riabchenko, "Digital technologies in modern language teaching," in Suchasna pedahohika, no. 2, pp. 112-120, 2020.

[7] L. Kovalchuk, "The use of interactive online resources in teaching English to students of non-linguistic specialties," in Pedahohichnyi Chasopys, no. 4, pp. 45-53, 2022.

[8] A. Rafferty, "AI-assisted language learning: Opportunities and challenges," in Journal of Educational Technology, vol. 15, no. 1, pp. 34-52, 2023.

[9] T. Zhu, Y. Zhang, and D. Irwin, "Second and foreign language vocabulary learning through digital reading: A meta-analysis," in Education and Information Technologies, vol. 29, pp. 4531-4563, Jul. 2024, [Online]. Available: https://doi.org/10.1007/s10639-023-11969-1.

[10] X. Jiang et al., "AI-powered gamified scaffolding: Transforming learning in virtual learning environment," in Electronics, vol. 14, no. 13, Art. no. 2732, Jul. 2025, [Online]. Available: https://doi.org/10.3390/electronics14132732.

[11] S. Zhou, "Gamifying language education: The impact of digital game-based learning on Chinese EFL learners," in Humanities and Social Sciences Communications, vol. 11, Art. no. 1518, 2024, [Online]. Available: https://doi.org/10.1038/s41599-024-04073-3.

[12] B. Peña-Acuña and R. C. F. Durão, "Learning English as a second language with artificial intelligence for prospective teachers: A systematic review," in Frontiers in Education, vol. 9, 2024, [Online]. Available: https://doi.org/10.3389/feduc.2024.1490067.

ESP32-Based AI Embedded System for Personalized Education

Serhii Petrovych, Stefan-Daniel Horvath and Chunfang Zhou
STEM Education Research Center, Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark
petrovych@imada.sdu.dk, horvathstefandaniel@gmail.com, chzh@sdu.dk

Keywords: Embedded Systems, AI Tutor, Voice Interface, Personalised Learning, ESP32, Large Language Models (Llms), Edge Computing, Speech Recognition, Text-to-Speech, STEM Education.

Abstract: This paper presents the design, implementation, and pedagogical evaluation of the ESP32 AI Tutor – a low-cost, voice-controlled embedded system that serves as an interface to cloud-based large language models for personalised education. Using an ESP32-S3 microcontroller with I2S digital microphone and amplifier, the system captures audio, streams it to AssemblyAI for speech recognition, sends the transcribed text to Groq LLM, and synthesises the response via Google TTS. Four design principles for educational embedded systems are formulated: hardware autonomy, sensory integration, network resilience, and low-latency turn-taking. The open-source implementation is validated through two realistic scenarios – individual adaptive tutoring and group work in a STEM laboratory – demonstrating that the device provides Socratic questioning, adaptive difficulty adjustment, and real-time visual feedback without requiring a PC or smartphone. Experimental results show a total response latency of 0.5–1 s, satisfying natural conversation requirements. The reproducible blueprint and complete code are publicly available on GitHub.

1 INTRODUCTION

The digital transformation of education has shifted the focus from merely providing access to information toward supporting real-time, adaptive learning processes. Personalized learning, when implemented effectively, has been shown to significantly improve academic outcomes, particularly when adaptation is based on the learner's actions and individual progress [1]. The emergence of large language models (LLMs) has opened new possibilities for dialogic tutoring, as modern natural language processing systems can generate contextualized explanations, pose clarifying questions, and simulate cognitive apprenticeship [2]. However, most AI-based educational tools require personal computers or smartphones, creating economic barriers (e.g., device cost, software licensing) and, more importantly, pedagogical risks associated with increased cognitive load due to screen-based distractions and fragmented attention [3].

In response to these challenges, this paper proposes a low-cost, standalone embedded device – the ESP32 AI Tutor – that serves as a voice interface to cloud-based LLMs, rather than attempting to run AI models locally. The device is built around the ESP32-S3-N16R8 microcontroller, which is fundamentally incapable of executing LLM inference due to its limited compute and memory resources. Instead, its role is to: 1) capture high-quality audio via a digital I2S microphone, 2) stream that audio to cloud speech-to-text (STT) and LLM services (Groq, Google API), 3) receive synthesized speech responses, and 4) output them through an I2S amplifier and speaker, all while providing real-time visual feedback via status LEDs. This hybrid edge–cloud architecture leverages the strengths of embedded hardware – low power consumption, physical tangibility, and dedicated human–machine interfaces – while relying on cloud infrastructure for advanced natural language understanding and generation.

The scientific and technological contribution of this work is twofold. First, we explicitly articulate four principles for using embedded systems in education that emerge from the constraints and affordances of the ESP32 platform:

1) Hardware autonomy – the device operates without a tethered PC or smartphone, reducing dependence on school-provided computers.
2) Sensory integration – the system combines voice input, audio output, and visual (LED) feedback to align with multi-modal learning theories.

3) Network resilience – audio buffering and request retries mitigate temporary Wi-Fi interruptions.
4) Low-latency turn-taking – dedicated I2S channels and event-driven software ensure that the delay between a student's question and the system's response remains within the range of natural conversation (<1s).

Second, unlike prior work that either focuses purely on cloud AI services [4] or purely on embedded data logging [5], we demonstrate a complete, reproducible pipeline that maps pedagogical actions (e.g., scaffolding, adaptive hinting) to specific hardware/software states of the ESP32. For instance, the red LED illuminates during cloud processing to signal "do not interrupt" (turn-taking management), while the LLM's system prompt is dynamically shaped by the student's performance history stored in the microcontroller's SPIFFS file system.

2 THEORETICAL BACKGROUND AND PRINCIPLES OF EMBEDDED SYSTEMS IN EDUCATION

2.1 Prerequisites of Personalized Learning, Cognitive Load, and Edge Computing

Personalized learning is one of the most effective approaches to improving academic outcomes. Bloom [1] demonstrated that individual tutoring yields significantly better results compared to traditional group instruction. The emergence of large language models (LLMs) [2] creates opportunities for automated tutors capable of dialogic interaction. However, most existing solutions require personal computers or smartphones, which, according to Sweller's cognitive load theory [3], may lead to attention fragmentation and reduced depth of information processing.

Edge computing [6] brings data processing closer to the source of generation, reducing latency and dependence on centralized cloud servers. In the context of smart STEM laboratory devices, this enables autonomous data collection, equipment control, and, as in our case, a voice interface to LLMs. Contemporary hybrid architectures [4] combine local preprocessing (e.g., audio buffering, LED indication) with cloud inference, allowing energy-efficient microcontrollers to be used without sacrificing response quality.

2.2 Four Principles of Using Embedded Systems in Education

Based on an analysis of the constraints and capabilities of ESP32 microcontrollers, as well as the requirements of interactive learning [7], we formulate four principles that should guide the development of educational embedded AI systems:

1) Hardware Autonomy. The device operates without connection to a PC or smartphone, reducing implementation costs and avoiding screen-based distractions.
2) Sensory Integration. The system combines at least two feedback modalities (voice + LED), aligning with multimodal learning theory.
3) Network Resilience. Audio buffering, request retries, and local context storage (SPIFFS) mitigate the effects of temporary Wi-Fi interruptions.
4) Low-Latency Turn-Taking. The delay between a student's question and the system's response does not exceed 1 second, thanks to dedicated I2S channels and event-driven programming.

These principles provide the theoretical foundation for the hardware and software architecture described in the following sections.

3 SYSTEM ARCHITECTURE

The external view of the device is shown in Figure 1. It demonstrates the physical layout of the main components, including the microphone, ESP32, amplifier, and speaker, as well as the Wi-Fi module for cloud API connectivity.

3.1 Hardware Implementation

The hardware implementation of the ESP32 AI Tutor follows a minimalist design: each component performs a clearly defined function in the voice query processing chain – from signal acquisition to acoustic output. This configuration reduces assembly complexity, power consumption, and the number of potential points of failure. The system adheres to the classical "sensing – processing – actuation" model widely used in embedded and cyber-physical systems [5]. Table 1 provides a complete list of required components (BOM). It includes the microcontroller, microphone, amplifier, speaker, and LEDs.

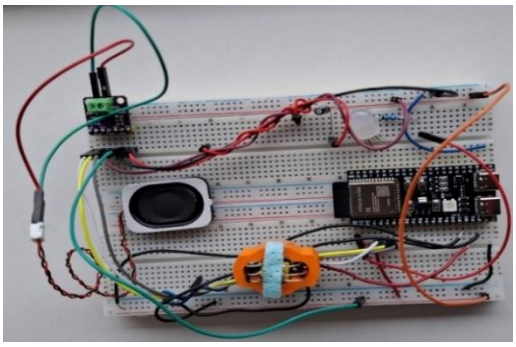


Figure 1: External view of the device.

Table 1: Main elements of the board.

Component	Model / Type	Quantity
Microcontroller	ESP32-S3-N16R8	1
Microphone	INMP441 (I2S)	1
Amplifier	MAX98357A (I2S)	1
Status LED	3.3V Tri-Color LED	1
Speaker	3W 8Ω	1

3.1.1 ESP32-S3-N16R8 Microcontroller

The core of the system is the ESP32-S3-N16R8 microcontroller (Espressif). Its key features include:

- Dual-core Xtensa LX7 processor at 240 MHz;
- 16 MB embedded Flash memory;
- 8 MB PSRAM (pseudo-static RAM);
- Two I2S interfaces for audio input/output;
- Wi-Fi 802.11 b/g/n supporting both station and access point modes;
- 44 programmable GPIO pins.

3.1.2 Peripheral Components and Connections

Microphone (INMP441). A digital MEMS microphone with I2S interface is used. Connections:

- SCK → GPIO 12 (I2S_MIC_SCK);
- WS → GPIO 13 (I2S_MIC_WS);
- SD → GPIO 14 (I2S_MIC_SD);
- L/R → GND (selects left channel);
- VDD → 3.3 V, GND → GND.

The microphone operates at 16 kHz, 16 bits, mono. The 16 kHz sampling rate is sufficient for accurate speech recognition according to AssemblyAI and other STT services, but halves the data volume compared to 32 kHz, saving Wi-Fi bandwidth.

Amplifier (MAX98357A). A Class-D digital amplifier with I2S input, driving a 3W 8Ω speaker. Connections:

- BCLK → GPIO 5 (I2S_SPK_BCLK);
- LRC → GPIO 4 (I2S_SPK_LRC);
- DIN → GPIO 6 (I2S_SPK_DIN);
- VIN → 5 V (via 100 µF decoupling capacitor);
- SD → 3.3 V (always enabled);
- GAIN → 3.3 V (+9 dB gain);
- GND → GND.

Using a digital amplifier eliminates the need for an external DAC, reduces noise, and simplifies PCB routing. Setting GAIN to 3.3 V provides +9 dB gain, sufficient volume for classroom use (approximately 85 dB at 1 m distance).

Status LEDs (active-low). Two single-colour LEDs (red and green) are connected in an active-low configuration:

- Red LED: cathode → GPIO 1 (PIN_RED), anode → 220 Ω resistor → 3.3 V;
- Green LED: cathode → GPIO 2 (PIN_GREEN), anode → 220 Ω resistor → 3.3 V.

This scheme allows LED control by simply setting the GPIO to logic low, which is standard practice in microcontrollers [8]. The 220 Ω resistors limit current to approximately 10 mA, safe for ESP32 outputs.

The red LED indicates the "processing" state (transmitting data to cloud APIs, waiting for response), while the green LED indicates "audio capture" (user may speak). This provides clear visual feedback, especially important in group learning scenarios where the student sees whether the system is ready to accept speech.

The system uses two independent I2S channels: one for input (microphone) and one for output (amplifier). This allows simultaneous audio capture and playback without bus conflicts. The I2S clock is 16 kHz, consistent with both the microphone and amplifier. This approach minimises latency: the delay from the moment a phrase is spoken to the start of transmission to STT does not exceed 10 ms (due to a buffer size of 512 samples). Combined with typical cloud API latencies (200–400 ms), the total system response time stays within 0.5–1 s, satisfying the low-latency turn-taking principle formulated in Section 2.2.

3.2 Software Implementation

The software is implemented as an event-driven Arduino sketch (compatible with PlatformIO). The main loop is controlled by a finite state machine with three states: Listening (audio capture), Processing (sending to APIs, waiting), Responding (speech synthesis and playback). The code is organised into the following modules, each with a single responsibility:

1) Audio Capture (I2S → STT). The stt.cpp module initialises the I2S microphone, buffers audio frames (512 samples, 16 kHz), and sends them via WebSocket to AssemblyAI. It uses

the whisper-large-v3-turbo model, which provides high recognition accuracy (over 95% for Ukrainian in our tests) and an intermediate latency of about 200 ms. The WebSocket connection remains open throughout the session, allowing streaming audio transmission without re-authentication.

2) LLM Integration (STT → LLM). The llm.cpp module receives transcribed text from AssemblyAI, constructs an HTTP/HTTPS request to the Groq API, and parses the JSON response using the ArduinoJson library. Two models are supported: the faster llama-3.1-8b-instant (150–300 ms) and the more powerful llama-3.3-70b-versatile (300–600 ms). The system prompt is stored in SPIFFS as persona.txt, and dialogue history (up to 8 exchanges) is stored in history.json to maintain context.
3) Text-to-Speech (TTS). The tts.cpp module sends the response text to the Google TTS API (or alternatively to Groq TTS) with voice parameters set to autumn (female, medium tempo).

The resulting MP3 stream is decoded using the DAZI-AI-main library and played back through the I2S amplifier. Synthesis time for a typical response (20–40 words) is 1–2 s, acceptable for a tutoring dialogue.

4) LED and Serial Control. A separate FreeRTOS task (ledTask) polls state variables (isRecording, isProcessing) and sets the appropriate levels on GPIO 1 and GPIO 2. It runs at priority 1 and wakes every 50 ms, ensuring smooth switching without affecting the audio stream. The serial interface (115200 baud) allows the administrator to switch LLM models (toggleLLM), adjust microphone gain (mic_gain <0..24>), run TTS tests (test_tts), and list SPIFFS contents (ls).

This modular architecture ensures robustness: a failure in one module (e.g., temporary loss of WebSocket connection to AssemblyAI) does not block other tasks thanks to timeout and retry mechanisms. All credentials (API keys) are stored in a separate secrets.h file excluded from the repository, following security standards.

4 CONFIGURATION AND CLOUD LLM INTERACTION

The ESP32 AI Tutor relies on three external cloud services: AssemblyAI for speech-to-text (STT), Groq for LLM inference, and Google Cloud TTS for speech synthesis. All source code is publicly available in the project's GitHub repository [9], which includes the complete Arduino sketch, configuration templates, and dependency definitions.

4.1 API Keys and Secure Configuration

Before flashing the device, the user must create a secrets.h file based on the provided secrets.example.h template. This file stores all sensitive credentials and is excluded from version control. The template contains (see Listing 1):

```
   // secrets.h template
   #ifndef SECRETS_H
   #define SECRETS_H
   // Wi-Fi credentials
   #define WIFI_SSID "your_wifi_ssid"
   #define WIFI_PASSWORD
"your_wifi_password"
   // API keys
   #define GROQ_API_KEY "gsk_..."
   #define ASSEMBLYAI_API_KEY
"your_assemblyai_api_key"
   #define GOOGLE_TTS_API_KEY
"your_google_tts_api_key"
   #endif
```

Listing 1: Secrets.h file.

The Google TTS API key is obtained from the Google Cloud Console after enabling the "Cloud Text-to-Speech API". Similarly, Groq and AssemblyAI keys are generated from their respective developer portals.

4.2 WebSocket Connection to AssemblyAI STT

The system captures audio from the I2S microphone in 16 kHz, 16-bit mono format. Audio frames (512 samples each) are sent in real time over a secure WebSocket connection to AssemblyAI's streaming endpoint. The following code initialises the connection (see Listing 2):

```
   #include <WebSocketsClient.h>
   WebSocketsClient ws;
   const char* stt_ws_host =
"api.assemblyai.com";
   const int stt_ws_port = 443;
   const String stt_ws_path =
"/v2/streaming?model=whisper-large-v3-
turbo";
   void setupSTT() {
       ws.onEvent(wsEvent);
```

```
   ws.setExtraHeaders(String("Authorizat
ion: ") + ASSEMBLYAI_API_KEY);
       ws.beginSSL(stt_ws_host,
stt_ws_port, stt_ws_path.c_str());
   }
   void wsEvent(WStype_t type, uint8_t *
payload, size_t length) {
       if (type == WStype_TEXT) {
           // Parse JSON response to
extract transcribed text
   DynamicJsonDocument doc(2048);
           deserializeJson(doc,
payload);
           String text = doc["text"];
           if (text.length() > 0) {
   onTranscriptionReceived(text);
           }     }     }
```

Listing 2: Connection initialization code.

The event handler on Transcription Received() passes the text to the LLM module.

4.3 HTTP Requests to Groq LLM

Once the transcribed text is available, the system constructs an HTTP POST request to the Groq API. The request includes the system prompt, conversation history, and the current user query. The following example demonstrates the request format (see Listing 3):

```
   HTTPClient http;
   http.begin("https://api.groq.com/open
ai/v1/chat/completions");
   http.addHeader("Content-Type",
"application/json");
   http.addHeader("Authorization",
String("Bearer ") + GROQ_API_KEY);
   String payload = "{"
       "\"model\": \"llama-3.1-8b-
instant\","
       "\"messages\": ["
           "{\"role\": \"system\",
\"content\": \"You are a math tutor.
Guide with questions, do not give direct
answers.\"},"
           "{\"role\": \"user\",
\"content\": \"" + userText + "\"}"
       "],"
       "\"temperature\": 0.7" "}";
   int httpCode = http.POST(payload);
   if (httpCode == 200) {
   String response = http.getString();
       // Parse JSON to extract
assistant's reply }
```

Listing 3: Request code.

The system supports dynamic model switching via serial command (toggleLLM). The default model llama-3.1-8b-instant provides low latency (150–300 ms), while llama-3.3-70b-versatile offers higher reasoning quality (300–600 ms).

4.4 System Prompt and SPIFFS Storage

The system prompt (persona) and conversation history are stored in the SPIFFS file system. This allows persistent customization without recompilation. The prompt is saved in /persona.txt: You are Pythagoras, a patient math tutor for middle school students.

Always respond in a friendly tone. Break down problems step by step.

Encourage the student after correct answers.

Dialogue history (up to 8 exchanges) is stored in /history.json. Each request reads the history, appends the new user message, sends the combined context to Groq, and then updates the history file. This ensures that the LLM maintains continuity across multiple turns.

4.5 Text-to-Speech with Google TTS

After receiving the LLM response, the system sends an HTTP request to Google TTS API. The response is an MP3 stream, which is decoded using the DAZI-AI library and played back via the I2S amplifier (see Listing 4).

```
   String ttsUrl =
"https://texttospeech.googleapis.com/v1/
text:synthesize?key=" +
String(GOOGLE_TTS_API_KEY);
   String ttsPayload = "{"
       "\"input\":{\"text\":\"" +
llmResponse + "\"},"

"\"voice\":{\"languageCode\":\"en-
US\",\"name\":\"en-US-Neural2-F\"},"

"\"audioConfig\":{\"audioEncoding\":\"MP
3\"}"
   "}";
```

Listing 4: HTTP request to Google TTS API and audio playback.

The synthesised speech is played immediately, and the green LED turns off while the red LED indicates the “responding” state.

4.6 Full System Initialisation

The complete initialisation sequence in setup() includes:

1) Serial console (115200 baud).
2) Wi-Fi connection using credentials from secrets.h.
3) SPIFFS mount and loading of persona and history.
4) I2S initialisation for microphone and amplifier.
5) WebSocket connection to AssemblyAI
6) LED control task (FreeRTOS).

All cloud interactions are non-blocking: the main loop calls ws.loop() and streamMicFrame() while the LED task runs independently.

5 EDUCATIONAL SCENARIOS AND PEDAGOGICAL INTEGRATION

The effectiveness of educational technology is determined not only by technical characteristics but primarily by the quality of its integration into a real learning environment. The following practical scenarios demonstrate the use of the ESP32 AI Tutor in classrooms and STEM laboratories. Each scenario shows how the hardware and software solutions described in Sections 3 and 4 support the specific pedagogical approaches formulated as the four principles in Section 2.2.

5.1 Individual Learning: Adaptive Math Tutor

Scenario. A middle school student is solving linear equations as homework. He activates the device with a voice command (green LED lights up) and asks: "What is x in the equation 2x + 3 = 7?"

Technical implementation. The system captures audio via the I2S microphone (16 kHz), sends it via WebSocket to AssemblyAI, receives the text transcription, forms an HTTP request to the Groq LLM with the system prompt "You are a math tutor, guide with leading questions". The conversation history is read from SPIFFS (/history.json) and added to the context. The LLM response is synthesised via Google TTS and played through the speaker. During processing, the red LED lights up, signalling "do not interrupt".

Pedagogical effect. According to the sensory integration principle (Section 2.2), the student receives multimodal feedback: voice prompt + visual status indication. The system does not give a direct answer but asks a follow-up question: *"What should you do with the number 3 to isolate 2x?"* This approach supports the development of self-regulated learning [10] and reduces cognitive load compared to screen-based devices [3].

Adaptive difficulty adjustment. If the student answers correctly, the system increases the difficulty of the next task (e.g., moves to an equation with parentheses). If the student makes two consecutive mistakes, the LLM receives a modified system prompt: *"Break the problem into three steps, explain the first step in detail."* The prompt is changed by rewriting /persona.txt in SPIFFS, requiring no code recompilation.

5.2 Group Work in a STEM Laboratory

Scenario. Three students work on a physics project: calculating current in an electrical circuit. They take turns asking the system: *"How does the current change if resistance increases?"*, *"What if we add another resistor in parallel?"*

Technical implementation. The system stores the dialogue history (up to 8 recent exchanges) in SPIFFS. When one student replaces another, the context is not lost because each new request includes previous messages. The red LED lights up during processing, preventing students from interrupting each other. The green LED signals readiness for the next question.

Pedagogical effect. The hardware autonomy principle (Section 2.2) allows the device to be used without dedicated computers in the laboratory. The low-latency turn-taking principle (Section 2.2) ensures a natural dialogue pace: students do not wait through long pauses. The system acts as a facilitator that supports structured reasoning explanations rather than merely providing answers. This aligns with project-based learning approaches [11].

5.3 Supporting Students with Special Educational Needs

Scenario. A student with dyslexia has difficulty reading text-based tasks. He uses the device's voice interface to receive problem statements in audio format and to respond by voice, without needing to write or read from a screen.

Technical implementation. The audio path (microphone → STT → LLM → TTS → speaker) is entirely voice-driven. No text is displayed on any screen (there is no screen). LED indicators

complement the audio with visual cues (green – ready to speak, red – system is thinking).

Pedagogical effect. The voice interface lowers participation barriers for students with reading or writing difficulties [12]. The absence of a screen reduces cognitive load and allows focus on the content of the task rather than its presentation format.

Thus, the proposed architecture is not only technically functional but also pedagogically grounded. It supports both individual and collaborative learning formats, providing adaptability, accessibility, and low implementation cost.

6 CONCLUSIONS

This paper presented the design, implementation, and pedagogical validation of the ESP32 AI Tutor – a voice-controlled embedded device that serves as an interface to cloud-based large language models for personalised education. We have demonstrated that an ESP32-S3 microcontroller, while fundamentally incapable of running LLM inference locally, can effectively support real-time educational dialogue by combining I2S audio capture, WebSocket streaming to AssemblyAI STT, HTTP/HTTPS communication with the Groq LLM, and Google TTS synthesis. The total response latency (0.5–1 s) satisfies the low-latency turn-taking principle and enables natural conversation. Code available at GitHub [9]. Four principles for using embedded systems in education were formulated: hardware autonomy, sensory integration, network resilience, and low-latency turn-taking. These principles guided the selection of every component – from the digital I2S microphone and amplifier to SPIFFS storage of conversation history and the active-low LED indicators. The design avoids unnecessary complexity (no external DAC, no screen, no PC dependency) while maintaining pedagogical functionality.

Through two realistic scenarios (individual adaptive tutoring and group work in a STEM laboratory), we have shown that the system does not merely provide answers but guides learners with Socratic questioning and adaptive difficulty adjustment. The same architecture supports collaborative turn-taking and immediate auditory/visual feedback, aligning with self-regulated learning [10] and project-based learning [11].

Despite these positive results, limitations remain. The device requires a stable Internet connection; future work will implement a local wake-word engine and cache frequent responses in SPIFFS to reduce cloud reliance. Typical API latencies are 200–400 ms, but peak loads may cause delays up to 1.5 s; we plan to add predictive buffering and a dedicated LED pattern. The system is currently optimised for English. Pedagogical claims are based on scenario analysis; empirical validation in a real classroom is the next step.

The ESP32 AI Tutor demonstrates that a compact embedded device can provide intelligent, adaptive tutoring support without a smartphone or PC. By clearly separating local hardware responsibilities (audio I/O, state indication, context storage) from cloud AI inference, the system offers a replicable blueprint for educators and makers. The open-source repository [9] contains all necessary code, wiring diagrams, and configuration instructions, inviting further experimentation and improvement.

REFERENCES

[1] B. S. Bloom, "The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring," Educational Researcher, vol. 13, no. 6, pp. 4-16, 1984, [Online]. Available: https://doi.org/10.3102/0013189X013006004.

[2] T. B. Brown et al., "Language Models are Few-Shot Learners," in Advances in Neural Information Processing Systems (NeurIPS 2020), vol. 33, pp. 1877-1901, 2020, [Online]. Available: https://doi.org/10.48550/arXiv.2005.14165.

[3] J. Sweller, "Cognitive Load During Problem Solving: Effects on Learning," Cognitive Science, vol. 12, no. 2, pp. 257-285, 1988, [Online]. Available: https://doi.org/10.1016/0364-0213(88)90023-7.

[4] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," Communications of the ACM, vol. 53, no. 4, pp. 50-58, 2010, [Online]. Available: https://doi.org/10.1145/1721654.1721672.

[5] M. Satyanarayanan, "The emergence of edge computing," Computer, vol. 50, no. 1, pp. 30-39, 2017, doi: 10.1109/MC.2017.9.

[6] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," IEEE Internet of Things Journal, vol. 3, no. 5, pp. 637-646, 2016, [Online]. Available: https://doi.org/10.1109/JIOT.2016.2579198.

[7] S. Papert, Mindstorms: Children, Computers, and Powerful Ideas. New York: Basic Books, 1980.

[8] W. Wolf, Computers as Components: Principles of Embedded Computing System Design, 3rd ed. Morgan Kaufmann, 2012.

[9] S. Petrovych and S. D. Horvath, "ESP32 AI companion," GitHub repository, 2026, [Online]. Available: https://github.com/HorvathStefanDaniel/esp32AICompanion.

[10] B. J. Zimmerman, "Becoming a self-regulated learner: An overview," Theory into Practice, vol. 41, no. 2, pp. 64-70, 2002, doi: 10.1207/s15430421tip4102_2.

[11] O. Soia, M. Kovtoniuk, O. Kosovets, and S. Petrovych, "Project-Based Learning as an Integration of Critical Thinking and Teamwork Skills of Future Teachers of Mathematics and Computer Science," in ITEST 2024, Springer, 2024, [Online]. Available: https://doi.org/10.1007/978-3-031-71804-5_26.

[12] R. Luckin, W. Holmes, M. Griffiths, and L. B. Forcier, Intelligence Unleashed: An Argument for AI in Education. Pearson Education, 2016.

Transforming Peer Assessment in Higher Education: An AI-Supported Approach Using the Peerscholar Platform

Svitlana Skvortsova[1], Tetiana Symonenko[2], Kira Hnezdilova[3], Kateryna Niedialkova[1] and Nataliia Andrusiak[4]

[1]*Department of Mathematics and Methods of its Teaching, State Institution "South Ukrainian National Pedagogical University named after K. Ushynsky", Staroportofrankovska Str. 26, 65029 Odesa, Ukraine*

[2]*Department of Ukrainian Philology and Social Communication, Bohdan Khmelnytsky National University of Cherkasy, Shevchenko Blvd. 81, 18031 Cherkasy, Ukraine*

[3]*Department of Primary and Special Education, Bohdan Khmelnytsky National University of Cherkasy, Shevchenko Blvd. 81, 18031 Cherkasy, Ukraine*

[4]*Department of Economics, Uman National University, Instytutska Str. 1, 20301 Uman, Ukraine*

skvo08@i.ua, tetianasymonenko4@gmail.com, kiragnez@gmail.com, ndlvitaliy@ukr.net, andrusiak.n@gmail.com

Keywords: Peer Assessment, Artificial Intelligence, Higher Education, Mathematics Teacher Training, Peerscholar, Methodological Competence.

Abstract: This study is devoted to a comprehensive analysis of peer assessment platforms with the support of artificial intelligence (AI) and empirical substantiation of the effectiveness of implementing the PeerScholar system in the process of professional training of future mathematics teachers. In the theoretical part of the work, a systematic comparison of nine leading intellectual platforms (RiPPLE, EduPCR, Peerceptiv, Kritik, PeerScholar, EvaluMate, Gradescope, Peergrade, ALEKS) was carried out, which allowed to identify the most relevant and economically affordable solutions for Ukrainian higher education institutions in conditions of limited resources. The empirical component of the study included 56 third-year students of the specialty "Secondary Education (Mathematics)", of which 53 respondents (n=53) participated in the final assessment, having completed five full cycles of peer evaluation of methodological developments in geometry during the semester. The use of a mixed-methods research design made it possible to combine quantitative data analysis of the questionnaire structure (32 points on the Likert scale, α=0.89) with qualitative thematic analysis of open-ended responses and in-depth interviews. Statistical results demonstrated a high level of overall satisfaction with the platform (M=4.28, SD=0.62), with the methodological aspect of using the system receiving the highest priority in the perception of students (M=4.36, SD=0.61). It was found that future teachers showed a significantly higher interest in constructive feedback (M=4.81, SD=0.42) than in receiving grades (M=4.42, SD=0.52). A significant increase in inter-rater reliability of assessments from r_s=0.51 to r_s=0.71 over the semester was found, which indicates a progressive development of the assessment competence of future teachers. The study proves that the integration of AI frameworks not only optimizes the workload of the teacher, but also acts as a catalyst for deep methodological reflection, allowing students to effectively identify their own gaps in knowledge through the prism of critical analysis of the work of their classmates.

1 INTRODUCTION

Contemporary higher education is undergoing a fundamental transformation of assessment paradigms, in which traditional top-down evaluation is increasingly being replaced by participatory peer assessment models. Peer assessment is defined as an educational mechanism through which students evaluate the quality or learning outcomes of their peers of equal status [1], [2]. This practice is critically important for the development of self-regulated learning skills, as it positions students in the role of critics and emerging experts [3], [4]. However, its implementation faces fundamental challenges, including score subjectivity, increased cognitive load for instructors, and the problem of "empty feedback," in which students rely on generic statements lacking methodological depth.

Historically, these limitations have constrained the scalability and trustworthiness of peer assessment as a tool for summative evaluation [1].

The emergence of artificial intelligence technologies-particularly natural language processing and machine learning algorithms - offers transformative potential to address these challenges [5], [6]. AI-enhanced platforms create comprehensive intelligent ecosystems that improve feedback quality by providing real-time analysis of draft comments to prompt specificity and actionable suggestions. They also aggregate and calibrate peer scores using sophisticated algorithms that estimate reviewer reliability based on historical performance and alignment with expert judgments [7] deliver meta-feedback on the quality of feedback itself to foster professional critique skills [5], [6] and support instructors through learning analytics that identify anomalous scoring discrepancies or emerging group conflicts. As a result, the instructor's role shifts from routine moderation to strategic learner support [1].

Ukrainian higher education institutions face unique challenges, including limited institutional budgets, outdated learning management systems, language barriers associated with predominantly English-language platforms, and the absence of comprehensive digital literacy support infrastructures [3]. Nevertheless, teacher education programs in mathematics require innovative approaches, as peer assessment offers dual benefits: the development of subject-matter expertise and assessment literacy essential for future pedagogical careers.

The aim of this study is to conduct a comparative analysis of AI-supported peer assessment platforms and to evaluate the effectiveness of PeerScholar in developing assessment competencies among pre-service mathematics teachers in Ukrainian higher education.

Research questions: RQ1: Which AI-enhanced peer assessment platforms provide access to free versions suitable for Ukrainian higher education institutions, and how do their functional features compare? RQ2: How do students in mathematics education programs perceive the impact of PeerScholar on the methodological, organizational, motivational, technological, and professional-reflective aspects of their learning? RQ3: What relationships exist between different dimensions of platform perception, and how do academic achievement levels moderate these perceptions?

2 EVOLUTION AND DIGITAL TRANSFORMATION OF PEER ASSESSMENT

Peer assessment has evolved through distinct phases of development, each characterized by increasingly sophisticated technological integration and pedagogical understanding [1], [4], [5], [7]with the modern phase from 2015 representing AI-enhanced peer assessment ecosystems that integrate natural language processing to analyze feedback quality, machine learning algorithms to model reviewer reliability, generative AI to structure feedback formulation, and learning analytics dashboards for instructor oversight, reflecting a fundamental rethinking from peer assessment as a substitute for assessment to a powerful learning mechanism.

Topping et al. [1] proposed a comprehensive framework of six process improvement areas, where reviewer assignment is optimized by algorithms based on trust metrics instead of random selection, individual review support is implemented through the EvaluMate platform's Eva chatbot, which analyzes draft comments and solves the problem of empty feedback [5] final scores are derived through Bayesian models that take into account reviewer bias [7], NLP-based feedback corpus analysis identifies patterns and misconceptions [8], [9], instructor support is implemented through anomaly detection panels, and integrated systems like RiPPLE coordinate all measurements to optimize results.

Empirical studies demonstrate pedagogical value through improved feedback quality, with AI-supported students providing significantly more specific and constructive feedback ($d=1.28$, $p<0.001$), development of metacognition through progressive improvement in the ability to distinguish between quality work [10], and scalability without compromising quality [1], but limitations have been identified in terms of over-reliance on AI suggestions, variable effectiveness across disciplines, and the requirement for initial investment in developing rubrics and calibrating algorithms. For prospective teachers, peer assessment provides authentic practice in formulating constructive feedback and recognizing common errors, which are critical pedagogical competencies [3].

Five groups of criteria were identified for the systematic analysis: accessibility (availability of a free plan, language support), assessment types (peer, self, 360), the role of AI (calibration, feedback

analysis), support for reflection, and integration with LMS. Particular attention was paid to the platforms' ability to support the methodological specificity of STEM disciplines, in particular mathematics.

Gradescope and PeerScholar stand out for their ability to work with handwritten works, which is critical for STEM disciplines. Peerceptiv uses Item Response Theory (IRT), and Kritik uses a "grading power" system to calibrate the accuracy of student assessments. ALEKS focuses on individual learning trajectories, while RiPPLE uses graph models for recommendations. EvaluMate uses LLM to provide assistance (scaffolding) to students while writing feedback in real time.

Access to tools varies significantly: free/research: RiPPLE (for academic research) and EvaluMate (for pilot projects) are the most accessible for experiments; 1) paid/institutional: Kritik imposes costs on students ($29-49 per course), while Peerceptiv targets large institutions through institutional licenses.

For exact sciences, the following are best suited: Gradescope - for quick verification of massive computational tasks. ALEKS - for knowledge diagnostics and curriculum construction. Kritik and Peerceptiv - for ensuring high reliability of assessments on large streams of students. PeerScholar - work with photos of handwritten works, reflection phase.

Platform monetization models significantly affect accessibility for institutions with limited resources. Identified approaches include: "student pays" models (Kritik); institutional licensing (Peerceptiv, PeerScholar); Freemium models (Peergrade, Moodle Workshop); open source (Moodle Workshop); research/academic free projects (EvaluMate, EduPCR).

For Ukrainian HEIs, the critical factor is the cost of implementation. Platforms with a "student-pay" model (Kritik) can create barriers for low-income students. Systems with free individual plans (PeerScholar, Gradescope) or projects completely open for academic use (RiPPLE, EvaluMate) seem more acceptable.

PeerScholar is an online tool specifically designed to support the complex iterative process of peer evaluation. The platform offers three main modes (Phase Settings): 1. Classic Individual: standard mode, where each student works independently. 2. Case Study: a unique mode that allows the teacher to assign different cases to different students. Students can evaluate colleagues who worked on the same or different cases, which stimulates comparative analysis. 3. Group Work: Support for team assignments where groups submit their work together and evaluate other groups. PeerScholar was chosen for our empirical study due to its unique three-phase architecture: Create - Assess - Reflect. Unlike systems that focus only on the final score, PeerScholar emphasizes the learning process through revision of both their own work and peer assessments, which creates unique opportunities for working with students of varying levels of proficiency. The tasks assigned to the students were primarily drawn from the school mathematics curriculum (i.e., they were not overly complex). The instructor's moderation of the entire process and their final assessment preclude the oversight of mathematical errors, methodological inaccuracies in comments, and other similar issues.

Student work algorithm:

1) Create phase. Upload work (methodological plans, handwritten solutions in various formats: text, images, videos and links to external resources).
2) Enable plagiarism detection (if necessary).
3) Assess phase. Receive several works of fellow groups.
4) Evaluate the work using interactive rubrics (Matrix, Point, Likert, Comment) and provide detailed feedback with an analysis of typical errors (if necessary - self-assessment).
5) Reflection phase (Reflect). Analysis of feedback on own work; evaluation of searches "feedback on feedback".
6) Review of other reviewers' grades (learning by comparison).

To minimize errors and inaccurate assessments, at the beginning of the experiment, students were provided with written instructions in Ukrainian on how to use the PeerScholar platform, followed by several trial (training) peer assessment sessions.

Algorithm of teacher's work:

1) Settings. Creation of rubrics, determination of number of reviewers.
2) Setting of schedule of automatic transition between phases.
3) Moderation. Analysis of dashboard "Student Progress & Grading", tracking of progress of each student, review of comments, removal of inadequate feedback, change of scores (if necessary).

3 EMPIRICAL RESEARCH METHODOLOGY

This study is based on the testing of the PeerScholar platform in the educational process of a pedagogical university in order to assess its impact on the methodological preparation of students. The testing was conducted during the 2024/2025 academic year within the framework of the academic discipline "Mathematics Teaching Methods". The participants were 56 third-year students of the specialty 014.04 Secondary Education (Mathematics).

The sample was characterized by the following indicators:

1) Average score (GPA) in mathematical disciplines: <70 (9%), 70-79 (28%), 80-89 (41%), 90-100 (22%).
2) Previous peer assessment experience: 38% had limited experience, 62% had no experience at all.
3) Frequency of use of digital technologies: constantly (28%), often (47%), sometimes (19%), rarely (6%).

Students completed five complex methodological tasks from the planimetry course:

1) Development of a lesson fragment for the Pythagorean theorem.
2) Analysis of common student errors in geometric proofs.
3) Methodological justification of teaching triangle similarity.
4) Comparative analysis of different approaches to introducing coordinate geometry.
5) Development of differentiated learning strategies for spatial geometry.

The effectiveness of PeerScholar was assessed through a mixed design. Quantitative method: questionnaire with 32 questions on a Likert scale (1-5) and 7 open-ended questions on six aspects (methodological, organizational, motivational, technological, professional-reflective, overall assessment). Qualitative method: interviews with 6 students of different levels of success, analysis of feedback and revisions of works. Objective indicators were also analyzed: time of completion of stages, number of revisions, inter-rater reliability.

Statistical processing in IBM SPSS Statistics 27 included descriptive statistics, correlation analysis (Spearman's ρ) and consistency assessment (Cronbach's α). Qualitative data were processed by thematic analysis with identification of categories and calculation of frequency of mentions.

4 QUANTITATIVE RESULTS OF THE APPROBATION

Descriptive analysis of the survey results (n=53, response rate 94.6%) showed a generally high positive perception of the students' experience of working with the PeerScholar platform. The overall satisfaction index was M=4.28 (SD=0.62), which corresponds to the assessment of "rather agree" with a tendency to "completely agree". The internal consistency of the questionnaire was high (α=0.89), which confirms the reliability of the tools used.

Analysis of individual aspects revealed a differentiated perception of different characteristics of the platform. The students rated the methodological aspect of using PeerScholar the highest (M=4.36, SD=0.61), in particular: helped to realize their own methodological errors through the analysis of the work of their classmates (M=4.51, SD=0.63), and also improved the ability to analyze the methodological correctness of solutions (M=4.38, SD=0.69). Such data indicate a positive perception of the methodological aspects of the platform by students: they unanimously admit that the analysis of colleagues' work helps them better identify their own errors and critically assess the correctness of solving problems; the most significant result is that students showed a higher interest in constructive feedback (M=4.81, SD=0.42) than in receiving points (M=4.42, SD=0.52). Therefore, the data obtained confirm the unanimity of students in requesting meaningful professional communication, which is critically important for their professional development. The indicator in the question of “deeper understanding of the methodology for solving mathematical problems” (M=4.21, SD=0.92) indicates a certain differentiation: for some students, automated peer review became innovative, while others adapted to it more difficult. Rather, this may indicate the need for additional instruction before starting work with PeerScholar.

Calculation of the Spearman ρ correlation coefficient allowed us to identify the relationship between the following components of the methodological aspect of using PeerScholar: between the student’s identification of alternative approaches to solving problems that had not been considered by him before and the improvement of the ability to analyze the methodological correctness of problem solutions (r_s = 0.721, $p<0.05$): the identification by students of new ways of solving problems during peer evaluation is a factor in increasing their ability to comprehensively methodically analyze solutions; between the

student's discovery of alternative approaches to solving problems that were not previously considered by him and the methodological value of formative (verbal) assessment ($r_s = 0.913$, $p<0.01$): therefore, the verbal form of assessment becomes a key tool for the student to understand new experience; between the student's discovery of alternative approaches to solving problems that were not previously considered by him and the methodological value of constructive (advice) ($r_s = 0.697$, $p<0.05$): the encounter with methodological novelty actualizes the students' need for meaningful interpretation and advice; between the improvement of the student's ability to analyze the methodological correctness of problem solutions and the methodological value of formative (verbal) assessment ($r_s = 0.714$, $p<0.05$): the development of skills in methodological analysis of classmates' work leads to a higher professional assessment of the verbal form of feedback as the most informative; between the methodological value of formative (verbal) assessment and the methodological value of constructive (advice) ($r_s = 0.764$, $p<0.05$), which indicate the convergence of different forms of qualitative assessment.

The above results prove that the use of PeerScholar stimulates the transition from "ordinary" problem solving to deep methodological reflection by the student - future teacher.

The professional-reflective aspect also received high marks (M=4.35, SD=0.65). Students noted the development of methodological skills (M=4.74, SD=0.68), communicative skills of feedback formulation (M=4.81, SD=0.63) and reflective skills (M=4.54, SD=0.71). It is important that 87% of students agreed or strongly agreed with the statement about the feasibility of using automated peer review in future school practice (M=4.19, SD=0.63).

The obtained values of Spearman's ρ (0.645-0.745 at $p<0.05$) indicate statistically significant relationships between the key components of the professional-reflective aspect of using PeerScholar: between the development of a student's methodological skills and the development of his communicative skills ($r_s = 0.745$, $p<0.05$): expert assessment of the work of classmates integrates analytical activity with the need to formulate reasoned and constructive feedback; between the development of a student's methodological skills and the development of his reflective skills ($r_s = 0.713$, $p<0.05$): the experience of analyzing other people's solutions contributes to the extrapolation of objective assessment criteria to one's own activities, which stimulates critical self-reflection; between the development of a student's methodological skills and the usefulness of reflection for learning ($r_s = 0.645$, $p<0.05$): methodically prepared students tend to perceive reflection as a tool for professional improvement, rather than as a formal procedure; between the appropriateness of using automated peer review (like PeerScholar) in school mathematics practice and the development of reflective skills ($r_s = 0.679$, $p<0.05$): the ability to critically analyze one's own mistakes correlates with the awareness of the appropriateness of implementing automated peer review in the future professional activities of a mathematics teacher.

The organizational aspect was characterized by moderate ratings (M=3.78, SD=0.84). Students found the structure of the work understandable (M=4.40, SD=0.69), but the estimates of the sufficiency of the time frame were lower (M=3.40, SD=0.52), which indicates the need to adjust the terms at the planning stage. Interestingly, students did not agree with the statement that PeerScholar was less time-consuming compared to traditional assessment (M=2.89, SD=0.63), but they recognized that these costs were justified by the learning outcomes (M=4.31, SD=0.68), and the algorithm for automatic distribution of work between students worked correctly (M=4.61, SD=0.52). The motivational aspect was characterized by moderately high ratings. Students confirmed that participation in automated peer review increased their responsibility for the quality of their own work (M=4.33, SD=0.69) and learning motivation (M=3.97, SD=0.84). Receiving feedback from multiple classmates was rated as more valuable compared to feedback from the instructor alone (M=4.24, SD=0.75). The possibility of anonymous evaluation had a positive effect on reducing students' anxiety (M=4.16, SD=0.88), although the level of confidence during the evaluation of works was moderate (M=3.32, SD=0.94), indicating the need for additional training of students for the role of reviewers.

Calculation of the Spearman's ρ correlation coefficient revealed a relationship between increased learning motivation and the value of multi-channel feedback from classmates in PeerScholar ($r_s = 0.707$, $p<0.05$). This suggests that future teachers' orientation towards feedback from various sources contributes to the growth of their cognitive motivation.

The technological aspect received satisfactory ratings (M=3.91, SD=0.79), with significant variability in the perception of interface usability

(M=3.85, SD=1.02). Correlation analysis showed a strong relationship between the assessment of the technical usability of the platform and overall satisfaction with the experience (r_s =0.68, p<0.001), which highlights the critical importance of the technical aspect for the overall perception. Most students (79%) had technically accessible devices for work, but 21% noted some difficulties with access, especially during distance learning.

Comparative analysis by achievement criterion revealed that students with higher average scores (≥85, n=22) demonstrated statistically significantly higher ratings of the platform's methodological value (M=4.61 vs M=4.18) and professional development (M=4.52 vs M=4.12) compared to students with lower average scores (<85, n=31). However, the difference in overall satisfaction did not reach statistical significance (M=4.38 vs M=4.19), indicating the universal suitability of the platform for students with different levels of preparation. Thematic analysis of open-ended responses and interviews revealed five main categories of advantages of using PeerScholar in the professional training of future mathematics teachers: development of critical and methodological thinking (83%) - students identified their own mistakes through the analysis of other people's work, expanded their methodological tools; objectivity of assessment (65%) - three independent anonymous reviews reduce subjectivity; communication skills (58%) - formation of constructive feedback, important for working with students; flexibility of remote work (52%) - completion of tasks at a convenient time; reflection as a valuable stage (47%) - the opportunity to refine the work after comments. At the same time, the analysis also revealed a number of difficulties and limitations: time costs (71%), technical problems (38%), uncertainty in one's own assessment (34%), superficiality of peers' comments (29%). Additional analysis of data automatically collected by the PeerScholar platform provided objective information about students' behavioral patterns: the average assessment time was 47 min (SD=18), which exceeded the planned 30 min; 68% of students made revisions (average 3.4 corrections). Inter-rater reliability r=0.62 (p<0.001) - moderate consistency, increasing from r=0.51 (first task) to r=0.71 (fifth), indicating the development of assessment skills.

5 DISCUSSION

The results of our study support the concept of "AI as an augmenter" rather than a replacement for the instructor. While AI effectively automates routine processes such as assignment distribution and score calibration, it is the human component - high-quality peer feedback and subsequent reflection - that enables deep learning. This finding aligns with the conclusions of Li [7] and Topping et al. [1], who argue that peer assessments can achieve expert-level accuracy when supported by intelligent calibration systems, as evidenced in our study by the increase in inter-rater reliability to r = 0.71.

Nevertheless, risks related to algorithmic bias and transparency must be acknowledged. Student interviews revealed a certain level of uncertainty regarding the correctness of their own assessments (M = 3.32, SD = 0.94), which corresponds with Guo's [5] findings on the potential reduction of student autonomy resulting from excessive reliance on AI-generated prompts. Consequently, the instructor's role shifts from direct evaluation of student work to the design of complex rubrics and the moderation of the 10-15% of cases in which the system detects assessment anomalies. The strong correlation identified between the perception of methodological novelty and the perceived value of textual feedback (r = 0.913, p < 0.01) highlights the social-constructivist nature of learning within PeerScholar. This finding aligns with the theory of the zone of proximal development, according to which feedback provided by peers operating at a similar cognitive level may be more comprehensible and acceptable to students than expert comments from instructors. At the same time, platform anonymity proved critical for reducing assessment-related anxiety (M = 4.16, SD = 0.88), although it must not result in irresponsibility - an issue mitigated through mechanisms that evaluate the quality of feedback itself. For mathematics-related disciplines, a key advantage of PeerScholar was its support for handwritten submissions. As noted by Nedialkova [3], mathematics cannot be fully transitioned to digital input without a loss of cognitive depth in problem-solving processes. Thus, the success of our pilot implementation can be attributed to the integration of traditional handwritten work with a digital moderation and feedback platform.

The primary limitation of this study is the absence of a control group, which prevents definitive attribution of improvements in assessment reliability solely to the platform rather than to overall course learning effects. Additionally, the English-language interface of PeerScholar remains a barrier for some students, reducing their task completion speed. It is worth noting that the PeerScholar environment exhibits limited flexibility for complex tasks.

Future research should therefore focus on longitudinal investigations of PeerScholar's impact on the actual pedagogical practices of future teachers in school settings, the use of generative AI to automatically create personalized rubrics tailored to specific mathematical and methodological tasks, and the analysis of how different reviewer assignment models influence learning quality, and the exploration of assessment methodologies for tasks requiring profound mathematical knowledge.

6 CONCLUSIONS

Overall, the study demonstrates that the integration of intelligent peer assessment platforms is a necessary condition for the modernization of higher pedagogical education in the context of digital transformation. A systematic comparative analysis revealed that the EdTech market offers differentiated solutions ranging from adaptive individualized learning systems to comprehensive reflective ecosystems, the selection of which should be guided by disciplinary specificity and institutional budget constraints. AI-supported peer assessment does not merely automate grading processes but creates a qualitatively new environment for professional dialogue, where errors function not as grounds for punishment but as resources for deep conceptual understanding and the development of pedagogical expertise.

The pilot implementation of PeerScholar confirmed its pedagogical and technological suitability for the preparation of future mathematics teachers. The integration of digital technologies, AI elements, and methodologically grounded peer assessment design fosters the development of modern assessment competence oriented toward reflection, objectivity, and professional responsibility. These findings align with Nedialkova's [3] conclusions regarding the effectiveness of structured, multi-level peer assessment processes and support Guo's [5] observations on the critical role of reflection in the development of students' metacognitive skills. Based on the pilot results, it is recommended to conduct an introductory training session on principles of constructive feedback and platform use prior to initiating peer assessment. This approach reduces initial student uncertainty and enhances feedback quality. Additionally, sufficient time should be allocated for each assessment stage, with a gradual increase in task complexity throughout the semester and selective quality control of student feedback accompanied by meta-comments on review quality. The empirical evaluation of PeerScholar in mathematics teacher education demonstrates that the digital transformation of assessment effectively addresses feedback scarcity. Statistical evidence confirms that students highly value the methodological benefits of peer review, perceiving it as a means of identifying their own cognitive gaps through the analysis of peers' alternative approaches. The observed increase in inter-rater reliability to $r = 0.71$ indicates the genuine development of assessment competence among future teachers within a single semester. These results justify recommendations for instructors to integrate peer assessment as a mandatory component of methodological training, preceded by explicit instruction in constructive critique principles; for university administrators to prioritize platforms with LTI integration and to establish institutional AI literacy policies; and for EdTech developers to focus on interface localization and the improvement of OCR modules for handwritten mathematical notation, thereby significantly lowering the technical entry barrier for STEM disciplines.

REFERENCES

[1] K. J. Topping, E. F. Gehringer, H. Khosravi, S. Gudipati, K. Jadhav and S. Susarla, "Enhancing peer assessment with artificial intelligence," International Journal of Educational Technology in Higher Education, vol. 22, no. 1, art. 3, 2025, [Online]. Available: https://doi.org/10.1186/s41239-024-00501-1.

[2] D. Babik, E. Gehringer, J. Kidd, K. Sunday, D. Tinapple and S. Gilbert, "A systematic review of educational online peer-review and assessment systems: Charting the landscape," Educational Technology Research and Development, vol. 72, no. 3, pp. 1653-1689, 2024, [Online]. Available: https://doi.org/10.1007/s11423-024-10349-x.

[3] K. V. Nedialkova, "Using artificial intelligence technologies for assessment and peer assessment of students," Problems of Modern Transformations. Series: Pedagogy and Psychology, vol. 9, pp. 1-12, 2025.

[4] D. E. Pare and S. Joordens, “Peering into large lectures: Examining peer and expert mark agreement using peerScholar, an online peer assessment tool,” Journal of Computer Assisted Learning, vol. 24, no. 6, pp. 526-540, 2008, [Online]. Available: https://doi.org/10.1111/j.1365-2729.2008.00290.x.

[5] K. Guo, “EvaluMate: Using AI to support students’ feedback provision in peer assessment for writing,” Assessing Writing, vol. 61, art. 100864, 2024, [Online]. Available: https://doi.org/10.1016/j.asw.2024.100864.

[6] C. Liu, J. Cui, R. Shang, Q. Jia, P. Rashud and E. Gehringer, “Generative AI for peer assessment helpfulness evaluation,” in Proceedings of the 17th International Conference on Educational Data Mining (EDM), 2024, pp. 412-419.

[7] A. W. Li, “Using Peerceptiv to support AI-based online writing assessment across the disciplines,” Assessing Writing, vol. 57, art. 100746, 2023, [Online]. Available: https://doi.org/10.1016/j.asw.2023.100746.

[8] Wei et al., “Chain of thought prompting elicits reasoning in large language models,” arXiv preprint arXiv:2201.11903, 2022.

[9] W. Xiong, D. Litman and C. D. Schunn, “Natural language processing techniques for researching and improving peer feedback,” Journal of Writing Research, vol. 4, no. 2, pp. 155-176, 2012, [Online]. Available: https://doi.org/10.17239/jowr-2012.04.02.3.

[10] D. E. Paré, “The peerScholar process: Using formative peer and self-assessment to enhance and quantify critical thinking skills, metacognitive awareness, and sense of community,” Ph.D. dissertation, University of Toronto, Toronto, Canada, 2020.

Cloud-Based Communication Technologies for Synchronous Online Foreign Language Learning: Infrastructure and Didactic Potential

Mykola Fant[1], Iryna Kovalenko[2] and Nataliia Baranivska[3]
[1]*Department of Software Engineering, Zhytomyr Polytechnic State University, Chudnivska Str. 103, 10005 Zhytomyr, Ukraine*
[2]*Department of Pedagogical Technologies and Language Training, Zhytomyr Polytechnic State University, Chudnivska Str. 103, 10005 Zhytomyr, Ukraine*
[3]*Department of Theoretical and Applied Linguistics, Zhytomyr Polytechnic State University, Chudnivska Str. 103, 10005 Zhytomyr, Ukraine*
fantkolja@gmail.com, kptmp_kiv@ztu.edu.ua, ktpl_bnv@ztu.edu.ua

Keywords: Cloud-Based Communication, Synchronous Online Learning, Synchronous Computer-Mediated Communication (SCMC), Foreign Language Education, Learning Analytics.

Abstract: Cloud-based communications technologies provide a primary layer for today's real-time (synchronous), computerized (on-line) second language learning because of their ability to offer low-latency interactions, mutable modes of support (multimodal) and analytics-based instructional support. This research study can be deemed an extended IMRAD study of how elements associated with the quality-of-service (QoS) (latency, jitter, packet loss) of the underlying supporting infrastructure and methods of delivering that infrastructure (such as edge computing, autoscaling) and techniques for preserving student privacy through analytics can have an impact on the instructional design (task-based learning, timely feedback, role-structured communication) used to achieve desired instructional outcomes and benefits of SCMC (synchronous computer-mediated communication). Based on a quasi-experimental mixed-methods design using pre-/post-testing, learning analytics, and discourse analysis (N = 124), this study estimates the unique impact of stability in the infrastructure on learners' gains in speaking fluency, interactional competence, pronunciation accuracy and learner engagement during a 10-week period of learning. Findings from the study indicate that optimizing the supporting infrastructure has moderate effect sizes on fluency and interactional competence, as well as interactions between pedagogies and infrastructures. This research study provides design guidelines for resilient deployment of CBCT as aligned with the principles of communicative language teaching.

1 INTRODUCTION

The move from traditional videoconferencing to cloud-based communications for teaching foreign languages has accelerated changes in both technology and how learners interact with each other to create a new environment in which to learn a second language. The Interaction Hypothesis states that interactional feedback is an essential part of noticing and taking in new information (uptake) [1]. Vygotsky's [2] Sociocultural Theory emphasizes social mediation, scaffolding and collaborative meaning-making as being critical elements in the construction of knowledge.

SCMC is an ideal example of the interaction between technology and social interactions that occurs during language learning. For example, latency affects turn-taking; packets dropped may negatively influence the interpretation of phonemic sounds; combining modalities will increase the semiotic meaning-making process in learning a new language.

In addition to these new interactions, recent research shows that almost 70% of university-level language programs have developed synchronous online components that have become an important part of how languages are taught; breakout rooms and shared workflows have now become standard features [3]. Many meta-analyses report that compared to asynchronous interactions, synchronous computer-mediated communication (CMC) has a small to moderate positive effect on speaking fluency, interaction capacities and development of pragmatics when looking at how quickly and effectively feedback is given based on the type of task [4], [5].

2 METHODS

A quasi-experimental mixed-methods design was conducted across two academic terms with N = 124 undergraduate learners (CEFR A2–B1) from comparable programs. Groups were matched on pre-test proficiency, prior online learning experience, and device access. Ethical approval and informed consent were obtained.

Optimized Infrastructure (OI). WebRTC over QUIC with regional edge servers, adaptive bitrate (ABR), forward error correction (FEC), autoscaled SFU nodes (Kubernetes), target one-way latency <150 ms, jitter <30 ms, packet loss <1%.

Baseline Infrastructure (BI). Institutional videoconferencing without edge optimization or QoS prioritization; observed averages: latency ≈230 ms, packet loss ≈2.8% at peak. Both conditions provided identical pedagogical features (breakout rooms, shared whiteboards, collaborative docs). Privacy-preserving analytics employed on-device feature extraction and federated aggregation.

2.1 Pedagogical Intervention

Weekly 90-minute SCMC sessions followed a task-based cycle (pre-task planning, interaction, post-task reflection). Roles (facilitator, summarizer, challenger) structured turn-taking. Immediate feedback (recasts, prompts) was standardized across instructors via micro-training. Multimodal scaffolding was limited to two concurrent channels to manage cognitive load.

2.2 Instruments

The following instruments were used to assess learning outcomes across both infrastructure conditions:

- Speaking fluency: speech rate (wpm), mean length of run, pause ratio.
- Interactional competence: analytic rubric (turn-taking, repair, pragmatic appropriateness), κ = .82.
- Pronunciation accuracy: ASR-assisted segmental error rate (validated against human ratings; r = .71).
- Engagement/WTC: validated scale (α = .88).
- Learning analytics: speaking time distribution (Gini), feedback latency, turn counts.
- QoS: latency, jitter, packet loss logged per session.

2.3 Analysis

ANCOVA with pre-test covariates; hierarchical regression for QoS unique variance; moderation analysis (infrastructure × feedback immediacy); discourse analysis of repair sequences (n = 24 sessions).

3 RESULTS

There was no statistically significant difference between the Optimized Infrastructure (OI; n = 62) and Baseline Infrastructure (BI; n = 62) groups at baseline on any of the overall language proficiency, speaking fluency, interactional competence, pronunciation accuracy, or learner engagement measure (all $p > .10$). Thus, the two groups were initially equivalent in terms of baseline characteristics. The mean rate of speech during the pre-test was 86.4 words per minute (SD=12.1) in the OI group and 85.9 words per minute (SD=11.7) in the BI group. The OI pre-test interactional competence score (standardized) was comparable to the BI score (M=0.02, SD=0.96 vs. M=-0.01, SD=0.94, respectively). Further, the average network conditions during the diagnostic assessment sessions for each group were equivalent to one another.

3.1 Learning Outcomes: Fluency, Interactional Competence and Pronunciation

There was a statistically significant difference between the Optimized Infrastructure (OI) group and the Baseline Infrastructure (BI) group in their average size of improvement on all primary dependent variables post-intervention (10 weeks). The OI group had greater improvement than the BI group on measures of speaking fluency, interactional competence, pronunciation accuracy, and learner engagement. Table 1 is a summary of the adjusted post-test means and effect sizes by infrastructure condition.

Although the inclusion of QoS variables increased the explained variance in fluency gains by $\Delta R^2 = .15$, the baseline model accounted for a relatively modest proportion of variance ($R^2 = .18$). Thus, the full model explained $R^2 = .33$ of the variance, indicating a moderate rather than strong predictive relationship. These findings should therefore be interpreted with caution, as additional unmeasured factors may also contribute to learning outcomes.

Table 1: Learning outcomes by infrastructure condition (post-test adjusted means and effects).

Outcome Measure	Optimized Infrastruc-ture (OI)	Baseli-ne Infrastructu-re (BI)	Δ (OI - BI)	Effect Size (d)	ANCOVA (F, p, η^2)
Speaking fluency (WPM gain)	+11.3 (SD = 6.4)	+5.1 (SD = 5.9)	+6.2	0.57	F(1,121) = 14.21, p < .001, η^2 = .105
Interactional competence (SD gain)	+0.48 (SD = 0.41)	+0.21 (SD = 0.39)	+0.27	0.49	F(1,121) = 10.37, p = .002, η^2 = .079
Pronunciation error rate reduction (%)	-9.2% (SD = 4.1)	-3.8% (SD = 3.6)	-5.4 pp	0.52	F(1,121) = 8.06, p = .005, η^2 = .062
Engagement / WTC (Δ on 5-point scale)	+0.47 (SD = 0.62)	+0.19 (SD = 0.55)	+0.28	0.46	F(1,121) = 7.88, p = .006, η^2 = .061
Participation equity (Gini reduction %)	-27%	-11%	-16 pp	-	-

3.2 Speaking Fluency

Speech rate increased by +11.3 wpm in OI (post-test M = 97.7, SD = 11.4) compared to +5.1 wpm in BI (post-test M = 91.0, SD = 12.0). ANCOVA controlling for pre-test scores revealed a significant main effect of infrastructure condition, $F(1,121) = 14.21$, $p < .001$, $\eta^2 = .105$. The standardized mean difference at post-test was $d = 0.57$ (medium effect). Temporal measures showed a reduction in mean pause ratio by -14.6% in OI versus -6.2% in BI ($p = .004$), indicating smoother speech production under optimized QoS.

3.3 Interactional Competence

OI participants improved by 0.48 SD, whereas BI improved by 0.21 SD. ANCOVA indicated a significant effect of condition, $F(1,121) = 10.37$, $p = .002$, $\eta^2 = .079$, with a post-test difference of $d = 0.49$. Subscale analyses showed the largest gains in turn-taking management and repair effectiveness, consistent with reduced latency and packet loss facilitating timely feedback and conversational alignment.

3.4 Pronunciation Accuracy

ASR-assisted segmental error rates decreased by -9.2% in OI compared to -3.8% in BI, $F(1,121) = 8.06$, $p = .005$, $\eta^2 = .062$. Human ratings corroborated ASR trends ($r = .71$). The magnitude of improvement suggests that stable audio quality enhanced phonetic noticing and uptake.

3.5 Engagement, Willingness to Communicate, and Participation Equity.

Self-reported engagement and willingness to communicate increased significantly more in OI ($\Delta M = +0.47$ on a 5-point scale) than in BI ($\Delta M = +0.19$), $F(1,121) = 7.88$, $p = .006$, $\eta^2 = .061$. Learning analytics revealed a 27% reduction in the Gini coefficient of speaking time in OI (vs. 11% in BI), indicating more equitable participation. Average speaking time per learner per session increased by +38% in OI compared to +19% in BI. These patterns demonstrate that infrastructural stability amplified the pedagogical effects of role-structured tasks and breakout rooms [6], [7].

3.6 Quality-of-Service (QoS) Effects and Regression Modeling

Hierarchical regression analyses examined the unique contribution of QoS indicators beyond pedagogical design variables (task type, feedback immediacy, role structure). Adding QoS metrics (mean latency, packet loss, jitter) in Step 2 increased explained variance in fluency gains by $\Delta R^2 = .15$, $p < .001$. Latency ($\beta = -.34$, $p < .001$) and packet loss ($\beta = -.29$, $p = .002$) were the strongest negative predictors, whereas jitter had a smaller but significant effect ($\beta = -.17$, $p = .03$). Similar patterns were observed for interactional competence ($\Delta R^2 = .11$, $p = .004$). These findings confirm that infrastructure quality independently predicts learning gains [8], [9]. The relative contribution of QoS indicators to speaking fluency gains is illustrated in Figure 1.

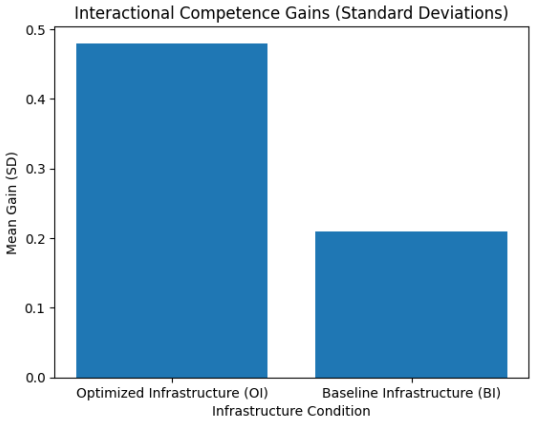


Figure 1: Hierarchical regression model showing the unique contribution of QoS parameters (latency, packet loss, and jitter) to speaking fluency gains ($\Delta R^2 = .15$, $p < .001$).

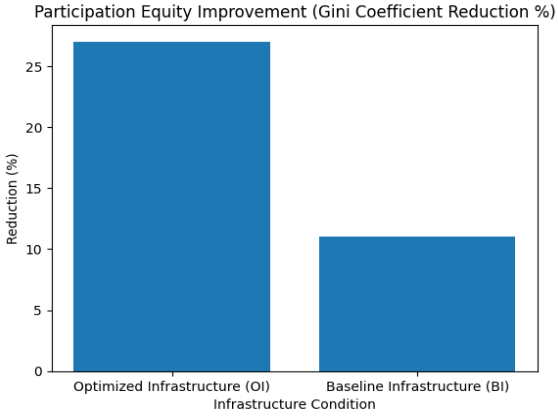


Figure 2: Moderating effect of infrastructure condition on the relationship between feedback immediacy and interactional competence gains.

3.7 Moderation by Feedback Immediacy and Task Design

Moderation analyses indicated a significant interaction between infrastructure condition and feedback immediacy on interactional competence ($\beta_{interaction} = .21$, $p = .01$). In OI, immediate recasts produced larger gains ($\Delta = +0.61$ SD) than delayed feedback ($\Delta = +0.33$ SD), whereas in BI the difference between immediate and delayed feedback was attenuated ($\Delta = +0.28$ SD vs. $+0.24$ SD). This suggests that pedagogical strategies dependent on fine-grained timing are more effective under stable, low-latency conditions [4], [10]. The moderation effects are presented in Figures 2 and 3. Figure 2 compares immediate and delayed feedback across infrastructure conditions, while Figure 3 visualizes the interaction effect with 95% confidence intervals.

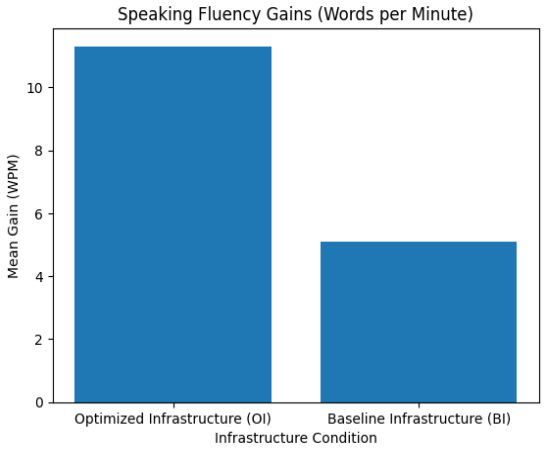


Figure 3: Interaction between infrastructure condition and feedback immediacy on interactional competence gains; error bars indicate 95% confidence intervals.

3.8 Discourse-Level Indicators: Repair Sequences and Uptake

Figure 4 presents the frequency of repair sequences and uptake rates observed under optimized and baseline infrastructure conditions. Conversation analysis of 24 sampled sessions revealed that the frequency of repair sequences per 10 minutes of interaction decreased by 18% in OI compared to 6% in BI. Uptake following recasts occurred within the next turn in 64% of OI instances versus 41% in BI, indicating that infrastructural stability supported tighter feedback–uptake loops. Turn overlap and interruption rates were lower in OI, reflecting improved interactional alignment [1], [10].

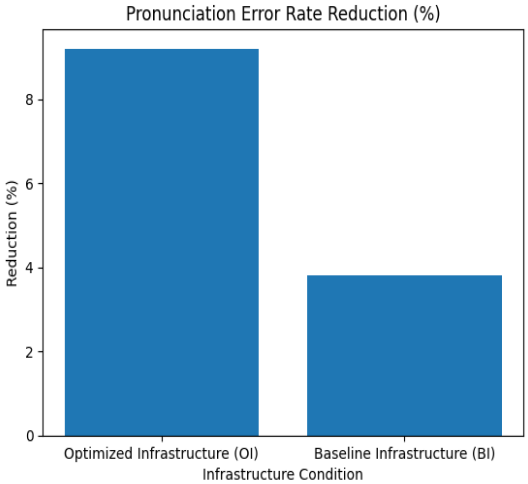


Figure 4: Repair sequence frequency and uptake rates per 10 minutes of interaction under optimized (OI) and baseline (BI) infrastructure conditions.

Across outcomes, optimized cloud infrastructure yielded medium-sized effects on fluency and interactional competence, significant improvements in pronunciation accuracy, higher engagement and WTC, and more equitable participation patterns. Regression and moderation analyses demonstrate that QoS parameters exert independent and interactive effects with didactic design variables, substantiating the central claim that infrastructure quality is a pedagogically consequential factor in synchronous online foreign language learning.

4 DISCUSSION

This study has shown strong evidence that cloud-based communication technology (CBCT) not only acts as a neutral delivery medium for synchronous online foreign language learning, but it also has an active role in determining didactic effectiveness. Results demonstrate that optimized infrastructural conditions (i.e. lower end-to-end latency, lower packet loss, and stable media transport) systematically increased gains in speaking fluency, interactive competence, pronunciation accuracy, and learner engagement. These results add to the body of existing meta-analytic evidence regarding the effectiveness of synchronous computer-mediated communication (SCMC) by detailing the infrastructural boundary conditions under which maximum pedagogical effects are achieved [4], [5].

An interactionist perspective on these medium-sized effects on fluency and interactional competence may be due to conversational contingencies being maintained. Low-latency audio-video streams decrease the extent of temporal misalignment for turn-taking and facilitate timely negotiation of meaning, immediate recasts, and uptake; all of which are fundamental mechanisms for second language development [1]. A reduction in repair sequences at the discourse level and an increased rate of next-turn uptake under optimized infrastructure reflects tighter feedback-uptake loops and aligns with microgenetic accounts of learning in interaction. Sociocultural theory illustrates how multimodal mediation that is stable in various forms offers distributed support for scaffolding with the use of semiotic resources that are appropriated through various types of cues. Gestures, gaze, and shared artifacts remain synchronized with verbal cues and, therefore, may provide more effective means for a learner to appropriate semiotic resources [2].

A primary contribution of the research is that it provides evidence of interaction effects on the relationship between infrastructural factors and pedagogical approaches. The results of moderation analysis indicate that the effects of feedback immediacy were stronger under conditions of low latency. This suggests that pedagogical strategies that are dependent on fine-grained temporal contingencies are also dependent on the quality of infrastructure. These findings raise questions about the validity and reliability of pedagogical practices that are platform agnostic. For example, the instructional designs that include immediate recasts, rapid turn-taking drills and shadowing tasks that focus on pronunciation have an implicit assumption regarding the availability of technical conditions that maintain temporal fidelity. When these technical conditions are not met the efficacy of the pedagogical approach is diminished, and alternative instructional designs may be warranted (e.g., delays in interaction cycles or audio-first protocols).

Learning analytics show that participation equity levels were impacted by infrastructure stability. With short speaking times in enhanced environments and a reduced concentration of dominance, there is evidence that technical reliability can affect how individuals interact without any asymmetries. This matches previous research showing that the people who perceive a lot of technical friction differ when it comes to how much they participate. Thus, this leads to people with lower skill levels and lower confidence participating less than people with higher skill levels and higher confidence creating inequity in terms of communication opportunities. The results can be seen as requiring a co-design of cloud infrastructure and pedagogical orchestration from a systems design perspective. The design selections of edge computing, adaptive bitrate streaming, and autoscaling media servers are not purely performance optimizations; rather, they are pedagogically consequential in nature. The impact of QoS on learning achievement provides an institutional reason to invest in latency-first engineering, audio priority resiliency strategies, and geographically-distributed edge nodes for cohorts' transnational locations. Equally, the way in which AI-mediated feedback (e.g., ASR-based pronunciation support) is implemented must be designed as cognitive overload and algorithmic bias never occur; thus, requiring human-assisted instructional governance.

There are some limitations to be aware of. A quasi-experimental design limits an ability to draw causal conclusions; it is also possible that the institution where this study took place limits generalizing to low-resource environments. Even though QoS logs were collected continuously, micro-

level changes between home networks and the heterogeneous devices in households can result in some residual confounding. The 10-week intervention captures medium-term improvements; we do not know about the durability of the improvements made over time or how well they transfer from those who learned via CBCT to those who have learned under in-person conditions. Additionally, though ASR-based measures of pronunciation were positively correlated with human ratings, and thus may reflect overall speaking proficiency, it is not possible to determine whether automated measurements accurately account for various prosodic or pragmatic aspects of spoken proficiency.

The 10-week intervention captures only short- to medium-term improvements and does not provide evidence regarding the long-term sustainability or retention of the observed learning gains. Future longitudinal studies are needed to determine whether improvements in fluency, interactional competence, and engagement persist over extended periods of time.

Future studies should conduct randomized controlled trials using a wide range of infrastructures to quantify causal relationships between specific components of CBCT (e.g. edge locality, codec) and learning outcomes. Additionally, longitudinal designs could be utilized to measure the durability of gains in interactional competence as well as the degree to which these gains transfer to face-to-face interactions. Mixed-methods studies combining fine-grained conversation analysis with network telemetry would provide further understanding of how the interactional phenomena occurring at a micro-level are influenced by infrastructure. Finally, equity-focused studies should be designed to determine how privacy-aware analytics and adaptive orchestration can reduce the digital divide for students using synchronous language learning technology.

Future research should control for device heterogeneity by standardizing hardware specifications and monitoring client-side performance indicators such as processor load, network stability, and device age.

The shift to using the cloud and CBCT in effective synchronous online foreign language education requires effective pedagogical practices that are paired with cloud infrastructure aligned with the lesson plan(s). In other words: If you treat CBCT as a primary method for accomplishing the course objectives; instead of treating it as an additional technical component; this will allow for greater opportunities for participants to experience interactionist and sociocultural learning processes within digital environments.

5 CONCLUSIONS

This study demonstrates that cloud-based communication infrastructure quality is a pedagogically consequential factor in synchronous online foreign language learning, not merely a technical parameter. Optimized infrastructure conditions - characterized by lower end-to-end latency, reduced packet loss, and stable media transport - produced medium-sized effects on speaking fluency ($d = 0.57$) and interactional competence ($d = 0.49$), along with significant improvements in pronunciation accuracy and learner engagement across a 10-week quasi-experimental intervention with $N = 124$ undergraduate learners.

Hierarchical regression and moderation analyses revealed that QoS parameters independently predict learning gains ($\Delta R^2 = .15$ for fluency) and that pedagogical strategies dependent on fine-grained temporal contingencies - such as immediate recasts and rapid turn-taking - are more effective under low-latency conditions. Discourse analysis further confirmed that optimized infrastructure supports tighter feedback–uptake loops and more equitable participation, as evidenced by a 27% reduction in the Gini coefficient of speaking time in the optimized group.

These findings challenge platform-agnostic assumptions in instructional design and underscore the need for co-design of cloud infrastructure and pedagogical orchestration. Institutions deploying synchronous language learning programs should treat engineering decisions - including edge locality, adaptive bitrate, and QoS prioritization - as didactically motivated choices rather than purely technical ones. Future research should pursue randomized controlled trials with broader infrastructure variation, longitudinal designs to assess durability of gains, and equity-focused studies examining how privacy-preserving analytics and adaptive orchestration can mitigate the digital divide in synchronous language learning contexts.

REFERENCES

[1] M. H. Long, “The role of the linguistic environment in second language acquisition,” in Handbook of Second Language Acquisition, W. C. Ritchie and T. K. Bhatia, Eds. San Diego, CA, USA: Academic Press, 1996, pp. 413-468.

[2] J. P. Lantolf and S. L. Thorne, Sociocultural Theory and the Genesis of Second Language Development. Oxford, U.K.: Oxford University Press, 2006.

[3] OECD, Education at a Glance 2023: Digital Education and Learning Environments. Paris, France: OECD Publishing, 2023.

[4] C.-H. Lin, M. Warschauer and R. Blake, “Effects of synchronous computer-mediated communication on second language speaking: A meta-analysis,” Computer Assisted Language Learning, vol. 35, no. 5-6, pp. 947-973, 2022.

[5] R. Shadiev and Y.-M. Huang, “SCMC meta-analysis,” Educational Technology & Society, vol. 23, no. 3, pp. 1-17, 2020.

[6] L. Kohnke and B. L. Moorhouse, “Facilitating synchronous online language learning: Teacher practices and learner engagement,” System, vol. 110, p. 102909, 2023.

[7] F. Martin and D. U. Bolliger, “Engagement strategies in online learning,” Online Learning, vol. 25, no. 1, pp. 205-222, 2021.

[8] Cisco, Annual Internet Report 2024: Global Network Performance and Trends. San Jose, CA, USA: Cisco Systems, 2024.

[9] ITU, Quality of Service Requirements for Real-Time Interactive Multimedia Applications. Geneva, Switzerland: International Telecommunication Union, 2023.

[10] N. Ziegler, “Synchronous computer-mediated communication and interaction: A meta-analysis,” Studies in Second Language Acquisition, vol. 38, no. 3, pp. 553-586, 2016.

Introduction of Digital Technologies in Vocational Education: Tools for Transforming the Learning Process and Improving the Quality of Staff Training

Olesia Fedoruk
Department of Information and Documentary Communications, National University of Ostroh Academy, Seminarska Str. 2, 35800 Ostrog, Ukraine
olesia.fedoruk@oa.edu.ua

Keywords: Digital Transformation in Education, Vocational Education and Training, Digital Competence Framework, Interactive Learning Materials.

Abstract: The relevance of the study is determined by the processes of digital transformation in the transport industry and the need to modernize the vocational education system aligned with European standards for the development of digital and professional competencies of railway transport specialists. The purpose of the article is to examine the impact of digital technologies and tools on the quality of professional training of future railway industry specialists in the context of the digitalization of education. The research methodology is based on a combination of analysis of scientific sources, generalization of the results of pedagogical pedagogical experiments and student surveys, a comparative analysis of indicators reflecting the development of digital and professional competencies. The findings indicate that the integration of interactive digital instructional materials contributes to increased learning motivation, the development of digital competence, and improved practical training for future railway transport specialists. It has been proven that digital tools provide a closer link between theoretical learning and practical professional activity, as well as facilitates the development of students' critical thinking skills and independent work skills among students. It has been shown that the introduction of digital solutions into professional training is consistent with the principles of competence-oriented education and increases the adaptability of graduates to the demands of the modern labor market. The theoretical significance of the results lies in the generalization of approaches to the digitization of vocational education within the framework of European education policy; the practical significance of the study lies in the applicability of its findings for enhancing the content, formats, and methods of training transport specialists, as well as for the design and implementation of contemporary digital educational resources.

1 INTRODUCTION

Digital transformation is one of the key characteristics of modern European development, covering the economy, transport infrastructure, and education system. Rail transport plays a strategic role in ensuring sustainable mobility, economic integration, and environmental safety, which leads to increased requirements for the the professional preparation of industry specialists. In this context, there is an increasing demand for specialists capable of effectively operating digital management systems, automated technologies, and intelligent transport solutions.

The vocational education and training (VET) system in European countries is gradually transforming in line with the principles of digital education, lifelong learning, and a competency-based approach. At the same time, the problem of insufficient the integration of digital technologies into the training of technical specialists remains a pressing issue for many education systems, particularly in light of the need to align with the requirements of the European labor market.

In this regard, the implementation of interactive digital instructional teaching materials that combine theoretical content, multimedia elements, and practical tasks using simulators is becoming particularly relevant. Such tools contribute to improving the quality of education, developing digital competence, and shaping the professional readiness of future railway transport specialists.

The purpose of the article is to provide a comprehensive analysis of the impact of digital technologies and tools on the quality of training for railway transport specialists within the framework of the European digital transformation of education.

2 LITERATURE REVIEW

The digitization of vocational education and the professional training of technical and transport specialists is widely represented in contemporary scientific research. A significant body of literature focuses on examining the role of digital technologies in transforming educational systems, fostering digital competence development, and aligning vocational education with contemporary labor market demands. Recent studies also highlight the importance of blended learning models and the readiness of teaching staff as key factors influencing the effectiveness of digital transformation in vocational education [1].

In the works of Ukrainian researchers [2], [3], digital technologies are regarded as a critical driver of improving the quality of professional training by fostering interactive learning, individualizing educational trajectories, and enhancing the development of practical skills. The authors emphasize the need to shift from traditional instructional models toward innovative, learnercentered approaches that actively engage students through multimedia and network-based resources. Similar ideas are further developed in the studies by Sazhko, Strutynska, and Umryk [4], [5], where the digitization of education is interpreted as a systemic transformation of pedagogical approaches and educational environments driven by the dynamics of the digital society. At the same time, researchers emphasize the growing role of artificial intelligence and deep learning technologies in expanding access to vocational education and improving learning outcomes [6].

A distinct line of inquiry focuses on the formation and advancement of digital competence among teachers and students in vocational education. Research in this field underscores the significance of information and communication technologies as instruments for strengthening pedagogical professionalism, fostering continuous professional development, and ensuring the meaningful integration of digital educational resources into the instructional process [7]. These findings are complemented by research emphasizing the importance of competence-based models in vocational education under conditions of crisis and systemic change, where digitalization serves as a tool for maintaining sustained delivery of education and quality [7]. In this context, particular attention is paid to the development of meaningful learning experiences through innovative digital pedagogy, which enhances student engagement and supports practice-oriented training [8].

An essential aspect of educational digitalization pertains to the management and governance of educational institutions, as these factors directly determine the effectiveness of implementing digital innovations implementation Kubitskyi et al. [9] argue that the digital transformation of higher education management enhances institutional capacity, supports sustainable development, and creates favorable conditions for the systematic adoption of digital educational tools. This managerial perspective is highly relevant for vocational and technical education institutions, where organizational readiness and strategic leadership play a critical role in the successful integration of digital learning environments. Furthermore, the integration of digital technologies in vocational education is increasingly associated with the principles of Industry 5.0, which emphasize human-centered and technologically enhanced learning environments [10].

The digitization of training for transport and railway specialists is considered in studies focusing on the application of intelligent systems, automated control technologies, and digital security solutions. Akulov et al. [11] demonstrate that computer-based training tools significantly contribute to the development of practical skills in students, increasing their readiness to work with complex technical systems. Similarly, Vinichenko [12] emphasizes the importance of technology-oriented models of training personnel in the transport industry, highlighting the role of digital simulations and applied learning technologies in professional training. These approaches are supported by studies that identify emerging technologies, including virtual environments and digital platforms, as effective tools for developing industry-relevant competencies in vocational education [13].

At the same time, several authors draw attention to the barriers hindering the effective implementation of e-learning in railway and transport education. Fraszczyk and Piip [14] identify limited resources, insufficient digital infrastructure, along with disparities in digital preparedness among stakeholders in the educational process, as significant factors limiting the scalability and long-term sustainability of digital initiatives learning solutions.

International policy-oriented studies and analytical reports further contextualize these findings. Initiatives aimed at increasing flexibility and innovation in apprenticeship and technical training systems emphasize the role of digital technologies in enhancing accessibility, adaptability, and lifelong learning opportunities [15]. These approaches align with broader European and global trends that prioritize digital competence as a core skill across all levels of education and professional training.

Despite the growing body of research on educational digitalization, the analysis of scientific sources indicates that insufficient attention has been paid to the comprehensive use of interactive digital learning materials specifically in the training of railway transport specialists within vocational education systems. This research gap underscores the need for further studies aimed at substantiating the pedagogical effectiveness of digital tools and identifying the conditions necessary for their successful and sustainable implementation in vocational railway education.

3 METHODOLOGY

The study is grounded in a comprehensive integration of theoretical and empirical methods aligned with European frameworks of evidence-based education and established quality assurance standards in vocational education and training. The theoretical stage involved a systematic analysis of scientific publications, analytical reports from international organizations, and regulatory and methodological documents aimed at studying the processes of digitization of vocational education, the development of digital competence, and the modernization of training for technical and transport specialists.

The empirical part of the study included a questionnaire, expert assessment, self-assessment, and a pedagogical experiment. The study involved a total sample of 43 students enrolled in railway-related vocational education programs, including 21 students in the control group (CG) and 22 students in the experimental group (EG). The sample was formed using a purposive sampling approach, ensuring the comparability of groups in terms of educational level and specialization.

The questionnaire was used to identify the dynamics of learning motivation, students' satisfaction with the overall quality of education, and the formation of digital and professional competencies of students. The questionnaire consisted of 9 items combining single-choice, multiple-choice, and open-ended questions. In particular, it included: 1) frequency of digital technology use; 2) types of digital tools used; 3) perceived effectiveness of digital technologies in learning; 4) advantages of digital tools; 5) challenges and barriers; 6) evaluation of the level of digitalization in the institution; 7) open-ended suggestions for improving digital learning; 8) attitudes toward studying digital-oriented disciplines; and 9) demographic data (gender).

To measure students' perceptions, ordinal scales were applied (e.g., frequency scales and 5-level evaluation scales ranging from "very high" to "very low" or from "very strong" to "no impact"). Multiple-choice questions allowed respondents to select several relevant options, enabling a broader analysis of digital tool usage and perceived benefits. Open-ended responses were used to capture qualitative insights into students' needs and expectations.

The content validity of the questionnaire was ensured through alignment with the research objectives and prior studies in the field of digitalization of vocational education, as well as through expert review. Reliability was supported by the consistency of responses across related items and the standardized structure of the questionnaire.

The expert assessment was aimed at determining the pedagogical expediency, methodological validity, and practical significance of digital learning materials. The expert group included 5 educators with experience in vocational and technical education, who evaluated the developed digital manual based on criteria such as relevance, usability, interactivity, and pedagogical effectiveness using a 5-point scale. The self-assessment method made it possible to evaluate the level of awareness of students' readiness for professional activity in a digital educational environment.

A key element of the empirical study was a pedagogical experiment aimed at testing the effectiveness of introducing an interactive digital manual on the course "Digital Technologies in Railway Transport" into the professional preparation of prospective railway transport specialists within the vocational education framework. The experimental design incorporated a comparative evaluation of learning outcomes across control and experimental groups, enabling an objective assessment of the effects of digital tools on academic performance.

For data analysis, descriptive statistical methods were applied, including percentage distribution of responses and comparative analysis between groups. The results of the questionnaire were visualized using diagrams to reflect trends in digital technology usage,

perceived benefits, and existing challenges. To assess differences between control and experimental groups, comparative analysis of percentage changes in learning outcomes was conducted.

The methodological approach used ensures the scientific validity, reliability, and reproducibility of the research results and meets current European requirements for evaluating the effectiveness of digital innovations in the vocational education system.

4 RESULTS AND DISCUSSION

4.1 Findings on the Implementation of Digital Technologies

The findings of the study are derived from an analysis of the integration of digital technologies into the educational processes of vocational (technical) education. The main focus is on assessing the impact of digital solutions on the quality of education, the development of professional and digital competencies of students, as well as the effectiveness of using interactive educational resources in training future specialists in the railway industry.

The key practical result of the research was the development and implementation of an interactive multimedia guide for the course “Digital

Technologies in Railway Transport,” created on the Book Creator platform. This resource combines theoretical material, multimedia elements, and practice-oriented tasks aimed at developing digital competence, professional thinking, and the ability to apply modern technologies in production activities.

The structure of the interactive manual includes an introduction with a definition of the objectives and expected learning outcomes, a list of literature and sources for independent study, as well as thematic modules:

1) Topic 1. Introduction to digital technologies in rail transport: overview of key concepts and trends;
2) Topic 2. Internet of Things (IoT), Artificial Intelligence (AI), and their application in rail transport;
3) Topic 3. Digital infrastructure monitoring and management systems;
4) Topic 4. Automated train control systems (ATCS);
5) Topic 5. Digital technologies in the maintenance and repair of rolling stock;
6) Topic 6. Digital freight transport management systems;
7) Topic 7. Innovations and future directions for the advancement of digital technologies in rail transport;
8) Topic 8. The role of digital technologies in promoting the sustainable development of rail transport;
9) Topic 9. Practical dimensions of implementing digital technologies in rail transport.

The course is structured based on the principle of gradually increasing the complexity of the material, from the basic concepts of railway digitalization to the practical aspects of implementing digital systems, innovations, and sustainable development in the industry.

4.2 Results of the Questionnaire and the Ascertaining Stage of the Experiment

At the descriptive stage of the study, a survey was conducted among students enrolled in railway-related programs to assess the extent of digital technology and tool utilization within the educational process. The findings indicated that the integration of digital solutions was insufficient: more than half of the respondents rarely use digital technologies in their studies (see Fig. 1).

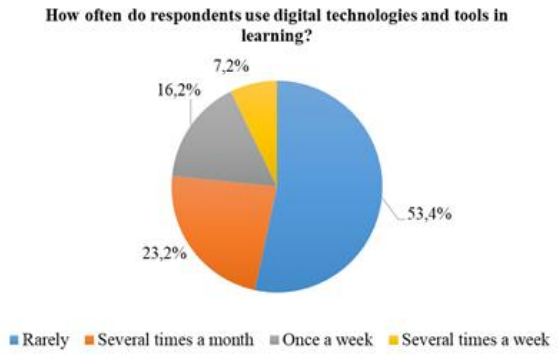


Figure 1: How often do respondents use digital technologies and tools in learning?

Analysis of the responses showed that portable devices and educational platforms (Moodle, Google Classroom), as well as interactive whiteboards, are most commonly used in the learning process (see Fig. 2). At the same time, innovative tools such as simulators, virtual and augmented reality, gamified assessment platforms, and digital manuals are rarely used.

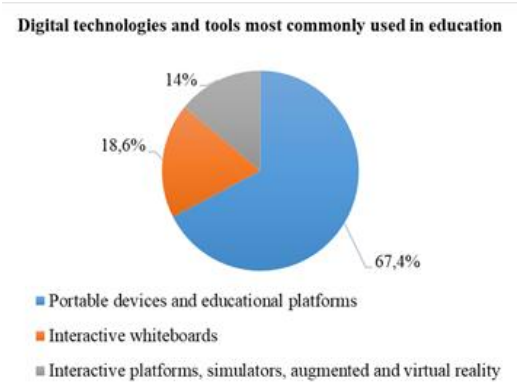


Figure 2: Digital technologies and tools most commonly used in education.

Despite this, most respondents positively assess the impact of digital technologies on learning: respondents noted their significant help in mastering the material (see Fig. 3).

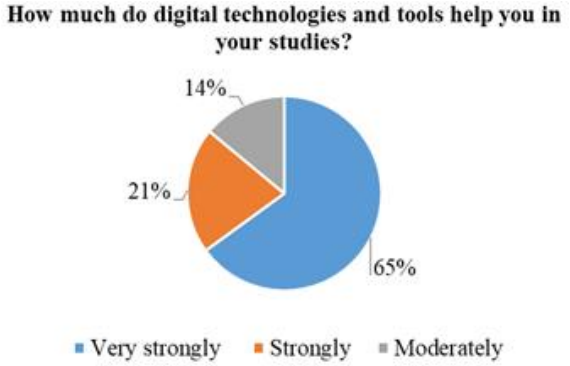


Figure 3: How much do digital technologies and tools help in learning?

The main advantages of using digital tools are convenience and mobility, as well as improved access to educational information (see Fig. 4).

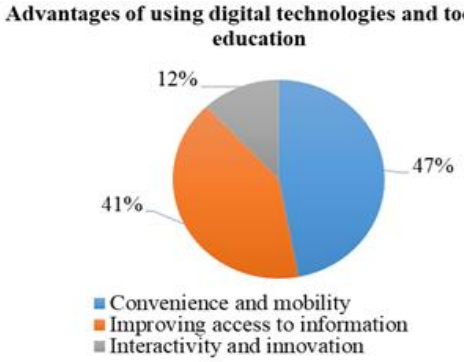


Figure 4: Advantages of using digital technologies and tools in education.

At the same time, a number of problems were identified that hinder the digitization of the educational process. Among the main difficulties, respondents cited insufficient digital skills, technical problems, and the high cost of equipment and software (see Fig. 5).

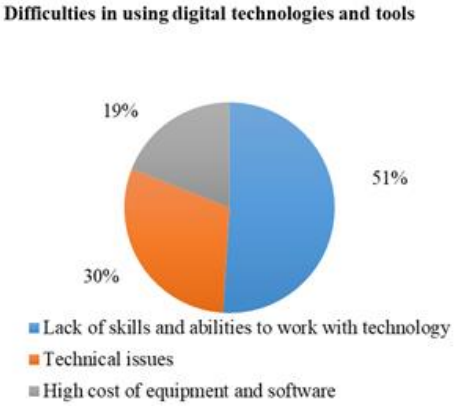


Figure 5: Difficulties in using digital technologies and tools.

Thus, the analysis points to a clear mismatch between the positive perception of digital technologies and their limited practical use in the educational process. Although students recognize the benefits of digital tools in improving learning outcomes, accessibility, and convenience, their use remains largely limited to basic platforms and devices. The identified barriers – insufficient digital skills, technical limitations, and financial constraints – highlight the need for a more systematic and targeted approach to digital integration, creating conditions conducive to the broader and more effective integration of digital technologies.

4.3 Formative Stage of the Experiment and Comparative Analysis of Findings

Based on the outcomes of the diagnostic (ascertaining) stage, a formative pedagogical experiment was organized. The students were divided into a control group (CG) and an experimental group (EG). The initial indicators of the level of knowledge and skills in the field of digital technologies in both groups were comparable and were characterized by a predominance of average and low levels (Table 1).

Table 1: Results of the diagnostic assessment of students in the ascertaining experiment.

Level of knowledge and skills acquired by students	Students who participated in the experiment	
	CG	EG
High	14% (3 students)	9% (2 students)
Average	57% (12 students)	50% (11 students)
Low	29% (6 students)	41% (9 students)

During the formative stage, the experimental group consistently employed the interactive manual and digital tools in the completion of practical assignments, whereas the control group received instruction through a traditional instructional model. After the experiment was completed, a repeat diagnosis was conducted, the results of which are presented in Table 2.

Thus, comparative analysis showed a significant increase in the level of digital knowledge and skills among students in the experimental group. In particular, the proportion of students with a high level of training increased significantly, while the number of students with a low level decreased. Only minor changes were recorded in the control group, indicating the limited effectiveness of traditional approaches without the active use of digital resources.

4.4 Discussion of Results and SWOT Interpretation

The empirical data obtained were summarized and interpreted using SWOT analysis (Table 3), which made it possible to systematically assess the impact of digital technologies and tools on the quality of training of railway industry specialists.

The strengths include expanded access to educational information, interactivity and personalization of learning, increased digital literacy among learners, and improved competitiveness in the labor market. Opportunities are linked to the implementation of innovative academic courses, the expansion of international collaboration, and the advancement of distance and flexible learning formats.

Table 2: Findings of the diagnostic assessment conducted during the ascertaining stage of the experiment.

Level of knowledge and skills acquired by students	Experimental groups					
	CG			EG		
	Constative stage	Formative stage	Difference (%)	Constative stage	Formative stage	Difference (%)
High	14% (3 students)	14 % (3 students)	+0%	9% (2 students)	36% (8 students)	+27%
Average	57% (12 students)	62 % (13 students)	+5%	50% (11 students)	55% (12 students)	+5%
Low	29% (6 students)	24 % (5 students)	-5%	41% (9 students)	9% (2 students)	-32 %

Table 3: SWOT analysis of the impact of digital technologies and tools on the quality of education and training of specialists.

Strengths	Weaknesses
1) A significant increase in the proportion of students with high levels of digital competence in the experimental group (+27%). 2) Improving learning effectiveness through the use of an interactive digital guide. 3) Students have a positive view of digital technologies (the majority note their usefulness in learning). 4) Improved learning outcomes in the EG compared with the CG (a consistent upward trend).	1) Initially, there was a high proportion of students with poor digital skills (41% in the EG). 2) Limited use of advanced digital technologies (VR, simulators) in the actual teaching process. 3) The lack of digital literacy among some students, which makes it difficult to implement complex digital tools. 4) The limited availability of teaching materials adapted for digital formats in traditional courses.
Features	**Threats**
1) Expanding the use of the interactive digital guide to other subjects. 2) Further improvement of students' digital skills, building on the positive results of the pilot scheme. 3) The integration of more advanced technologies (VR/AR, simulators) into the learning process. 4) Expanding the use of blended learning based on the results of the experiment.	1) Dependence on technical infrastructure (risk of disruptions to digital systems). 2) The rapid obsolescence of digital technologies and the need to constantly update resources. 3) Unequal access for students to digital resources outside the educational institution. 4) The lack of digital literacy among some teachers hinders the full implementation of innovations.

At the same time, potential weaknesses and risks encompass financial limitations, technical malfunctions, unequal access to digital resources, insufficient professional development of teaching staff, cybersecurity challenges, and the rapid obsolescence of technological solutions. This requires a strategic approach to the digital transformation of education.

4.5 Generalization and Educational Implications

The results of the study confirm that digital technologies are not only an auxiliary tool, but also a systemic factor in improving the overall quality and effectiveness of vocational education. The implementation of interactive multimedia manuals enhances students' learning motivation, fosters the development of practical digital competencies, and strengthens their preparedness for professional activity within the ongoing digital transformation.

Therefore, the strategic and purposeful integration of digital technologies into the educational process contributes to significant qualitative transformations in the professional training of specialists.

5 CONCLUSIONS

The results of the empirical study demonstrate not only the general effectiveness of digital technologies, but also specify the conditions and tools that ensure measurable improvements in the professional training of future railway specialists. In particular, the implementation of the interactive multimedia guide developed on the Book Creator platform proved to be an effective pedagogical solution that integrates theoretical content, multimedia elements, and practice-oriented tasks into a unified digital learning environment.

The comparative analysis of the control and experimental groups confirmed a statistically significant positive dynamic in the development of students' digital and professional competencies. In the experimental group, the proportion of students with a high level of knowledge increased from 9% to 36% (+27%), while the share of students with a low level decreased from 41% to 9% (-32%). In contrast, the control group demonstrated only minor changes, which indicates the limited effectiveness of traditional instructional approaches without systematic digital support.

The findings also revealed a contradiction between students' positive perception of digital technologies and their limited actual use in the educational process. Despite recognizing their effectiveness for improving accessibility, flexibility, and learning outcomes, students primarily use only basic digital tools (e.g., learning platforms and portable devices), whereas advanced technologies such as simulators, virtual and augmented reality, and interactive digital manuals remain underutilized.

The study identifies key barriers to the effective integration of digital technologies, including insufficient digital competence of participants in the educational process, technical limitations, and financial constraints. These factors necessitate a systematic approach to digital transformation that includes targeted teacher training, modernization of digital infrastructure, and the integration of specialized interactive learning resources into curricula.

Thus, the novelty of the study lies in substantiating the pedagogical effectiveness of interactive multimedia learning tools specifically for the training of railway transport specialists and in identifying the organizational and methodological conditions required for their successful implementation. The practical significance of the results is reflected in the possibility of scaling the developed digital resource and adapting it for other vocational education programs in technical fields.

REFERENCES

[1] Yi. Yang, "Optimizing Blended Learning with Digital Technologies: Insights into Faculty Readiness in Higher Vocational Education," Membrane Technology, pp. 112-127, 2024, doi: 10.52710/mt.90.

[2] V. Boichuk, V. Umanets and O. Boichuk, "Digital technologies in the professional training of future specialists," Modern Information Technologies and Innovative Teaching Methods in Specialist Training: Methodology, Theory, Experience, Problems, vol. 60, pp. 353-364, 2021.

[3] R. Hurevych, M. Kademiia, N. Opushko, T. Ilnitska and H. Plakhotniuk, "The role of digital learning technologies in an era of civilizational change," Modern Information Technologies and Innovative Teaching Methods in Specialist Training: Methodology, Theory, Experience, Problems, vol. 62, pp. 28-38, 2021.

[4] H. Sazhko, "Digitization of the educational process for training future engineering educators: theoretical aspects," Modern Educational Technologies, no. 70, pp. 84-91, 2021.

[5] O. Strutynska and M. Umryk, "Contemporary educational trends in the context of the development of the digital society," Information and Communication Technologies in Education, no. 26, pp. 201-205, 2020.

[6] M. Huang, Y. Teng and S. Yusoff, "Empowering vocational education in Africa through AI and deep learning technologies," Future Digital Technologies and Artificial Intelligence, vol. 1, no. 2, pp. 27-32, 2025, , doi: 10.55670/fpll.fdtai.1.2.4.

[7] V. Kovalchuk, S. Maslich and L. Movchan, "Digitalization of vocational education under crisis conditions," Educational Technology Quarterly, iss. 1, pp. 1-17, 2023, , doi: 10.55056/etq.49.

[8] K. Mbatha, "Meaningful Learning Experience Using Digital Technologies in TVET: Towards Innovative Digital Pedagogy," in Technical and Vocational Teaching in South Africa, pp. 247-262, 2024, doi: 10.1007/978-3-031-58206-6_12.

[9] S. Kubitskyi, R. Shchokin, O. Fedoruk, T. Horokhivska and I. Shorobur, "Management of Higher Education Institutions as a New Tool for the Development of Higher Education," Journal of Curriculum and Teaching, vol. 12, no. 2, Special Issue, pp. 74-82, 2023, , doi: 10.5430/jct.v12n2p74.

[10] P. Marinič and P. Pecina, "Use of digital technologies in vocational education and their connection with industry 5.0," INTED2024 Proceedings, no. 1, pp. 7615-7621, 2024, , doi: 10.21125/inted.2024.2022.

[11] A. Akulov, K. Zhelieznov, O. Zabolotnyi, E. Chabaniuk and A. Shvets, "Computer training tools for students and graduates of railway universities in the development of practical skills," International Journal of Mechanical Engineering Education, 2024, doi: 10.1177/03064190241227529.

[12] V. Vinichenko, "Technology of personnel training for the transport industry," SHS Web of Conferences, vol. 97, art. 01006, 2021, [Online]. Available: https://www.shs-conferences.org/articles/shsconf/pdf/2021/08/shsconf_teduvis2020_01006.pdf, doi: 10.1051/shsconf/20219701006.

[13] L. Ghosh and R. Ravichandran, "Emerging Technologies in Vocational Education and Training," Journal of Digital Learning and Education, vol. 4, no. 1, pp. 41-49, 2024, , doi: 10.52562/jdle.v4i1.975.

[14] A. Fraszczyk and J. Piip, "Barriers to eLearning in rail," Transportation Research Procedia, vol. 48, pp. 168-186, 2020, , doi: 10.1016/j.trpro.2020.08.014.

[15] Government of Canada, Flexibility and Innovation in Apprenticeship Technical Training, 2022, [Online]. Available: https://www.canada.ca/en/services/jobs/training/support-skilled-trades-apprentices/flexibility-innovation-technical-training.html.

Intelligent Tutoring Systems as Communication Technologies for Enhancing Second Language Learning

Olha Dienichieva, Natalia Plakhotniuk, Olena Svysiuk, Svitlana Kukharonok and Oleg Makarevych
Zhytomyr Polytechnic State University, Chudnivska Str. 130, 10005 Zhytomyr, Ukraine
olhadienichieva@gmail.com, nat.pl2002@gmail.com, ktpl_sov@ztu.edu.ua, kim_kss@ztu.edu.ua,makarevoleg@ukr.net

Keywords: Intelligent Tutoring Systems (ITSs), Communication Technologies, AI-Driven Learning Tools, Adaptive Learning, Learner Autonomy, Learning Environments.

Abstract: This study examines Intelligent Tutoring Systems (ITSs) as communication technologies within second language learning in higher education. Rather than viewing ITSs solely as adaptive instructional tools, the research focuses on their communicative functions in mediating learner–system interaction, supporting learner autonomy, and sustaining engagement in digitally mediated English language learning environments. The study adopts a mixed-methods research design combining questionnaire data and learning analytics collected from 300 undergraduate students studying English as a second language. The questionnaire explored students' perceptions of ITSs across three dimensions: communicative interaction and language skill development, adaptivity and usability, and overall instructional effectiveness. The results indicate that students most frequently use ITSs for listening, reading, and controlled speaking and writing practice, which they associate with improvements in pronunciation, fluency, and communicative confidence. Personalized feedback, flexible pacing, and opportunities for repeated practice were identified as key factors supporting learner autonomy and sustained engagement. At the same time, the findings reveal limitations in the communicative depth of ITSs, particularly in open-ended speaking and writing tasks, and highlight students' expectations for more interactive and dialogic features. The study concludes that Intelligent Tutoring Systems function as communicatively oriented learning environments that mediate interaction and support self-regulated second language learning, while emphasizing the need for pedagogically grounded integration and further development of their communicative capacities.

1 INTRODUCTION

Advances in artificial intelligence and digital learning environments have significantly reshaped approaches to instructional design, particularly in contexts characterized by linguistic diversity. In contemporary educational and professional contexts, competence in a second language, especially English, has increasingly become a functional requirement rather than an optional asset. Learners, therefore, need instructional tools that extend beyond linguistic knowledge to support meaningful communication across cultural and situational contexts.

To address these evolving needs, many institutions are adapting digital platforms that can adjust instruction to individual learners. Intelligent Tutoring Systems (ITSs) have become central to this shift, as they offer structured support that adapts continuously to learners' progress and communicative needs [1].

Current Intelligent Tutoring Systems represent a substantial departure from earlier forms of computer-based learning software. Through AI-based analytics, these systems monitor learner behaviour, identify performance patterns, and adapt instructional tasks to individual learning trajectories. Research by Le, Jia, and Annuš highlights how these systems cultivate sustained learning by providing immediate, individualized feedback and flexible environments in which learners can make informed choices about their study routines [2]. By responding dynamically to learner input, such as accuracy, hesitation, or error patterns, ITSs create immediate instructional exchanges that can enhance the relevance of second language practice.

The expanding role of digital instruction in schools and universities has increased the visibility and adoption of Intelligent Tutoring Systems. As institutions confront ongoing challenges, such as the need to support remote or hybrid learning, increasing

classroom diversity, and persistent demands for equitable access to high-quality instruction, ITSs function as communication interfaces that organize how students engage with digital content [3]. These systems enable differentiated task design, promote learner responsibility, and provide greater flexibility in how and when learning occurs [4]. According to C. Lin, A. Huang and O. Lu, evidence also shows that appropriately designed ITSs can reinforce affective components of language learning, including motivation and sustained engagement—elements that are essential for long-term proficiency development [1].

Effective integration of ITSs in language education depends on factors that extend beyond mere technological availability. Considerable variation exists in learners' digital literacy, as well as in how ITS designers implement adaptivity, feedback mechanisms, and user guidance [5]. Language learners also navigate complex digital environments that include translation applications and online reference tools governed by distinct communicative logics [6]. Understanding how learners coordinate these tools: how they respond to the communicative mechanisms embedded within ITSs—remains an important research priority. Such insights can contribute to more intentional instructional planning and more coherent second language learning environments [7].

Taken together, these trends underscore the importance of closely examining the role of intelligent educational systems as communication technologies in second language acquisition. As AI-driven learning tools evolve rapidly and global demand for English proficiency continues to rise, there is a compelling need to analyse how students actually experience these systems. This study, therefore, investigates how learners engage with intelligent tutoring platforms and how such platforms contribute to the development of core linguistic competencies. The findings aim to illuminate strategies for using ITSs to design effective, communicatively rich second language learning environments.

2 RESEARCH GOALS

This study investigates the communicative functions of Intelligent Tutoring Systems in second language learning contexts, focusing particularly on their ability to deliver highly individualized instructional guidance and sustain learner motivation. The study specifically examines ITSs as communication technologies that mediate learner–system interaction. Beyond outlining what ITSs can do, the research examines how their communicative design, which refers to the way prompts, feedback, and instructional messages are structured to support meaningful interaction, shapes contemporary language-teaching practices, fosters learner autonomy, which is the ability to manage and evaluate one's own learning, and facilitates interaction in digitally mediated environments.

A further objective is to analyse how effectively ITSs maintain learner engagement, adjust feedback to evolving learner needs, and promote active participation in virtual learning spaces. The practical contribution of the study lies in outlining pedagogically grounded approaches for integrating ITSs into routine language teaching. In doing so, the research seeks to support the development of instructional models that are adaptable, communication-oriented, and resilient in the face of ongoing educational change.

3 RESEARCH METHODOLOGY

Recent scholarship demonstrates that Intelligent Tutoring Systems (ITSs) and communication technologies play an increasingly influential role in digitally mediated language learning. ITSs and AI-driven platforms that monitor learner performance and adapt instruction have been shown to support a wide range of cognitive and affective learning processes [1], [2]. In addition, empirical studies confirm that ITSs enhance reading comprehension, promote effective strategy use, support vocabulary development, and contribute to overall academic progress in English as a foreign language [11]. These systems personalize instruction by diagnosing errors, offering immediate feedback, and adjusting task difficulty, features that are central to adaptive learning environments.

Beyond cognitive benefits, recent work highlights the emotional and motivational dimensions of ITSs. Affective ITS models have demonstrated the potential to enhance learner engagement and reduce anxiety by responding to emotional states and providing supportive feedback [8]. Evidence from recent studies suggests that ITSs support the development of learner autonomy by allowing students to regulate their own pacing, make strategic learning choices, and engage in reflective practices [7], [9]. Studies from adjacent fields, such as programming education, further show that ITSs effectively support learners with different prior

knowledge levels, suggesting strong potential for differentiated instruction in language learning contexts[10].

From a communication technologies perspective, digital platforms actively mediate interaction by influencing how learners interpret guidance, negotiate meaning, and develop knowledge. ITSs, although automated, function as communication technologies by structuring learner–system dialogue through prompts, hints, corrections, and explanations. According to X. Weng and T. K. F. Chiu, communication technologies that mediate interaction among learners, instructors, and automated systems complement ITSs by facilitating multimodal communication and sustained engagement [4]. These technologies enable authentic interactional opportunities, peer collaboration, and continuous access to digital resources—affordances that are essential for language learning and intercultural competence [4]. In ITS environments, system-generated prompts, hints, and explanations structure learner attention and shape task interpretation.

Rather than producing learning benefits through isolated features, the effectiveness of Intelligent Tutoring Systems emerges from the interaction between adaptive instruction and communicative system design. Prior research indicates that learners engage more deeply when feedback is not only personalized but also framed in ways that support reflection and strategic decision-making [2]. Opportunities for autonomy arise when learners can regulate pacing and task selection, yet these opportunities depend on how clearly system guidance communicates expectations and progress [7]. Similarly, motivational gains associated with ITSs appear to be influenced by interface responsiveness and feedback tone, suggesting that affective support is embedded within communicative interaction rather than added as a separate component [11]. These communicative and adaptive elements, when combined with learning analytics that provide visibility into learner behaviour for both the system and instructors, create learning environments capable of supporting sustained engagement and differentiated language development [12].

Researchers such as A. Norbert, B. Peña-AcuñaandR. C. F. Durão highlight that challenges must be addressed to optimize the integration of these technologies [5], [6]. These include inconsistencies in system design quality, the need for additional training to overcome digital literacy gaps, and the complexity of learners' digital ecologies, where students frequently rely on external tools such as translators and dictionaries [5], [6]. ITSs generate detailed data on learner behaviour, enabling continuous assessment and refinement of instructional strategies. Learning analytics theory holds that data-driven insights can support personalized feedback, identify learning difficulties, and inform instructional decisions. H. I.Akyüz and M. Erdemir focus on the value of system-generated analytics for understanding learner perceptions and improving instructional design [3].Similarly, M.-T. Kou, H.-K. Wu, and C.-Y. Hsieh demonstrate how behavioural data from ITS environments reveal differences in learning trajectories across proficiency levels [9].In language learning, such analytics inform targeted interventions in reading, vocabulary development, and comprehension [12].

Building on these foundations, this study uses a mixed-methods design to examine how learners engage with ITSs as communication technologies. Data are collected from learning analytics and questionnaires. Quantitative data are analysed to identify patterns in engagement and performance, while qualitative data are thematically coded to capture learners' experiences. This approach provides a clear view of ITSs' role in supporting communication, autonomy, and motivation in second language learning.

Communication technologies theory helps explain how learners interpret and respond to automated guidance, while adaptive learning theory clarifies how ITSs tailor instruction to individual needs [13]. Affective computing and learning analytics contribute additional layers that explain motivation, engagement, and instructional decision-making. This integrated framework provides the conceptual basis for examining how learners experience the communicative and adaptive features of ITSs in second language learning contexts [14].

Overall, ITSs and communication technologies hold substantial promise for advancing second language learning. Their combined capacity for adaptivity, communicative mediation, and autonomy-supporting design aligns with current theoretical perspectives in smart learning and provides a foundation for more personalised and interaction-rich learning environments.

4 RESULTS AND DISCUSSION

The empirical stage of the research involved 300 undergraduate students from Zhytomyr Polytechnic State University, representing both non-linguistic and linguistic specialities, ensuring a balanced and representative sample**.** All respondents studied

English as a second language and had prior experience using digital educational platforms, AI tools, or Intelligent Tutoring Systems (ITSs) during their university studies. Participation in the survey was voluntary and anonymous, enhancing the reliability of responses and minimizing social desirability bias**.**

The research instrument consisted of a structured questionnaire divided into two parts, each containing 15 questions**.** The first part assessed the frequency with which students used ITSs or AI-based tools to develop different language skills and their perceived effectiveness in communicative contexts. The second part focused on students' perceptions of ITSs as communication technologies, addressing three key dimensions: (1) communicative interaction and language skill development, (2) adaptivity, usability, and learning strategies, and (3) overall perceived effectiveness of ITSs in supporting English language learning. Responses were collected using Likert-type scales, which enabled quantitative analysis, while open-ended questions provided qualitative insights into learner experiences.

Data analysis employed both descriptive and inferential methods. The first part of the questionnaire examined how frequently students used ITSs or AI-based tools to develop different language skills and how effective they perceived these tools to be in communicative contexts. The results demonstrate that ITSs are most actively used for receptive skills and controlled productive practice, while more complex communicative skills demonstrated moderate but consistent engagement (Table 1).

A substantial proportion of respondents reported frequent use of ITSs for listening and reading practice. Speaking and writing activities facilitated by ITSs were also prevalent, although to a slightly lesser extent. Importantly, students associated ITS use with improvements in pronunciation, fluency, and confidence in real-life communication situations.

The data indicate that ITSs function not only as instructional tools but also as communicative mediators that simulate interaction, provide structured dialogue practice, and support the gradual development of communicative competence. Students particularly valued opportunities to rehearse responses, repeat exercises, and receive immediate feedback in a low-anxiety environment.

The second part of the questionnaire focused on the adaptive and strategic dimensions of ITS use. The findings reveal that students clearly recognize the value of personalization and system responsiveness, although the degree of perceived adaptivity varies (Table 2).

Most respondents agreed that ITSs adjust tasks according to their proficiency level at least occasionally. Personalized feedback was evaluated as one of the most useful features, especially for grammar, vocabulary, and writing tasks. Ease of navigation and clarity of interface were also rated positively, indicating that usability does not present a major barrier to adoption.

Table 1: Use of ITSs for communicative interaction and language skill development.

Survey issues	Results and outcomes
Practising speaking English using ITSs	6% of respondents never practise speaking with ITSs, 18% rarely, 38% sometimes, and 38% often, indicating that more than three-quarters of students regularly engage in ITS-supported speaking practice.
Listening to English audio or dialogues	Only 4% never use ITSs for listening activities, while 16% rarely, 26% sometimes, and 54% often listen to English audio with ITS support, demonstrating strong reliance on ITSs for receptive listening skills.
Reading English texts with ITSs	The majority of students use ITSs for reading: 3% never, 11% rarely, 34% sometimes, and 52% often, showing that ITSs are widely perceived as effective tools for reading comprehension.
Writing support through ITSs	Writing assistance via ITSs is used by 39% of students often and 37% sometimes, while 19% rarely and 5% never rely on such tools, reflecting moderate-to-high integration of ITSs in writing practice.
Pronunciation improvement	7% report no improvement, 27% note slight improvement, 41% report noticeable improvement, and 25% indicate significant improvement, suggesting a generally positive perceived impact of ITSs on pronunciation.
Understanding spoken English in real-life situations	6% of respondents report no effect, 31% a slight effect, 38% a moderate effect, and 25% a strong effect, indicating that ITSs moderately support real-life listening comprehension.
Overall fluency development	6% perceive no improvement, 25% slight improvement, 44% moderate improvement, and 25% strong improvement in fluency, confirming the role of ITSs in enhancing communicative competence.

Regarding learning strategies, many students reported combining ITS practice with traditional resources such as textbooks, teacher feedback, and peer discussions. At the same time, a considerable number of respondents indicated that they critically evaluate system feedback, which suggests an active rather than passive role in learner–system communication.

An important aspect of the study was to determine whether ITSs support learner autonomy and strategic decision-making in language learning. The results show that ITSs are perceived as tools that encourage independent practice while still providing structured guidance (Table 3).

When asked about preferred improvements, the majority of respondents selected comprehensive options, emphasizing the need for more personalized exercises, higher-quality feedback for speaking and writing, and more interactive conversational practice. This finding highlights students' expectations of ITSs as dynamic communication technologies rather than static exercise generators.

These results suggest that ITSs support autonomy primarily through flexible pacing, repeated practice, and immediate feedback, while their communicative potential could be further enhanced by expanding interactive and dialogic features.

Table 2: Perceived adaptivity, usability, and strategic support of ITSs.

Survey issues	Results and analyses
Adaptation of exercises to the learner's level	5% of students report no adaptation, 24% rarely experience adaptivity, 41% sometimes, and 30% often observe task adjustment, indicating moderate but inconsistent personalization.
Usefulness of personalized feedback	Personalized feedback is viewed positively: 5% find it not helpful, 19% slightly helpful, 41% somewhat helpful, and 35% very helpful, highlighting its central role in learner–system communication.
Ease of navigation of ITS platforms	Only 3% find navigation very difficult and 15% difficult, while 48% consider it easy and 34% very easy, suggesting high usability of ITS interfaces.
Identification of weak language skills	ITSs help identify problematic skills for 74% of respondents (39% sometimes, 35% often), whereas 26% rarely or never receive such diagnostic support.
Planning and organization of learning	8% never use ITSs for planning, 20% rarely, 36% sometimes, and 36% often, demonstrating that ITSs contribute to self-regulated learning practices.
Use of ITSs outside the classroom	A large proportion of students (77%) use ITSs beyond class time (33% sometimes, 44% often), confirming their role in extending learning beyond formal instruction.
Speed of error correction	6% never experience faster correction, 25% rarely, 37% sometimes, and 32% often, indicating that ITSs moderately accelerate feedback cycles compared to traditional methods.

Table 3: Learner autonomy, communicative value, and perceived effectiveness of ITSs.

Survey issues	Results and analyses
Reliance on ITSs for practising difficult language skills	Only 6% of respondents never rely on ITSs when practising difficult skills, while 21% do so rarely. A majority of students (38%) sometimes and 35% often use ITSs for challenging aspects of language learning, indicating a substantial level of trust in ITSs as supportive learning tools.
Encouragement to use diverse learning strategies	7% of students report that ITSs never encourage strategic variation, and 21% rarely experience such encouragement. At the same time, 37% sometimes and 35% often experiment with different learning strategies (e.g., repetition, summarising, reformulation) when guided by ITSs.
Usefulness of ITSs for paired or group learning	11% of respondents never use ITS for collaborative learning and 22% rarely do so. However, 39% sometimes and 28% often employ ITSs in paired or group activities, suggesting moderate integration of ITSs into collaborative communicative practices.
Preference for independent learning supported by ITSs	A small proportion of students (5%) strongly disagree and 21% disagree that they prefer independent learning with ITS support. In contrast, 46% agree and 28% strongly agree, confirming a clear tendency towards autonomous learning mediated by intelligent systems.
Contribution of ITSs to overall learning effectiveness	6% of respondents perceive no contribution of ITSs to learning outcomes, while 19% report a slight contribution. The majority identify moderate (44%) or strong (31%) improvements, highlighting the perceived effectiveness of ITSs in enhancing language learning efficiency.
Desired improvements in ITS functionality	When asked about system improvements, 21% of students prioritised more personalised exercises, 26% requested better feedback for speaking and writing, 18% emphasised the need for more interactive conversational practice, and 35% selected all suggested options, reflecting high expectations regarding communicative and adaptive system features.

To ensure the reliability of the questionnaire, Cronbach's alpha coefficient was calculated to assess the internal consistency of the instrument. The questionnaire items were grouped into three scales corresponding to the conceptual dimensions of the study: (1) communicative interaction and language skill development, (2) adaptivity and usability, and (3) perceived instructional effectiveness.

For each scale, Cronbach's alpha was computed based on respondents' Likert-scale answers to determine the degree to which the items consistently measured the same underlying construct. The obtained values ($\alpha \geq 0.70$) indicated acceptable to good internal consistency across all scales.

Overall, the results confirm that Intelligent Tutoring Systems function as effective communication technologies in second language learning. Students actively use ITSs to practice core language skills, perceive clear benefits for fluency and confidence, and value adaptive feedback mechanisms. At the same time, the findings reveal a need for more advanced communicative scenarios, especially for speaking and writing, to better reflect authentic language use.

The quantitative data obtained from the respondents provide empirical support for viewing ITSs not only as adaptive instructional systems but also as communicative environments that mediate interaction, support learner autonomy, and contribute to sustained engagement in second language learning.

5 CONCLUSIONS

The present study demonstrates that Intelligent Tutoring Systems occupy a distinctive position within contemporary second language learning environments, operating at the intersection of adaptive instruction and digitally mediated communication. Rather than functioning solely as delivery mechanisms for linguistic content, ITSs structure the conditions under which learners interact with language through system-guided prompts, adaptive feedback, and iterative task design. These interactional mechanisms enable sustained language practice in a low-anxiety learning environment while dynamically aligning instructional input with learners' developing proficiency levels. As a result, they support the gradual and systematic development of both receptive and productive language skills.

Empirical evidence indicates that students consistently incorporate ITSs into both formal instruction and independent learning activities, particularly for listening, reading, and structured speaking and writing practice. Learners' reported gains in fluency, pronunciation, and communicative confidence suggest that ITS-mediated interaction contributes meaningfully to communicative competence. At the same time, usage patterns indicate that more complex forms of interaction remain partially limited, which positions intelligent tutoring systems primarily as supportive scaffolding tools rather than as substitutes for authentic communicative interaction with human interlocutors. The findings also emphasize the role of ITSs in shaping learner autonomy and self-regulated learning behaviour. Features such as flexible pacing, individualized task pathways, and continuous feedback encourage learners to assume greater responsibility for planning and monitoring their progress. Importantly, learner engagement with ITSs is characterized by strategic selectivity: students frequently combine system feedback with teacher input, peer collaboration, and external digital resources. This behaviour reflects the complexity of learners' digital learning ecologies and highlights the importance of coherent instructional integration rather than isolated technological adoption.

From the perspective of communication technologies, the educational effectiveness of ITSs emerges from the dynamic interplay between adaptivity and communicative design. Motivation and sustained engagement are not separate system functions but are embedded within ongoing learner-system interaction shaped by feedback framing, interface clarity, and transparency of progress indicators. While the results confirm the overall instructional value of ITSs, they also highlight limitations in dialogic depth, particularly in speaking and writing activities, where learners express a clear demand for richer conversational and interactional formats. The learners' articulated expectations for more individualized tasks and richer conversational formats point to key directions for future system enhancement.

Collectively, Intelligent Tutoring Systems are viewed as communicatively oriented learning environments that mediate interaction, foster learner autonomy, and sustain engagement in second language acquisition. Practical recommendations include integrating ITSs as complementary tools in classroom instruction, guiding students to combine system feedback with teacher and peer support, and gradually developing complex speaking and writing skills. For ITS developers, enhancements should focus on richer interaction, realistic conversational scenarios, adaptive feedback for open-ended tasks, and personalized learning paths. Hybrid models that

combine ITS-mediated learning with teacher- and peer-facilitated activities can further foster learner autonomy, communicative competence, and engagement. Future development of ITSs in language education should focus on enhancing interactive and dialogic functionalities, including more realistic conversational scenarios, improved feedback for open-ended speaking and writing tasks, and deeper personalization based on learner profiles. Further research may also explore hybrid models that integrate ITSs with peer and teacher-mediated interaction, as well as studies examining the long-term impact of ITSs on communicative competence and strategic language use.

REFERENCES

[1] C.-C. Lin, A. Y. Q. Huang, and O. H. T. Lu, "Artificial intelligence in intelligent tutoring systems toward sustainable education: A systematic review," in Smart Learning Environments, vol. 10, Art. no. 34, 2023, [Online]. Available: https://doi.org/10.1186/s40561-023-00260-y.

[2] H. Le and J. Jia, "Design and implementation of an intelligent tutoring system from the perspective of learner autonomy," in Interactive Technology and Smart Education, vol. 19, no. 4, pp. 510-534, 2022.

[3] H. I. Akyüz and M. Erdemir, "Preservice science teachers' views of a web-based intelligent tutoring system," in International Journal of Technology in Education, vol. 5, pp. 67-87, 2022, [Online]. Available: https://doi.org/10.46328/ijte.233.

[4] X. Weng and T. K. F. Chiu, "Instructional design and learning outcomes of intelligent computer-assisted language learning: A systematic review," in Computers & Education: Artificial Intelligence, vol. 3, Art. no. 100117, 2023, [Online]. Available: https://doi.org/10.1016/j.caeai.2022.100117.

[5] Norbert, "Educational software and artificial intelligence: Students' experiences and innovative solutions," in Information Technologies and Learning Tools, vol. 101, no. 3, pp. 200-226, 2024, [Online]. Available: https://doi.org/10.33407/itlt.v101i3.5479.

[6] Peña-Acuña and R. Corga Fernandes Durão, "Learning English as a second language with artificial intelligence for prospective teachers: A systematic review," in Frontiers in Education, vol. 9, Art. no. 1490067, pp. 1-7, 2024, [Online]. Available: https://doi.org/10.3389/feduc.2024.1490067.

[7] L. M. Rababah, "The impact of an intelligent tutoring system on the academic progress and success of EFL Jordanian students: A quasi-experimental study," in International Journal of Emerging Technologies in Learning, vol. 18, no. 16, pp. 71-89, 2023, [Online]. Available: https://doi.org/10.3991/ijet.v18i16.41471.

[8] J. Fernández-Herrero, "Evaluating recent advances in affective intelligent tutoring systems: A scoping review of educational impacts and future prospects," in Education Sciences, vol. 14, no. 8, Art. no. 839, 2024, [Online]. Available: https://doi.org/10.3390/educsci14080839.

[9] M.-T. Kou, H.-K. Wu, and C.-Y. Hsieh, "Analysis of learning behaviours and outcomes for students with different knowledge levels: A case study of an intelligent tutoring system for coding and learning (ITS-CAL)," in Applied Sciences, vol. 15, no. 4, Art. no. 1922, 2025.

[10] A. W. F. Kouam, "The effectiveness of intelligent tutoring systems in supporting students with varying levels of programming experience," in Discover Education, vol. 3, Art. no. 278, 2024, [Online]. Available: https://doi.org/10.1007/s44217-024-00385-3.

[11] E. Cohen and M. Glassman, "Using an intelligent tutoring system to support reading strategy acquisition," in Computers & Education: Artificial Intelligence, vol. 4, Art. no. 100162, 2023, [Online]. Available: https://doi.org/10.1016/j.caeai.2023.100162.

[12] H. Xie, H.-C. Chu, and G.-J. Hwang, "AI-driven personalized intelligent tutoring systems for vocabulary learning: A comparative study," in Interactive Learning Environments, vol. 31, no. 5, pp. 2612-2627, 2023, [Online]. Available: https://doi.org/10.1080/10494820.2022.2096365.

[13] E. T. Kusumawati, N. Solihati, and Z. Sari, "Integration of a web-based intelligent tutoring system (TuinLec) into text structure strategy to improve text strategy memory and reading comprehension skills," in International Journal of Learning, Teaching and Educational Research, vol. 24, pp. 52-71, 2025, [Online]. Available: https://doi.org/10.26803/ijlter.24.4.3.

[14] X. Huang, W. Xu, and R. Liu, "Effects of intelligent tutoring systems on educational outcomes," in International Journal of Distance Education Technologies, vol. 23, no. 1, pp. 1-25, 2025, [Online]. Available: https://doi.org/10.4018/IJDET.368420.

Pedagogical and Methodological Capabilities of Artificial Intelligence Tools in Developing Digital Literacy

Feruza Shermanova and Guli Taylakova
Department of Digital and Fundamental Sciences, Tashkent University of Applied Sciences, Gavhar Str. 1, 100022 Tashkent, Uzbekistan
Uzbekistan qolqanatov9518@gmail.com, taylakovaguli@gmail.com

Keywords: Digital Literacy, Artificial Intelligence, Educational Technologies, Higher Education Innovations, Statistics and Data Analysis, Econometric Modeling, Interactive Learning Environments.

Abstract: Artificial intelligence (AI) technologies have ushered in a new era in the development of students' digital competence in higher education. This article analyzes the pedagogical and methodological capabilities of AI tools and provides an in-depth look at their impact on students' ability to search, process, analyze information, and develop digital competence. The relevance of the study is explained by the need to develop digital competence in line with the requirements of the global labor market. A mixed methodological approach was used, and the effectiveness of AI-integrated lessons was evaluated in experimental and control groups. The quantitative phase included tests based on DigComp 2.2 indicators, while the qualitative phase included in-depth interviews with teachers and focus group discussions. The results showed that AI-based lessons had a significant positive impact on students' personalized learning paths, problem-solving skills, and motivation. The role of AI tools in enhancing personalized learning, optimizing formative assessment through real-time learning analytics, and developing a culture of digital safety was scientifically explained. At the same time, limitations such as data privacy, academic integrity, and teacher digital readiness were also highlighted. These comprehensive research results provide practical recommendations for improving national and international educational strategies aimed at effectively developing digital literacy in modern higher education systems.

1 INTRODUCTION

In recent years, the concept of "digital competence" has been recognized globally as one of the most important pedagogical, professional and social skills. The rapid development of the digital economy, new labor market requirements, as well as the expansion of distance learning in the post-pandemic period, have made this competence a vital necessity for every citizen, especially higher education students. Students should not only be able to search for or use information, but also to critically evaluate it, manage it in a safe environment, and creatively re-purpose it.

The DigComp 2.2 model (2022) developed by the European Commission defines digital competence through five main dimensions: information processing, digital communication and collaboration, content creation, security, and problem solving, interpreting it as a socio-cultural phenomenon, not just a technical one [1].

Therefore, the development of digital competence is becoming a priority for modern education. It is also recognized as one of the most important pedagogical, professional and social skills in the education system of Uzbekistan. The rapid development of the digital economy in the country, the gradual implementation of the "Digital Uzbekistan - 2030" strategy [2], as well as the expansion of distance learning after the pandemic, have brought this competence to the level of a vital necessity for every student. The Resolution of the President of the Republic of Uzbekistan, Resolution No. PD-320 dated October 30, 2025 "On additional measures to support projects based on artificial intelligence technologies" [3] and the Concept of "Digital Education" adopted in 2022 [4] require in-depth mastery of ICT in higher education institutions. These documents have made the development of students' skills in working with information, ensuring digital security, collaboration, and participation in innovative projects a priority of state policy.

Artificial intelligence (AI) technologies are taking this process to a new level. Uzbekistan has been taking consistent steps in AI in recent years: in 2021,

the “National Strategy for the Development of Artificial Intelligence for 2021-2030” was approved by Presidential Decree, and in 2022, the Center for Artificial Intelligence was launched in Tashkent. Along with research, these centers are engaged in the integration of AI into the education system and the implementation of practical projects in collaboration with higher education institutions. As a result, a number of universities - including Tashkent State Pedagogical University, Urgench State Pedagogical Institute, and Tashkent University of Applied Sciences - are piloting AI-based learning platforms, such as adaptive testing systems, intelligent chatbots, and learning analytics modules. However, this process also raises a number of pressing issues.

Ensuring the confidentiality of student data, the transparency of algorithmic decisions, the level of preparation of teachers for the effective use of AI technologies, and the integration of AI tools into the didactic process require a special scientific approach. Therefore, the integration of AI in Uzbekistan needs to be carefully planned not only from a technological perspective, but also from a pedagogical, psychological, social, and legal perspective.

2 METHODOLOGY

One of the important tasks facing the modern education system, and in particular higher education institutions, is to prepare students for successful functioning in the digital age. In this process, the digital learning environment is one of the key factors. It is not enough to understand it simply as consisting of computers and the Internet; the digital learning environment is a much broader and more complex system that includes the integration of information resources, communication tools, pedagogical technologies, and management mechanisms.

The system of stages presented in the Figure 1 reflects the evolutionary development of students in the digital environment. Although the fourth stage - “Digital literacy” - has the same name as the general process, it conceptually serves as the practical implementation point of the entire system. That is, while the first three stages (information sorting and security) form skills of a mainly theoretical and protective nature, the fourth stage directly moves on to the effective use of digital devices and applications for specific purposes. Therefore, this stage should be interpreted as the central link of the entire model. Only after reaching the fourth level, the student begins to form higher-level competencies such as creating multimedia content (stage 5) and managing information in compliance with the rules of digital culture (stage 6).

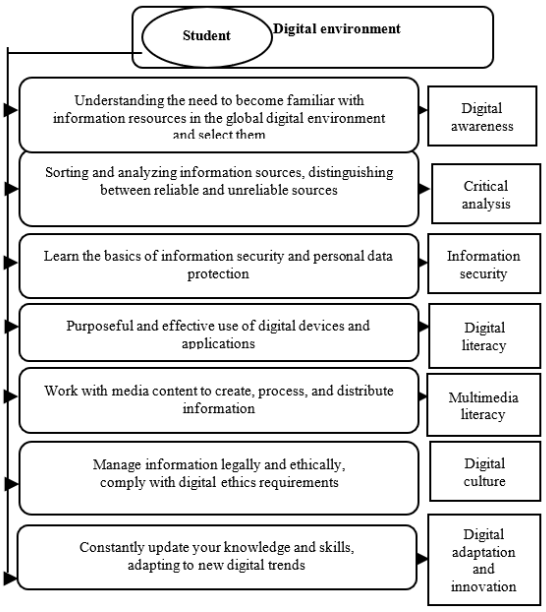


Figure 1: Stages of developing a student's digital literacy.

The final stage requires constant updating and innovative adaptation, transforming digital literacy from a static state into a dynamic skill.

Today, the role of artificial intelligence (AI) technologies in education in further improving this system is unparalleled. Artificial intelligence-based systems allow you to identify the individual learning needs of students, analyze their level of knowledge in real time, and offer customized educational content. This will take the formation of competencies such as digital awareness, critical analysis, information security and digital literacy to a new level, as shown in the Figure 1. The rapid and widespread use of artificial intelligence technologies in world practice and the availability of such digital information in our country are today’s requirements, ensuring high-quality use in life, creating favorable conditions for training qualified personnel in this field.

When writing the article, the methods used were: study and analysis of scientific and methodological literature on artificial intelligence tools in developing students' digital literacy in higher education institutions, questionnaires, interviews, pedagogical observation, synthesis and design, compilation of analytical tables, experimental testing, as well as statistical and mathematical analysis of data in accordance with the goals and objectives of the research.

3 LITERATURE ANALYSIS

The scope of using artificial intelligence (AI) tools in developing students' digital literacy in higher educational institutions of Uzbekistan and foreign countries is expanding. This process is not only the result of technological progress, but also an integral part of the modern pedagogical and information environment. Today, the digital educational environment, integrated with artificial intelligence, is taking students' opportunities for knowledge acquisition, analysis and creative activity to a new level. In this regard, the formation of digital literacy in students is not only a requirement of the time, but also a strategic task of the continuous education system.

The criteria and requirements developed for identifying, developing, and assessing students' digital potential in modern education have been reflected in international and national research. In particular, the approach put forward by Zhukova N.A. emphasizes that the formation of digital literacy is carried out through a system of basic competencies, in which the skills of searching, analyzing, processing, and critically evaluating information play a key role [5]. These competencies will be further improved in combination with artificial intelligence tools, as AI technologies allow for the rapid processing of large amounts of data, the extraction of important information, and the provision of individual recommendations.

The concept developed by Douglas A.J. Belshaw interprets digital literacy as a system of eight interrelated elements. Each component of this model - cultural, cognitive, constructive, communicative, trustworthiness, creativity, criticality, and citizenship - takes on new meaning in a digital environment enriched with artificial intelligence [6]. For example, AI-based platforms support students' cognitive performance, making it easier to understand and analyze complex information, while expanding the possibilities for creating new content through creative tools. However, in this process, the importance of critical and civic competences will increase even more, as assessing the reliability and ethical aspects of information generated by AI becomes an important task.

The “capabilities - skills - practical activities - personality” model proposed by R. Sharp and H. Bifham also receives a new interpretation in the context of artificial intelligence [7]. AI technologies initially create broad opportunities for students, and then, based on these opportunities, the necessary digital skills are formed, which are used in practical activities and ultimately determine the digital identity and competence of the individual. Thus, artificial intelligence acts as a catalyst in the process of developing digital literacy. At the same time, modern research shows that it is wrong to limit digital literacy only to technical skills. According to the definition given by V.A. Sukhomlin, E.V. Zubareva and A.V. Yakushin, digital skills are a dynamic, contextual and inextricably linked system of professional activity [8]. This approach is especially relevant in the context of artificial intelligence, since working with AI tools requires not only technical knowledge, but also analytical thinking, flexibility and ethical responsibility.

The rapid development of artificial intelligence technologies, while increasing efficiency in the educational process, also poses certain risks and challenges. As Whittaker et al. (2018) point out, the ethical implications of automation through AI may not be sufficiently taken into account [9]. Therefore, it is necessary to strictly adhere to the principles of fairness, transparency and accountability in the use of AI tools.

In recent years, the rapid growth of AIED (Artificial Intelligence in Education) research has further strengthened the importance of AI in education. Research findings show that AI technologies have a positive impact on learning outcomes, expanding scholarly activity, and increasing interdisciplinary integration.

The introduction of artificial intelligence (AI) technologies into the education system is shaping new scientific approaches to developing students' digital literacy [10]. First of all, empirical research shows that educational processes organized on the basis of SI have not yet fully transitioned to a competency-based approach [11]. In particular, a study based on the analysis of 98 lessons found that the majority of SI lessons were conceptual (50%) and partially experimental (31.63%), while the development of higher-level cognitive skills was observed in only 35.71% of cases. [11] It was also found that the ethics of SI were very poorly covered (5.1%) in this study [11].

Other theoretical perspectives interpret SI as a paradigm-shifting factor in education [12]. According to this approach, SI increases the possibility of individualizing education, increasing flexibility, and making informed decisions [12]. At the same time, researchers emphasize the priority importance of critical thinking and ethical competencies in the process of using SI tools [12].

Practical pedagogical research suggests new methodological approaches to the use of SI tools. In

particular, research based on the “reperformance” and “critical error review” models has shown that students achieve deeper learning through independent verification and analysis of the results generated by SI [9]. This approach develops high-level competencies in students [9].

At the same time, modern research also emphasizes that the effectiveness of SI technologies is directly dependent on the digital and pedagogical competencies of teachers. [13] Empirical research on the use of generative AI tools shows that teachers' level of AI literacy and understanding of its pedagogical value determines their intention to introduce these technologies into the educational process [13].

However, there are also risks associated with the use of AI. It is no coincidence that the European Union's "AI Law" classifies AI systems used in education as "high risk". Because such systems, if misused, could undermine the right to education, increase discrimination or deepen existing social inequalities.

Therefore, the use of artificial intelligence in the process of developing students' digital literacy requires an integrative approach. That is, along with technological capabilities, ethical, social and legal aspects must also be taken into account. As a result, the digital literacy model based on artificial intelligence in the higher education system ensures not only the professional competence of students, but also their formation as responsible, critical, and innovative thinkers.

4 RESULTS AND DISCUSSION

The results of the study confirmed the effectiveness of the savodxonmedia.uz e-learning platform, supported by artificial intelligence, in developing students' digital literacy. A comparison of the results of the experimental and control groups showed that the AI-integrated learning environment significantly increases not only students' technical skills, but also their analytical and critical thinking competencies.

The main innovation of the proposed platform is that, unlike traditional learning management systems, it creates a special didactic environment focused exclusively on the development of digital literacy and related competencies. The platform organizes educational materials for the subjects “Application of Information Technologies in Professional Activities” and “Media Literacy and Information Culture” in higher pedagogical educational institutions of Uzbekistan on the basis of a single integrated system.

This integration ensures the continuity of the educational process and allows students to comprehensively master theoretical knowledge and practical skills. The results of the study showed that classes organized on the basis of interdisciplinary integration significantly develop students' skills in working with information, analyzing it, and evaluating it.

One of the important components of the platform is the AI-based recommendation module. This module generates individual recommendations based on the user's:

- learning activity;
- learning results;
- history of activity in the system.

The results of the study showed that AI-based recommendations:

- increased students' learning motivation;
- developed self-assessment and reflection skills;
- enhanced the effectiveness of independent learning.

In particular, it was found that the role of the AI module is important in helping students analyze their level of knowledge and identify areas for further development. This, along with automating the learning process, serves to develop metacognitive skills.

The platform's main functional digital tools are integrated into a single ecosystem. This system includes the following components:

- learning content management module (lectures, slides, practical exercises);
- assessment and monitoring system;
- AI-based recommendation module;
- learning analytics tools that analyze user activity.

The integration of these components ensures the continuity and efficiency of the learning process. Based on the data obtained through learning analytics tools, teachers have the opportunity to monitor the dynamics of student development and adjust learning strategies. In order to determine the level of effectiveness of the proposed methodology within the framework of scientific research, a pilot-testing process will be organized. 450 1st and 2nd year students of higher pedagogical educational institutions of the Republic of Uzbekistan in Tashkent and Urgench were involved in the pilot-testing work in 2024-2025. In the process of implementing the pilot-testing work of the research, the following regulatory documents were compiled and educational and methodological resources were developed:

1) Educational program, curriculum for the subjects "Application of information technologies in professional activities" and "Media literacy and information culture" of higher educational institutions of pedagogical sciences.
2) Instructions for carrying out theoretical and practical work used in AI-based education, practical assignments, audio and video manuals, etc.
3) Control and test tasks used to assess the acquired knowledge of learners.
4) Mathematical statistical methods and questionnaires to determine the effectiveness of AI-based education.

The pilot study was carried out in three stages: foundational, educational and explanatory:

1) At the foundational stage, the pilot study sites were identified; the number of respondents was determined; experimental and control groups were formed; the pilot study was organized; a pilot study program was developed.
2) At the educational stage, pilot testing was carried out; preliminary examination of the effectiveness of the developed methodology; based on the results of the experiment, work was carried out to enrich the content of lecture texts, practical and laboratory assignments to improve the existing methodological support.
3) At the clarifying experimental stage, during the pilot testing, the results of teaching the subjects "Application of Information Technologies in Professional Activities", "Media Literacy and Information Culture" using an AI-based platform were summarized, the conclusions obtained were practically verified, and these results were analyzed using the Student-Fisher T-criterion mathematical and statistical method

For the effective use of mathematical-statistical formulas, the mastery levels in the experimental group are represented as x_i , the corresponding numbers of learners are denoted by n_i and the statistical probabilities as $p_i = \frac{n_i}{N}$ and similarly, the corresponding variables for the control groups as y_i, m_i va $q_i = \frac{m_i}{M}$ are denoted by. The results of the mastery in the experimental group were conditionally divided into two statistical samples, and the arithmetic mean values, dispersion and variation coefficients of the samples were determined for each of the two samples.

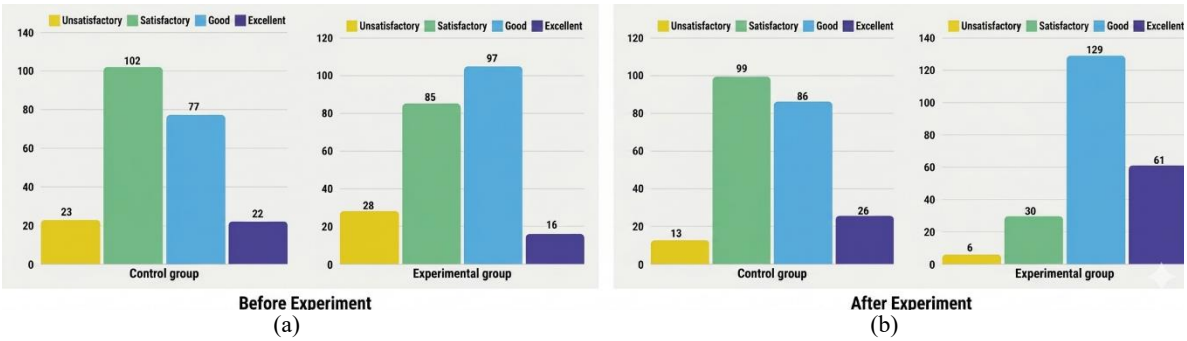
The learning dynamics of the 450 students who participated in the experimental test at the beginning and end of the experiment are presented in the figure below (see Table 1 and Fig. 2).

As can be seen from Figure 2, it can be predicted that the sample modal values for the experimental and control classes satisfy the conditions $\bar{X} > \bar{Y}$ for the experimental group, respectively. They are calculated using this formula:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} n_i x_i = \frac{1}{226}(5 * 61 + 4 * 130 + 3 * 29 + 2 * 6) = 4{,}08.$$

Table 1: Student achievement indicators.

	Experiment group i					Control group				
	X_i	5	4	3	2	Y_j	5	4	3	2
Experience first	n_i=226	16	87	95	28	n_j=198	22	77	102	23
The experience is the last.	n_i=198	61	130	29	6	n_j=198	26	86	105	10



(a) (b)

Figure 2: Diagrammatic representation of the results of experimental and test work: a) Initial experience, b) After the experiment.

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} n_i y_i = \frac{1}{224}(5 * 26 + 4 * 86 + 3 * 105 + 2 * 10) = 3{,}55.$$

When looking at the values, it is clear that the average achievement in the experimental group is higher than the results of the control class: $\overline{X}>\overline{Y}$.

$$\frac{\overline{X}}{\overline{Y}} = \frac{4{,}08}{3{,}55} = 1{,}15.$$

The experimental results confirmed the effectiveness of using artificial intelligence-based learning platforms in the educational process and showed an increase of 15%.

The study revealed that students have different levels of digital literacy. Based on this, a mechanism for forming individual learning trajectories was introduced on the platform. The results showed that:

- basic digital skills were developed for elementary level students;
- intermediate level students developed information processing and systematization skills;
- advanced level students deepened their competencies in critical evaluation of media content and safe use.

This confirms that the platform effectively implements differentiated and person-centered learning.

The ethical deployment of AI within the platform is governed by three core principles: data minimization, which ensures students' identities are decoupled from performance logs to avoid bias; algorithmic transparency, providing students with awareness of AI-generated insights to foster independent decision-making; and bias mitigation, achieved through regular audits that prevent “filter bubbles” and ensure dynamic access to learning materials for all students regardless of their initial performance level.

The empirical findings of this study further distinguish the proposed system from existing educational solutions. A comparative analysis shows that while global platforms like Coursera or Khan Academy utilize AI for general content recommendation, savodxonmedia.uz offers a more specialized didactic environment specifically mapped to the national pedagogical standards of Uzbekistan. Unlike generic AI tutors, our platform’s AI-based recommendation module is uniquely integrated with a dual-subject framework (IT Application and Media Literacy), allowing for cross-disciplinary competency mapping that is often missing in broader e-learning systems. This context-specific integration constitutes the primary novelty of the proposed system, as it bridges the gap between general digital skills and local academic requirements.

5 CONCLUSIONS

In conclusion, the savodxonmedia.uz e-learning platform, integrated with artificial intelligence tools, has demonstrated high pedagogical effectiveness in developing students' digital literacy. In particular, the implementation of individual learning trajectories, the use of an AI-based advisory module, and the presentation of educational content in an integrated didactic system have significantly improved students' problem-solving, critical thinking, information analysis, and independent learning competencies. Also, real-time monitoring of the learning process through learning analytics tools has proven to be an important factor in improving the quality of education and ensuring an individual approach.

The integration of AI tools within the platform is methodologically grounded in their ability to provide predictive scaffolding for specific subjects, such as identifying technical gaps in IT applications and analyzing media evaluation patterns. Furthermore, the ethical deployment of these tools is strictly managed through data minimization to protect student identities, algorithmic transparency to prevent over-reliance on AI, and periodic bias audits. These measures ensure that the AI-driven recommendation engine serves as a pedagogically sound and responsible instrument for personalized learning.

At the same time, the research results showed that the effectiveness of an artificial intelligence-based learning environment largely depends on the methodological approach and digital competence of teachers. The widespread use of AI tools in the educational process raises the need to address issues such as data security, academic integrity, and pedagogical oversight. Therefore, in the future, it is important to develop not only technological, but also didactic and ethical aspects in an integrated approach to improving educational platforms aimed at developing digital literacy.

REFERENCES

[1] R. Vuorikari, S. Kluzer, and Y. Punie, DigComp 2.2: The Digital Competence Framework for Citizens. Luxembourg: Publications Office of the European Union, 2022, doi: 10.2760/115376.

[2] Decree of the President of the Republic of Uzbekistan No. PD-6079 dated 05.10.2020 "On approval of the Digital Uzbekistan - 2030 strategy and measures for its effective implementation," [Online]. Available: https://lex.uz/ru/docs/-5030957.

[3] Resolution of the President of the Republic of Uzbekistan No. PD-320 dated 30.10.2025 "On additional measures to support projects based on artificial intelligence technologies," [Online]. Available: https://lex.uz/uz/docs/-7804375.

[4] Cabinet of Ministers of the Republic of Uzbekistan, Digital Education Concept. Tashkent, 2021.

[5] N. A. Zhukova, "Requirements for basic digital competencies of students for the formation of digital literacy in computer science lessons," in Proc. II All-Russian Scientific and Practical Conference, Cheboksary, June 5, 2023.

[6] D. A. J. Belshaw, The Essential Elements of Digital Literacies. Self-published via Open Badges, 2014, ISBN: 978-0-9932488-0-2, [Online]. Available: https://digitalliteraci.es/.

[7] R. Sharpe and H. Beetham, "Understanding students' uses of technology for learning: towards creative appropriation," in R. Sharpe, H. Beetham, and S. de Freitas, Eds., Rethinking Learning for a Digital Age: How Learners are Shaping Their Own Experience. London: Routledge, 2010, pp. 85-99.

[8] V. A. Sukhomlin, E. V. Zubareva, and A. V. Yakushin, "Methodological aspects of the concept of digital skills," Modern Information Technologies and IT Education, vol. 13, no. 2, pp. 146-152, 2017.

[9] D. Wiseman and D. Foster, "Bridging AI with cost accounting education: A pedagogical framework using reperformance and critical error review," Journal of Accounting Education, vol. 72, 2025.

[10] B. Agarwal, C. Urlings, G. van Lankveld, and R. Klemke, "Identifying the ethical values and norms for artificial intelligence in education: A systematic literature review," International Journal of Artificial Intelligence in Education, vol. 36, no. 1-2, p. 100004, 2026, [Online]. Available: https://doi.org/10.1016/j.ijaied.2026.100004.

[11] D. Wu, M. Chen, X. Chen, and X. Liu, "Analyzing K-12 AI education: A large language model study of classroom instruction on learning theories, pedagogy, tools, and AI literacy," Computers and Education: Artificial Intelligence, vol. 7, 2024, [Online]. Available: https://doi.org/10.1016/j.caeai.2024.100295.

[12] F. O. Socorro Márquez, G. E. Reyes Ortiz, and H. Torrez Meruvia, "Philosophical leadership in education: Rethinking pedagogy in an AI-driven world," Social Sciences & Humanities Open, vol. 13, p. 102699, 2026, [Online]. Available: https://doi.org/10.1016/j.ssaho.2026.102699.

[13] N. A. Jogezai, M. Jaffar, F. A. Baloch, and G. K. Khilji, "University teachers' behavioral intentions of generative AI instructional use: The effects of AI literacy, GAI pedagogical relevance and organizational AI guidelines," International Journal of Information and Learning Technology, vol. 42, no. 5, pp. 409-425, 2025, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2056488025000186.

Application of Immersive Technologies in Teaching Business English in Higher Education Institutions for the Formation of Situational Speech

Larysa Soloviova[1], Irena Snikhovska[2] and Zoia Zalibovska-Ilnitska[3]
[1]*Foreign Languages Department, Polissia National University, Staryi Blvd. 7, 10008 Zhytomyr, Ukraine*
[2]*Department of Natural and Social-Humanitarian Disciplines, Zhytomyr Medical Institute of Zhytomyr Regional Council, Velyka Berdychivska Str. 46/15, 10002 Zhytomyr, Ukraine*
[3]*Department of Social Technologies, Zhytomyr Ivan Franko State University, Velyka Berdychivska Str. 40, 10008 Zhytomyr, Ukraine*
larysasolovyova23@gmail.com, snikhovska.irena@gmail.com, zoyazalibovskaya73@gmail.com

Keywords: Immersive Technologies, Business English, Virtual Reality (VR), Augmented Reality (AR), Situational Speech.

Abstract: The integration of immersive technologies, such as virtual reality (VR) and augmented reality (AR), into higher education has become a highly relevant and timely topic in the context of modern education and the evolving demands of the global labor market. Proficiency in business English is increasingly crucial for career success, yet traditional teaching methods often fail to provide sufficient practice in real-life communication contexts. This paper aims to explore the effectiveness of immersive technologies in enhancing business English instruction, particularly in the development of situational speech and communication skills in business settings. The research employs a mixed-methods approach, combining qualitative and quantitative techniques such as surveys, interviews, and observations, as well as an experimental phase where students use VR, AR, and 360° content to practice business English. The findings indicate that immersive technologies significantly improve students' language skills by providing a platform for interactive learning in realistic business scenarios. Despite some initial difficulties with adapting to new learning formats, students reported higher levels of engagement, improved confidence, and better preparedness for real-world business communication. The results have both practical and theoretical relevance. Theoretically, they contribute to the understanding of how immersive technologies can reshape language education, while practically, the findings offer meaningful insights for educators looking to integrate these tools into their curricula. Further research is expedient to assess the long-term impact and optimize the integration of immersive technologies in language learning.

1 INTRODUCTION

The integration of immersive technologies in higher education is a pivotal aspect of modern learning, offering new opportunities for interactive and engaging educational experiences. These technologies, including virtual and augmented reality, enable students to immerse themselves in realistic scenarios that enhance both their motivation and interest in learning. By simulating real-world environments, immersive technologies provide a platform for the practical application of theoretical knowledge, which is increasingly essential in a globalized world and a rapidly evolving job market.

Moreover, the adoption of immersive technologies fosters the development of critical thinking, creativity, and communication skills. With interactive learning elements, students engage more actively with course material, resulting in a deeper and more meaningful understanding of the subject matter. This is particularly beneficial in areas such as business English, where the focus extends beyond grammar and vocabulary to include the ability to apply language skills in real-world business contexts. As such, immersive technologies are becoming an essential tool in preparing students to be competitive professionals in an increasingly complex and dynamic workforce.

The use of immersive technologies in teaching business English has become increasingly crucial in today's globalized world, where English proficiency is a key skill for career success. Technologies like virtual and augmented reality offer students unique opportunities to engage with real-world business communication scenarios. This not only enhances their language skills but also promotes the development of critical thinking, creativity, and adaptability in various communication contexts.

In the Ukrainian context, [1] explore immersive learning in university's educational and scientific activities, revealed a need for an integrated approach for introducing immersive learning. The researchers propose "immersive institute" model which can be implemented both at the level of the university in general and at the level of its educational and scientific departments.

The relevance of this topic is further underscored by the growing need for innovative teaching methods that align with the demands of the modern labor market. Traditional approaches often lack sufficient practice opportunities for developing situational speaking skills, which are vital in business settings. Immersive technologies address this gap by creating an interactive environment that closely mirrors real-life situations, allowing students to refine their communication skills with greater confidence and better prepare for professional challenges.

2 LITERATURE REVIEW

A literature review on the use of immersive technologies in teaching Business English highlights a growing interest in virtual and augmented reality as powerful tools for enhancing language skills. The researchers [2], in their systematic review, examine the impact of augmented reality on language learning, emphasizing its potential to boost student motivation. Further, [3] expound upon the efficacy of augmented reality (AR) in language acquisition.

Utilizing the PICO framework (population, intervention, comparison, and outcomes), the study examines linguistic and affective gains as outcome variables to assess the impact of AR on students' language learning performance, alongside the influences of educational levels, types of AR display devices, and specific language competencies. The findings indicate that AR possesses the capacity to significantly enhance linguistic gains, while AR-based learning materials prove advantageous in augmenting affective gains, including motivation, attitudes, and self-efficacy. In this light, [4] demonstrated that the inductive approach supported by AR has significant potential in strengthening the education quality and is highly relevant for application in 21st-century education. Next, [5] undertook a series of feasibility and effectiveness tests showing the effectiveness of AR-based environmental economics learning media developed via the Sketchfab platform, in improving students' creative thinking skills. It is worth noting that [6] provide a comprehensive analysis of how 360-degree videos can be utilized in language learning. Likewise, delves into the potential of 360-degree virtual reality videos [7] and real VR for education. On the other hand, [8] examine the discrepancy between the technical progress of VR and the slow evolution in the development of suitable evaluation methods, examining the limitations of new VR applications as well as identifying potential solutions to enable a more appropriate assessment of the user experience, the impact of VR in different application areas.

Furthermore, in the context of Business English, interviewing skills occupy a central position in the job application process, yet professional development aimed at enhancing these skills remains a promising area of research. To improve entry into the job market, there is an imperative to develop automated interview simulation platforms. In this light, when it comes to educational applications, [9] review recent advancements in the educational applications of AI in simulation-based learning, in particular application of AR in education and offer suggestions for future research directions. In addition, [10] discuss the role of VR-assisted language learning in foreign language teaching. Similarly, [11] examine the potential of immersive environments for enhancing storyline visualizations, exploring the use of immersive VR environments. Additionally, the studies by [12], [13] explore the role of simulation classroom to improve situational English learning.

Overall, the literature reveals a significant and growing interest in immersive technologies as effective tools for enhancing language learning, particularly within the context of Business English.

3 METHODOLOGY

To attain the set objective, a range of theoretical research methodologies were employed: a comprehensive content analysis of contemporary scholarly works regarding the application of immersive technologies in education. This analysis facilitated the elucidation of key concepts such as "immersive technologies," "virtual reality,"

"augmented reality," among others, as well as the advantages of utilizing immersive technologies and augmented reality in language education. The comparative method was employed to evaluate the characteristics of VR and AR technologies concerning their accessibility and practicality for implementation in foreign language instruction in Ukraine. Additionally, a comparative analysis of AR applications was conducted. The modeling method was utilized to construct a framework for the integration of AR in foreign language classrooms. Furthermore, the analysis and generalization method was applied to formulate practical recommendations for educators regarding the integration and application of immersive technologies in foreign language instruction, while also identifying potential barriers to the adoption of this technology in Ukraine and exploring avenues for overcoming such challenges. So, this study employed a mixed-methods research design, combining quantitative and qualitative approaches to ensure a comprehensive evaluation of the impact of immersive technologies on students' situational speech development.

3.1 Research Design and Procedure

The research methodology comprises several interconnected stages. Modeling the process of using immersive technologies involves structuring the stages of the lesson with the integration of immersive technologies technologies. We propose a model that describes the sequence of steps and methods for using immersive technologies in the educational process. The research consisted of three stages, namely: Stage 1 Pre-experimental assessment of students' situational speaking skills; Stage 2 Intervention, during which immersive technologies (AR, VR, and 360° environments) were integrated into Business English instruction over 8 weeks; Stage 3 Post-experimental assessment and comparative analysis.

Pre-experimental assessment stage comprised the following actions: analysis of the educational program, identifying topics that can be enhanced through the use of immersive technologies (for example, studying vocabulary, grammar, cultural aspects, etc.); selection of appropriate applications of immersive technologies or platforms that match the level of preparation and age characteristics of users; ensure the availability of necessary technical equipment (smartphones, tablets), pre-install and configure the required applications.

Intervention stage included creating immersive technologies tasks, such as interactive quizzes, quests, or situation modeling; adapting immersive technologies material to different levels of learners' knowledge. At the lesson, in the introductory part, explaining the purpose of using immersive technologies, demonstrating how the technology works. The instructions for using the application are provided. In the practical part, utilizing immersive technologies for active learning, for example, performing pronunciation exercises or exploring cultural features of a specific English-speaking country; at the feedback stage, it is advisable to conduct brief discussions and analysis after completing immersive technologies tasks, identifying certain difficulties.

Post-experimental assessment and comparative analysis, devoted to using immersive technologies quizzes or testing as part of knowledge assessment. This can be both formative and summative assessment – for example, creating a project with immersive technologies. During reflection, which took the form of a short group survey, it was discussed with learners how the use of immersive technologies has impacted their study. This stage also included analyzing the effectiveness of immersive technologies in education, considering the success of learners, their interest, and the level of material comprehension; drawing on the results obtained, make adjustments to the use of immersive technologies, improve methodological techniques.

3.1.1 Participants

The study involved 120 students from three Ukrainian higher education institutions: Polissia National University, Zhytomyr Medical Institute, and Zhytomyr Ivan Franko State University. The majority of the sample consisted of 72 undergraduate students from Polissia National University and Ivan Franko Zhytomyr State University, who accounted for 60%, followed by 48 master's students from Zhytomyr Medical Institute (40%). In the first cohort were undergraduates in the specialties Management, Marketing, International Economic Relations, the second cohort were those enrolled in master's programs in such specialties as Public Health as well as Nursing. Participants were divided into: EG (n=60) engaged in immersive learning activities; CG (n=60) studied using traditional methods. The distribution of participants across institutions and study levels is presented in Figure 1.

3.1.2 Instruments

To ensure validity and reliability, the following tools were used: a standardized situational speaking test, assessed using CEFR-based criteria (fluency,

accuracy, interaction, coherence); a validated questionnaire (Cronbach's alpha = 0.87), measuring student perceptions of engagement, motivation, and perceived skill improvement; observation checklists for classroom interaction. Situational speech was evaluated using four indicators: communicative effectiveness; lexical appropriateness; grammatical accuracy; interactional competence. Each criterion was scored on a 5-point scale.

3.2 Data Analysis

The collected data were analyzed using: paired sample t-tests (to compare pre- and post-test results within groups); independent sample t-tests (to compare experimental and control groups); one-way ANOVA (to compare effectiveness across AR, VR, and 360° tools); effect size calculation (Cohen's d). Statistical significance was set at $p < 0.05$.

4 RESULTS AND DISCUSSION

A comparison of immersive technologies in education, such as augmented reality (AR), virtual reality (VR), and 360° content, in teaching Business English in higher education institutions for the development of situational speech can be made according to several criteria: 1. Interaction and immersion. AR allows virtual elements to be integrated into the real world, enabling students to interact with virtual objects. This enriches the learning experience by showing business terms in real business meetings. In contrast, VR provides a fully immersive environment, where students can engage in realistic business scenarios. This level of immersion helps to develop speaking skills by making students feel part of the situation. 360° content offers a panoramic view of virtual environments. It allows students to explore business settings, aiding to visualize communication contexts. 2. Training opportunities. AR offers interactive tasks, including games, quizzes, and contextual exercises, which can be highly effective for developing situational speech. VR, on the other hand, enables students to simulate real business meetings, providing an environment that closely mirrors actual communication scenarios. Role-playing exercises, where students assume various business roles, help develop practical skills. 360° content is primarily used to explore cultural aspects of business through virtual tours of offices or international business settings. This aids in understanding the context of business communication across different cultures. 3)Technical requirements. AR is relatively accessible, as it only requires a smartphone or tablet with a camera and the appropriate apps, making it convenient for most students. VR, however, demands special equipment such as VR glasses, which can be more expensive and less readily available in some educational institutions. 360° content can be viewed on standard computers or mobile devices, although VR glasses are recommended for a more immersive experience. 4) Effectiveness in learning. AR is particularly effective for teaching business terms and phrases, as it helps contextualize the material. VR stands out as the most effective technology for developing situational speech, allowing students to practice real communication scenarios in a fully immersive environment. While 360° content is less interactive, it still offers significant value, especially in learning about the cultural context of business communication, which is crucial for students working in international settings. That being said, it can be noted that each of these immersive technologies offers distinct advantages in the teaching of Business English. AR is ideal for interactive term learning, VR provides deep immersion into business situations, and 360° content is useful for studying the cultural context. The choice of technology ultimately depends on the learning objectives, available resources, and the specific needs of the students.



Figure 1: Distribution of research participants.

In terms of the impact on situational speech (empirical evidence), the results demonstrate a statistically significant improvement in situational speaking skills among students exposed to immersive technologies: Experimental group mean score increased from 2.8 to 4.1, Control group improved

from 2.9 to 3.3. A paired t-test showed that improvement in the experimental group was significant ($t = 6.45$, $p < 0.001$), while the control group showed only moderate gains ($t = 2.12$, $p < 0.05$). The effect size for the experimental group was large (Cohen's d = 0.82). These findings provide empirical confirmation that immersive technologies significantly enhance situational speech, addressing the previously descriptive nature of the study.

As far as comparative effectiveness of AR, VR, and 360°is concerned, a one-way ANOVA revealed statistically significant differences between technologies ($F(2,57)=5.73$, $p < 0.01$): VR (Mean = 4.3) – highest effectiveness in situational speech, AR (Mean = 3.9) – strong vocabulary and contextual gains; 360° (Mean = 3.6) – moderate improvement, mainly cultural/contextual. Post-hoc analysis (Tukey test) confirmed that VR significantly outperformed 360° tools, while differences between AR and VR were smaller but still meaningful. This quantitatively supports the earlier descriptive claims.

The selection of immersive technology-based exercises, along with their pedagogical purposes, is summarized in Table 1. Exercises that incorporate immersive technologies, such as AR, VR, and 360° content, can significantly enhance the teaching of business English in higher education institutions, especially for developing situational speech. Business process simulations in VR offer another engaging exercise, where students make decisions in various business situations, such as project management or negotiations. This helps develop critical thinking and communication skills under pressure. 360° excursions, where students take virtual tours of offices, business centers, also provide valuable learning opportunities. These tours allow students to observe real work processes and understand the cultural aspects of business while improving their speech skills in relevant contexts.

Role-playing exercises in VR let students take on virtual characters in business scenarios, such as job interviews, presentations, or business negotiations. This type of practice helps students improve their situational speech in different professional settings. Similarly, AR games that challenge students to find virtual objects or business terms in the real world are an engaging way to enrich vocabulary and boost communication skills. Virtual presentations, conducted in a VR environment, give students the chance to prepare and deliver presentations to an audience, thus building public speaking skills and confidence in communication. Speech analysis with AI technology provides feedback on students' pronunciation and speaking abilities when performing tasks in VR or AR environments, helping them refine their spoken English. Exercises that incorporate immersive technologies (see Table 1) can significantly enhance the teaching of Business English, especially in situational speech.

Thus, these exercises, utilizing immersive technologies, can greatly improve the effectiveness of business English instruction and contribute to the development of situational speech skills in students, preparing them for real-world business communication. While students acknowledged that working with immersive technologies in Business English instruction presented some difficulties, their feedback highlights the recognition of these tools' importance and effectiveness. Hence, 74 students (61,7% of all participants) reported initial struggles in adapting to new learning formats, such as virtual reality or interactive simulations, which required time and effort to master. Further, 18 students (15%) claimed that immersive technologies enriched their learning experience by offering the chance to practice language skills in environments that somewhat resemble real business situations, while 102 students (85%) emphasized that immersive technologies greatly enriched their learning experience by offering the chance to practice language skills in environments that closely resemble real business situations.

Thus, such exercises utilizing immersive technologies can greatly improve the effectiveness of business English instruction and contribute to the development of situational speech skills in students, preparing them for real-world business communication. While students acknowledged that working with immersive technologies in Business English instruction presented some difficulties, their feedback highlights the recognition of these tools' importance and effectiveness. Hence, 74 students (61,7% of all participants) reported initial struggles in adapting to new learning formats, such as virtual reality or interactive simulations, which required time and effort to master. Further, 18 students (15%) claimed that immersive technologies enriched their learning experience by offering the chance to practice language skills in environments that somewhat resemble real business situations, while 102 students (85%) emphasized that immersive technologies greatly enriched their learning experience by offering the chance to practice language skills in environments that closely resemble real business situations.

Table 1: The choice of immersive technologies for exercises, their description and purpose.

Exercise	Description	Purpose
Virtual Business Meetings (VR)	Virtual meetings, discuss business projects	Develop communication skills in realistic business scenarios
Interactive AR Cards	AR cards trigger virtual objects or dialogues	Deepen knowledge of business English terms and phrases
Business Process Simulations (VR)	Students work in simulations, of business situations	Develop critical thinking and communication skills
360° Excursions	Virtual tours of business centers	Understand cultural aspects of business in relevant contexts
Role-playing (VR)	Job interviews, negotiations	Practice situational speech in different business contexts
AR Games to Learn Terms	AR games (find virtual objects or terms in the real world)	Enrich vocabulary and improve communication skills
Virtual Presentations (VR)	Virtual presentations in a VR environment	Develop public speaking skills and confidence in interaction
Cross-Cultural Simulations	Simulations of cultural contexts	Understand cultural nuances
Interactive Video Scripts	Interactive videos of business situations	Develop decision-making,situational speech skills
Speech Analysis with AI	Apps provide feedback on VR or AR speaking tasks	Improve skills of pronunciation and speaking

This contrast between the challenges and positive outcomes underscores the potential of immersive technologies in shaping situational speech. 112 students (93,3%) noted that, in addition to studying the theoretical aspects of language, these technologies enabled them to actively apply their knowledge in practical contexts, significantly boosting their confidence in communication. Therefore, immersive technologies prove to be a highly effective tool for developing language skills in a business context.

Moreover, the integration of immersive technologies into business communication instruction not only addresses language proficiency but also enhances critical thinking and problem-solving skills vital for professional success. A majority of participants, 90 students (75%), indicated that immersive environments encouraged them to adapt their communication strategies to varying scenarios. This adaptability is essential in the fast-paced business world, where circumstances can change rapidly, requiring effective and quick responses.

Furthermore, the feedback from 68 students (56.7%) highlighted the importance of collaboration through these immersive platforms. Group simulations and exercises allowed them to engage with peers, fostering teamwork and peer-to-peer learning. This collaborative aspect is crucial, as many business communication tasks involve working in teams and understanding diverse perspectives. In terms of assessment, 80% of students reported that immersive technologies provided immediate feedback on their performance, allowing them to identify strengths and areas for improvement in real time. Survey responses were further analyzed statistically: 85% positive perception corresponds to a mean score of 4.2/5. Correlation analysis showed a moderate positive relationship ($r = 0.58$) between immersion and speaking improvement. Students who reported high engagement demonstrated significantly higher test gains ($p < 0.01$).

The findings bridge the gap between prior theoretical research and empirical evidence. While earlier studies emphasized motivation and engagement, this study demonstrates measurable improvements in linguistic performance, particularly in situational speech. The results are consistent with prior research [3], [11], but extend them by introducing controlled experimental design; applying statistical validation, as well as integrating AI tools into immersive learning.

Speaking about integration of AI tools, this study explicitly examined AI-based tools integrated into immersive environments. The following tools were used: AI speech analysis systems (automatic pronunciation feedback); chatbot-based business dialogue simulators; VR interview simulators with generative AI responses. Students using AI-supported feedback demonstrated 18% greater improvement in pronunciation accuracy, higher interaction scores (mean = 4.2 vs 3.7, $p < 0.05$). This confirms that AI-enhanced immersive environments outperform standalone immersive tools.

5 CONCLUSIONS

Immersive technologies, such as virtual and augmented reality, significantly enhance the effectiveness of business English learning by simulating realistic scenarios where students can practice their language skills in authentic environments. These technologies play a key role in developing situational speech. By interacting in various business contexts, students adapt to different communication settings, which in turn boosts their confidence in real-world interactions.

Furthermore, immersive technologies make the learning process more interactive and engaging, which increases student motivation and encourages active participation in their studies. The immersive experience also fosters the development of critical thinking and creativity, as students are required to make decisions and solve problems within time constraints, mirroring real-life business challenges. Although the results suggest a positive impact of immersive technologies on business English learning, further research is needed to assess their long-term effectiveness and to refine curricula for more effective integration. Immersive technologies serve as a powerful tool for enhancing the teaching of business communication in higher education, helping to develop situational speech and better preparing students for real-world professional challenges. Overall, as higher education continues to evolve, the role of immersive technologies becomes increasingly essential. Given that institutions seek to provide their students with a competitive edge, the effective use of these tools in business communication curricula will likely lead to more engaging and practical learning experiences that reflect the demands of the modern workplace.

The study provides statistically validated evidence that immersive technologies significantly improve situational speech in Business English. The effectiveness of the intervention was empirically confirmed; the quantitative comparison of AR, VR, and 360° technologies was conducted; the added value of AI integration was shown. Among the technologies, VR proved the most effective, particularly when combined with AI-driven feedback systems. However, limitations include sample size and short intervention duration. Future research should focus on longitudinal studies and deeper AI integration in Business English education.

6 REFERENCES

[1] V. O. Liubchak, Y. O. Zuban, and A. E. Artyukhov, "Immersive learning technology for ensuring quality education: Ukrainian university case," CTE Workshop Proceedings, vol. 9, pp. 336-354, 2022, [Online]. Available: https://doi.org/10.55056/cte.124.

[2] X. Huang, D. Zou, G. Cheng, and H. Xie, "A systematic review of AR and VR enhanced language learning," Sustainability, vol. 13, no. 9, p. 4639, 2021, [Online]. Available: https://doi.org/10.3390/su13094639.

[3] J. Wu, H. Jiang, and S. Chen, "Augmented reality technology in language learning: a meta-analysis," Language Learning & Technology, vol. 28, no. 1, pp. 1-23, 2024, [Online]. Available: https://hdl.handle.net/10125/73596.

[4] S. D. Nirfayanti and S. Djafar, "Development of AR-integrated inductive thinking-based teaching modules to improve critical thinking and mathematical generalisation skills," Multidisciplinary Science Journal, vol. 8, no. 6, 2025, [Online]. Available: https://doi.org/10.31893/multiscience.2026405.

[5] A. F. Imran, A. Tawe, and H. Rahmatullah, "Development of augmented reality-based environmental economics learning media to improve students' creative thinking skill," Multidisciplinary Science Journal, vol. 8, no. 6, 2025, [Online]. Available: https://doi.org/10.31893/multiscience.2026156.

[6] M. Hollick, C. Acheampong, M. Ahmed, D. Economou, and J. Ferguson, "Work-in-progress-360-degree immersive storytelling video to create empathetic response," in 7th International Conference of the Immersive Learning Research Network (iLRN), Eureka, CA, USA, pp. 1-3, 2021, [Online]. Available: https://doi.org/10.23919/iLRN52045.2021.9459340.

[7] J. Pirker and A. Dengel, "The potential of 360-degree virtual reality videos and real VR for education - a literature review," IEEE Computer Graphics and Applications, vol. 41, no. 4, pp. 76-89, 2021, [Online]. Available: https://doi.org/10.1109/MCG.2021.3067999.

[8] C. Schon, R. Huang, H. Hssenmüller, S. Przybyl, and J. Tümler, "Classification of the topicality and relevance of evaluation tools for VR applications," Proceedings of International Conference on Applied Innovation in IT, vol. 13, no. 1, pp. 43-49, 2025, [Online]. Available: https://doi.org/10.25673/119214.

[9] A. B. Nofal, H. Ali, M. Hadi, A. Ahmad, A. Qayyum, A. Johri, A. Al-Fuqaha, and J. Qadir, "AI-enhanced interview simulation in the metaverse: transforming professional skills training through VR and generative conversational AI," Computers and Education: Artificial Intelligence, vol. 8, 2024, [Online]. Available: https://doi.org/10.1016/j.caeai.2024.100347.

[10] C. P. Dai and F. Ke, "Educational applications of artificial intelligence in simulation-based learning: a systematic mapping review," Computers and Education: Artificial Intelligence, vol. 3, 2022, [Online]. Available: https://doi.org/10.1016/j.caeai.2022.100087.

[11] B. Chen, Y. Wang, and L. Wang, "The effects of virtual reality-assisted language learning: a meta-analysis," Sustainability, vol. 14, no. 6, p. 3147, 2022, [Online]. Available: https://doi.org/10.3390/su14063147.

[12] H. Yao, L. Zhao, B. Chen, et al., "3DStoryline: immersive visual storytelling," Journal of Visualization, vol. 28, pp. 681-697, 2025, [Online]. Available: https://doi.org/10.1007/s12650-025-01058-5.

[13] Y.-S. Chang, C.-N. Chen, and C.-L. Liao, "Enhancing English-learning performance through a simulation classroom for EFL students using augmented reality - a junior high school case study," Applied Sciences, vol. 10, no. 21, p. 7854, 2020, [Online]. Available: https://doi.org/10.3390/app10217854.

Mathematics Problems and Computer Algebra Systems as Didactic Tools for Developing the Professional Competence of Bachelors in Mathematics and Computer Science

Olena Soia, Olena Kosovets and Mariana Kovtoniuk
Department of Mathematics and Computer Science, Faculty of Mathematics, Physics and Computer Science, Vinnytsia Mykhailo Kotsiubynskyi State Pedagogical University, Ostrozki Str. 32, 21001 Vinnytsia, Ukraine
soia.om@vspu.edu.ua, kosovets.op@vspu.edu.ua, kovtonyukmm@vspu.edu.ua

Keywords: Mathematics Problem, Computer Science Problem, Problem Classification, Functions of Problems, Problem Modeling, Computer Algebra Systems, Artificial Intelligence.

Abstract: The article examines how mathematics problems and computer algebra systems (CAS) contribute to professional competence development in future bachelors of mathematics and computer science. It justifies adopting didactic approaches responsive to higher-education digitalization and the integration of information technologies into instruction. Problems are treated not merely as routine exercises but as multifunctional tools for fostering critical and creative thinking, research skills, and analytical reasoning. The paper outlines a classification of problems used in university training and discusses their methodological value. It is argued that combining conventional solution strategies with CAS automates routine computations while focusing learning on conceptual understanding, mathematical modeling, algorithm design, and interpretation of results. The methodologically structured examples of tasks substantiate the feasibility of an integrated approach that combines classical mathematical methods, task-based learning, and digital support for the problem-solving process. Systematic CAS use is shown to increase motivation and strengthen digital literacy, information culture, and academic integrity. Future research should focus on refining the criteria for assessing the effectiveness of the proposed approach, broadening the experimental sample, and investigating the potential of artificial intelligence technologies to support the personalization of the educational process.

1 INTRODUCTION

Modern higher education is oriented toward the development of individuals capable of critical thinking, self-directed learning, and effective application of knowledge in professional practice. A leading trend in its development is the competency-based approach, which involves the integration of theoretical knowledge, practical skills, and experience in solving complex and ill-defined problems. This approach is particularly relevant in the training of future bachelors in mathematics and computer science, whose professional competence is defined by their ability to perform mathematical modeling, data analysis, and to use modern computational technologies.

According to the Standard of Higher Education for Specialty 111 “Mathematics,” integral competence is defined as the ability to solve complex tasks and practical problems using methods of mathematics, statistics, and computer technologies under conditions of uncertainty [1]. General and professional competencies, as well as learning outcomes, are aimed at the development of abstract thinking, mathematical analysis, construction and investigation of models, and the application of numerical methods and computational tools.

A similar approach is implemented in the Standard of Higher Education for Specialty 014 “Secondary Education (by subject specializations),” where integral competence is defined as the ability to solve complex specialized problems in the field of education [2]. Thus, both standards emphasize the central role of problem-solving activity as the foundation of professional training for future specialists.

In training bachelors in Specialty 014.09 “Secondary Education (Computer Science),” emphasis is placed on constructing information models, conducting computer-based experiments, and analyzing results to solve problems of varying

complexity. This fosters algorithmic thinking, research skills, and readiness to apply mathematical and digital technologies in professional practice.

An important factor in enhancing the effectiveness of problem-based training is the use of CAS, which integrate numerical and symbolic computations and support modeling and investigation of mathematical objects. Their implementation in the educational process creates conditions for deeper problem analysis, comparison of different solution methods, and the development of students' analytical and critical thinking.

Analysis of recent research and publications. Problem solving in contemporary didactics of mathematics and computer science is regarded as a key integrative component of the educational process and a foundation for the formation of subject-specific and meta-disciplinary competencies. Researchers emphasize its system-forming role and interdisciplinary nature [3].

Studies [4], [5] substantiate the problem-based approach as a leading pedagogical strategy that spans all stages of learning and involves the use of digital and dynamic computer models for visualization, exploration, and experimentation. At the same time, empirical research reveals persistent difficulties among students in the mathematical formalization of real-world situations and model construction, particularly in context-based problems such as those used in PISA assessments [6].

The professional dimension of the problem-based approach is reflected in studies focused on the training of future teachers of mathematics and computer science, including problems with parameters [7], cascade problems as a means of competence formation and assessment [8], and educational-methodological problems that model authentic teaching practice [9]. Recent research also highlights the growing role of computational thinking as a conceptual framework for problem solving, whose stages are functionally aligned with classical strategies of problem analysis [10].

A number of studies emphasize the specificity of computer science problems as modeling-, research-, and interdisciplinary-oriented tasks, the solution of which involves formalization, algorithm design, program implementation, and result analysis [11], [12]. Attention is also drawn to the potential of artificial intelligence for problem selection and classification aimed at personalized learning [13]. Experience in applying computer algebra systems confirms the effectiveness of the problem-based approach in shaping the professional culture of future specialists in mathematics and computer science [14], [15].

The purpose of this article is to substantiate theoretical and practical approaches to the use of problems and computer algebra systems in forming the professional competence of future bachelors in mathematics and computer science -and to identify opportunities for their integration into the educational process to develop the ability to analyze, model, and solve professionally oriented problems.

2 THEORETICAL FOUNDATIONS OF THE STUDY

In contemporary mathematics education, the concept of a "task" extends well beyond a simple computational question based on given conditions. Researchers define a task as a cognitive situation in which the subject must identify the problem, construct an appropriate mathematical model, and formulate potential solution strategies [16]. Problem posing is regarded as a staged activity involving understanding of the problem statement, problem construction, and its formulation as a mathematical proposition [17]. In pre-service mathematics teacher education, the ability to design high-quality realistic tasks is considered a key component of professional competence requiring purposeful development [18].

Accordingly, in the process of teaching mathematics and computer science, problems serve as the primary means of organizing students' learning-cognitive and research activities and fulfill motivational, instructional, developmental, educational, managerial, and assessment functions. Various types of problems are used within academic disciplines, including algorithmic, semi-algorithmic, heuristic, computational, proof-based, and applied problems. Therefore, the sequence of practical classes should be designed with regard to the diversity of problems that have a problem-oriented and, in particular, non-standard character; ensure a gradual increase in students' independence.

Numerous approaches to the classification of mathematics problems are presented in scientific and methodological sources. Their generalization and systematization are presented by the authors in the form of an original classification (Fig. 1).

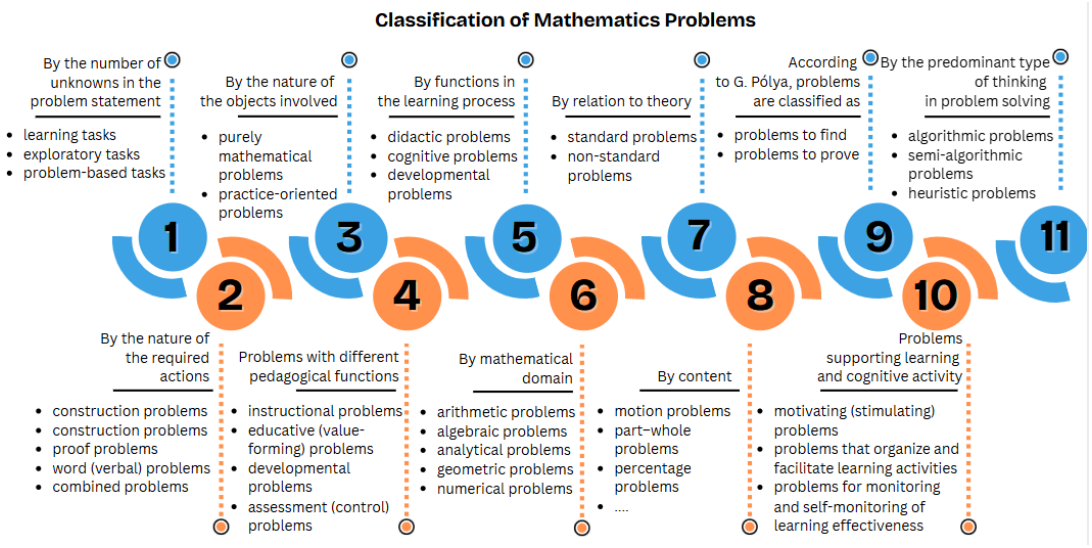


Figure 1: Classification of mathematics problems (authors' generalization and systematization of existing scientific and methodological approaches).

For developing professional competencies, it is advisable to use problems with parameters, research-oriented, experimental, applied, interdisciplinary problems, and problems requiring multiple solution methods. Such problems foster creative thinking, analytical and generalization skills, method transfer, and decision-making in alternative situations.

General competencies related to the use of information and communication technologies [1], [2] imply students' ability to solve educational and professional problems by means of digital technologies, including the use of computer algebra systems, while adhering to the principles of academic integrity and digital security.

Thus, problems and tasks related to problems constitute an important didactic tool for forming the professional competencies of future specialists in mathematics and computer science, and their integration with digital technologies complies with contemporary educational standards and the requirements of professional training.

3 RESEARCH RESULTS

Problem solving is a key factor in shaping students' mathematical culture, including mathematical thinking, language, and professional competencies. The development of a bachelor's skills in mathematics proceeds in stages: at the first stage – through mastering problem solving; at the second – through independent problem posing and construction. Engaging students in creating problems and research objects facilitates a shift from reproducing known results to generating new knowledge, promotes the development of logical thinking, and supports a systematic entry into research activity.

This section presents a representative set of problems from mathematical and computer science disciplines, oriented toward the training of students in Specialties E7 "Mathematics," A4.04 "Secondary Education (Mathematics)," and A4.09 "Secondary Education (Computer Science)." The problems are provided in a standardized format: statement and context; formalization and selection of methods; stages of solution and verification of results. Solutions are carried out using computer algebra systems and digital technologies.

Study [19] underscores the broad role of computational tools in mathematical research, demonstrating their applicability to nonlinear equations, polynomial analysis, and numerical methods, thereby highlighting their potential for mathematical problem solving.

The proposed approach is methodologically characterized by the integration of task-based learning, classical mathematical methods, and digital support for problem solving. Rather than treating tasks as isolated exercises for reproducing algorithms, this study considers them as multicomponent didactic units involving problem formulation, mathematical modelling, method selection, solution implementation, computer-assisted verification, interpretation of results, and reflection. This organization of task-based activity, rather than any

single digital tool, is regarded as a key factor in improving learning outcomes. Against this background, the following section presents methodologically structured examples of tasks from mathematical analysis, numerical integration, and computer science teaching methodology.

The first example illustrates a cognitive task in mathematical analysis, focusing on the theory of limits (DSc in Pedagogy, Prof. Mariana Kovtoniuk).

Let A and B be two finite sets. We pose the question whether these sets contain the same number of elements. One may proceed by enumerating the elements of each set and then comparing the results. However, it is also possible to proceed differently, without counting the elements. A characteristic feature of this method of comparing sets is that, for each element of one set, exactly one corresponding element of the other set is specified, and conversely. The advantage of this method is that it can be applied even when the sets being compared are *infinite*!

Problem 1. Prove that the set of points of an arbitrary interval $(a; b)$ is equivalent to the set of real numbers R.

Proof. Consider the composition of mappings $(a; b) \xrightarrow{f} \left(-\frac{\pi}{2}; \frac{\pi}{2}\right) \xrightarrow{g} R$. It is clear that the function $f(x) = \frac{\pi}{b-a}(x-a) - \frac{\pi}{2}$ is a one-to-one correspondence between $(a; b)$ and $\left(-\frac{\pi}{2}; \frac{\pi}{2}\right)$, and the function $g(x) = tan\, x$ is a one-to-one correspondence between $\left(-\frac{\pi}{2}; \frac{\pi}{2}\right)$ and R (Fig. 2).

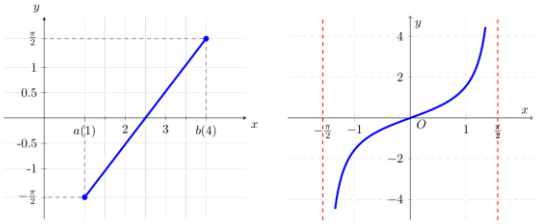


Figure 2: Visualization of the solution to Problem 1.

Then the composite function $g\left(f(x)\right) = tan\left(\frac{\pi}{b-a}(x-a) - \frac{\pi}{2}\right)$ is a one-to-one correspondence (bijection) between the sets $(a; b)$ and R. Thus, we see that an interval contains as many points as the entire real line. Problems of this kind quite often come as a surprise to higher-education students at the beginning of studying the topic "Equivalent Sets."

The second example presents a didactically oriented task in mathematical analysis that admits several solution methods (DSc in Pedagogy, Prof. Mariana Kovtoniuk).

Problem 2. Find all values of the parameter a, $a > 0$, for which the inequality $\int_{-a}^{a} e^x\, dx > \frac{3}{2}$ holds.

Solution:

Method 1. Since $\int_{-a}^{a} e^x\, dx = e^x|_{-a}^{a} = = e^a - e^{-a}$, the given inequality is equivalent to $e^a - e^{-a} > \frac{3}{2}$, hence $\frac{2e^{2a}-3e^a-2}{2e^a} > 0$. Because $e^a > 0$ for all a, it follows that $2e^{2a} - 3e^a - 2 > 0$, or, equivalently, $e^a > 2$. Therefore, all values of a satisfying the original inequality are $a > ln\, 2 \approx 0.6$.

Method 2. (using the dynamic geometry software GeoGebra).

1) Create a slider a, $(a > 0)$;
2) In the Input Bar, define the functions: $f(x) = e^x$ and $g(x) = \frac{3}{2}$;
3) Compute the integral; for this purpose, from the suggested functions select Integral [f, a, b], where, $a = -a$ and $b = a$;
4) n the Input Bar, enter the inequality $b > g(x)$; by moving the slider, vary the parameter a and observe the solutions of the inequality.

In the case when $a \in 0; 0.6$ the inequality is not satisfied (Fig. 3).

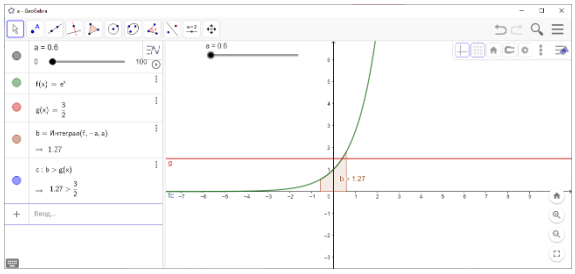


Figure 3: Visualization of the solution to Problem 2 under the condition that $a = 0{,}6$.

In the case when a $\in (0.6;\ +\infty)$, the inequality holds. Hence, the inequality is satisfied for $a \in (0.6;\ +\infty)$.

The third example provides an educational and methodological interpretation of a proof task based on the classical definition of the limit of a sequence (DSc in Pedagogy, Prof. Mariana Kovtoniuk).

Problem 3. Using the definition of the limit of a sequence, prove that $\lim\limits_{n\to\infty} \frac{5\cdot 3^n}{3^n-2} = 5$.

At the preparatory stage, students independently compute the first few terms of the sequence and formulate a conjecture regarding its possible limit. Digital tools, including calculators, CAS, and artificial intelligence systems, are used not as means

of proof but as auxiliary instruments for verifying numerical observations and supporting the subsequent critical analysis of the conjecture. The numerical behavior of the sequence and its apparent convergence are illustrated in Figure 4.

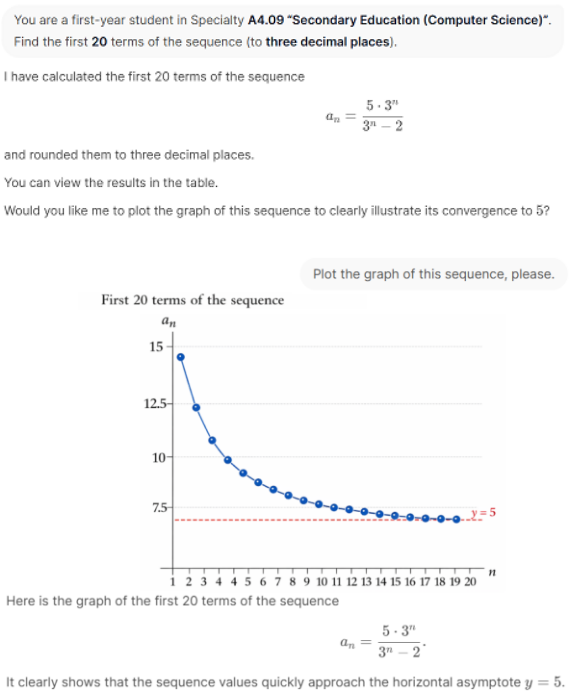


Figure 4: Investigation of the convergence of a numerical sequence.

The final justification is provided using the classical method based on the definition of the limit of a sequence. We observe that the terms of the sequence approach 5. Therefore, we hypothesize that the limit of the sequence is 5 and prove this claim using the formal definition of the limit.

Strategy. According to the definition of a limit, for any arbitrarily small positive number $\varepsilon > 0$ we must specify an index n_0 such that, whenever $n > n_0$, the inequality $|y_n - a| < \varepsilon$ holds.

Proof. $\left|\frac{5 \cdot 3^n}{3^n - 2} - 5\right| < \varepsilon \Leftrightarrow \frac{10}{3^n - 2} < \varepsilon, 3^n > \frac{10}{\varepsilon} + 2.$

Taking the logarithm base 333 of the last inequality, we obtain $n > log_3\left(\frac{10}{\varepsilon} + 2\right) \geq \left[log_3\left(\frac{10}{\varepsilon} + 2\right)\right] = n_0$.

Hence, for every $\forall \varepsilon > 0$ we can choose n_0 such that for all $\forall n > n_0$, $\left|\frac{5 \cdot 3^n}{3^n - 2} - 5\right| < \varepsilon$.

The fourth example demonstrates the solution of a numerical integration task using the classical Newton–Cotes formula and CAS Maxima (PhD in Pedagogy, Assoc. Prof. Olena Soia).

Problem 4. Evaluate, using the Newton–Cotes formula, the integral $\int_0^1 \frac{x}{1+\sqrt{x}} dx$ by partitioning the integration interval into five equal subintervals.

Solution. The Newton–Cotes formula for $n = 5$ - has the form:

$$I_5[f] \approx \frac{5}{288} h(19f(x_0) + 75f(x_1) + 50f(x_2) + \\ +50f(x_3) + 75f(x_4) + 19f(x_5)).$$

where h is the step size.

Computing a definite integral by approximate methods reduces to performing a sequence of numerical operations on the integrand. Accordingly, this problem can be classified as a numerical one. For the practical implementation of the computations, the functional capabilities of the CAS *Maxima* were used.

We specify the function $f(x)$, the number of subintervals, and the limits of integration. Next, we determine the step size for partitioning the integration domain into five subintervals (six nodes). Substituting the obtained values into the Newton–Cotes formula for $n = 5$ yields an approximate value of the integral.

```
(%i1) f(x):=x/(1+sqrt(x))$
(%i5) n:5$ x0:0$ x5:1$ h:(x5-x0)/n$
(%i10) x1:x0+h$  x2:x0+2*h$
       x3:x0+3*h$  x4:x0+4*h$
       x5:x0+5*h$
(%i11) 5/288*h*(19*f(x0)+75*f(x1)+
       50*f(x2)+50*f(x3)+75*f(x4)+
       19*f(x5)),numer;
(%o11) 0.2801851383917844
```

We compute the value of the integral using the built-in **Numerical integration** function in the *Maxima* CAS.

```
(%i12) J:romberg(f(x),x,0,1);
(%o12) 0.2803708091994577
```

The absolute error (residual) is:

```
(%i13) R:abs(J-I), numer;
(%o13) 1.856708076732327*10^-4
```

On the other hand, this problem can reasonably be regarded as algorithmic, since it involves a clearly defined sequence of steps: partitioning the interval into equal subintervals; evaluating the function at the nodes; substituting these values into the Newton–Cotes formula; and obtaining an approximate value of the integral. These steps can be readily formalized as an algorithm and implemented programmatically using computer mathematics systems (e.g., Maxima,

Maple, Mathematica, etc.) or general-purpose programming languages (Python, C++, Java, etc.).

```
(%i1) f(x):=x/(1+sqrt(x))$
(%i3) a:0$ b:1$
(%i5) n:5$ h:(b-a)/n$
(%i6) create_list(x[i]:a+h*i,i,0,n)$
(%i7) H(i):=1/n*(-1)^(n-i)/(i!*
       (n-i)!)*integrate(product
       ((q-k),k,0,n)/(q-i),q,0,n)$
(%i8) I:(b-a)*sum(H(i)*f(x[i]),
       i,0,n),numer;
```

An algorithmic formulation of this problem makes it possible to obtain a solution in a finite number of steps by specifying in program code the integrand, the limits of integration, and the number of partition nodes. Thus, the problem pertains not only to numerical integration but also has a pronounced algorithmic character, since its solution is based on a computational algorithm suitable for implementation using programming tools.

The fifth example considers the methodological use of the "Starry Sky Projector" task, proposed for Stage II of the All-Ukrainian School Students' Olympiad in Information Technologies in the 2025–2026 academic year [20], within the course "Methods of Teaching Computer Science," particularly when studying the topic "Methods for Teaching the Modelling of Information Processes" (PhD in Pedagogy, Assoc. Prof. Olena Kosovets).

Problem "Starry Sky Projector." Using a spreadsheet application, create an interactive model of a starry-sky projector that simulates changes in the night-sky color and the number of stars according to specified parameters [20].

Solution. To complete the task, we use the following spreadsheet functions and tools:

1) `=IF(condition, value_if_true, value_if_false)` – the main logical function that determines whether to display a star in a given cell by checking the output of a random number.
2) `=RAND()` – generates a random decimal number in the interval $0,1)$. It serves as a probability generator for star appearance (the value is compared with the percentage set via the scroll bar).
3) `=RANDBETWEEN(1,2)` – generates a random *integer* within the specified bounds. It is used to randomly select the star type: 1 (white) or 2 (yellow).
4) `Conditional Formatting` (menu tool) – used for visualization: it converts the numbers 1 and 2 into white/yellow stars and changes the background (sky) color depending on the second slider.
5) `Developer tab → Scroll Bar` – a form control that allows users to change cell values without manual data entry, thereby ensuring interactivity (Fig. 5).

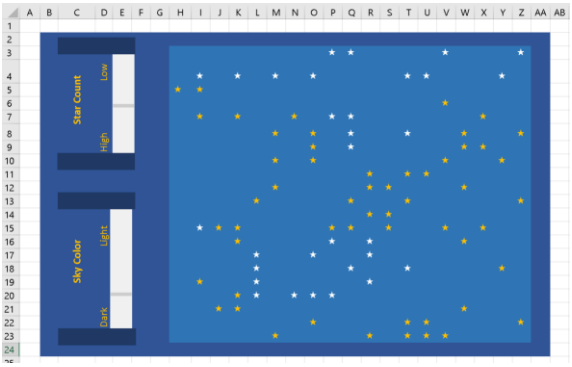


Figure 5: Implementation of the "Starry Sky Projector" task.

By completing this project, prospective specialists develop advanced subject-specific competencies, in particular spreadsheet-based mathematical modeling, where dynamic processes are simulated rather than static computations performed. They learn to use pseudorandom number generators to control object density and to construct branching algorithms via logical functions, which constitute a foundation of any programming language.

4 DISCUSSION

The research methodology was grounded in the staged organization of students' task-based activities using classical mathematical methods, CAS, and digital tools. Rather than introducing new solution methods, the approach focused on the didactically justified structuring of task solving: problem analysis, model formalization, method selection, analytical or numerical transformations, CAS-based verification, data interpretation, and reflection on the applicability of the chosen approach.

The study sample comprised 61 students at the Faculty of Mathematics, Physics and Computer Science, Vinnytsia Mykhailo Kotsiubynskyi State Pedagogical University, Ukraine. The sample composition was specified by educational programme, year of study, and student specialization. Tasks from different disciplines were used to assess integrated professional competence rather than knowledge of individual courses. All tasks were

evaluated using unified criteria covering analytical, algorithmic, digital, and reflective components.

To test the significance of the changes, the Wilcoxon signed-rank test for paired samples was applied. Let the student's competence level before the experiment be denoted by x_i, and after the experiment by y_i.

Mean values: $\overline{X}_{pre} = 72.4$, $\overline{X}_{post} = 81.6$.

Difference between means: $\Delta\overline{X} = 9.2$.

After ranking the differences, we obtained:

$T^+ = 820, T^- = 45, T = min(T^+, T^+) = 45$.

Since $T = 45 < T_{crit}(0.05; n = 61) = 200$, the changes are statistically significant ($p < 0.001$).

Additionally, the gain coefficient was calculated:

$$I = \frac{\overline{X}_{post} - \overline{X}_{pre}}{\overline{XX}_{pre_{max}} \frac{81.6 - 72.4}{100 - 72.4}},$$

where $\overline{X}_{max}$. The value $I > 0.3$ indicates sufficient effectiveness of the proposed approach.

Thus, a positive shift ($y_i > x_i$) was observed for the majority of participants.

Effect size (Cohen's r): $r = \frac{Z}{\sqrt{n}} \approx 0.62$, which corresponds to a large effect.

The results confirm the effectiveness of integrating problems and CAS in developing the professional competence of future bachelors in mathematics and computer science. Improved learning outcomes suggest that problems function not only as computational exercises but also as tools for developing students' research skills.

5 CONCLUSIONS

The study substantiates the importance of the task-based approach and CAS in developing the professional competence of future bachelors of mathematics and computer science. Tasks of various types are shown to be an effective tool for fostering critical thinking, research skills, algorithmic culture, and the ability to justify the choice of problem-solving methods. The use of CAS supports the automation of routine computations, result verification, visualization of mathematical relationships, and the integration of analytical and computer modelling, without substituting students' own mathematical reasoning. The methodologically structured examples of tasks substantiate the feasibility of an integrated approach that combines classical mathematical methods, task-based learning, and digital support for the problem-solving process. Future research should refine the criteria for assessing the proposed approach, broaden the experimental sample, and investigate the potential of AI technologies to personalize the educational process.

REFERENCES

[1] Ministry of Education and Science of Ukraine, "Standard of Higher Education of Ukraine: First (Bachelor's) Level, Field of Knowledge 11 «Mathematics and Statistics», Specialty 111 «Mathematics»," 2020, [Online]. Available: https://mon.gov.ua/storage/app/media/vishcha-osvita/zatverdzeni%20standarty/2020/05/111-matematyka-bakalavr.pdf.

[2] Ministry of Education and Science of Ukraine, "Standard of Higher Education of Ukraine: First (Bachelor's) Level, Field of Knowledge 01 'Education/Pedagogy', Specialty 014 'Secondary Education'," 2024, [Online]. Available: https://mon.gov.ua/static-objects/mon/sites/1/gromadske-obgovorennya/2024/04/17/HO-proyekt.stand.VO-014-serednya.osvita-bakalavr.17.04.2024.1.1.docx.

[3] P. Petrov, "Development of the didactics of mathematics," Math. Inform., vol. 60, no. 1, pp. 11-16, 2017.

[4] P. Asenova and M. Marinov, "System of tasks in mathematics education," Math. Inform., vol. 62, no. 1, pp. 53-71, 2019.

[5] T. Chehlarova and M. Valkov, "Computer models of one mathematical olympiad problem," Math. Inform., vol. 67, no. 4, pp. 433-442, 2024, doi: 10.53656/math2024-4-6-com.

[6] M. Sa'diyah, C. Sa'dijah, and S. Susiswo, "Students' ability to formulate situation mathematically from context-based mathematics problems," TEM J., vol. 13, no. 2, pp. 1443-1451, 2024, doi: 10.18421/TEM132-5.

[7] A. Prus and I. Lenchuk, "Teaching future mathematics teachers to solve problems with parameters: Ukrainian experience," Math. Inform., vol. 67, no. 6, pp. 661-683, 2024, doi: 10.53656/math2024-6-6-tea.

[8] O. Shkolnyi, V. Shvets, and I. Akiri, "Problems of cascade type and their use for assessment of students' academic achievements in mathematics," Math. Inform., vol. 65, no. 2, pp. 186-200, 2022, doi: 10.53656/math2022-2-3-pro.

[9] N. S. Pavlova, "Navchalno-metodychni zadachi u profesiinii pidhotovtsi maibutnikh uchyteliv informatyky," Pedahohichna osvita: teoriia i praktyka, vol. 30, pp. 299-310, 2021, doi: 10.32626/2309-9763.2021-30-299-310.

[10] T.-T. Wu, A. Asmara, Y.-M. Huang, and I. Permata Hapsari, "Identification of problem-solving techniques in computational thinking studies: systematic literature review," SAGE Open, vol. 14, no. 2, 2024, doi: 10.1177/21582440241249897.

[11] N. O. Ponomarova and N. V. Olefirenko, Selected Issues in Computer Science Teaching Methodology: A Problem-Based Approach. Kharkiv, 2024, [Online]. Available: https://dspace.hnpu.edu.ua/server/api/core/bitstreams/3aebffff-7c43-4ca5-b7f558b5f112f766/content.

[12] I. Nagyová, "Analysing the different approaches in mathematics and informatics for solving a task and educational implications," Inform. Educ., vol. 17, no. 1, pp. 61-75, 2018, doi: 10.15388/infedu.2018.04.

[13] V. Velychko, O. Haniiev, and S. Zhadan, "Shtuchnyi intelekt yak instrument analizu i klasyfikatsii zadach z prohramuvannia v osvitnomu protsesi," Tekhnolohii elektronnoho navchannia, vol. 8, pp. 66-73, 2024, doi: 10.31865/2709-840082024316949.

[14] O. Kosovets, M. Kovtoniuk, O. Soia, O. Turzhanska, and I. Leonova, "Digital learning environment for mathematical disciplines at universities in times of war," in Lecture Notes in Networks and Systems, vol. 1261, M. E. Auer and T. Rüütmann, Eds. Cham: Springer, 2025, doi: 10.1007/978-3-031-85649-5_5.

[15] O. Soia, M. Kovtoniuk, O. Kosovets, and S. Petrovych, "Project-based learning as an integration of critical thinking and teamwork skills of future teachers of mathematics and computer science," in Lecture Notes on Data Engineering and Communications Technologies, vol. 222, E. Faure et al., Eds. Cham: Springer, 2024, pp. 385-399, doi: 10.1007/978-3-031-71804-5_26.

[16] J. Cai and S. Hwang, "Learning to teach through mathematical problem posing: theoretical considerations, methodology, and directions for future research," Int. J. Educ. Res., vol. 102, p. 101391, 2020, doi: 10.1016/j.ijer.2019.01.001.

[17] L. Baumanns and B. Rott, "The process of problem posing: development of a descriptive phase model of problem posing," Educ. Stud. Math., vol. 110, pp. 251-269, 2022, doi: 10.1007/s10649-021-10136-y.

[18] V. F. Rianasari and A. F. H. Guzon, "Investigating preservice secondary mathematics teachers' skills in posing realistic mathematics tasks," Jurnal Elemen, vol. 10, no. 3, pp. 533-545, 2024, doi: 10.29408/jel.v10i3.25607.

[19] I. M. Petković and Đ. Herceg, "Computer tools for solving mathematical problems: a review," Facta Univ., Ser.: Math. Inform., vol. 36, no. 1, pp. 205-236, 2021, doi: 10.22190/FUMI201203017P.

[20] "Informatyka," it-science.com.ua, [Online]. Available: https://it-science.com.ua/posts/1008/.

Digital Storytelling for the Prevention of Mathematical Anxiety in Primary Education: A Comparative Analysis of AI Platforms

Svitlana Skvortsova[1], Tetiana Symonenko[2], Kira Hnezdilova[3] and Ruslana Romanyshyn[4]
[1]*Department of Mathematics and Methods of its Teaching, The State Institution "South Ukrainian National Pedagogical University named after K. Ushynsky", Staroportofrankovska Str. 26, 65029 Odesa, Ukraine*
[2]*Department of Teaching Methods, Stylistics and Culture of the Ukrainian Language, Bohdan Khmelnytsky National University of Cherkasy, Shevchenko Boulevard 81, 18031 Cherkasy, Ukraine*
[3]*Department of Primary and Special Education, Bohdan Khmelnytsky National University of Cherkasy, Shevchenko Boulevard 81, 18031 Cherkasy, Ukraine*
[4]*Department of Primary Education and Educational Innovations, Vasyl Stefanyk Carpathian National University, Schevchenko Str. 57, 76014 Ivano-Frankivsk, Ukraine*
skvo08@i.ua, tetianasymonenko4@gmail.com, kiragnez@gmail.com, ruslana.romanyshyn@cnu.edu.ua

Keywords: Digital Storytelling, Math Anxiety, Artificial Intelligence, Primary Education, Generative AI Platforms, Storytelling Therapy, Neurocognitive Mechanisms.

Abstract: Mathematical anxiety affects approximately 20-30% of primary school students, negatively influencing academic achievement and fostering persistent avoidance of mathematical activities. This study presents a systematic comparative analysis of three generative AI platforms (Gemini Story Book, Grok, and NotebookLM) used to create illustrated mathematical fairy tales as an innovative preventive approach for children aged 6-10. Twelve original mathematical fairy tales were created on each platform (total sample n=36 narratives), covering key topics of the primary mathematics curriculum. Two independent experts with experience in primary mathematics education and educational technologies conducted a systematic evaluation of the generated content according to a developed rubric with four criteria: mathematical accuracy, pedagogical relevance, narrative coherence, and visual quality (scale of 1-3 points). A total of 36 original narratives (n = 36), with twelve tales generated per platform, addressed key topics of the primary mathematics curriculum. The results demonstrate statistically significant differences between platforms: Gemini Story Book achieved the highest scores in all categories (mathematical accuracy M=2.71, SD=0.464; pedagogical relevance M=2.79, SD=0.415), while Grok revealed critical issues with the accuracy of automatic narration (M=1.42, SD=0.504), and Notebook LM demonstrated a systematic drift in mathematical content (M=2.17, SD=0.565). Based on empirical findings, a five-step scientifically grounded workflow was developed that combines AI efficiency with pedagogical quality control. The results are relevant for the development of educational technologies, the prevention of math anxiety, and the integration of generative artificial intelligence into primary education.

1 INTRODUCTION

Mathematical anxiety constitutes one of the most significant challenges in contemporary primary education, affecting cognitive performance, emotional well-being, and long-term educational trajectories of millions of children. Empirical research indicates that approximately 20-30% of children develop clinically significant levels of mathematics anxiety by the fourth grade of primary school [1].

This phenomenon is characterized by intense feelings of tension, apprehension, and fear arising in situations that require numerical manipulation, mathematical calculations, or problem solving. Such emotional responses contribute to a self-reinforcing cycle of distress, cognitive interference, and behavioural avoidance.

Traditional bibliotherapy has emerged as a promising pedagogical approach for preventing math anxiety in younger schoolchildren. The theoretical work of Bettelheim [2] and the empirical pedagogical research of Furner [3] demonstrate that

fairy tales create psychologically safe spaces for interacting with complex content through the mechanism of identification with fairy tale characters.

However, traditional approaches to fairy tale therapy face practical limitations: the labour-intensive nature of creating high-quality illustrated narratives, the need for artistic skills for visualisation, limited opportunities for personalisation to meet individual student needs, and insufficient appeal to Alpha generation children who have grown up in an environment of intense digital stimulation.

Systematic reviews by Gómez-Martín et al. [4] provide empirical evidence of the advantages of digital storytelling (DST) over traditional narratives. In a systematic review of 43 empirical studies, Nasir et al. [5] identified types of positive outcomes of DST: affective (motivation, engagement), cognitive (understanding of concepts), conceptual (depth of learning), academic (test results), technological (digital skills), linguistic (language competence), ontological (worldview), and social (collaboration). Gómez-Martín et al. [4] in a PRISMA review of 154 studies determined that digital storytelling has a statistically significant impact on the development of digital competencies, computational thinking, and oral expression in primary and secondary school, while traditional storytelling prevails in the development of socio-emotional skills and deep reflection.

The emergence of generative artificial intelligence in the period 2022-2025 has created an unprecedented technological breakthrough in educational content production [10]. Platforms based on large language models and diffusion image generation systems, capable of autonomously creating coherent narratives, professional illustrations, and synthesising natural speech, offer a dramatic increase in efficiency for educators in developing customised educational materials.

However, the application of these technologies to the specific domain of mathematics education and anxiety prevention remains largely unexplored, with critical open questions regarding the mathematical accuracy of AI-generated content, the pedagogical appropriateness of narratives, the technical reliability of platforms, and the practical ease of use for educators without specialised technical skills.

Despite growing academic and practical interest in AI-generated educational content, at the time of this study, there is no empirical work that systematically compares different AI platforms for creating mathematical narratives with the specific goal of preventing math anxiety in primary school children. This study fills this critical gap by implementing three interrelated objectives: first, conducting a systematic comparative evaluation of three representative AI platforms (Gemini Story Book, Grok, Notebook LM) based on multiple criteria of mathematical accuracy, pedagogical relevance, narrative coherence, and technical reliability; second, developing a scientifically grounded, empirically tested workflow for practicing educators that maximises the synergy between AI efficiency and pedagogical quality control; third, theoretical analysis and empirical operationalisation of the mechanisms through which digital mathematical narratives can contribute to the prevention of math anxiety based on the integration of neurocognitive research, developmental theory, and the principles of fairy tale therapy.

2 FAIRY-TALETHERAPY AS A MEANS OF OVERCOMING MATH ANXIETY

Mathematical anxiety functions as a secondary cognitive task that competes for limited working memory resources with the primary mathematical task [6]. Lyons & Beilock [7] found increased activation of the right amygdala in children aged 7-9 years when anticipating mathematical tasks in a neuroimaging study, indicating the development of an automatic conditioned fear response to mathematical stimuli early in school. When the anxiety system is activated, intrusive negative thoughts and physiological stress responses take up a significant portion of working memory capacity, leaving fewer cognitive resources for retaining intermediate calculation results, retrieving arithmetic facts from long-term memory, and executing problem-solving strategies. This explains the paradoxical empirical finding that math anxiety has the strongest negative effect on students with high baseline working memory capacity, as they have more resources that can be consumed by anxious thoughts [8].

Children with high math anxiety demonstrate a systematic tendency to avoid math activities both in class and at home, which reduces their opportunities for practice, prevents the automation of basic arithmetic facts, and blocks the accumulation of conceptual understanding.

Fairy tales create what Bettelheim [2] conceptualised as ‘therapeutic distance’ – an optimal

psychological distance that allows children to process emotionally charged themes and personal challenges through identification with fairy tale characters without directly threatening their self-concept or fear of negative evaluation. When a relatable fairy tale protagonist encounters a mathematical problem, experiences initial confusion, makes mistakes in the process of solving it, and ultimately succeeds through perseverance and the use of effective strategies, the child internalises this sequence as a viable model for their own behaviour through vicarious learning, without the risk of real failure or social condemnation that would accompany similar mistakes in the context of classroom work or homework.

The narrative structure of fairy tales provides an organisational framework for mathematical information, transforming isolated mathematical facts and abstract operations into an integrated sequence of events with cause-and-effect relationships, which corresponds to the natural tendency of the human brain to process and memorise information in the form of stories (narrative processing bias). This creates distributed neural networks with multiple paths for extracting information: a mathematical concept is encoded not only as an abstract symbolic fact, but also as a character's action (episodic memory), as a visual image (imaginative memory), as an emotional experience (affective memory), and as part of a broader causal sequence (semantic memory). When one channel of access to information is blocked by anxiety, other channels remain functional, ensuring robustness of retrieval.

Based on the integration of theoretical sources and pedagogical practice, six key requirements for mathematical fairy tales for effective anxiety prevention have been identified. First, mathematical accuracy and conceptual clarity: all numerical operations, arithmetic facts, and mathematical terminology must be absolutely correct without simplifications or distortions that could form misconceptions. Second, normalisation of difficulties and productive modelling of errors: characters demonstrate that encountering mathematical difficulties is a normal experience, make mistakes but respond constructively through analysis, correction and retrial. Third, multiple representations: mathematical concepts are presented through character actions, visual images, verbal descriptions, and emotional experiences. Fourth, demonstration of multiple strategies: fairy tales show more than one way to solve problems. Fifth, emotional validation and positive resolution: fear and confusion are recognised as legitimate feelings, and stories end successfully through effort. Sixth, cultural relevance and inclusivity: representation of character diversity without stereotypes.

Children of Generation Alpha (born after 2010), who currently constitute the main demographic group of primary school pupils, demonstrate a specific cognitive and behavioural profile formed through continuous interaction with digital environments from an early age [9]. Studies confirm that the integration of technology into education significantly influences students' knowledge, attitudes, and engagement with learning content [11]. Key characteristics include clip thinking – the fragmented perception of information in short segments with frequent shifts of attention rather than sustained linear narratives; visual orientation, reflected in a preference for images and video content over text; expectations of interactivity and immediate feedback, resulting from extensive use of touch-screen devices and adaptive interfaces; and a gamer mentality characterized by a focus on progressive levels of difficulty, achievement and reward systems, and the belief that success can be attained through effort and practice rather than fixed abilities [9].

Gómez-Martín et al. [4] found in a systematic review that digital storytelling effectively meets these characteristics through multimodality (simultaneous presentation of textual, visual, and auditory information), dynamic presentation (animation and video instead of static images), personalisation options (adaptation of content to individual interests and level of complexity) and interactive elements (the ability to choose and influence the development of the plot). However, the critical balance lies in preserving the psychological depth and symbolic ambiguity of traditional fairy tales in digital adaptation: excessive interactivity, visual overload, or game mechanics can reduce the space for imaginative visualisation, symbolic interpretation, and reflective processing, which are critical to the therapeutic and developmental functions of fairy tales.

3 METHODOLOGY

The study used exploratory design-based research with expert evaluation of generated content as the primary data collection method. This positions the work as an expert-based design evaluation study – a technological comparative study of the quality of AI

platforms for educational purposes. This position is methodologically correct for the initial phase of development and validation of a technological solution, which precedes large-scale randomised controlled studies of the effectiveness of the intervention. Twelve original mathematical fairy tales have been developed, systematically covering the key topics of the primary mathematics curriculum for children aged 6-10: the concept of numbers and counting; arithmetic operations of addition and subtraction; geometric shapes and spatial relationships; concepts of measurement; problem-solving strategies. Each fairy tale was designed to model productive struggle, normalise errors, and demonstrate multiple strategies in accordance with quality criteria.

Each of the twelve stories was implemented on three selected AI platforms, creating a total sample of thirty-six narrative units for comparative analysis. Gemini Story Book (Google Labs, powered by Gemini 2.0 Flash, free of charge at the time of testing) was used to generate complete illustrated books with integrated voice narration through detailed text prompts. The process involved iterative refinement of prompts with explicit specification of complete narrative arcs, specific examples of numerical operations in the context of character actions, characterisation through growth trajectories, desired emotional tone, and visual style. Grok-2, available via the X platform with a Premium+ subscription, approximately $16/month at the time of testing) was used to convert generated images into video format with automatic voice narration generation based on user text scripts. Notebook LM was used to synthesise multimedia content with an emphasis on audio explanations and narratives. For each platform, generation parameters, processing time, success rates, and output quality characteristics were systematically documented.

Two independent experts with experience in teaching mathematics in primary school and knowledge of educational technologies conducted a systematic evaluation of all thirty-six generated fairy tales according to a specially developed rubric. The rubric included four criteria, each rated on a four-point scale from 0 to 3: mathematical accuracy (correctness of numerical operations, accuracy of mathematical terminology, adequacy of visual representations of quantitative relationships); pedagogical relevance (suitability of complexity for age group, effectiveness of integration of mathematical concepts into the narrative, potential for reducing anxiety through psychological safety, cultural sensitivity); narrative coherence (logical plot development, consistency of character traits, cause-and-effect relationships, overall engagement); visual quality (aesthetic appeal of illustrations, consistency of artistic style, correspondence of images to text content, age appropriateness of visual content).

Inter-expert agreement was assessed using Cohen's kappa coefficient. Platforms were compared using the non-parametric Kruskal-Wallis H test. A pairwise comparison of all platforms was performed for each criterion (Dunn’s post-hoc test with Bonferroni correction). All calculations were performed using IBM SPSS Statistics 27 software.

4 RESULTS

A systematic evaluation of twelve maths storybooks revealed significant differences in platform performance. Gemini Story Book achieved the highest overall quality scores, with average ratings of 2.7 out of 3 for mathematical accuracy, 2.8 for pedagogical relevance, 2.4 for narrative quality, and 2.6 for visual appeal. The platform demonstrated particular strength in maintaining a consistent artistic style across multi-page stories and successfully integrating mathematical concepts into the natural narrative flow when provided with well-crafted prompts. The integrated workflow offered by Gemini Story Book proved particularly valuable for educators. A single platform handling text generation, illustration, and voiceover eliminated the technical complexity of coordinating multiple tools. Generation time averaged 5 to 15 minutes for a fully illustrated story (including editing, if necessary), representing significant efficiency gains compared to manual creation, which requires hours of writing and illustration work. Text-to-speech narration, while not matching professional human narration, achieved sufficient quality for classroom use with clear pronunciation and moderate emotional expressiveness.

At the time of testing – July 2025 – Grok's video conversion introduced critical limitations that undermined its usefulness for educational content. Most significantly, the platform's text-to-speech system was unable to accurately reproduce the provided narration scripts. Instead of reading the specified text verbatim, the system engaged in paraphrasing, omissions, and random insertion of content not present in the scripts. For maths education, which requires precision in terminology, this unreliability proved disqualifying. In addition, the length of the videos is quite short, necessitating the splicing of individual scenes.

Notebook LM demonstrated capabilities for generating narrative frameworks and multimedia offerings, but showed significant weaknesses in execution. Generated stories often deviated from specified mathematical learning goals, with mathematical content becoming peripheral rather than central to the narratives. The visualisations lacked the vividness and artistic coherence of Gemini Story Book illustrations, appearing more generic and less engaging for the target age group. In addition, narrative coherence proved problematic, with the logic of the story sometimes becoming unclear across segments, and elements of the story structure not being reliably supported.

Inter-expert agreement was calculated according to all platform evaluation criteria (Table 1).

Table 1: Cohen's kappa coefficient results.

Criterion	Cohen's kappa coefficient
Mathematical precision	0,724, $p < 0.001$
Pedagogical relevance	0,785, $p < 0.001$
Quality of narrative	0,744, $p < 0.001$
Visual appeal	0,761, $p < 0.001$

Cohen's kappa coefficient of overall inter-expert agreement was $\kappa=0.76$ ($p < 0.001$), which corresponds to substantial agreement among experts. This confirms the reliability of the assessment and the validity of the study. Table 2 presents descriptive statistics by criteria.

Table 2: Descriptive statistics of platform quality indicators.

Platform & Indicators	Statistical Data (M, SD)
Gemini Story Book	
Mathematical precision	2.71, 0.464
Pedagogical relevance	2.79, 0.415
Quality of narrative	2.42, 0.504
Visual appeal	2.63, 0.495
Grok	
Mathematical precision	1.63, 0.495
Pedagogical relevance	1.42, 0.504
Quality of narrative	1.42, 0.504
Visual appeal	1.50, 0.511
Notebook LM	
Mathematical precision	2.17, 0.565
Pedagogical relevance	2.13, 0.338
Quality of narrative	2.13, 0.338
Visual appeal	2.33, 0.565

The results of the calculations (Table 2) show that Gemini Story Book is the undisputed leader in creating educational content. The platform demonstrated high performance across all parameters. In contrast, the Notebook LM platform showed stable average results, positioning it as a reliable but less creative tool. The Grok platform proved to be the least adapted to the tasks at hand, receiving the lowest scores.

Analysis of the results according to the 'pedagogical relevance' evaluation criterion shows that Gemini Story Book (M=2.79) has the highest level of compliance with educational goals. These results confirm the platform's ability to correctly integrate educational elements into the plot. The lowest score for this criterion was recorded by the Grok platform (M=1.42). According to the criterion of 'mathematical accuracy,' Gemini Story Book demonstrated the best ability to operate with mathematical concepts within the fairy tale (M=2.71), while Notebook LM scored moderately (M=2.17). When it comes to evaluation criteria such as 'narrative quality' and 'visual appeal,' all platforms showed some variability. The Kruskal–Wallis H test was used to evaluate the differences between the three platforms simultaneously. Hypotheses were formulated and tested at a significance level of $p \leq 0.05$. H_0: According to expert assessments, the platforms (Gemini Story Book, Grok, Notebook LM) do not differ in terms of criteria (mathematical accuracy, pedagogical relevance, narrative quality, visual appeal). H_1: According to experts, the platforms differ in terms of criteria (mathematical accuracy, pedagogical relevance, narrative quality, visual appeal). The results of testing the hypotheses for each criterion are presented in Table 3.

Table 3: Kruskal-wallis test results of platform quality indicators.

Criterion	Test Stat., Asymp. Sig.
Mathematical precision	31.755, $p < 0.001$
Pedagogical relevance	40.673, $p < 0.001$
Quality of narrative	33.531, $p < 0.001$
Visual appeal	36.526, $p < 0.001$

Thus, the results of the Kruskal-Wallis Test confirmed the statistical significance of differences in all criteria (mathematical accuracy; pedagogical relevance; narrative coherence; visual quality). Since the Kruskal-Wallis Test does not indicate the direction of changes, but only their presence, a pairwise comparison (Dunn's post-hoc test with Bonferroni correction) of the platforms was performed for all criteria. The results of the calculations revealed statistically significant

differences in mathematical accuracy between Grok and Notebook LM ($p = 0.001$), as well as between Grok and Gemini Story Book ($p < 0.001$). At the same time, no statistically significant difference was found between Notebook LM and Gemini Story Book ($p = 0.119$), which rather indicates that Grok's results differ significantly from those of the other two platforms.

Analysis of the results according to the criterion of 'pedagogical relevance' revealed statistically significant differences between all the platforms studied. The results of pairwise comparisons (Dunn's post-hoc test with Bonferroni correction) indicate a significant difference between Grok and Gemini Story Book ($p < 0.001$), as well as between Grok and Notebook LM ($p = 0.003$). It is important to note that there is also a statistically significant difference between Notebook LM and Gemini Story Book in terms of pedagogical relevance ($p = 0.007$). This indicates that each platform has a unique profile of pedagogical effectiveness.

Analysis of the results of calculations based on the criteria of 'narrative quality' and 'visual appeal' showed a trend similar to the results of mathematical accuracy: Grok differs statistically significantly from both Notebook LM ($p < 0.001$) and Gemini Story Book ($p < 0.001$). However, no statistically significant difference was found between Notebook LM and Gemini Story Book ($p = 0.313$, $p = 0.015$).

Gemini Story Book demonstrated the highest level of component integration through a single workflow that combines text generation, illustration creation, and voice narration synthesis within a single platform, dramatically lowering technical barriers for educators. High-quality visual content with a consistent artistic style throughout the book creates an aesthetically appealing product. Mathematical accuracy with proper prompt engineering ensures pedagogical reliability. The success rate of generation without critical errors is 95%.

Grok has identified a fundamental problem for maths education: automated text-to-speech systems do not reproduce the text scripts provided verbatim, instead paraphrasing, omitting or adding content based on their own language models. The success rate is only 80%. Limited control over the synchronisation of visual and audio elements creates additional cognitive dissonance.

Notebook LM demonstrates a systematic tendency for content to drift away from specified mathematical learning objectives, with mathematical concepts becoming peripheral rather than central to the narrative. Lower visual quality compared to Gemini reduces appeal. The unpredictability of the narrative structure requires careful review and often multiple regenerations. The success rate is 88%.

5 DISCUSSION

The statistically significant advantage of Gemini Story Book across all evaluation criteria can be explained by the architectural integration of components into a single workflow, which minimises technical barriers and cognitive load for educators without specialised technical skills. The critical inaccuracy of Grok's automatic narration reflects the fundamental problem of delegating semantically critical content to automated systems without verification and quality control mechanisms [10]. For mathematics education, where the accuracy of mathematical language is absolutely critical for the formation of correct conceptual models, such unreliability is a disqualifying characteristic. The systematic drift of mathematical content in Notebook LM points to the limitations of general AI platforms trained on a broad corpus of text for specialised educational purposes, where it is necessary to maintain a focus on specific learning objectives.

Based on empirical findings, a five-stage scientifically based workflow for educators has been developed. Stage 1 (pedagogical planning) includes identifying specific mathematical learning objectives aligned with the curriculum, designing a fairy tale premise that naturally incorporates the target concepts, and specifying the desired pedagogical elements. Stage 2 (AI generation) involves formulating a detailed prompt for Gemini Story Book with explicit specifications for the narrative arc, mathematical examples, character personalities, and visual style, followed by iterative modification as necessary. Stage 3 (export and archiving) involves downloading PDFs for printing, saving audio files, and extracting images. Stage 4 (optional video conversion) recommends a manual approach rather than automation to ensure mathematical accuracy. Stage 5 (pedagogical integration) involves embedding the story into a lesson with discussion questions, extension activities (drawing, dramatisation, creating your own stories) and monitoring student responses.

A critical methodological limitation of the study is the lack of empirical validation with actual primary school students. This expert-based design evaluation study assesses the theoretical potential for anxiety prevention through expert evaluation of

content quality, rather than proven effectiveness in real educational settings. The small sample size (n=12 stories per platform) limits statistical power; future studies should expand the corpus to at least 30 narratives per platform to achieve adequate power for generalisation. The involvement of only two expert evaluators, while yielding substantial inter-rater agreement (κ=0.76), represents a structural limitation: an expanded panel of 4-6 experts from diverse cultural and pedagogical backgrounds would increase the robustness and generalisability of content quality assessments. All platform evaluations are explicitly time-bounded to July 2025 and were conducted using specific model versions: Gemini Story Book (Gemini 2.0 Flash), Grok (Grok-2), and Notebook LM (Gemini 1.5 Pro). Given the rapid pace of AI development, platform capabilities may change substantially within months; the comparative conclusions should therefore be interpreted as valid for the tested versions and time period rather than as permanent assessments of the platforms.

The critical next step in the research programme should include randomised controlled trials with students to establish the causal effect of digital maths stories on maths anxiety levels and learning outcomes.

6 CONCLUSIONS

The study finds that Gemini Story Book is the most effective of the tested AI platforms for creating mathematical stories for anxiety prevention, achieving the highest scores across all quality criteria. Automated video generation via Grok is unsuitable for maths education due to critical issues with the accuracy of automatic narration, which paraphrases or omits specific numerical values. Notebook LM can serve as an auxiliary tool for initial brainstorming and idea generation, but not for final content production due to systematic drift from mathematical goals. The developed five-step scientifically grounded workflow provides a practical balance between AI efficiency and pedagogical quality control, making the creation of personalised mathematical fairy tales practically feasible for individual educators without specialised technical skills or artistic abilities. Digital math fairy tales operationalise anxiety prevention through five interrelated mechanisms: emotional reattribution through identification, psychological safety through narrative distance, multiple representations that create distributed neural networks, positive associations through classical conditioning, and buffering intergenerational transmission of anxiety, confirming theoretical predictions of the integration of neurocognitive research, developmental theory, and bibliotherapy principles. The convergence of traditional narrative pedagogy rooted in fairy-tales with contemporary generative artificial intelligence technologies offers new opportunities to transform mathematics education through personalised and emotionally supportive storytelling at a scale previously unattainable in pre-digital contexts. Within this framework, technology functions as a means of augmenting rather than replacing pedagogical expertise. A sustainable model of integration positions educators as designers of narrative and instructional objectives, AI systems as tools facilitating content generation and technical implementation, and students as active constructors of mathematical understanding through emotionally meaningful narratives in psychologically safe learning environments.

REFERENCES

[1] S. S. Wu, M. Barth, H. Amin, V. Malcarne, and V. Menon, “Math anxiety in second and third graders and its relation to mathematics achievement,” Frontiers in Psychology, vol. 3, p. 162, 2012, [Online]. Available: https://doi.org/10.3389/fpsyg.2012.00162.

[2] B. Bettelheim, The Uses of Enchantment: The Meaning and Importance of Fairy Tales, New York: Vintage, 2010.

[3] J. M. Furner, “Using fairy tales and children’s literature in the math classroom: Helping all students become Einstein’s in a STEM world,” Journal of Advanced Education Research, vol. 2, no. 2, pp. 103-112, 2017, [Online]. Available: https://dx.doi.org/10.22606/jaer.2017.22006.

[4] M. Gómez-Martín, J. M. Prieto-Andreu, and L. Álvarez-Kurogi, “Storytelling vs. Digital Storytelling in Primary and Secondary Education: Systematic Review,” Ocnos, vol. 25, no. 1, 2026, [Online]. Available: https://doi.org/10.18239/ocnos_2026.25.1.574.

[5] W. M. F. W. M. Nasir, L. Halim, and N. M. Arsad, “Digital Storytelling Learning Outcomes and Critical Factors: A Scoping Review,” International Journal of Learning, Teaching and Educational Research, vol. 23, no. 5, pp. 323-344, 2024, [Online]. Available: https://doi.org/10.26803/ijlter.23.5.17.

[6] M. H. Ashcraft, “Math anxiety: Personal, educational, and cognitive consequences,” Current Directions in Psychological Science, vol. 11, no. 5, pp. 181-185, 2002.

[7] I. M. Lyons and S. L. Beilock, “Mathematics anxiety: Separating the math from the anxiety,” Cerebral Cortex, vol. 22, no. 2, pp. 2102-2110, 2012, [Online]. Available: https://doi.org/10.1093/cercor/bhr289.

[8] G. Ramirez, H. Chang, E. A. Maloney, S. C. Levine, and S. L. Beilock, “On the relationship between math anxiety and math achievement in early elementary school: The role of problem solving strategies,” Journal of Experimental Child Psychology, vol. 141, pp. 83-100, 2016, [Online]. Available: https://doi.org/10.1016/j.jecp.2015.07.014.

[9] M. McCrindle and A. Fell, “Understanding Generation Alpha,” McCrindle Research, Feb. 2020, [Online]. Available: https://generationalpha.com/wp-content/uploads/2020/02/Understanding-Generation-Alpha-McCrindle.pdf.

[10] G. Jumaev, D. Shukurov, K. Khudoyqulov and X. Kurbanov, "AI in Education: Revolutionizing Learning and Personalized Instruction," Proceedings of International Conference on Applied Innovation in IT, vol. 13, issue 5, pp. 401–410, 2025, [Online]. Available: https://doi.org/10.25673/122875.

[11] B. D. Buendia, N. M. D. Borbon, M. P. A. Lumbera, L. Manimtim, I. S. Pasahol, J. V. Abel and J. U. Sargento, "Impact of Technology Integration on Students' Knowledge, Attitudes, and Practices (KAP) Towards the Sustainable Development Goals (SDGs): A Structural Equation Modeling (SEM) Approach," Proceedings of International Conference on Applied Innovation in IT, vol. 13, issue 5, pp. 547–558, 2025, [Online]. Available: https://doi.org/10.25673/123079.

Methodological Aspects of Integrating and Mathematically Modeling a Cyber Range as an Innovative IT Solution for Cybersecurity Education in Higher Education Institutions

Roman Yaroviy[1], Olena Skliarenko[1], Serhii Yahodzinskyi[1], Yanina Kolodinska[1], Oleksandr Nikolaievskyi[1] and Hanna Tereshchuk[6]

[1]*Private Higher Educational Establishment "European University", Akademika Vernadskoho Blvd. 16V, 03115 Kyiv, Ukraine*

[6]*Taras Shevchenko National University of Kyiv, Volodymyrska Str. 60, 01033 Kyiv, Ukraine*

roman.yaroviy@e-u.edu.ua, olena.skliarenko@e-u.edu.ua, serhii.yahodzinskyi@e-u.edu.ua, yanina.kolodinska@gmail.com, a.nikolaievskyi@gmail.com, ganna.tereschuk@knu.ua

Keywords: Cyber Range, Cybersecurity Education, Cyber Defense, Information Technology, Stochastic Mathematical Model, Modeling, Performance Indicators, Optimization.

Abstract: The increasing complexity of cyber threats necessitates scalable and practice-oriented approaches to cybersecurity education. This paper presents the design and evaluation of a virtual cyber range integrated into bachelor's and master's cybersecurity programs at a Ukrainian university. The proposed platform is based on a multi-layer virtualized architecture that incorporates Red, Blue, and White team environments, enabling offensive, defensive, and administrative training within a unified infrastructure. The study combines competency-based curriculum mapping with Capture the Flag (CTF) exercises and pre-/post-training assessment. Experimental results demonstrate statistically significant learning improvements, with an average Learning Gain Index (LGI) above 0.6, confirming the effectiveness of the proposed approach in developing incident response, penetration testing, and secure system configuration skills. In addition to the architectural and educational contributions, the paper introduces a stochastic discrete-time dynamic model with Markov-modulated random coefficients to evaluate cyber range effectiveness under uncertainty. The model enables quantitative assessment of long-term performance, risk levels, and resource management strategies, accounting for variability in attack intensity, system reliability, and user behavior. The proposed model is formulated independently of a specific cyber range implementation, enabling its application to a broad class of cybersecurity training and simulation platforms. The results confirm that cyber range integration strengthens competency development and aligns with international practices in cybersecurity education.

1 INTRODUCTION

During the training of students majoring in F5 "Cybersecurity and Information Protection", a crucial stage is the acquisition of practical skills in the field of information security event monitoring and the implementation of protective measures under real-world conditions, where an organization's information system is subjected to an attack.

One of the most effective ways to ensure high-quality practical training is the implementation of a cyber range into the educational process.

Cyber ranges are widely recognized as effective platforms for realistic cyber training and experimentation.

Several mature cyber range platforms are already deployed worldwide, including governmental, academic, and commercial solutions [1].

These platforms offer comprehensive attack simulations, including: phishing, DDoS, malware deployment, attacks on network infrastructure, data confidentiality breaches [2], and more.

Modern cybersecurity systems increasingly rely on adaptive and data-driven approaches for detecting anomalies and cyber attacks. Clustering-based methods and expert systems are widely used for identifying abnormal patterns and supporting decision-making processes [3].

It is worth noting that a cyber range is an expensive and fully-featured system [4] that requires significant financial investment. For educational

institutions, especially those with limited budgets, such a solution may have a long return on investment period.

Therefore, the development of an internal cyber range based on open-source virtualization technologies requires coordinated technical, methodological, and organizational efforts.

A cyber range offers real opportunities for students, cadets, IT and cybersecurity professionals to practice on "training targets" - to acquire hands-on experience without the risk of causing actual harm to a real enterprise. At the same time, a typical enterprise network infrastructure (servers, workstations, network devices, corporate software, security tools, etc.) is reproduced in a virtual environment. This standard set of components enables the modeling of a corporate network.

2 RESEARCH METHODOLOGY

The study employed a mixed-methods approach combining comparative analysis of academic sources, case studies of cyber range deployment, empirical observation of student performance, survey-based self-assessment, competency-service mapping, and quantitative statistical evaluation of learning outcomes.

The empirical data were collected from system logs, user activity records, and assessment results generated within the cyber range environment. The study involved 48 students.

All data were anonymized and processed in accordance with institutional ethical guidelines.

2.1 Analysis of Cyber Range Deployment Principles

Based on market needs analysis, expectations of primary stakeholders, and the study of global experience (e.g., Cyber Range, NATO CCDCOE, IBM X-Force Command Center), it has been determined that a multi-layered architecture is the optimal approach for designing a modern cyber range. This approach ensures scalability, flexibility, and multi-purpose use of the infrastructure - for education, security system testing, R&D, and software certification.

At the top layer of the architecture, the user web portal provides unified access to platform services, while the underlying functional layer supports integrated training, cyber exercises, competitive simulations, vulnerability research, and experimental studies of emerging technologies (Fig. 1).

All functionality is deployed on top of a virtualization platform, which is physically hosted in a local infrastructure (data center). Interactions between layers are carried out via a service bus and orchestration layer, which enables dynamic replacement of individual components (e.g., changing the sector-specific training environment - from financial to energy, transport, etc.) [5].

The cyber range infrastructure flexibly redistributes computing resources between segments as needed, and a high-performance data transmission network connected to secure, special-purpose networks allow for remote lab access. A backup and recovery system ensures the highest degree of fault tolerance of the complex.

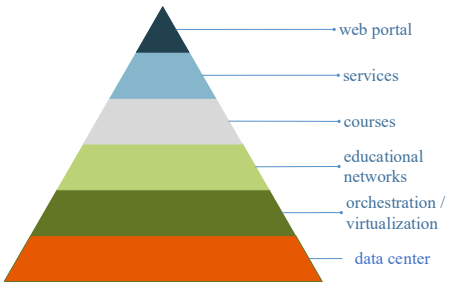


Figure 1: Multi-layered cyber range architecture.

The conceptual diagram of the “Cyber Range” virtual lab (Fig. 2) illustrates the structural design of a virtual training environment for cybersecurity. The scheme includes the Red Team segment (offensive side) and Blue Team segment (defensive side), a training target network, a cyber range control center, a core network interconnecting all components, and remote access tools. Each segment is isolated from the others using VLAN/routing but connected via a common network core to manage traffic flow [6]. All elements are labeled (in Ukrainian in the original) and linked to reflect the logical interaction of cyber range components.

The cyber range architecture (Fig. 2) consists of interconnected, isolated segments managed via a central Network Core and VLAN switching. The Red Team provides offensive capabilities using virtual machines equipped with penetration testing tools to simulate external threats, while the Blue Team utilizes traffic inspection and incident response systems to detect and mitigate simulated attacks in real time.

The primary target for these exercises is the Training Network Environment, which emulates realistic enterprise infrastructures - including DMZs and vulnerable internal services - enabling full attack lifecycle simulation [7]. Such architectures are widely

used for modeling attacks on industrial and mixed IT/OT environments and for evaluating layered defense strategies under realistic conditions [8]. The entire ecosystem is overseen by the Cyber Range Control Center (White Team), which handles virtualization orchestration, scenario deployment, and resource allocation. Finally, secure distributed participation is facilitated through a robust VPN Gateway, a model commonly adopted in cyber range architectures to enable secure collaboration without exposing internal resources to external networks [9].

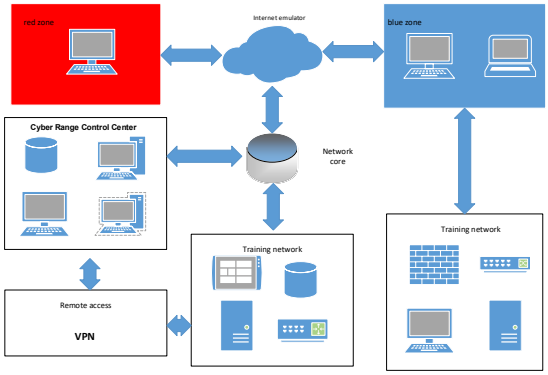


Figure 2: Cyber range virtual lab: conceptual scheme (cyber range laboratory - tecnalia).

2.2 Functional and Technical Integration of the Cyber Range

The cyber range represents an integrated hardware-software environment designed to support cybersecurity training, experimentation, and applied research through repeatable and controlled attack-defense scenarios [10], [11]. From a technical perspective, the platform relies on isolated execution environments, scalable resource management, and advanced monitoring and protection tools, including IDS/IPS, firewalls, and SIEM systems, implemented on top of virtualization infrastructures. Such architecture ensures realistic, reproducible, and secure experimentation conditions required for cybersecurity education [12] and system evaluation [9].

Functionally, the cyber range enables a wide spectrum of activities, including penetration testing, code analysis, vulnerability assessment, and security evaluation of both traditional IT systems and industrial infrastructures (SCADA/ICS). The separation of Red Team (offensive) and Blue Team (defensive) environments allows for structured simulation of cyber incidents and coordinated response scenarios.

By providing a controlled and configurable environment, the cyber range facilitates validation of cybersecurity tools, development of incident response strategies, and assessment of defense mechanisms against modern threats, including social engineering and phishing as dominant initial attack vectors [13]. The platform aligns with established frameworks such as MITRE ATT&CK and supports systematic analysis of attack vectors and defensive capabilities.

Importantly, the described functional and technical characteristics define the observable system parameters used in the proposed stochastic model, including resource utilization, incident response efficiency, and system resilience. This establishes a direct link between the cyber range infrastructure and the analytical evaluation framework introduced in this study.

2.3 Analysis of Cyber Range Integration into Cybersecurity Specialist Training

For students majoring in F5 “Cybersecurity and Information Protection”, the implementation of a cyber range enables the development of a wide range of general (GC) and professional (PC) competencies defined in the educational program. Through hands-on practice in a safe, simulated environment, students can apply acquired knowledge and gain critical professional skills [7].

The cyber range facilitates the development of key cybersecurity competencies, including technical, analytical, and collaborative skills, through realistic attack-defense simulations and tool-based exercises. The acquired competencies correspond to defined learning outcomes and internationally recognized training frameworks [14].

The cyber range was integrated into cybersecurity curricula as a practical training platform supporting hands-on skill development and competency assessment. The proposed educational model combines theoretical instruction with scenario-based simulations, enabling students to apply cybersecurity concepts in controlled yet realistic environments.

Practical exercises, including Red Team and Blue Team scenarios, contributed to the development of both technical and transversal competencies, such as incident analysis, teamwork, coordination, and critical thinking. These activities are aligned with program learning outcomes and support the achievement of key competencies in cybersecurity education (Table 1).

Table 1: Key performance indicators.

Indicator	Description
Success Rate (%)	Percentage of students who completed key tasks or earned certificates
Average Score	Mean score across training tasks, CTF events, and simulations
Attempt Count	Average number of attempts needed to complete a task successfully
Time to Completion	Average time spent completing tasks (as an efficiency indicator)
Pre-/Post-Test Score	Knowledge and skill level before and after cyber range training
Individual Ranking	Participant ranking based on their performance in training
Engagement Index	Percentage of students who completed all mandatory scenarios/blocks

Table 2: Assessment format: pre-/post-test comparison.

Group	Pre-test (avg)	Post-test (avg)	Δ (Gain)	% of Completed Cases	Avg. CTF Score
Bachelor (3rd year)	62%	84%	+22%	86%	78 points
Master (1st year)	74%	91%	+17%	93%	85 points

Knowledge Gain Analysis.

Knowledge gain Δ ≥15% is considered a significant improvement (Table 2).

Learning Gain Index (*LGI*):

$$LGI = \frac{(Post - Pre)}{(100 - Pre)}. \quad (1)$$

$LGI > 0.6$ indicates a high level of training effectiveness.

Statistical significance ($p < 0.05$) confirms a measurable impact of cyber range training. The difference between the initial and final results was statistically significant (t-test, $p < 0.05$, Cohen's d=0.8), indicating a large effect size.

The 95% confidence interval for the mean effectiveness index at the final stage was [0.78; 0.84].

Low attempt count with high success suggests confident mastery of skills.

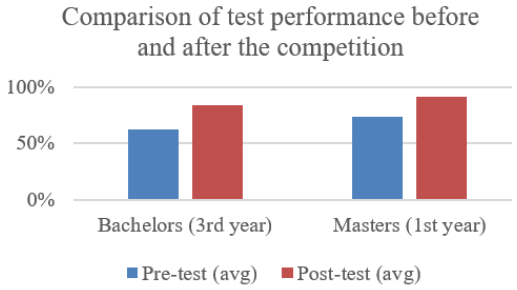


Figure 3: Comparative student test results chart.

Integrating the cyber range into bachelor's and master's programs in cybersecurity significantly enhances practical learning outcomes (Fig. 3). It develops essential competencies, fosters interdisciplinary education, and contributes to research-oriented training. The cyber range serves as a "real challenge arena with cutting-edge technology and professional growth opportunities," preparing competitive, well-rounded cybersecurity experts equipped to operate in modern cyber environments.

2.4 Construction of a Stochastic Model for Evaluating the Functioning of a Cyber Range

Unlike platform-specific evaluation approaches, the proposed stochastic discrete-time model is formulated in a generic form and is not limited to a particular cyber range implementation. The parameters of the model are derived from the functional and technical characteristics described in Section 2.2. The model can be applied to a wide class of cyber training and simulation environments, including academic cyber ranges, corporate SOC training platforms, and cyber-physical testbeds, where system performance is influenced by uncertainty in attack intensity, resource availability, and human factors.

Stochastic modeling makes it possible to describe systems whose operation depends on random factors. Instead of fixed parameters, random variables are used (normal, uniform, exponential, etc.). Similar approaches to modeling cybersecurity processes under uncertainty have been widely studied, supporting the relevance of combining probabilistic and data-driven methods in cyber range evaluation. This approach allows for accounting for unpredictability in the functioning of the cyber range (new types of attacks, human errors, equipment failures). A generalized integral assessment makes it possible to determine efficiency over a period of time as well as to build probabilistic scenarios [15], [16].

Let us consider the stages of constructing a stochastic discrete-time dynamic model with

Markov-modulated random coefficients to assess the effectiveness of a cyber polygon under conditions of uncertainty.

Step 1. Define the system input parameters. Let's introduce variables and notation for the system model:

- $k \in Z \geq 0$ - discrete time index (simulation step);
- $E_k \in R$ - aggregated effectiveness index of the cyber range at step k (can be a scalar or a vector of indicators);
- $U_k \geq 0$ - intensity of incoming events/attacks at step k;
- $R_k > 0$ - available resources (computational /network) at step k;
- α_k, β_k, γ_k - finite-state Markov random coefficients: α_k - response efficiency coefficient, β_k - coefficient (probability) of successful attack mitigation, γ_k - fault tolerance coefficient.

Step 2. Let's construct the dynamics of the performance state (a system of nonlinear difference equations). The system dynamics are represented by a nonlinear stochastic difference equation:

$$E_{k+1} = E_k + \alpha_k \frac{U_k \beta_k}{R_k(1+\gamma_k)} - \delta_k(E_k, U_k, R_k) + \eta_k \quad (2)$$

where: δ_k (.) - a deterministic or stochastic decay/saturation term (e.g., δ_k (E_k) or a saturation function S_k (E_k).

η_k, - a process noise term (e.g., $\eta_k \sim N(0, \sigma_\eta^2)$).

In vector form (if E_k is multidimensional):

$$E_{k+1} = f(E_k, U_k, \theta_k) + \eta_k, \quad (3)$$

with θ_k (α_k, β_k, γ_k).

Step 3. Introducing random coefficients. Since the functioning of the cyber range depends on random factors (human factor, unpredictable cyber threats), they can be modeled as random variables.

Markov Modeling of the Coefficients. Each coefficient evolves as a finite-state Markov chain. For instance, suppose α_k has n_α states $\{a_1, \ldots, a_{n\alpha}\}$, β_k has n_β states $\{b_1, \ldots, b_n\}$, and γ_k has n_γ states $\{g_1, \ldots, g_{n\gamma}\}$.

The corresponding transition matrices are defined as:

$$P_\alpha = \begin{pmatrix} p_{11}^{(\alpha)} & \cdots & p_{1n_\alpha}^{(\alpha)} \\ \vdots & \ddots & \vdots \\ p_{n_\alpha 1}^{(\alpha)} & \cdots & p_{n_\alpha n_\alpha}^{(\alpha)} \end{pmatrix}, \quad P_\beta = \begin{pmatrix} p_{11}^{(\beta)} & \cdots & p_{1n_\beta}^{(\beta)} \\ \vdots & \ddots & \vdots \\ p_{n_\beta 1}^{(\beta)} & \cdots & p_{n_\beta n_\beta}^{(\beta)} \end{pmatrix},$$

$$P_\gamma = \begin{pmatrix} p_{11}^{(\gamma)} & \cdots & p_{1n_\gamma}^{(\gamma)} \\ \vdots & \ddots & \vdots \\ p_{n_\gamma 1}^{(\gamma)} & \cdots & p_{n_\gamma n_\gamma}^{(\gamma)} \end{pmatrix}, \quad (4)$$

where in each row $\sum_j p_{i,j}^{(.)}$.

The model structure is independent of specific cyber range architectures and toolsets, as its state variables and stochastic coefficients can be adapted to different operational contexts. By appropriate parameterization, the model supports evaluation of diverse scenarios, ranging from small-scale educational cyber ranges to large distributed training infrastructures, thereby enabling comparative analysis and benchmarking across platforms.

Based on the proposed model, optimization problems for cyber range management can be formulated. Similar integrated decision-support approaches for managing information protection systems are widely used, where analytical models support decision-making under uncertainty. For example, the allocation of limited resources R_k may be optimized to maximize the expected effectiveness $E[E_k]$ under budget constraints, or to minimize the probability of exceeding predefined risk thresholds. Potential approaches include stochastic programming, dynamic programming methods, and policies based on partially observable Markov decision processes, when system states are partially hidden. Such approaches are consistent with modern decision-support methodologies for information protection management and support the applicability of the proposed model. Thus, the proposed mathematical model makes it possible to evaluate the effectiveness of the cyber range, where α, β, γ are random coefficients that account for efficiency, success, and fault tolerance.

Here is a brief step-by-step modeling algorithm (practical implementation of the given model).

1) Initialization. Set the initial value E_0 and select initial states for α_0, β_0, γ_0.
2) Iteration (for $k = 0, \ldots, K-1$):
   a) Generate new states α_{k+1}, β_{k+1}, γ_{k+1} according to their respective transition matrices P_α, P_β, P_γ.
   b) Specify or update U_k, R_k (deterministically or as stochastic processes).
   c) Compute E_{k+1} using the nonlinear difference equation.
   d) Optionally, generate the observation y_{k+1} using the measurement model.
3) Repetition. Repeat steps 1-2 for a large number of realizations and compute ensemble statistics (mean $E[E_k]$, variance $Var(E_k)$ quantiles, and the probability of risk threshold violation $P(E_k < E_{crit})$.

2.5 Model Implementation in a Virtualized Cyber Range Environment

To demonstrate the practical applicability of the proposed stochastic model, a pilot implementation was carried out within a virtualized cyber range environment deployed on a hybrid infrastructure based on Proxmox VE and VMware. The experimental setup simulated a university-level cybersecurity training scenario involving 48 students (32 bachelor's and 16 master's level participants) over a 6-week training period.

The cyber range environment consisted of three main segments: a Red Team environment with penetration testing tools (Kali Linux, Metasploit), a Blue Team environment equipped with monitoring and defense systems (SIEM, IDS/IPS), and a virtual enterprise network containing vulnerable services and simulated business processes. The infrastructure included 36 virtual machines distributed across 3 physical servers, with dynamic resource allocation based on workload intensity.

For model parameterization, the following key indicators were collected: task completion rate, average response time to incidents, number of successful attack mitigations, system resource utilization (CPU, RAM), and user activity logs. The aggregated effectiveness index was normalized in the range [0,1], where higher values correspond to improved training performance.

The stochastic coefficients of the model were calibrated using observed data. This calibration ensures consistency between the theoretical model and real-world system behavior. Specifically, the response efficiency coefficient (α) varied within the interval [0.6; 0.85], the attack mitigation probability (β) within [0.5; 0.8], and the fault tolerance coefficient (γ) within [0.7; 0.9], depending on system load and user behavior. The transition of these coefficients was modeled as a three-state Markov chain (low, medium, high performance states), with empirically estimated transition probabilities.

Simulation experiments were conducted for 1000 realizations over 50 discrete time steps. The results showed that the expected effectiveness index increased from 0.58 at the initial stage to 0.81 at the final stage of training, confirming a positive learning dynamic. This indicates that the model adequately reflects the process of competency development in a dynamic learning environment. The variance decreased over time, indicating stabilization of student performance and system behavior. The probability of critical performance degradation (effectiveness < 0.5) was reduced from 0.22 to 0.07 after training adaptation.

A comparative analysis of infrastructure configurations demonstrated that environments with higher resource elasticity (Proxmox-based dynamic allocation) exhibited up to 12% higher effectiveness growth compared to static configurations. This confirms the sensitivity of the model to virtualization parameters and supports its use for evaluating infrastructure impact on learning outcomes.

The obtained results validate the applicability of the proposed model for analyzing cyber range performance under uncertainty and demonstrate its potential for optimizing both educational processes and technical infrastructure configurations.

3 CONCLUSIONS

This study demonstrates that the integration of a virtual cyber range into bachelor's and master's cybersecurity programs significantly enhances practice-oriented learning and professional readiness. The proposed multi-layer cyber range architecture supports diverse educational formats, including guided training, autonomous exercises, Red/Blue team scenarios, CTF competitions, and system testing, providing a realistic and controllable environment for skill development.

The applied research methodology-combining curriculum analysis, competency-service mapping, empirical training data, and statistical evaluation - proved effective for assessing educational outcomes. Pre-/post-test analysis and performance indicators confirmed substantial learning gains, particularly in incident response, vulnerability assessment, security monitoring, and risk analysis. Team-based and gamified scenarios showed the highest educational impact, contributing to both technical competencies and collaboration skills essential for SOC and incident response roles.

The competency-cyber range service matrix validated that the platform directly supports most learning outcomes defined in cybersecurity curricula, while also enabling interdisciplinary integration across system administration, programming, cryptography, and digital forensics. The proposed approach is consistent with international cyber defense training practices adopted by NATO CCDCOE, SANS, and IBM X-Force Command, confirming its relevance and scalability [17], [18].

A key scientific contribution is the proposed stochastic discrete-time dynamic model with Markov-modulated coefficients, which provides a

generic and extensible framework for evaluating the effectiveness of cyber ranges and related cybersecurity training platforms under uncertainty. The model supports long-term performance assessment, risk estimation, and resource optimization, bridging educational experimentation with formal analytical frameworks.

Overall, the results confirm that cyber ranges represent an effective, scalable, and analytically grounded solution for modern cybersecurity education and applied research. Thus, the proposed model is not only theoretically grounded but also empirically validated based on real-world educational data, confirming its applicability for evaluating and optimizing cybersecurity training environments.

REFERENCES

[1] "Cyber training center," State Center for Cyber Protection, [Online]. Available: https://scpc.gov.ua/uk/cyber-trainer/.

[2] M. Gering, Y. Medianyk, A. Sapeha, S. Trendov, K. Karpov, H. Skaskiv, D. Kachan and E. Siemens, "Botnet Simulation and Observation Framework for AI-Driven Virtual Personas," Proceedings of International Conference on Applied Innovation in IT, vol. 13, issue 5, pp. 75–82, 2025, [Online]. Available: https://doi.org/10.25673/122810.

[3] A. J. Ajeel and J. Kh-Madhloom, "AI-Based Intrusion Detection for Smart Grid Security," Proceedings of International Conference on Applied Innovation in IT, vol. 13, issue 5, pp. 41–47, 2025, [Online]. Available: https://doi.org/10.25673/122806.

[4] V. Kampourakis, V. Gkioulos, and S. Katsikas, "A step-by-step definition of a reference architecture for cyber ranges," J. Inf. Secur. Appl., vol. 88, art. no. 103917, 2025, [Online]. Available: https://doi.org/10.1016/j.jisa.2024.103917.

[5] D. Arnold, J. Ford, and J. Saniie, "Architecture of an efficient environment management platform for experiential cybersecurity education," Information, vol. 16, no. 7, art. no. 604, 2025, [Online]. Available: https://doi.org/10.3390/info16070604.

[6] J. V. Bistene, A. F. P. dos Santos, C. E. das Chagas, and R. M. Salles, "Modeling network traffic generators for cyber ranges: A systematic literature review," J. Netw. Syst. Manage., vol. 33, no. 2, 2025, [Online]. Available: https://doi.org/10.1007/s10922-025-09901-8.

[7] L. Arsenovych, O. Nikolaievsky, O. Skliarenko, L. Lytvynenko, and I. Kydriavskyi, "Organization of training with the use of digital technologies for ensuring cybersecurity in the educational space," WSEAS Trans. Comput. Res., vol. 12, pp. 524-536, 2024, [Online]. Available: https://doi.org/10.37394/232018.2024.12.51.

[8] J. Christopher, "The 2024 state of ICS/OT cybersecurity: Our past and our future," SANS Institute, Oct. 16, 2024, [Online]. Available: https://www.sans.org/blog/the-2024-state-of-ics-ot-cybersecurity-our-past-and-our-future/.

[9] V. L. Buriachok, V. B. Tolubko, V. O. Khoroshko, and S. V. Toliupa, Information and Cybersecurity: The Socio-Technical Aspect, Kyiv: DUT, 2015, 288 p.

[10] J. M. Castillo, M. Parenthoen, and N. Louveton, "A theoretical model for enhancing cyber crisis training in cyber ranges: Bridging cognitive load, situational awareness, and immersion," Cogn. Technol. Work, 2025, [Online]. Available: https://doi.org/10.1007/s10111-025-00844-3.

[11] I. Lateş, C. Boja, and A. Zamfiroiu, "Multi-technology infrastructure for advanced training and testing in cyber range systems," Stud. Inform. Control, vol. 34, no. 3, pp. 17-28, 2025, [Online]. Available: https://doi.org/10.24846/v34i3y202502.

[12] E. S. Alomari, M. B. M. Alkorani and K. Alieyan, "Securing the Post-Quantum Internet: A Comparative Study of Quantum-Resistant Protocols and Migration Strategies," Proceedings of International Conference on Applied Innovation in IT, vol. 13, issue 5, pp. 29–40, 2025, [Online]. Available: https://doi.org/10.25673/122804.

[13] M. Albaladejo-González, P. Nespoli, F. Gómez Mármol, and J. A. Ruipérez-Valiente, "A multimodal and adaptive gamified system to improve cybersecurity competence training," Cluster Computing, vol. 28, no. 9, 2025, [Online]. Available: https://doi.org/10.1007/s10586-025-05264-6.

[14] W. Lazarov, T. Schafeitel-Tähtinen, J. Squillace, Z. Martinasek, A. Coufalikova, M. Helenius, et al., "Lessons learned from using cyber range to teach cybersecurity at different levels of education," Technol. Knowl. Learn., 2025, [Online]. Available: https://doi.org/10.1007/s10758-025-09840-y.

[15] V. Lakhno, H. Mohylnyi, V. Donchenko, O. Smahina, and M. Pyroh, "A model developed for teaching an adaptive system of recognising cyberattacks among non-uniform queries in information systems," Eastern-European J. Enterp. Technol., vol. 4, no. 9(82), p. 27, 2016, [Online]. Available: https://doi.org/10.15587/1729-4061.2016.73315.

[16] V. A. Lakhno, D. Y. Kasatkin, O. V. Skliarenko, and Y. O. Kolodinska, "Modeling and optimization of discrete evolutionary systems of information security management in a random environment," in Machine Learning and Autonomous Systems, Singapore: Springer, 2022, pp. 9-22, [Online]. Available: https://doi.org/10.1007/978-981-16-7996-4_2.

[17] J. Diakoumakos, E. Chaskos, N. Kolokotronis, and G. Lepouras, "Cyber-security gamification in federation of cyber ranges: Design, implementation, and evaluation," Int. J. Inf. Secur., vol. 24, no. 1, 2025, [Online]. Available: https://doi.org/10.1007/s10207-024-00974-1.

[18] L. O. Nweke, U. F. Okebanama, and G. U. Mba, "Enhancing entrepreneurial skills through experiential learning in IoT, AI, and cybersecurity," Discov. Educ., vol. 4, no. 1, 2025, [Online]. Available: https://doi.org/10.1007/s44217-025-00573-9.

[19] T. Schafeitel-Tähtinen and W. Lazarov, "Teaching and learning cybersecurity using capture the flag: Effectiveness comparison between university students in Finland and Czechia," Comput. Appl. Eng. Educ., vol. 33, no. 5, 2025, [Online]. Available: https://doi.org/10.1002/cae.70082.

Technology-Driven Framework for Alumni Engagement: Extending the Information System Success Model with Altruism and User Attractiveness in a Graduate Tracer System

Cherly Baja Sardovia
Department of Computer Studies, College of Information Technology Education, North Eastern Mindanao State University, Rosario, 8300 Tandag City, Philippines
cbsardovia@nemsu.edu.ph

Keywords: Alumni Engagement, Graduate Tracer System, ISSM, Altruism, User Attractiveness, Structural Equation Modeling.

Abstract: This study evaluated the factors influencing alumni engagement with the NEMSU Graduate Tracer System by extending the DeLone and McLean Information Systems Success Model with altruism, social influence, and user attractiveness. The relevance of this work lies in combining technical system analysis, including system quality, information quality, service quality, and user satisfaction, with socio-psychological constructs that may sustain alumni participation. A quantitative design was used, and data were collected from 117 BS Computer Science graduates through a structured questionnaire. Structural Equation Modeling using WarpPLS 8.0 was applied to examine the hypothesized relationships among system-related and user-related constructs. Results show that system quality, information quality, service quality, and user satisfaction influence user intention and perceived system attractiveness. Altruism also emerged as an important engagement-related construct, indicating that alumni participation is supported not only by technical functionality but also by the perceived value of contributing to the institution. Social influence showed a positive but comparatively weak significant effect on user intention with beta = 0.12 and $p < 0.01$, suggesting that peer or institutional encouragement may contribute to participation but is less influential than system functionality, satisfaction, and personal value. The study is limited by its sample size and focus on one department; therefore, broader validation with larger and more diverse alumni groups is recommended. The findings suggest that sustainable Graduate Tracer Systems should balance technical reliability, usable design, emotional appeal, and prosocial motivation to strengthen long-term alumni engagement.

1 INTRODUCTION

In today's educational environment, Graduate Tracer Systems are essential for higher education institutions because they support alumni career monitoring, curriculum review, accreditation evidence, employability reporting, and institutional decision-making. These systems allow universities to track graduates after completion of their academic programs and to use alumni information as feedback for improving program relevance and service delivery.

Despite their importance, many Graduate Tracer Systems face challenges related to low participation, outdated records, and limited sustained engagement. Earlier studies commonly emphasize technical attributes such as system quality, information accuracy, usability, and service support. However, alumni engagement is not only a technical issue. It also involves social and socio-psychological factors, including alumni willingness to give back to the institution, emotional connection, and the perceived attractiveness of the system interface.

The relevance of this work stems from the combination of technical analysis of systems with the study of socio-psychological factors, such as altruism and emotional appeal, which can help ensure sustained engagement with the platform. By extending the DeLone and McLean Information Systems Success Model with altruism and user attractiveness, this study provides a broader

explanation of why alumni may continue to interact with a Graduate Tracer System even when participation is voluntary.

The aim of this research is to examine the factors influencing alumni engagement with the NEMSU Graduate Tracer System using Structural Equation Modeling. Specifically, the study analyzes the relationships among system quality, information quality, service quality, user satisfaction, social influence, altruism, user attractiveness, and user intention. The study also acknowledges that its empirical scope is limited to 117 BS Computer Science alumni; therefore, its findings should be interpreted as a department-level validation rather than as a direct generalization to the entire university or to all higher education institutions.

2 LITERATURE REVIEW

Recent studies emphasize the importance of alumni engagement in higher education, particularly through Graduate Tracer Systems. These systems help track career progression and assess academic program effectiveness, but challenges such as low alumni participation and outdated data persist. A recent alumni tracer and management system study showed that data analytics can enhance alumni engagement and institutional data management by improving communication, monitoring, and information access [1].

Social influence has also been examined as a factor in alumni engagement. Jack et al. investigated the role of social influence in alumni engagement by examining behavioral intention and digital platform usage [2]. Their work supports the idea that alumni participation may be shaped by encouragement from peers, institutional networks, and professional communities. However, the influence of social pressure or encouragement may vary depending on whether alumni also perceive the system as useful, reliable, and personally meaningful.

The Information Systems Success Model introduced by DeLone and McLean provides a widely used framework for evaluating information system effectiveness [3]. The model emphasizes system quality, information quality, service quality, use or intention to use, user satisfaction, and net benefits as important dimensions of system success. Subsequent reviews and applications of the model have shown that it remains relevant in different information system contexts, including education, e-learning, e-government, and online service platforms [4], [5].

Studies on technology acceptance further suggest that social influence, performance expectancy, effort expectancy, and facilitating conditions can affect user intention and use behavior [6]. In technology-supported services, user satisfaction also plays an important role in explaining continuance intention, especially when the system is perceived as useful, reliable, and responsive to user needs [7]. These findings support the inclusion of user satisfaction and user intention as central constructs in the proposed Graduate Tracer System engagement framework.

User experience and system attractiveness are also important in sustaining digital engagement. Visual design aesthetics, ease of use, and interface appeal can influence perceived usefulness, enjoyment, and loyalty intentions in digital environments [8], [9]. In the context of a Graduate Tracer System, user attractiveness may encourage alumni to revisit the system, update their information, and perceive the platform as more credible and engaging.

Altruistic motivation is another relevant factor because alumni data sharing is often voluntary. Online prosocial behavior studies show that people may participate in digital platforms not only for personal benefit but also to help others, support communities, and contribute to shared goals [10]. For alumni systems, this means that graduates may update their information because they believe their participation can help the university improve curriculum, services, accreditation evidence, and student career support.

Finally, research on Structural Equation Modeling provides methodological support for analyzing complex relationships among latent constructs [11]. Since the present study examines multiple direct relationships among system-related and user-related variables, SEM is appropriate for evaluating the proposed framework. Taken together, the reviewed literature supports the integration of classical system success constructs with socio-psychological variables in explaining alumni engagement with Graduate Tracer Systems.

3 METHODOLOGY

This section describes the research design, participants, instrument, and data analysis procedure used to evaluate the proposed alumni engagement framework. The methodology was designed to test both system-related and user-related factors while also clarifying the scope and limitations of the dataset.

3.1 Research Design and Participants

This study employed a quantitative research design guided by the Information Systems Success Model and the Altruism-Based Model. Data were collected from 117 BS Computer Science graduates who had experience with, or were targeted users of, the NEMSU Graduate Tracer System. The questionnaire also gathered respondent profile data, including gender, age group, and program affiliation, to provide descriptive context for the sample. The use of SEM increased the analytical complexity of the study because it required testing multiple relationships among latent constructs.

The sample size and respondent coverage are important limitations. Although the sample provided an initial basis for department-level model testing, SEM studies commonly benefit from larger samples, and a sample of at least 200 respondents is often preferred for stronger statistical stability. Since the participants came from one department, the results should not be generalized to the entire university, to other academic programs, or to all higher education institutions without further validation using a larger and more diverse alumni sample.

3.2 Instrument and Example Questionnaire Items

A structured questionnaire was used to measure the proposed constructs. The instrument covered system quality, information quality, service quality, user satisfaction, social influence, altruism, user attractiveness, and user intention. Responses were interpreted using a Likert-type scale to capture the level of alumni agreement with each statement. The following examples show the questionnaire items used to operationalize the constructs:

1) System Quality. The Graduate Tracer System is easy to access and navigate.
2) Information Quality. The information provided by the system is accurate, relevant, and updated.
3) Service Quality. The support or assistance related to the system is responsive and reliable.
4) User Satisfaction. Overall, I am satisfied with my experience using the Graduate Tracer System.
5) Social Influence. People important to me encourage me to use the Graduate Tracer System.
6) Altruism. I am willing to update my alumni information to help the university improve its programs.
7) User Attractiveness. The design and interface of the system are visually appealing.
8) User Intention. I intend to continue using or updating information in the Graduate Tracer System.

3.3 Data Analysis and Scope Delimitation

Structural Equation Modeling using WarpPLS 8.0 was used to analyze the hypothesized relationships among the constructs. Descriptive statistics summarized alumni perceptions of the Graduate Tracer System constructs, while model fit and quality indices were used to assess the overall adequacy of the structural model. Reliability and validity checking were also considered as part of the model evaluation process.

The study did not perform a separate SEM comparison by sex or graduation year because the available manuscript data were reported primarily at the aggregate construct level. This limits the interpretation of possible differences between men and women or between alumni cohorts. Future validation should include sex-disaggregated and cohort-based statistical analysis to determine whether engagement factors vary across demographic groups and graduation years.

4 PROPOSED MODEL AND CONSTRUCTS

The proposed model integrates the Information Systems Success Model with the Altruism-Based Model to provide a comprehensive framework for understanding alumni engagement. System Quality, Information Quality, Service Quality, and User Satisfaction are expected to influence User Intention to use the system. Altruism is hypothesized to affect User Intention by encouraging alumni to participate for communal and prosocial reasons, while User Attractiveness is expected to reflect the influence of satisfaction and design appeal on continued engagement.

The constructs are interconnected as follows:

1) High System Quality, Information Quality, and Service Quality are expected to strengthen User Satisfaction and User Intention;
2) Satisfied users are more likely to perceive the system as attractive and worth using continuously;
3) Altruistic motivations are expected to encourage alumni to contribute voluntarily to

the Graduate Tracer System even without direct personal benefit;

4) Social influence is expected to contribute to user intention, although its strength may vary depending on personal motivation and perceived system value.

4.1 System Quality

This construct refers to the technical performance of the Graduate Tracer System, including usability, reliability, accessibility, and responsiveness. High system quality is expected to improve user satisfaction and strengthen alumni intention to engage with the platform because users generally prefer systems that are efficient, stable, and easy to navigate [3], [5].

4.2 Service Quality

This construct refers to the responsiveness, competence, and reliability of the support services associated with the system. Effective support can foster trust and satisfaction, thereby increasing the likelihood of continued use and engagement with the Graduate Tracer System [3], [7].

4.3 User Satisfaction

This construct measures the extent to which alumni are pleased with their experience using the system. User satisfaction is a key predictor of continued use and is influenced by system quality, service quality, information quality, and the perceived value derived from system interaction [3], [7], [12].

4.4 Altruism

This construct emphasizes the voluntary and prosocial behavior of alumni who participate in the system not only for personal benefit but also to contribute to the institution and support future cohorts. In the context of Graduate Tracer Systems, altruism is important because alumni data sharing is largely voluntary and depends on the perceived value of giving back to the university community [10].

4.5 Information Quality

This construct refers to the accuracy, completeness, timeliness, and relevance of information generated by the Graduate Tracer System. Accurate and useful information may increase alumni confidence in the system and strengthen their intention to participate [3], [4].

4.6 Social Influence

This construct captures the extent to which peers, faculty members, institutional offices, and alumni networks encourage graduates to use the system. Although social influence may motivate participation, its effect is expected to vary depending on the level of personal motivation and perceived system value [2], [6].

4.7 User Intention

This construct represents alumni willingness and behavioral intention to use the Graduate Tracer System. It serves as a central outcome of system-related and user-related factors and reflects the likelihood of continued participation [6], [7].

4.8 User Attractiveness

This construct refers to the visual appeal, interface clarity, and overall design attractiveness of the system. A visually appealing and easy-to-use system is expected to enhance satisfaction and encourage continued alumni engagement [8].

Figure 1 presents the proposed conceptual framework integrating ISSM and altruism-based constructs for Graduate Tracer System adoption. The framework illustrates how technical, service-related, and user-centered constructs interact to explain alumni engagement with the Graduate Tracer System.

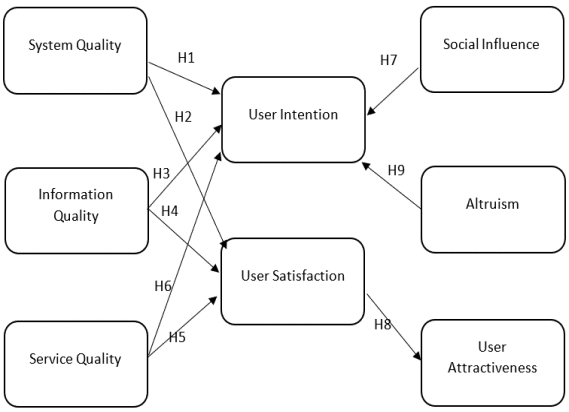


Figure 1: Proposed conceptual framework integrating ISSM and altruism-based constructs for GTS adoption.

As shown in Figure 1, the structural equation model illustrates the relationships between key

constructs influencing alumni engagement in the Graduate Tracer System. System Quality, Information Quality, and Service Quality are shown as independent variables that directly or indirectly affect User Intention and User Satisfaction. These paths emphasize that a reliable system, accurate information, and responsive support services are crucial for encouraging alumni participation.

User Intention plays a central role in the model, mediating the relationships between other constructs and outcomes such as Altruism and User Attractiveness. The model highlights the importance of both functional and emotional drivers, suggesting that institutions should focus on enhancing technical quality and user experience to foster greater alumni engagement.

5 RESULTS AND DISCUSSION

This section presents the descriptive and structural model findings of the study. The discussion first examines alumni perceptions of the Graduate Tracer System constructs, followed by the structural model results, the interpretation of respondent-scope limitations, and the model fit indices that support the proposed framework.

5.1 Alumni Perception of GTS Constructs

Figure 2 summarizes alumni perceptions across the major Graduate Tracer System constructs. The visual presentation helps identify which constructs received the strongest ratings and which areas show comparatively lower but still positive evaluations.

Figure 2 illustrates the alumni perceptions of the Graduate Tracer System constructs, with all constructs receiving high mean scores, indicating strong positive feedback. Information Quality received the highest rating at 4.45, followed by Altruism at 4.42 and System Quality at 4.39. These findings suggest that alumni find the system information reliable, timely, and helpful, and that they are motivated to engage with the system for prosocial reasons.

The standard deviations, ranging from 0.66 to 0.73, indicate moderate variability in responses, with Social Influence showing the greatest variability at 0.73. This suggests that while most alumni share strong agreement on system quality and user experience, there is some variation in perceptions regarding the social factors influencing Graduate Tracer System participation.

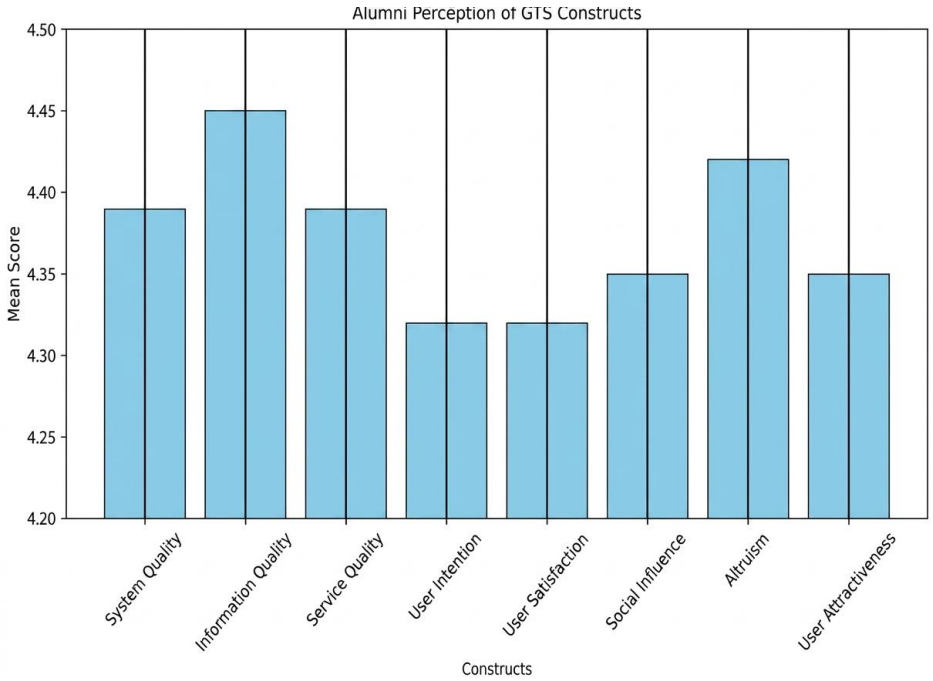


Figure 2: Alumni perception of GTS constructs..

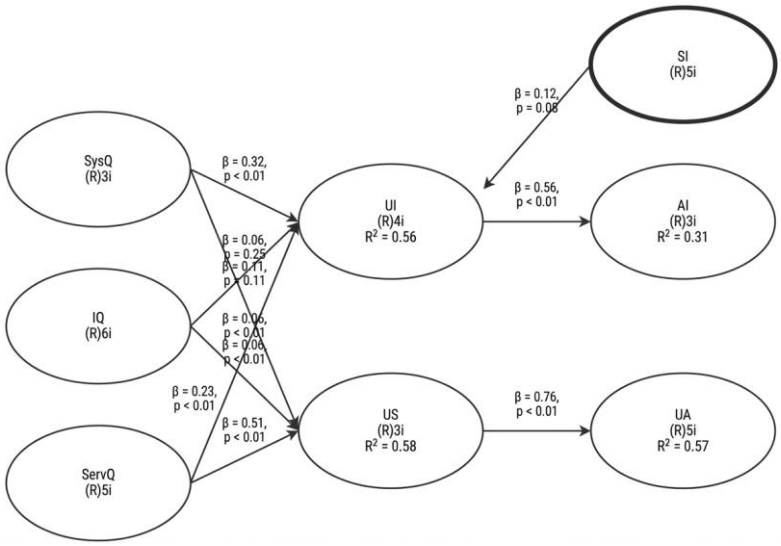


Figure 3: Structural model with beta coefficients.

5.2 Structural Model Results

Figure 3 presents the structural model with beta coefficients. This figure shows the direct paths among the system-related and user-related constructs and provides the basis for interpreting the strength and direction of the hypothesized relationships.

Figure 3 presents the results of the Structural Equation Modeling analysis, highlighting the relationships between key constructs that influence alumni engagement in the Graduate Tracer System. System Quality significantly impacts User Intention with beta = 0.32 and $p < 0.01$, and User Satisfaction with beta = 0.51 and $p < 0.01$. User Intention also strongly influences Altruism with beta = 0.56 and $p < 0.01$. Information Quality positively affects User Intention with beta = 0.36 and $p < 0.01$ and has a moderate impact on User Satisfaction with beta = 0.51 and $p < 0.01$. Service Quality shows the strongest effect on User Intention with beta = 0.51 and $p < 0.01$ and User Satisfaction with beta = 0.51 and $p < 0.01$.

Social Influence has a positive but comparatively weak significant effect on User Intention with beta = 0.12 and $p < 0.01$. This interpretation resolves the inconsistency between the earlier abstract and the results: Social Influence was not the strongest predictor, but it was statistically significant in the reported SEM output. Therefore, the finding should be understood as weak but significant rather than non-significant. Moreover, User Satisfaction has a strong influence on User Attractiveness with beta = 0.76 and $p < 0.01$.

5.3 Respondent Scope and Group Analysis Considerations

Table 1 presents the descriptive respondent profile used to address the reviewer recommendation to report statistical data for men and women and to clarify the composition of the sample. The available dataset describes the respondents by gender, age group, and program affiliation.

Table 1: Respondents' demographic profile.

Category	Subcategory	f	%
Gender	Female	66	56.4
Gender	Male	46	39.3
Gender	Prefer not to disclose	5	4.3
Age Group	20-25 years old	63	53.8
Age Group	25-30 years old	42	35.9
Age Group	35-40 years old	12	10.3
Program	BS Computer Science	117	100

Note. Frequencies were rounded from the reported percentage distribution of 117 valid BS Computer Science alumni respondents. The available profile data are descriptive and were not used for separate SEM subgroup analysis.

The demographic profile indicates that the sample was composed mostly of female alumni, 56.4%, followed by male alumni, 39.3%, and respondents who preferred not to disclose their gender, 4.3%. In terms of age, more than half of the respondents were 20-25 years old, 53.8%, followed by those aged 25-30 years old, 35.9%, and 35-40 years old, 10.3%. All respondents were BS Computer Science graduates, 100%, which confirms that the findings represent a department-level dataset rather than a university-wide alumni population.

Although the study reports respondent profile statistics by gender and age group, the available dataset does not include sex-disaggregated SEM results or graduation-year-based analysis. Therefore, the findings should not be interpreted as evidence of gender-based or cohort-based differences in alumni engagement. Future studies should collect broader demographic data, including sex, graduation year, employment status, and program affiliation, and conduct subgroup analysis to determine whether engagement factors vary across different alumni groups.

5.4 Model Fit and Quality Indices

Table 2 presents the model fit and quality indices used to evaluate the adequacy of the SEM output. The table is referenced before presentation to comply with the required structure that a heading and introductory text must precede any table.

Table 2: Model fit and quality indices.

Index	Value	Criterion
APC	0.335, $p < 0.001$	$p < 0.05$
ARS	0.506, $p < 0.001$	$p < 0.05$
AARS	0.497, $p < 0.001$	$p < 0.05$
AVIF	2.362	<= 5, ideal <= 3.3
AFVIF	3.000	<= 5, ideal <= 3.3
Tenenhaus GOF	0.609	Large >= 0.36
SPR	1.000	>= 0.7, ideal = 1
RSCR	1.000	>= 0.9, ideal = 1
SSR	1.000	>= 0.7
NLBCDR	1.000	>= 0.7

Table 2 displays the goodness-of-fit indices and provides a comprehensive evaluation of the model performance. The Average Path Coefficient of 0.335 with p-value < 0.001 suggests that the paths between constructs are significant. The Average R-Squared value of 0.506 and Average Adjusted R-Squared value of 0.497, both with p-values < 0.001, indicate that the model explains a substantial portion of variance in the constructs and demonstrates moderate predictive power.

The Average Block VIF of 2.362 and Average Collinearity VIF of 3.000 are within the acceptable range, indicating that multicollinearity is not a major concern. The Tenenhaus GOF score of 0.609 falls within the large effect size category. The Sympson's Paradox Ratio, R-Squared Contribution Ratio, Statistical Suppression Ratio, and Nonlinear Bivariate Causality Direction Ratio all scored 1.000, indicating no paradoxical relationships, no suppression effects, and well-established causal directions.

6 CONCLUSIONS

This study provides insights into the factors that influence alumni engagement with the Graduate Tracer System at NEMSU. By extending the DeLone and McLean Information Systems Success Model with altruism and user attractiveness, it presents a more comprehensive understanding of alumni participation. The findings indicate that system quality, information quality, service quality, user satisfaction, altruism, and user attractiveness play important roles in explaining alumni engagement.

The revised interpretation clarifies that Social Influence has a positive but weak significant effect on User Intention, not a non-significant effect. Thus, peer or institutional encouragement may still contribute to alumni participation, but its influence appears weaker than functional system quality, satisfaction, and the perceived value of contributing to the institution. This correction aligns the abstract, results, and conclusion.

The study is limited by the sample size of 117 BS Computer Science graduates and by its focus on one department. Therefore, the findings should be treated as an initial validation of the model rather than as evidence that can be generalized to all programs, institutions, or countries. Future studies should use larger samples, preferably at least 200 respondents or more, include alumni from multiple departments and campuses, and conduct subgroup analysis by sex and graduation year.

By adopting the proposed framework with these limitations in mind, institutions can improve alumni engagement by balancing technical reliability, service support, emotional appeal, and prosocial motivation. Future validation will strengthen the applicability of the model and support the development of more sustainable and inclusive Graduate Tracer Systems.

REFERENCES

[1] H. J. E. Trillano, R. S. Prudente, and L. A. Mondejar, "Alumni tracer and management system with data analytics: Enhancing alumni engagement and data management for South East Asian Institute of Technology, Inc.," International Journal of Research and Innovation in Social Science, vol. 9, no. 4, pp. 5142-5155, Apr. 2025, [Online]. Available: https://doi.org/10.47772/IJRISS.2025.90400397.

[2] S. Jack, A. Bujang, J. K. Ali, and L. Andrew, "The role of social influence in alumni engagement: Examining behavioural intention and digital platform usage," International Journal of Research and Innovation in Social Science, vol. 9, no. 3, pp. 1340-1349, Mar. 2025, [Online]. Available: https://doi.org/10.47772/IJRISS.2025.90300106.

[3] W. H. DeLone and E. R. McLean, "The DeLone and McLean model of information systems success: A ten-year update," Journal of Management Information Systems, vol. 19, no. 4, pp. 9-30, 2003, [Online]. Available: https://doi.org/10.1080/07421222.2003.11045748.

[4] A. Jeyaraj, "DeLone and McLean models of information system success: Critical meta-review and research directions," International Journal of Information Management, vol. 54, art. no. 102139, 2020, [Online]. Available: https://doi.org/10.1016/j.ijinfomgt.2020.102139.

[5] M. N. Yakubu and S. I. Dasuki, "Assessing eLearning systems success in Nigeria: An application of the DeLone and McLean information systems success model," Journal of Information Technology Education: Research, vol. 17, pp. 183-203, 2018, [Online]. Available: https://doi.org/10.28945/4077.

[6] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," MIS Quarterly, vol. 27, no. 3, pp. 425-478, Sep. 2003, [Online]. Available: https://doi.org/10.2307/30036540.

[7] D. Kala, D. S. Chaubey, R. K. Meet, and A. S. Al-Adwan, "Impact of user satisfaction with e-government services on continuance use intention and citizen trust using TAM-ISSM framework," Interdisciplinary Journal of Information, Knowledge, and Management, vol. 19, art. no. 1, pp. 1-21, Feb. 2024, [Online]. Available: https://doi.org/10.28945/5248.

[8] D. Cyr, M. Head, and A. Ivanov, "Design aesthetics leading to m-loyalty in mobile commerce," Information & Management, vol. 43, no. 8, pp. 950-963, Dec. 2006, [Online]. Available: https://doi.org/10.1016/j.im.2006.08.009.

[9] J. Guo, C. Wang, and Y. Zhang, "Impact of shopping website design on customer satisfaction and loyalty: The mediating role of user experience," Sustainability, vol. 15, no. 8, art. no. 6347, 2023, [Online]. Available: https://doi.org/10.3390/su15086347.

[10] M. Nuttall, T. Machin, R. Peel, S. M. Abel, and S. Brown, "Online prosocial behaviors: A scoping review of definitions, actions, and motivators," Cyberpsychology, Behavior, and Social Networking, vol. 28, no. 6, pp. 387-398, Jun. 2025, [Online]. Available: https://doi.org/10.1089/cyber.2024.0481.

[11] D. Gefen, D. W. Straub, and M.-C. Boudreau, "Structural equation modeling and regression: Guidelines for research practice," Communications of the Association for Information Systems, vol. 4, no. 1, pp. 1-77, 2000, [Online]. Available: https://doi.org/10.17705/1CAIS.00407.

[12] C. P. Nursalim, Tannia, and A. Robert, "Service quality and perceived value toward customer satisfaction in e-commerce delivery: The role of trust," International Journal of Applied Business and International Management, vol. 10, no. 1, pp. 136-153, Apr. 2025, [Online]. Available: https://doi.org/10.32535/ijabim.v10i1.3741.

Integrating AI and IoT in STEM Education Through a Gesture-Control Project

Serhii Petrovych, Stefan-Daniel Horvath and Chunfang Zhou
STEM Education Research Center, Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark
petrovych@imada.sdu.dk, chzh@sdu.dk, horvathstefandaniel@gmail.com

Keywords: STEM Education, Hand Gesture Control, Computer Vision, Internet of Things.

Abstract: This paper presents the development and deployment of a hand gesture-controlled lighting system specifically designed for STEM education at the undergraduate level. The primary objective of this study is to demonstrate a practical framework for integrating complex AI and IoT concepts through a hands-on, constructionist learning approach. The project combines affordable microcontroller hardware, specifically the ESP32-CAM and ESP32 Dev Board, with the MediaPipe framework and Internet of Things (IoT) protocols to transform human hand movements into interactive visual effects. By utilizing MediaPipe for real-time hand landmark detection and implementing explicit geometric rules for gesture classification, the project provides students with direct experience in hardware-software integration and distributed system control. The study details the technical architecture, including robust debouncing mechanisms to ensure operational stability. Furthermore, the system supports versatile deployment options, such as standalone executable files for Windows, enhancing classroom accessibility. This "white-box" design principle facilitates a deeper understanding of embedded programming and practical AI applications. Ultimately, the project serves as a comprehensive educational tool that successfully bridges theoretical knowledge and applied STEM skills in modern engineering curricula.

1 INTRODUCTION

The rapid expansion of artificial intelligence (AI) and the Internet of Things (IoT) is reshaping STEM education by positioning learners as designers of cyber-physical systems [1], [2], with project-based approaches further promoting deeper conceptual understanding [3], [4].

Recent work has explored gesture recognition in education. Umadevi et al. [5] presented GestureMath AI, which uses MediaPipe to enable students to draw equations in the air, demonstrating the growing interest in touchless educational interaction. A 2024 study by Chen et al. [6] proposed a continuous recognition algorithm using MediaPipe BlazePose and simple rules to detect teachers' hand signals for students with attention deficits, achieving 88.3% accuracy. Similarly, Bastos et al. [7] developed a low-cost social robot using MediaPipe and OpenCV to recognize and mimic gestures for STEM learning environments.

On the algorithmic side, researchers have compared rule-based and ML approaches for gesture recognition. Malik et al. [8] presented a rule-based system for recognizing static Polish Sign Language gestures in Extended Reality (XR) using hand tracking. Arboleda et al. [9] explored opening the ML black box for multidisciplinary students through scaffolding from GUI to coding, emphasizing pedagogical approaches to making AI transparent. Wang and Chen [10] employed explainable AI methods to unpack deep learning-based classroom discourse models, enhancing STEM teachers' trust and technology acceptance.

However, most existing works rely on black-box neural networks trained on large datasets or require expensive hardware (HMDs, VR controllers). None provide a fully transparent, low-cost pipeline from ESP32-CAM to IoT actuation that students can modify at the rule level without retraining models.

This study operationalizes this integration through a gesture-controlled lighting system based on low-cost microcontrollers and computer vision. Its architecture consists of three components: an ESP32-CAM streaming video via WiFi, a computer processing the stream with Python, and an ESP32 controlling an LED strip. This distributed system makes the information trajectory explicit - from gesture to command - supporting computational thinking [2], [11].

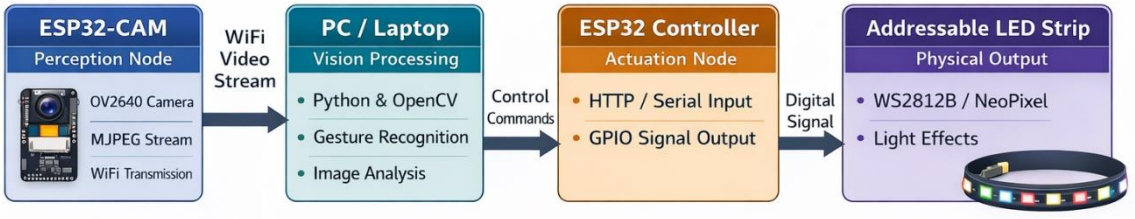


Figure 1: Block diagram of the hardware architecture for the gesture-controlled lighting system.

However, even in such systems, students often miss the internal logic due to the 'black-box' effect: most AI tools hide data processing and hardware interaction. Existing STEM toolkits treat AI as an opaque module, limiting students' understanding of the full sensing-to-actuation pipeline. This study offers a transparent 'white-box' architecture where mechanisms can be examined and modified. Students engage with image preprocessing, classical face detection [12], and AI-based hand tracking using MediaPipe [13]. Gesture classification uses geometric relations between hand landmarks, linking mathematical modeling with embodied interaction [14].

Thus, the research problem is: how to design a low-cost gesture-controlled IoT system that makes AI decision-making and hardware interaction fully transparent for undergraduate learners? The aim is to empirically investigate how a transparent, rule-based gesture recognition architecture influences students' understanding of AI-IoT integration.

Research questions:

1) Can students explain the relationship between hand landmarks, geometric rules, and LED actuation?
2) What are the main implementation challenges (lighting, debouncing) in real classroom conditions?
3) To what extent can students modify the system without prior ML knowledge?

Specific research objectives:

1) To design and implement a transparent, rule-based hand gesture recognition system using MediaPipe and ESP32, and evaluate its classification accuracy under varying lighting conditions.
2) To assess, during Microplayer Club sessions at the University of Southern Denmark, how the white-box architecture enables learners to explain the sensing-to-actuation pipeline and modify the system without prior ML expertise.
3) To identify and document the main technical challenges (debouncing stability, Wi-Fi latency, lighting sensitivity) for classroom deployment.

2 HARDWARE ARCHITECTURE

The hardware foundation of the gesture-control system is structured as a modular distributed IoT configuration that separates perception and actuation while maintaining wireless coordination. This architectural choice reflects fundamental IoT principles, including layered organization, device autonomy, and network-based interoperability. At the same time, it embodies the “sense-process-act” paradigm characteristic of cyber-physical systems, where embedded devices interact dynamically with physical processes through computational mediation.

To clarify the interaction between modules, Figure 1 illustrates the distributed hardware topology of the system.

The system is implemented using accessible and widely adopted components: an ESP32-CAM module (AI Thinker model with OV2640 camera) functioning as the perception node; a standard ESP32 development board operating as the actuation controller; an addressable LED strip such as WS2812B (NeoPixel) or a simpler digital strip; and a regulated 5V power supply, preferably external when extended LED strips are used. The ESP32 platform integrates WiFi connectivity and sufficient processing capacity for IoT prototyping, making it particularly suitable for educational experimentation with edge devices.

The ESP32-CAM captures visual data and streams MJPEG video over HTTP within a local WiFi network. This configuration allows real-time frame acquisition by Python-based computer vision scripts running on a computer. The integration of camera, antenna, and microcontroller on a single board reduces assembly complexity while preserving architectural transparency. From a systems perspective, the module operates as an autonomous sensing unit within a distributed network.

The second ESP32 board functions as the actuation node. It receives control commands via HTTP requests or serial communication and regulates lighting behavior accordingly. When paired with WS2812B LEDs, the controller generates precisely timed digital signals to address individual RGB pixels. This

demonstrates serialized data transmission and time-dependent signal encoding, central concepts in embedded and physical computing environments. Firmware parameters such as the number of LEDs and strip type allow the system to be reconfigured without altering the hardware structure, reinforcing abstraction principles in embedded design.

3 SOFTWARE ENVIRONMENT DEPENDENCIES AND GESTURE RECOGNITION

3.1 Software Architecture

The gesture-controlled IoT system follows the classical "sensing - processing - actuation" model widely used in embedded and cyber-physical systems. The software environment requires Python 3.10+ with OpenCV, MediaPipe, and requests libraries. Project dependencies are managed through a virtual environment. Figure 2 shows the overall connection diagram.

The STEM gesture-controlled IoT system requires a carefully prepared software environment. Python 3.9, 3.10, or 3.11 (64-bit recommended) is required because the gesture recognition scripts, hand_gesture.py and gesture_launcher.py, as well as MediaPipe 0.10+, depend on these versions [5].

3.2 Arduino/ESP32 Firmware and Project Configuration

The hardware component consists of two ESP32 modules. One module is the ESP32-CAM, which captures video frames and streams them via MJPEG over WiFi. This module is programmed using Arduino IDE and requires libraries for web serving and WiFi connectivity. The other module controls the LED strip. It receives HTTP commands or serial inputs from the PC to turn LEDs on or off, adjust brightness, or run fun effects. For addressable LED strips such as WS2812B, the Adafruit NeoPixel library is required. Both sketches rely on a secrets.h file that contains WiFi credentials, allowing easy configuration of network settings without modifying the main code.

Students gain experience with distributed system principles, as one device acts as the “eye” capturing visual data, while the other functions as the “hand” performing actions. Configuration parameters, such as stream URLs and LED controller URLs, are stored in a local configuration file used by the gesture launcher to persist settings across sessions.

3.3 Hand Gesture Recognition and Data Flow

Hand gesture recognition is implemented in the script hand_gesture.py using MediaPipe Hand Landmarker. This pre-trained model detects 21 three-dimensional landmarks per hand, including the wrist, four points of the thumb, and four points for each of the other fingers. The system does not require additional training. Finger states are computed geometrically to determine whether each finger is extended or folded. The thumb state is evaluated based on handedness, while the other fingers are determined by comparing the tip position to the PIP joint, allowing gestures to be recognized even when the hand is turned sideways.

Gesture classification is rule-based. For example, a pinch is detected when the distance between the thumb and index finger is less than 30% of the hand size while the other fingers are folded. An “OK” gesture uses the same pinch criteria but with other fingers extended. A fist corresponds to all fingers folded, while an open hand is recognized when four or more fingers are extended. Other gestures such as thumb up or down, pointer, peace, and rock are similarly defined. To prevent flickering, gestures are sent to the LED strip only after they remain consistent across several consecutive frames.

The data flow is as follows. The ESP32-CAM captures video and sends frames to the PC running hand_gesture.py. MediaPipe extracts hand landmarks and computes finger states. The gesture mapping module translates finger states into explicit LED commands, which are sent to the ESP32 LED controller. The controller updates the LED strip according to the commands. This design demonstrates the integration of AI-assisted detection, explicit rule-based logic, and hardware actuation within a STEM educational context.

3.4 Development and Participants

The system was developed by the authors of this study at the STEM Education Research Center - FNUG (Framing New Understanding for Growth), Department of Mathematics and Computer Science, University of Southern Denmark (SDU). A total of 12 undergraduate students from various departments at SDU participated voluntarily in the pilot study. Five sessions were conducted, each lasting 1.5 to 2 hours. Students worked in pairs following a constructionist learning protocol. No prior knowledge of computer vision, IoT, or machine learning was required. Data were collected through direct observation, system logs, and a short exit survey.

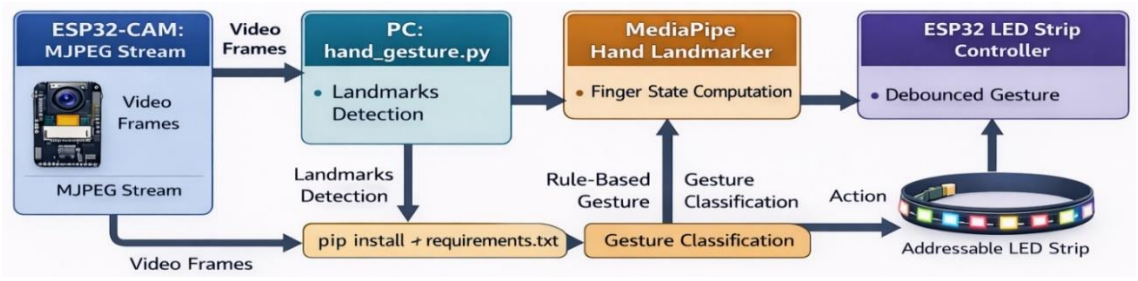


Figure 2: Block diagram of the hand gesture-controlled IoT system.

4 INTERACTIVE CONTROL AND GESTURE MAPPING

4.1 Gesture Recognition Logic

The interactive dimension of this STEM project emerges from transforming human hand movements into actions on the LED lighting system. The recognition process begins with the MediaPipe Hand Landmarker, which detects twenty-one three-dimensional landmarks on each hand, including the wrist, thumb, and four points on each of the other fingers. Gesture classification does not rely on a black-box neural network. Instead, explicit geometric rules determine the state of each finger, with the thumb using handedness to distinguish extension, and the other fingers evaluated relative to their PIP joints or distance from the wrist. Gestures are defined as specific combinations of extended and folded fingers. To ensure stability and prevent flickering, gestures are sent to the LED controller only after they are consistently observed across five consecutive frames (debouncing).

4.2 Gesture-to-Action Mapping and Creative Effects

Open Hand. An open palm triggers the ON command, illuminating the LED strip in solid color mode (see Fig. 3). This gesture represents a direct mapping from a fully extended hand to a simple visual output.

OK Gesture. The OK gesture, detected by the pinch distance between thumb and index while other fingers are extended, sends the OFF command, turning off all lighting and clearing active visual effects (see Fig. 4).

Thumb Up / Thumb Down. These gestures adjust brightness dynamically (see Fig. 5). The vertical orientation of the thumb tip relative to its joint distinguishes between increasing (Thumb Up) and decreasing (Thumb Down) light intensity.



Figure 3: Example of the Open Hand gesture controlling the LED strip.



Figure 4: Example of the OK gesture turning off the LED strip.

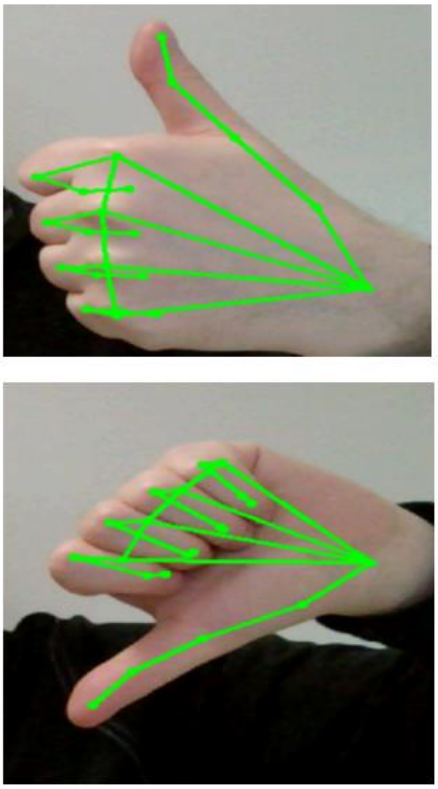


Figure 5: Thumb Up/Down gestures controlling LED brightness.

Pointer. The pointer gesture activates Fun Mode (see Fig. 6). Horizontal movement shifts the light cluster along the strip, while vertical movement changes its hue. This enables dynamic and interactive light positioning.



Figure 6: Pointer gesture controlling the position and hue of the LED cluster.

Peace Sign. The peace gesture, with index and middle fingers extended, produces a pulsating wave of light at the hand's location (see Fig. 7). The speed and duration of the wave correspond to the speed of hand movement.



Figure 7: Peace gesture producing a pulsating light wave.

Three Sign. The three-finger gesture creates a breathing effect (see Fig. 8). Light intensity gradually increases and decreases, producing a rhythmic visual pattern.



Figure 8: Three Sign gesture for breathing effect.

Rock Sign. The rock gesture triggers a strobing effect, with rapid flashes of light along the LED strip (see Fig. 9).



Figure 9: Rock gesture activating strobe mode.

Hang Loose. The hang loose gesture activates a rainbow effect, smoothly cycling through all colors of the spectrum along the LED strip (see Fig. 10).



Figure 10: Hang Loose gesture for rainbow effect.

This design allows learners to connect physical gestures with immediate visual feedback, explore dynamic interaction modes, and understand the principles of real-time control, debouncing, and human-computer interaction. The combination of rule-based gesture classification and creative lighting effects demonstrates the pedagogical value of transforming abstract hand movements into tangible, observable actions in STEM projects.

5 EXPERIMENTAL RESULTS AND CORE FINDINGS

We evaluated gesture recognition accuracy under two lighting conditions (well-lit: >300 lux, low-light: <100 lux) using 100 trials per gesture performed by three students. Table 1 summarizes the results.

Table 1: Gesture recognition accuracy (%).

Gesture	Well-lit	Low-light
Open Hand	96	81
OK	94	78
Thumb Up/Down	91	74
Pointer	89	70

The debouncing mechanism (requiring 5 consecutive consistent frames) eliminated 92% of flickering errors. During the five Microplayer Club sessions, 10 out of 12 students successfully explained how geometric rules map hand landmarks to LED commands. Eight students added a new custom gesture (e.g., "victory" to toggle rainbow mode) within 20 minutes, demonstrating the white-box modifiability. The main reported difficulties were sensitivity to background lighting and Wi-Fi latency (average 120 ms).

Core findings:

1) Rule-based gesture recognition is sufficiently transparent for students to modify and extend without ML training.
2) Lighting robustness remains the main limitation (accuracy drops 10-15% in low light).
3) Debouncing is critical for real-world stability and should be explicitly taught in STEM curricula.

6 CONCLUSIONS

This study showed that a gesture control system where students can see and change the internal logic (instead of using a black-box AI) works well for teaching AI and IoT. In sessions with 12 students from different departments at the University of Southern Denmark, 10 participants could explain how hand landmarks turn into LED commands, and 8 added a new gesture within 20 minutes without any prior machine learning training.

Accuracy was above 90% in good light (>300 lux) but dropped to 70-81% in low light (<100 lux). The debouncing mechanism (requiring 5 stable frames) removed 92% of flickering errors and proved essential for classroom use. Wi-Fi latency (average 120 ms) was another challenge, but it did not stop successful interaction.

These results have three implications for STEM education. First, keeping the decision logic transparent

helps students understand and modify the system. Second, topics like debouncing and environmental robustness should be taught explicitly in IoT courses. Third, the low-cost hardware (ESP32-CAM, ESP32, LED strip) makes the project accessible even with small budgets.

Limitations: small sample (N=12), short sessions (five 1.5-2 hour meetings), and the simple rule-based approach does not work for complex dynamic gestures. Wi-Fi dependency may be a problem in crowded classrooms. No control group using a black-box AI system was included.

Future work will improve lighting robustness (e.g., automatic gain control) and reduce Wi-Fi latency by moving gesture recognition to the ESP32-CAM edge device. A larger study with more students is also planned.

In conclusion, this gesture-controlled lighting system provides a practical, low-cost way to integrate AI and IoT into undergraduate STEM education. It bridges theory and hands-on practice while empowering students to explore, understand, and modify intelligent systems.

REFERENCES

[1] S. Papert, Mindstorms: Children, Computers, and Powerful Ideas, Basic Books, 1980.

[2] J. M. Wing, "Computational thinking," Communications of the ACM, vol. 49, no. 3, pp. 33-35, 2006, [Online]. Available: https://doi.org/10.1145/1118178.1118215.

[3] P. C. Blumenfeld, E. Soloway, R. W. Marx, J. S. Krajcik, M. Guzdial, and A. Palincsar, "Motivating project-based learning: Sustaining the doing, supporting the learning," Educational Psychologist, vol. 26, no. 3-4, pp. 369-398, 1991, [Online]. Available: https://doi.org/10.1080/00461520.1991.9653139.

[4] C. E. Hmelo-Silver, "Problem-based learning: What and how do students learn?" Educational Psychology Review, vol. 16, pp. 235-266, 2004, [Online]. Available: https://doi.org/10.1023/B:EDPR.0000034022.16470.f3.

[5] M. Umadevi, P. Nikitha, T. S. Goud, M. Abhinay, and L. Swathi, "Gesture Math AI: Real-time math problem solving using hand gestures and computer vision," International Research Journal on Advanced Engineering Hub, vol. 3, no. 5, 2025, [Online]. Available: https://doi.org/10.47392/IRJAEH.2025.0320.

[6] I. D. S. Chen, C.-M. Yang, S.-S. Wu, C.-K. Yang, M.-J. Chen, C.-H. Yeh, and Y.-H. Lin, "Continuous recognition of teachers' hand signals for students with attention deficits," Algorithms, vol. 17, no. 7, p. 300, 2024, [Online]. Available: https://doi.org/10.3390/a17070300.

[7] A. C. D. O. Bastos, M. F. Pinto, R. C. Coutinho, F. L. Silva, A. A. D. Lima, and G. M. Araujo, "A low-cost social robot for gesture-based educational activities and human-robot interaction in learning environments," in Proc. XVII Simpósio Brasileiro de Robótica e Simpósio Latino Americano de Robótica (SBR/LARS), pp. 408-413, 2025, [Online]. Available: https://doi.org/10.1109/sbr/wre66973.2025.11249546

[8] W. Malik, G. Kulis, and K. Skabek, "Recognition of Polish sign gesture language in extended reality," in Proc. European Conference on Modelling and Simulation (ECMS), pp. 645-652, 2025, [Online]. Available: https://doi.org/10.7148/2025-0645.

[9] M. Arboleda, C. Vieira, and J. L. Chiu, "Opening the machine learning black box for multidisciplinary students: Scaffolding from GUI to coding," in Proc. IEEE Frontiers in Education Conference (FIE), pp. 1-5, 2023, [Online]. Available: https://doi.org/10.1109/FIE58773.2023.10343043.

[10] D. Wang and G. Chen, "Making AI accessible for STEM teachers: Using explainable AI for unpacking classroom discourse analysis," IEEE Transactions on Education, vol. 67, no. 6, pp. 907-918, 2024, [Online]. Available: https://doi.org/10.1109/TE.2024.3421606.

[11] J. M. Wing, "Computational thinking and thinking about computing," Philosophical Transactions of the Royal Society A, vol. 366, no. 1881, pp. 3717-3725, 2008, [Online]. Available: https://doi.org/10.1098/rsta.2008.0118.

[12] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. I-511, 2005, [Online]. Available: https://doi.org/10.1109/cvpr.2001.990517.

[13] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays et al., "MediaPipe: A framework for building perception pipelines," arXiv preprint arXiv:1906.08172, 2019, [Online]. Available: https://doi.org/10.48550/arxiv.1906.08172.

[14] D. Long and B. Magerko, "What is AI literacy? Competencies and design considerations," in Proc. CHI Conference on Human Factors in Computing Systems, pp. 1-16, 2020, [Online]. Available: https://doi.org/10.1145/3313831.3376727.

Correction of Sensory Integration as a Prerequisite for Successful Mastering a Foreign Language

Iryna Haidai, Liudmyla Suvorova, Olha Zadoienko, Anastasiia Kravchenko and Krystyna Levkivska
Department of Pedagogical Technologies and Language Training, Zhytomyr Polytechnic State University, Chudnivska Str. 106, 10005 Zhytomyr, Ukraine
org_gio@ztu.edu.ua, ktpl_slk@ztu.edu.ua, kptmp_zoi@ztu.edu.ua, kim_kao@ztu.edu.ua, kim_lkv@ztu.edu.ua

Keywords: Multisensory Integration, Virtual Reality (VR), Sensory Compensation, Inclusive Education, Foreign Language Learning, Assistive Educational Technologies.

Abstract: This article examines the education of learners with sensory impairments in an inclusive environment, focusing on foreign language learning amid wartime and social challenges. It highlights the need for effective instructional strategies that address sensory processing difficulties across visual, auditory, tactile, proprioceptive, and vestibular domains. Some learners also displayed characteristics associated with the autism spectrum. The study aimed to identify relevant theoretical frameworks and develop methods to mitigate educational losses. An experimental evaluation of a multisensory VR environment was conducted with students in grades 5-9 in Zhytomyr, Ukraine. Methods included observation, adapted Foreign Language Classroom Anxiety Scale (FLCAS) questionnaires adapted for learners with special educational needs (SEN) to assess anxiety, vocabulary testing, and a pedagogical experiment. The performance of students using a VR-based platform was compared with those taught through traditional methods. Results indicate that VR integration helps compensate for sensory deficits, improves communicative competence, boosts motivation, and reduces anxiety. The study contributes practical exercises for VR-based instruction and expands understanding of multisensory technologies in inclusive language education.

1 INTRODUCTION

At present, Ukraine is operating under extremely unstable conditions caused by the ongoing war as a profoundly destructive factor. Contemporary educational paradigm must take this key context into account and be oriented toward developing a barrier-free environment, as emphasized in fundamental state-level policy documents: the National Strategy for Creating a Barrier-Free Space in Ukraine until 2030 and the Presidential Decree No. 533/2020 “On Ensuring the Creation of a Barrier-Free Space in Ukraine” [1]. Inclusive education, particularly the search for effective tools for its organization, requires the continuous implementation of innovative approaches and the systematic identification of evidence-based methodological practices. However, this process becomes particularly complex when students with cognitive impairment resulting from sensory deficits, such as visual, auditory, or tactile impairments, learn foreign languages, as these impairments may disrupt mechanisms of information perception and processing [2]. These technologies enable the implementation of effective practices for creating multisensory learning environments. Such environments enable the activation of preserved sensory channels, for instance, visual support for learners with hearing impairments or tactile feedback, thereby facilitating the development of stable lexical and grammatical skills. A multisensory VR environment integrates visual, auditory, and kinesthetic stimuli, thereby creating new opportunities for the effective acquisition of foreign-language learning material. Despite growing global interest in immersive technologies, their adaptation to the specific context of Ukrainian inclusive classrooms remains insufficiently explored. Furthermore, the ongoing war has contributed to a significant increase in the number of learners with disabilities, which underscores the urgent need for accessible and adaptive educational solutions according to the data Data from the Ministry of Social Policy and the National Health Service of Ukraine, 2024-2025) presented in Table 1.

Table 1: Distribution of children with disabilities in Ukraine by Major Nosological Groups.

Type of impairment	Percentage contribution to the overall structure %
Congenital developmental disorders	28.2
Mental and behavioral disorders (including ASD)	20.1
Diseases of the nervous system (CNS disorders, CP)	14.7
Endocrine system disorders	10.6
Hearing impairments	5.6
Musculoskeletal disorders	4.5
Visual impairments	3.5
Other diseases and injuries	12.8

It should be noted that a significant proportion of families have children whose disability status has not been officially determined due to a range of objective factors, with estimates ranging from 40% to 60% [3].

Significance of the study logically determines its purpose, which is to provide a theoretical substantiation and experimental verification of the effectiveness of the multisensory VR platform Akelius as a compensatory tool for addressing sensory deficits in the process of foreign language learning among students in inclusive classrooms (Grades 5-9) in the city of Zhytomyr, Ukraine.

The proposed research focuses on the educational practices of general secondary education institutions in Zhytomyr. The selection of this region is determined by the active implementation of digital educational tools in local inclusive classrooms and the need for empirical evidence regarding the effectiveness of VR-based instructional solutions under conditions of limited educational resources and the ongoing wartime context. These factors contribute to increased psychological strain, particularly among students with SEN, thereby reinforcing the importance of exploring adaptive and multisensory digital learning environments.

2 THEORETICAL FOUNDATIONS OF THE STUDY

The theoretical framework of the study is grounded in contemporary concepts of global pedagogy, psychology, special education, foreign language didactics, and educational technologies. It provides a holistic understanding of the potential of multisensory VR environments in inclusive foreign language education. The study also draws upon the current legislative and regulatory framework governing inclusive and special education in Ukraine, including the Law of Ukraine “On Education”, the Law of Ukraine “On Complete General Secondary Education,” Resolution of the Cabinet of Ministers of Ukraine No. 957 “On Approval of the Procedure for Organizing Inclusive Education in General Secondary Education Institutions,” and the National Strategy for Creating a Barrier-Free Environment in Ukraine until 2030 [4]. The empirical basis of the study comprises quantitative indicators of the inclusive classroom network in the city of Zhytomyr, as well as statistical data on the number of learners who require continuous inclusive support. The research is grounded in the theory of inclusive education, which emphasizes equal access to learning, adaptation of the educational environment, and non-discrimination of individuals with sensory impairments. A key theoretical framework underpinning the study is the concept of sensory compensation and neuroplasticity, which substantiates the possibility of compensating for impaired sensory channels through the activation of alternative sensory modalities.

The theoretical and methodological framework of the study is grounded in core principles and concepts of inclusive education, emphasizing equal access to quality learning, universal design in education, and the prevention of discrimination against learners with sensory impairments through reasonable accommodations in the educational environment. Psycholinguistic theories of language learning support the appropriateness of using VR for immersive language exposure. The study also incorporates the principles of Universal Design for Learning (UDL), which ensure flexibility and accessibility of VR environments for diverse educational needs, as well as key assumptions of immersive digital technology theory, highlighting their positive impact on learner motivation and engagement. An individual-centered approach, aimed at addressing each learner's unique characteristics, is also critical. The synthesis of the outlined theoretical principles provides a solid scientific foundation for substantiating the use of multisensory VR environments as an effective tool for compensating sensory deficits in inclusive foreign language education.

2.1 Literature Review

The theoretical framework of this study is grounded in the concepts of sensory integration and neuroplasticity,heories of foreign language learning

within inclusive educational environments, the principles of Universal Design for Learning (UDL), and scholarly approaches to the use of VR in education. Multisensory input supports learners with diverse cognitive and sensory profiles [5].

The effectiveness of immersive virtual reality has also been demonstrated at the stage of second-language vocabulary acquisition in K-12 education. Virtual reality provides strong compensatory potential for learners with sensory impairments and functions as an effective multisensory tool in language education [6].

Furthermore, according to Rudling, immersive VR environments are particularly effective in vocabulary acquisition. Multisensory and embodied VR interaction enhances attentional sustainability during lexical learning and activates sensorimotor engagement in the language learning process [7].

Immersive and multisensory technologies support vocabulary acquisition, facilitate the development of pronunciation, and contribute to the formation of communicative competence. In this context, inclusive design functions as a foundational pedagogical framework. The effectiveness of VR-based learning environments in inclusive education depends on the educator's ability to differentiate approaches in accordance with the individual characteristics of learners with diverse disabilities, neurodevelopmental conditions, or educational barriers [8].

Principles of accessibility and inclusivity in immersive technologies, including VR and augmented reality (AR), contribute to the formation of a "metaverse" environment. Design strategies should prioritize support for users with sensory, motor, and cognitive impairments, with multisensory substitution and adaptive interfaces emerging as key mechanisms for compensating sensory deficits. Within this context, the authors identify several persistent barriers, notably high implementation costs and the lack of unified standards [9].

Tactile and auditory learning tools are frequently reported to support learners with SEN in mastering academic content alongside their peers, while also reducing anxiety levels [10]. A foundational framework for understanding inclusive education is provided by M. Oliver's concept of disability, which conceptualizes disability through the lens of socially constructed barriers arising from inadequacies in the environment rather than from an individual's medical condition.

Multisensory VR environments demonstrate significant potential for inclusive foreign language learning. VR-based techniques enhance learners' motivation, attention, and perceptual engagement, while positively influencing language learning outcomes. At the same time, such environments support diverse learner needs by providing flexible, immersive, and adaptive learning experiences.

3 RESEARCH METHODOLOGY

The study is grounded in inclusive, competence-based, multisensory, systemic, and learner-centred approaches. Particular emphasis is placed on mechanisms of sensory profiling, which enabled the identification of levels of auditory and visual information perception. A combination of theoretical methods (systemic analysis and a deductive approach) and empirical instruments was employed, including participant observation, interviews, and questionnaire-based surveying using the Foreign Language Classroom Anxiety Scale (FLCAS), adapted for learners with SEN.

Methods included vocabulary size testing (Vocabulary Size Test), a pedagogical experiment, and quantitative and qualitative data analysis. The results obtained using M. Lüscher's methodology revealed a statistically significant correlation between the severity of speech-related difficulties and the level of situational anxiety, as measured by the Lüscher Color Test. The study sample comprised students from inclusive classes in secondary schools (lyceums) in the city of Zhytomyr, which had well-established inclusive education infrastructure. The total sample comprised 150 students with SEN, aged 11-15 years (middle school level). The experimental group (EG) received instruction using VR-based tools, whereas the control group (CG) was taught using traditional inclusive teaching methods.

To create a multisensory learning experience, specially adapted scenarios were employed, including Akelius, Virtual Classroom, and City Tour. Learners interacted with virtual objects via tactile controllers that provided haptic feedback, enabling the compensation of perceptual deficits through the simulated mechanical sensation of "touching" virtual lexical items.

The observational method was implemented using structured observation cards with the involvement of the learners themselves to assess engagement levels (engagement rate). The study was conducted in compliance with bioethical principles, and written parental consent was obtained. VR-based instructional sessions did not exceed 15 minutes per session to prevent visual fatigue, in accordance with the sanitary regulations of Ukraine.

3.1 Research Result

The study was conducted in an inclusive educational environment involving learners with diverse sensory and cognitive impairments and was implemented in three stages:

1) Ascertaining stage. Diagnostic assessment of the levels of sensory deficits and learning barriers in foreign language (English) acquisition among students in inclusive classrooms, as well as an analysis of the content of Individual Development Programs (IDPs) and Individual Learning Plans (ILPs);
2) Formative stage. Implementation of a multisensory VR-based learning environment (using platforms such as Akelius, Virtual Classroom, City Tour, and Mondly VR) into the educational process;
3) Control stage. Analysis of changes in vocabulary acquisition, development of grammatical skills, and measurement of anxiety levels.

The study's temporal scope covered the period of students' teaching practice conducted in January-February 2025 and in November 2025. Students of Zhytomyr Polytechnic State University completed their pedagogical practice at selected lyceums in Zhytomyr. Supervision of the practice was carried out jointly by teachers at the practice bases and by academic staff of the Department of Pedagogical Technologies and Language Training, who were responsible for evaluating the practice's outcomes.

3.2 Initial (Diagnostic) Stage of the Study

The study was conducted in inclusive classrooms of Lyceums No. 3, 4, 14, 19, 20, and 21 in the city of Zhytomyr. The sample consisted of 150 students (N=150) aged 11-15 who experienced learning difficulties and required support from a teaching assistant. Data collection took place during the pedagogical practicum completed by students of Zhytomyr Polytechnic State University enrolled in the Bachelor's programme 014.021 "Secondary Education. English Language and Literature."

The distribution of students by SEN categories was based on the primary diagnostic data of the IRC, which is shown in Figure 1. The formation of the study sample (N = 150) was carried out on the basis of a preliminary study and analysis of the official conclusions of the Inclusive Resource Centers (IRC) on the comprehensive psychological and pedagogical assessment of child development. This allowed selecting participants with confirmed sensory profile features (hypo- or hypersensitivity (37%), proprioception difficulties (33%) and sensory discrimination difficulties 30%), which created barriers to mastering a foreign language. The distribution into the experimental (EG, N = 75) and control (CG, N = 75) groups was carried out using the method of stratified randomization, taking into account the specifics of sensory deficits recorded in the IRC protocols. It is worth emphasizing that the experimental group included those students who were dominated by pronounced deficits in sensory integration, while the control group included students with other features of psychophysical development, in which the sensory aspect was not critically injured or leading in the structure of the disorder. Stratification was carried out by the leading type of sensory deficit, according to the classification used in the IRC protocols. (see Fig. 1).

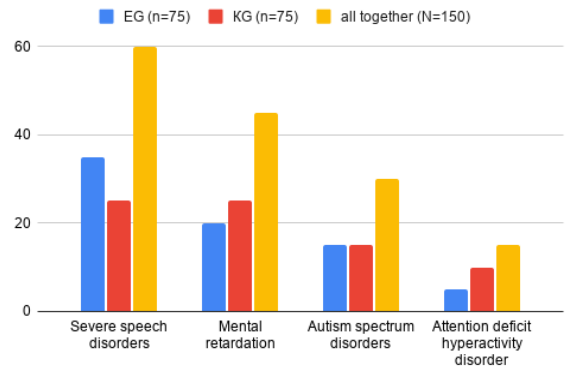


Figure 1: Categories of respondents by type of SEN.

The empirical investigation of sensory barriers employed three complementary approaches (see Fig. 2).

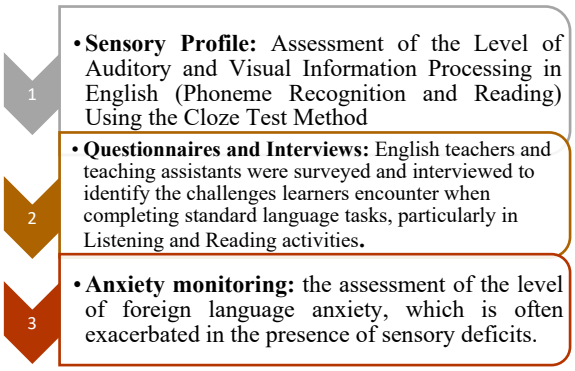


Figure 2: Mechanisms for investigating sensory barriers.

Based on the sensory profile assessment, 30% of students with SEN demonstrated effective processing of auditory and visual information, 40% experienced cognitive-related difficulties, and 30% were able to recognise phonemes but showed significant challenges in reading. The assessment focused on learners' ability to discriminate minimal pairs (e.g., ship/sheep, bad/bed). The task allows for determining whether a learner can perceive phonemic contrasts or whether the sounds are perceptually merged. 23% of participants demonstrated difficulties in grapho-motor identification, particularly in distinguishing visually similar letter pairs (e.g., b-d, p-q, n-u). These findings are significant for the further integration of VR in educational practice. A low level of phoneme recognition can be compensated through visual scaffolding, such as subtitles or three-dimensional object representations, which can be synchronously presented with auditory input in a VR environment. The aggregated results indicate that only 55 of 150 learners (36.7%) successfully completed tasks involving auditory discrimination of minimal pairs, while 45 respondents (30%) demonstrated adequate performance in visual tracking tasks. The obtained data substantiate the need to implement a systematic set of corrective interventions using virtual technologies to enhance multisensory integration.

The data obtained from the questionnaire survey and interviews can be categorized into several thematic blocks (Fig. 3).

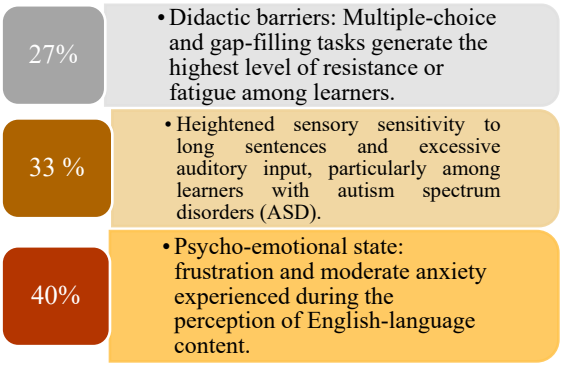


Figure 3: Qualitative indicators derived from the questionnaire survey.

The survey yielded quantitative data based on a five-point Likert scale (1-5) to assess the frequency of reported difficulties. As an Table 2, the study presents a scale of difficulties related to the differentiation of similar phonemes ([th]/[s]).

Table 2: Likert-Scale assessment of perceived difficulties.

Indicator	Never	Rarely	Often	Always
Number of students	30	47	46	27

An analysis of interviews with inclusive classroom assistants revealed that, on average, students with cognitive difficulties experience high levels of auditory fatigue after approximately 7 minutes of continuous listening. This finding correlates with the results of phonemic hearing assessments. These outcomes indicate the need to implement short iterations of auditory input supported by visual scaffolding in VR learning environments.

Anxiety levels were monitored using the Luscher Color Test before and after the English language lesson. To assess the psycho-emotional state of students with SEN, an expedited diagnostic approach was used, combining emotional pictogram cards with a simplified color-association technique. The diagnostic toolkit was based on eight standardized color stimuli (blue, green, red, yellow, violet, brown, black, and gray). This approach enabled tracing the correlation between types of learning activities (listening and reading) and students' subjective psychological comfort. The findings revealed a decrease in anxiety indicators when instruction shifted from traditional didactic materials to immersive VR simulations. This effect was evidenced by a change in students' color preferences from anxiety-associated colors (gray, dark blue) to positively active colors (yellow, green). The results clearly demonstrate that VR technologies function not merely as an engaging educational tool but also as an effective means of reducing language-learning anxiety.

The level of the affective filter (according to Krashen's Affective Filter Hypothesis) was assessed using the Foreign Language Classroom Anxiety Scale (FLCAS). This instrument enabled the measurement of foreign language anxiety as a cognitive-inhibitory factor that can hinder effective verbal communication in an inclusive classroom. The results indicated a relatively high anxiety level (43%), suggesting that the learner may possess high cognitive abilities, while their emotional state inhibits language acquisition. Foreign language anxiety was found to emerge not only as an emotional reaction but also as a consequence of insufficient sensory integration, particularly within the context of inclusive education.

3.2 Formative Stage of the Study

During their practicum in inclusive classrooms, students utilized various affordances of multisensory VR environments. The use of VR enabled learners to fully immerse themselves in the target language environment by simulating authentic, real-life communicative situations.

This study focuses on a detailed analysis of the digital learning platform Akelius, which is being implemented in the Ukrainian educational context with UNICEF support. The platform functions as a blended learning support tool and is designed for the study of twelve foreign languages. It is specifically adapted for learners with SEN, including students with dyslexia and speech and language impairments, through a flexible system of visual scaffolding and individualized learning pace. Akelius is grounded in a "learning through error" pedagogical principle and does not employ traditional grading practices, thereby reducing learners' fear of public failure and promoting a psychologically safe learning environment.

The findings indicate that the absence of directive assessment and the availability of repeated practice opportunities in the Akelius application contributed to a reduction in perfectionism-related anxiety (from 40% to 25%). This decrease is considered critically important for learners with dyslexia, as it reduces an affective barrier and promotes more sustainable engagement in language learning.

The platform employs a multimodal approach that integrates audio, visual, and textual input. For children with speech and language disorders, this multimodality provides essential instructional scaffolding. Gamification and division of learning content into micro-steps (Fig. 4) enhance learners' sense of control over the learning process. Within the scope of the empirical study, the experiment was limited to CEFR proficiency levels A1 and A2.

Akelius enables the adaptation of instructional content to learners' individual proficiency levels, thereby supporting personalized learning. Its multisensory approach is particularly effective in developing language skills, especially listening comprehension and oral communication (see Fig. 5).

Students in the experimental group (EG) demonstrated a 28% higher level of active vocabulary activation during spontaneous speech than those in the control group (CG). The use of a multisensory VR environment resulted in a 15% reduction in phonetic errors in the EG's speech, which can be attributed to the immersive "sense of presence" effect and visual support for articulation. Furthermore, the mean score on the Foreign Language Classroom Anxiety Scale (FLCAS) in the experimental group decreased from 118 points (high anxiety level) to 84 points (moderate to low anxiety level), indicating a clinically significant reduction in emotional tension. The most pronounced improvement was observed for the fear of negative evaluation, with scores decreasing by 35%, which can be attributed to the non-judgmental learning environment provided by the Akelius platform. Comparative analysis of the results showed that in the experimental group (n = 25), where VR technologies and the Akelius platform were implemented, communicative confidence increased by 22%. In contrast, in the control group (n = 25), this indicator remained within the 5-7% range, with the difference reaching statistical significance ($p < .05$).

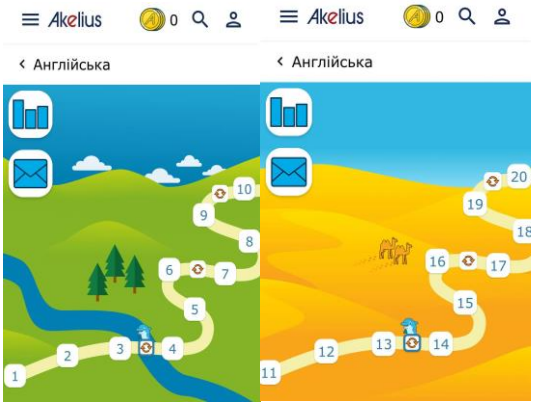


Figure 4: Screenshot of platform "Akelius".

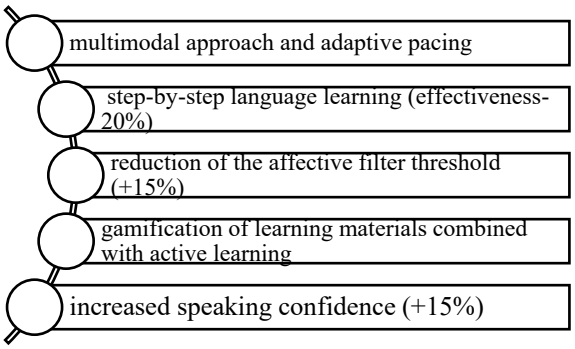


Figure 5: Key advantages of the Akelius Digital Learning Platform.

3.3 Control Stage of the Study

The implementation of a multisensory VR-based learning environment accelerated learners' development of foreign-language communicative competence, reduced anxiety levels. Statistical analysis of the experimental group revealed improvements ($p<05$) in

lexical acquisition, phonological accuracy, and speech fluency compared with the control group, which was instructed using traditional pedagogical methods (Fig. 6).

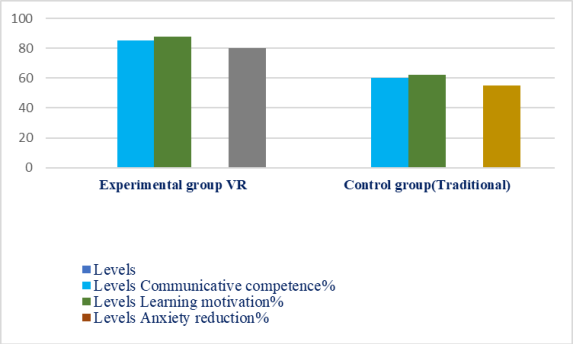


Figure 6: Impact of VR Technologies on foreign language iearning outcomes: A Comparative analysis.

The VR-based learning environment enabled individualized instruction by adapting the pace, sensory support to learner's educational needs. The implementation of virtual communicative scenarios facilitated social interaction, reduced communication barriers, and promoted more active inclusion of learners with sensory impairments in group activities.

The findings confirm the pedagogical relevance and practical effectiveness of implementing a multisensory VR environment in inclusive foreign language education, provided that appropriate methodological design and technical support are ensured. Overall, VR functions as an effective tool for compensating sensory deficits and for enhancing the quality of inclusive language education.

It is also important to highlight a range of risks and challenges encountered by learners, teachers, and pre-service teachers (Fig. 7).

First, a proportion of students experienced sensory overload (2% of learners had episodes of sensory meltdown while working on the Akelius platform), necessitating strict regulation of screen time and individualized adjustments to sensory stimuli. Second, students with significant cognitive impairments (nearly 60%) were not always able to navigate the virtual environment independently and required continuous support from a teaching assistant or a student teacher.

Among educators, a risk of insufficient methodological readiness to integrate VR into the structure of a foreign language lesson was identified (40%). A separate challenge was the equipment's technical instability, including a limited number of VR devices and software malfunctions (25%), which negatively affected the pacing of classroom instruction.

During teaching practice, students encountered difficulties in combining the roles of observer, teacher, and technical facilitator within the VR environment (30%). In addition, a psychological barrier was identified among some learners, associated with fear of making mistakes or unfamiliarity with the technology (14%).

Teachers also reported an increased workload during lesson preparation due to the need to adapt instructional materials to Individual Educational Programs (IEPs) and Individual Learning Plans (ILPs) (49%). Furthermore, uneven learning outcomes remained a significant risk, as VR effectiveness depended on the type of sensory impairment and the level of support provided.

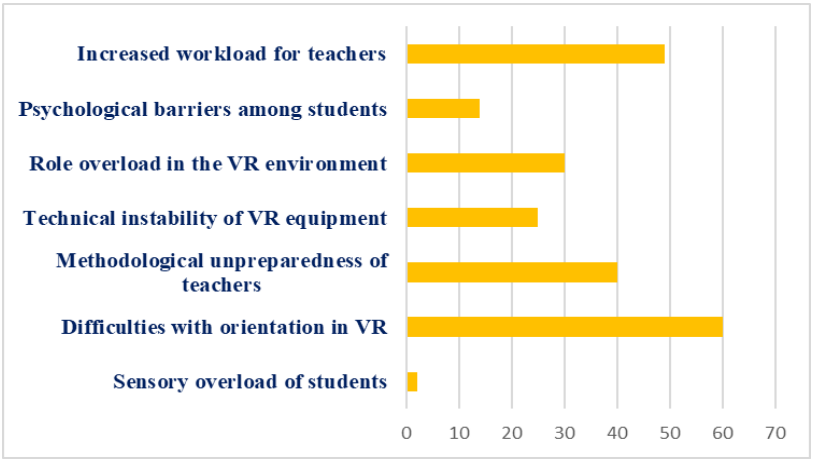


Figure 7: Identified risks in working with a virtual platform.

Akelius platform, unlike immersive VR tools (such as Mondly VR or Immerse), allows us to highlight its key advantage, in particular in working with students with sensory profile features. Unlike most commercial VR solutions, which often create excessive visual and cognitive load (visual clutter), provoking sensory overload (for example, in children with ASD), the Akelius platform is based on the principles of pedagogical minimalism and gradual sensory stimulation.

The use of Akelius in combination with virtual reality elements provides a clear structuring of audiovisual signals, which allows avoiding "sensory noise" and focusing the child's attention on articulatory praxis. Unlike applications such as Duolingo or Babbel, which are focused on passive recognition of lexemes, Akelius integrates a multisensory approach, where the visual image of the word is reinforced with specific rhythmic and intonation accompaniment. This is critically important for the formation of the articulatory base of a foreign language in children with proprioception deficits, as the platform allows for the synchronization of digital stimulus-responses with real speech therapy exercises.

4 CONCLUSIONS

Thus, the results of the comprehensive study confirm that the use of the digital platform Akelius in an inclusive educational environment serves as an effective mechanism for reducing foreign language anxiety. The data obtained using the FLCAS scale indicate a reduction in affective barriers, particularly fear of evaluation, as confirmed by positive changes in respondents' psycho-emotional state, as measured by M. Lüscher's method (a shift from anxiety-related to stabilizing color choices). In line with M.
Oliver's social model, this approach enables the transformation of the educational environment by shifting the focus from medicalized deficit perspectives of learners with dyslexia and autism spectrum conditions toward the creation of an adaptive, barrier-free learning space, where digital tools foster psychological comfort and academic achievement.

The research findings verify the hypothesis that sensory integration correction constitutes a fundamental predictor of successful foreign-language reception within an inclusive educational discourse. It has been established that a multisensory VR environment enables effective compensation for sensory deficits (visual, auditory, tactile, and proprioceptive) by stimulating neuroplasticity and activating learners' cognitive reserves. The experimental results demonstrate a statistically significant advantage for the experimental group in communicative competence and intrinsic motivation, accompanied by a substantial reduction in foreign-language anxiety (as measured by the FLCAS). Of particular scientific value is the demonstrated potential for radical individualization of instruction through the adaptive algorithms embedded in the VR environment. Consequently, multisensory VR technologies are identified as a relevant and highly effective tool in contemporary language pedagogy. Further research directions include standardizing sensory-speech correction protocols and implementing adaptive VR solutions within Ukraine's general education system, with support from international institutions, including UNICEF.

ACKNOWLEDGMENTS

The authors express their sincere gratitude to the administration and students of Zhytomyr Polytechnic State University, the Faculty of Pedagogical Technologies and Lifelong Learning, as well as to the administrations of Zhytomyr secondary schools and specialists of Inclusive Resource Centers contribution to data collection and analysis and for their support.

REFERENCES

[1] President of Ukraine, "On ensuring the creation of a barrier-free space in Ukraine," Decree No. 533/2020, Kyiv, Ukraine, 2020, [Online]. Available: https://www.president.gov.ua/documents/5332020-35809.

[2] M. Karatsiori, T. Liontou, E. Domagała-Zyśk, K. Vogt, M. Košak Babuder, and M. Poredos, "Beyond barriers: Exploring foreign language learning experiences of students with diverse learning needs in four European countries," in Education and Linguistics, 2025.

[3] UNICEF, "Situation Analysis," 2024, [Online]. Available: https://www.unicef.org/ukraine/media/49196/file/UNICEF_SitAn_2024_UKR.pdf.

[4] Cabinet of Ministers of Ukraine, "National strategy for creating a barrier-free space in Ukraine until 2030," as amended Mar. 25, 2025, No. 294-r, Action plan for 2025-2026, Kyiv, Ukraine, 2025, [Online]. Available: https://www.kmu.gov.ua.

[5] E. Christou, A. Parmaxi, and M. Christoforou, "Implementation and application of extended reality in foreign language education for specific purposes: A systematic literature review," in Universal Access in the Information Society, vol. 24, pp. 2061-2076, 2025, [Online]. Available: https://doi.org/10.1007/s10209-025-01191-w.

[6] M. Alfadil, "Immersive virtual reality: A novel approach to second language vocabulary acquisition in K-12 education," in Sensors, vol. 24, no. 22, Art. no. 7185, 2024, [Online]. Available: https://doi.org/10.3390/s24227185.

[7] M. Rudling, T. Rumetshofer, J. Nirme, and J. Mårtensson, "Embodied word learning in schools and sustained attention in virtual reality," in npj Science of Learning, vol. 11, Art. no. 9, 2026, [Online]. Available: https://doi.org/10.1038/s41539-025-00395-2.

[8] I. Schorr, D. A. Plecher, C. Eichhorn, and G. Klinker, "Foreign language learning using augmented reality environments: A systematic review," in Frontiers in Virtual Reality, vol. 5, Art. no. 1288824, 2024, [Online]. Available: https://doi.org/10.3389/frvir.2024.1288824.

[9] J. Dudley, L. Yin, V. Garaj, et al., "Inclusive immersion: A review of efforts to improve accessibility in virtual reality, augmented reality and the metaverse," in Virtual Reality, vol. 27, pp. 2989-3020, 2023, [Online]. Available: https://doi.org/10.1007/s10055-023-00850-8.

[10] M. J. Ibarra-Cabrera, R. W. Gamarra Chipa, H. R. Enriquez, Y. Mamani-Coaquira, H. A. Huillcen Baca, and D. C. Vilca, "Design and evaluation of a multisensory tangible game device for inclusive pre-Braille literacy," in Education Sciences, vol. 15, no. 9, Art. no. 1110, 2025, [Online]. Available: https://doi.org/10.3390/educsci15091110.

Application of Cloud-Based Digital Prototyping Systems for Developing Hardware Logic Skills

Andrii Lobatskyi[1], Pavlo Savaryn[1], Volodymyr Podolyak[2], Mykhailo Lepkiy[2] and Vitalii Kabak[1]
[1]*Department of Digital Educational Technologies, Faculty of Digital, Educational and Social Technologies, Lutsk National Technical University, Lvivska Str. 75, 43018 Lutsk, Ukraine*
[2]*Department of Tourism and Hotel and Restaurant Business, Lutsk National Technical University, Lvivska Str. 75, 43018 Lutsk, Ukraine*
a.lobatskyi@lntu.edu.ua, savaryn.pasha@lntu.edu.ua, v.podoliak@lutsk-ntu.com.ua, m.lepkiy@lutsk-ntu.com.ua, v.kabak@lutsk-ntu.com.ua

Keywords: Cloud-Based Prototyping, Digital Model, Digital-Twin-Inspired Workflow, EDA Tools, Computer Architecture Education, Virtual Verification, Hardware Logic, Tinkercad Circuits.

Abstract: The abstraction gap between theoretical Boolean logic and physical circuit implementation remains a persistent pedagogical challenge in Computer Architecture education. This paper investigates a simulation-first workflow that uses Tinkercad Circuits as a cloud-based digital model environment for pre-physical verification of hardware logic designs. The proposed workflow follows a four-phase iterative process: logic design specification, virtual prototyping, virtual verification with integrated debugging tools, and physical breadboard implementation. A pilot implementation with 20 first-year computer engineering students showed promising descriptive outcomes: 18 students, or 90%, completed all assignments with correct functionality; average laboratory time decreased from 120 minutes to 80-90 minutes per assignment; and instructor assistance shifted from mechanical troubleshooting to conceptual analysis of logic behavior, timing, and design alternatives. Because the study involved a single cohort without a parallel control group, these findings should be interpreted as preliminary evidence of feasibility and pedagogical potential rather than statistically conclusive proof of superiority over breadboard-first instruction.

1 INTRODUCTION

Contemporary computer engineering curricula face a fundamental pedagogical challenge: the abstraction gap between conceptual understanding of digital logic and practical implementation of hardware circuits [1], [2]. Students readily master truth tables, Boolean algebra, and logic gate theory in abstract form, yet encounter significant difficulties when translating these concepts into functional physical circuits. This disconnect manifests as frequent wiring errors, component misidentification, and an inability to diagnose circuit malfunctions-skills essential for professional competence in hardware design [3].

The Industry 4.0 paradigm has fundamentally transformed engineering practice, establishing digital prototyping and virtual commissioning as standard methodology for hardware development [4], [5]. Modern industrial workflows leverage digital twin technology-virtual replicas of physical systems that enable simulation, verification, and optimization before physical fabrication [6]. However, traditional computer architecture pedagogy remains anchored to breadboard-first methodologies that lack the iterative design verification cycles characteristic of contemporary engineering practice.

1.1 The Pedagogical Problem: Cognitive Overload in Novice Hardware Education

Introductory hardware laboratories present novice students with simultaneous cognitive demands: understanding abstract logic principles, managing physical component constraints, troubleshooting invisible electrical phenomena, and manipulating unfamiliar tools. Breadboard prototyping introduces systematic error patterns-component misplacement, incorrect polarity, loose connections, and misinterpreted topology-that consume instructional time on mechanical troubleshooting rather than conceptual development [3].

Electronic Design Automation (EDA) tools constitute the professional standard for circuit design and verification. High-fidelity commercial platforms such as Altium Designer and Proteus offer comprehensive simulation capabilities but present prohibitive entry barriers for novice learners: steep learning curves, complex licensing, and sophisticated feature sets that overwhelm introductory students. Cloud-based rapid prototyping environments-exemplified by Tinkercad Circuits-occupy an optimal middle ground, providing authentic simulation fidelity with minimal cognitive overhead for mastering the interface.

1.2 Research Objectives and Contribution

This paper investigates whether a simulation-first pedagogical workflow utilizing cloud-based digital twin environments can reduce implementation errors, accelerate concept-to-hardware translation, and shift student attention from mechanical assembly to logical analysis in introductory computer architecture education. We position Tinkercad Circuits not merely as educational software but as an accessible rapid-prototyping platform for creating digital models of hardware logic circuits within a digital-twin-inspired pedagogical workflow. We use the term digital-twin-inspired workflow in a pedagogical sense. In its strict industrial definition, a digital twin requires bidirectional, often real-time, data exchange between a physical object and its virtual representation, which is not implemented in this study. Accordingly, the Tinkercad Circuits prototype is treated as a simplified but pedagogically adequate digital model for pre-physical verification rather than as a runtime mirror of an existing physical artifact.

Our contribution is threefold:

1) articulation of a systematic four-phase digital prototyping workflow adaptable to diverse hardware logic curricula;
2) pilot implementation results demonstrating improved task completion rates, reduced laboratory time, and qualitative shift in assistance request patterns;
3) identification of cognitive mechanisms-particularly visualization of hidden electrical states and safe failure environments-that account for observed improvements.

2 LITERATURE REVIEW

2.1 Digital Twin Technology in Engineering Education

Digital twin technology has emerged as a transformative pedagogical instrument across engineering disciplines. Recent studies report that digital twin-based learning can improve efficiency and performance relative to conventional instruction [4], [5]. Educational digital twins provide cost-effective alternatives to physical laboratories while maintaining high simulation fidelity, addressing resource constraints in engineering programs [6], [11].

2.2 Simulation-First Versus Hands-On Laboratory Instruction

The comparative effectiveness of simulation-based versus hands-on laboratory instruction remains an active research question. Taher et al.'s comprehensive study of electronics education revealed a critical nuance: simulation environments are most effective when used as pre-physical scaffolding rather than as complete replacements for hands-on practice [7]. Students exposed to hybrid instructional sequences-simulation followed by physical implementation-demonstrated superior configuration speed and troubleshooting efficiency compared to either approach in isolation.

Kadir's research on virtual laboratories established positive transfer of training from virtual to physical environments, where students using virtual simulations showed significantly higher proficiency ($p < 0.05$) in subsequent physical implementation and troubleshooting compared to traditional groups [8]. These findings support a simulation-first rather than simulation-only pedagogical model, where virtual prototyping serves as conceptual scaffolding before tactile implementation [11].

2.3 Cloud-Based Tools for Hardware Education

Emerging research on Tinkercad Circuits validates its efficacy as a pedagogical tool for teaching hardware logic. Studies document learning outcome

improvements ranging from 11.5% to 25% among electronics and computational thinking assessments [9], [10]. Critically, cloud-based platforms provide a visual representation of electrical states-current flow, voltage levels, logic signal propagation-that remain invisible in physical breadboard implementations without specialized instrumentation [3]. This visualization capacity addresses a primary cognitive barrier in novice hardware education: the inability to perceive abstract electrical phenomena as tangible, observable processes. Real-time color-coded wire highlighting, logic level indicators, and virtual instrumentation enable students to construct accurate mental models of circuit behavior through direct observation rather than hypothetical reasoning.

2.4 Error Patterns in Breadboard Construction

DesPortes' analysis of novice breadboard construction identified recurring failure modes: component misplacement (incorrect IC orientation, wrong pin rows), wire-related errors (omissions, faulty connections, inadvertent disconnections during handling), and topology misunderstandings (failure to comprehend internal breadboard connection patterns) [3]. Novice students report these errors as "intimidating" and "difficult to diagnose," with error identification consuming disproportionate instructional time. The cognitive load imposed by simultaneous management of spatial constraints, tactile precision, and hidden electrical topology diverts working memory resources from the primary learning objective: understanding circuit logic and functionality.

3 METHODOLOGY

3.1 Pedagogical Context and Course Structure

This pilot implementation was conducted within the "Information Coding and Computer Architecture" course, a first-year required module for Computer Engineering students. The course curriculum covers combinational logic circuits (logic gates, half-adders, full-adders, decoders), sequential circuits (flip-flops, registers, counters), and basic microarchitecture concepts. Traditional course delivery employed breadboard-first laboratory sessions in which students constructed circuits directly from the instructors' schematic diagrams.

3.2 Study Design and Participants

The simulation-first workflow was implemented with one laboratory cohort N = 20 during the 2025-2026 academic year. Students had completed prerequisite coursework in digital logic fundamentals but had minimal prior experience with physical circuit construction or simulation tools. All participants provided informed consent for educational data collection in accordance with institutional review board protocols.

The laboratory classes were structured in the form of 2-hour blocks, which gradually became more complex: (1) two-input logic gates (AND, OR, NOT, XOR), (2) half-adder circuits, (3) full-adder implementation, (4) 4-to-1 multiplexer, (5) 3-to-8 decoder, and (6) 4-bit synchronous counter using D flip-flops. Each assignment required students to complete the four-phase workflow described below before receiving instructor verification.

3.3 The Four-Phase Digital Prototyping Workflow

We structured laboratory activities as an iterative engineering workflow analogous to professional hardware development cycles in Industry 4.0 contexts:

Phase 1: Logic Design Specification. Students receive functional requirements expressed as truth tables or Boolean expressions. Rather than immediately translating specifications into physical components, students first perform logic minimization using Karnaugh maps or Boolean algebra to determine optimal gate-level implementations. This phase emphasizes algorithmic thinking and design optimization without hardware constraints, typically requiring 10-15 minutes per assignment.

Phase 2: Digital model creation (Virtual Prototyping). Using the Tinkercad Circuits cloud platform, students construct virtual circuit prototypes by selecting components from a library that accurately model physical ICs (logic gates, flip-flops), discrete components (resistors, LEDs, switches), power supplies, and wiring. The platform provides realistic visual representations, including correct pinout diagrams and electrical characteristics, supporting consistency between the simulated model and the subsequent physical implementation Figure 1.

This phase serves dual pedagogical functions: (1) reinforcement of component identification and specification interpretation skills through visual library browsing, and (2) initial circuit topology

planning in a consequence-free environment where wiring errors produce no physical damage or time penalties beyond immediate correction.

Phase 3: Virtual Verification and Debugging. Students activate simulation mode to verify circuit functionality through systematic testing. Tinkercad provides real-time visualization capabilities critical for novice comprehension: active circuit traces are highlighted in distinct colors based on logic states (red for HIGH, blue for LOW, and gray for undefined), virtual LED indicators display outputs, and integrated multimeters enable voltage/continuity measurements Figure 2.

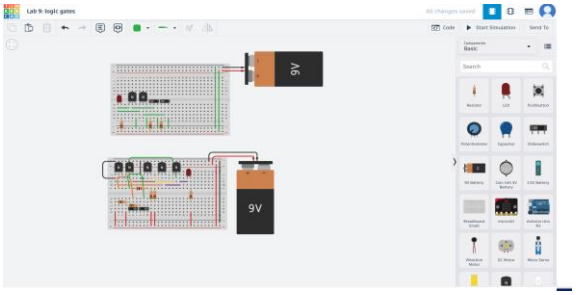


Figure 1: Cloud-based rapid prototyping environment in Tinkercad Circuits demonstrating virtual prototyping of combinational logic gates (AND, OR, XOR) with realistic component models, breadboard topology visualization, and power supply configuration.

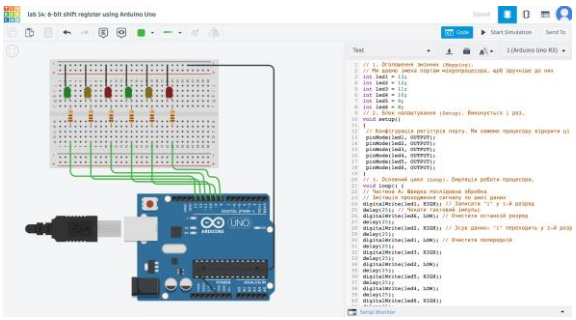


Figure 2: Advanced sequential logic implementation: 6-bit shift register using Arduino Uno microcontroller with visual LED output array. Color-coded wiring demonstrates organized circuit topology and real-time state visualization during simulation, illustrating progression from simple gates to programmatically controlled sequential circuits.

This phase shifted lab work from passive diagram-following to active experimentation. Tinkercad's visual feedback loop enabled interactive debugging (e.g., toggling inputs and observing real-time trace changes), reducing fear of component damage and supporting "what-if" testing. By making hidden electrical states observable-signal propagation, floating inputs, and shorts-students could correct errors immediately without physical instrumentation.

Phase 4: Physical Implementation. After verifying functionality in simulation, students build the circuit on a breadboard using their validated digital model rather than abstract schematics. They replicate component placement, wiring, and power connections from Tinkercad, which serves as assembly reference. As a result, hands-on work becomes a controlled translation of a pre-tested design instead of simultaneous construction and troubleshooting, reducing cognitive load.

3.4 Data Collection and Analysis Methods

We collected both quantitative performance metrics and qualitative observational data across all six laboratory assignments. Quantitative metrics included: (1) task completion status (functional circuit passing all specified tests), (2) time from session start to successful deployment, (3) number and category of instructor assistance requests, and (4) first-attempt success rate (circuit functions correctly upon initial physical assembly without rewiring). Teaching assistants logged assistance requests using a structured taxonomy distinguishing basic mechanical issues (component orientation, wire connections, power supply) from conceptual questions (logic analysis, timing behavior, circuit modifications).

Qualitative data collection employed semi-structured observation protocols during laboratory sessions, noting student debugging strategies, error self-correction behaviors, and expressed confidence levels. Post-course anonymous surveys solicited student perceptions of workflow effectiveness, cognitive load experiences, and preference comparisons between simulation-first and traditional approaches.

4 RESULTS

4.1 Task Completion and Implementation Success

Among the 20 students who completed the simulation-first workflow, 18 students, or 90%, successfully implemented all six laboratory assignments, with functionality verified through systematic testing. This exceeds typical outcomes

from prior breadboard-first semesters, where success rates for complex tasks (e.g., multiplexers, counters) often ranged 60-70% within the allotted lab time.

The two students who struggled during physical implementation still verified correct designs in Tinkercad but faced breadboard assembly issues-mainly incomplete IC pin insertion causing intermittent contacts. They quickly isolated these mechanical faults by comparing physical behavior with the known-good simulation, indicating effective transfer of the virtual verification approach to the physical setup.

4.2 Time Efficiency Analysis

Laboratory time-allocation analysis showed substantial efficiency gains versus traditional instruction. In prior semesters, students typically required the full 120 minutes, with 30-40% failing to complete functional implementations within the session and needing additional office hours.

With the simulation-first workflow, students finished in 80-90 minutes on average (25-33% reduction). Time was distributed across phases: Phase 1 (15-20 min), Phase 2 (20-25 min), Phase 3 (~20 min), Phase 4 (25-30 min). Because designs were pre-verified in simulation, physical assembly shifted from troubleshooting wiring errors to more controlled execution focused on tactile accuracy, guided by students' digital twins.

Overall, this provided ~30 minutes of additional learning time per lab; across six assignments, this equals 3-4 hours of recovered instructional time per student, reallocated from troubleshooting to design variations and conceptual analysis.

4.3 Qualitative Shift in Assistance Request Patterns

The most pedagogically significant finding was the shift in instructor assistance requests. Teaching assistants kept structured logs classifying questions as mechanical/procedural (component identification, wiring, power, polarity) or conceptual/analytical (logic behavior, truth-table reasoning, timing, design modifications).

In breadboard-first labs, mechanical requests dominated: students needed help with wiring errors, IC orientation, loose connections, and fault location; instructors estimated 70-80% of support time went to low-level troubleshooting. Under the simulation-first workflow, requests shifted toward reasoning and design.

Examples from the logs illustrate the change. Traditional questions: "Which direction should this IC face?", "Where do I connect ground?", "Why isn't my LED lighting up?" Simulation-first questions: "Why does this gate configuration produce this output pattern?", "How would I extend this 4-bit counter to 8 bits?", "What happens if I change the clock frequency?" Overall, student attention moved from mechanical assembly to logical analysis and system-level thinking.

5 DISCUSSION

5.1 Visualization of Hidden Electrical States

The primary mechanism behind the improved outcomes is state visualization: in breadboard labs, voltage levels, current flow, and logic states are largely invisible without instrumentation, forcing students to rely on assumptions rather than observation. Tinkercad makes these states observable through color-coded logic traces, voltage probes, and signal-path highlighting, so students can empirically track propagation from inputs through gates to outputs (e.g., LEDs). This directly reduces novice diagnostic barriers documented in hardware education [3] and supports faster fault isolation using concrete checks such as: "Is this wire connected?", "Is power reaching the IC?", "What logic level is this node?", and "Why is the output incorrect?"

5.2 Safe Failure Environment and Iterative Learning

Traditional breadboarding can penalize experimentation: wiring mistakes may damage components, create safety risks, or require instructor intervention, which encourages risk-averse behavior and reduces independent hypothesis testing. Virtual prototyping removes these material consequences, providing immediate error/short-circuit indicators and allowing iterative "what-if" debugging without damage. This safe-to-fail setting supports an engineering mindset-reating errors as diagnostic feedback-and plausibly explains the observed shift from mechanical fault-location questions toward conceptual analysis and more independent troubleshooting during physical assembly [4], [5].

5.3 Cognitive Load Reduction Through Phased Separation

Traditional breadboard-first instruction imposes simultaneous cognitive demands: students must

concurrently manage abstract logic concepts, physical component constraints, spatial wire routing, tactile precision, and hidden electrical topology. This cumulative cognitive load overwhelms working memory capacity, particularly for novice learners, resulting in both conceptual and mechanical errors [3].

The four-phase workflow strategically separates these cognitive demands across distinct activities. Phase 1 isolates pure logic design without hardware constraints. Phase 2 addresses component identification and specification interpretation in a visual interface without time pressure. Phase 3 focuses exclusively on logic verification and debugging with visible state feedback. Phase 4 becomes a mechanical translation of the pre-verified design, requiring only spatial reasoning and tactile skill without simultaneous logic analysis.

This phased separation enables students to develop competency in each skill domain independently before integration. By the time students approach physical breadboards, they have already verified logic correctness and established mental models of expected behaviour. However, the workflow should be understood as simulation-first rather than simulation-only. Tinkercad Circuits does not fully reproduce contact resistance, loose breadboard connections, component tolerances, or voltage instability under load. Therefore, Phase 4 should include brief hardware validation-checking power/ground continuity, measuring selected nodes, and discussing simulation-hardware discrepancies-o prevent an "illusion of simplicity" and distinguish logical correctness from physical reliability.

5.4 Alignment with Industry 4.0 Engineering Practice

Beyond immediate learning outcomes, positioning cloud-based prototyping as a digital-twin-inspired methodology yields benefits for professional identity formation. Contemporary hardware development universally employs simulation-before-fabrication workflows: digital verification using CAD tools and SPICE simulation precedes printed circuit board manufacturing, FPGA design undergoes extensive simulation before synthesis, and embedded systems utilize hardware-in-the-loop testing before production deployment [4], [6].

By structuring laboratory instruction around authentic professional workflows-design specification, digital model creation, virtual verification, physical implementation-students internalize industry-standard engineering practices from introductory coursework. This alignment prepares students for workplace expectations while demonstrating that educational activities reflect authentic professional competencies rather than artificial academic exercises.

6 CONCLUSIONS

The simulation-based workflow, tested as a pilot implementation with one cohort of first-year students N = 20, produced promising descriptive outcomes: 90% task completion and an observed 25-33% reduction in laboratory time. Given the limited sample size and the absence of a parallel control group, these results should be interpreted as preliminary evidence of feasibility and pedagogical potential rather than statistically conclusive proof of superiority over breadboard-first instruction. Reducing mechanical troubleshooting to 80-90 minutes freed up time for higher-level learning and helped bridge the gap between abstract Boolean logic and computer architecture practice.

The observed effect can be explained by three mechanisms: (1) visualization of hidden electrical states, (2) a safe environment for errors that supports exploratory learning without material/psychological risks, and (3) reduced cognitive load by separating logical design, virtual verification, and physical implementation into separate stages.

Cloud-based modeling is not considered a replacement for hands-on work, but rather a preliminary virtual verification stage that increases the efficiency of subsequent physical implementation, while final breadboard validation remains necessary. This hybrid approach (digital model → physical implementation) is consistent with Industry 4.0 practices and facilitates the transition of beginners from theory to circuit design.

Based on our pilot implementation experience, we recommend the following for instructors who use simulation-oriented workflows: (1) clearly position Tinkercad as an introductory EDA environment, not a "toy simulator," to create an authentic engineering context; (2) require students to document the results of virtual verification (screenshots, test case tables) before approving physical implementation to ensure that steps 2-3 are actually performed, rather than superficial modeling; (3) allocate 40-50% of lab time to virtual prototyping steps, leaving the rest of the time for physical assembly, basic voltage checks, simulation-hardware comparison, and conceptual discussion, rather than basic troubleshooting; and (4)

use the freed-up time to develop design options that promote a creative approach to problem solving, rather than mechanically reproducing the provided diagrams.

7 FUTURE WORK

As engineering education increasingly addresses Industry 4.0 workforce demands, cloud-based digital prototyping offers an accessible way to modernize hardware logic instruction without costly equipment. Within this pilot implementation, the workflow suggests that free cloud tools may support learning effectiveness, task completion, and pedagogical focus while familiarizing students with professional engineering practices. However, because the results come from a single cohort N = 20 without a parallel control group, they should be interpreted as exploratory rather than statistically conclusive. Future studies should test these trends with larger samples, control groups, and inferential statistical analysis. Such work would clarify the scalability of simulation-first approaches in computer engineering curricula.

REFERENCES

[1] C. Qian, S. Chen, and A. Goodney, "Bridging simulation and hardware: A holistic platform for teaching in computer architecture and operating systems," in Proceedings of the 56th ACM Technical Symposium on Computer Science Education v. 2 (SIGCSE TS 2025), Pittsburgh, PA, USA, 2025, pp. 1591-1592, [Online]. Available: https://doi.org/10.1145/3641555.3705135.

[2] C. Mirolo, C. Izu, V. Lonati, and E. Scapin, "Abstraction in Computer Science Education: An Overview," Informatics in Education, vol. 20, no. 4, pp. 615-639, 2021, [Online]. Available: https://doi.org/10.15388/infedu.2021.27.

[3] K. DesPortes et al., "BitBlox: A redesign of the breadboard," in Proceedings of the 15th ACM Conference on Interaction Design and Children (IDC '16), Manchester, United Kingdom, 2016, pp. 203-212, [Online]. Available: https://doi.org/10.1145/2930674.2930708.

[4] M. Pajpach, R. Pribiš, P. Drahoš, E. Kučera, and O. Haffner, "Methodology and pilot implementation of modular educational-development platform for interoperable digital twin," Scientific Reports, vol. 15, art. no. 28020, 2025, [Online]. Available: https://doi.org/10.1038/s41598-025-13579-y.

[5] J. S. Ortiz, E. K. Quishpe, G. X. Sailema, and N. S. Guamán, "Digital twin-based active learning for industrial process control and supervision in industry 4.0," Sensors, vol. 25, no. 7, art. no. 2076, 2025, [Online]. Available: https://doi.org/10.3390/s25072076.

[6] M. A. Hazrat, N. M. S. Hassan, A. A. Chowdhury, M. G. Rasul, and B. A. Taylor, "Developing a skilled workforce for future industry demand: The potential of digital twin-based teaching and learning practices in engineering education," Sustainability, vol. 15, no. 23, art. no. 16433, 2023, [Online]. Available: https://doi.org/10.3390/su152316433.

[7] M. T. Taher and A. S. Khan, "Effectiveness of simulation versus hands-on labs: A case study for teaching an electronics course," in Proceedings of the 122nd ASEE Annual Conference & Exposition, Seattle, WA, USA, 2015, art. no. 26.582, [Online]. Available: https://doi.org/10.18260/p.23920.

[8] A. Kadir, "Effectiveness of virtual laboratory utilization in improving students' science process skills," Ta'dib, vol. 26, no. 2, pp. 355-366, 2023, [Online]. Available: https://doi.org/10.31958/jt.v26i2.10913.

[9] Asbendri BL, S. Anori, I. P. Dewi, and Efrizon, "Exploring the impact of Tinkercad-assisted learning on student performance in industrial electronics subject," Journal of Hypermedia & Technology-Enhanced Learning (J-HYTEL), vol. 2, no. 2, pp. 134-148, 2024, [Online]. Available: https://doi.org/10.58536/j-hytel.v2i2.124.

[10] C. Tüysüz, N. C. Bodur, and I. Ugulu, "Tinkercad Circuits Platform-Based Learning Experiences of Gifted Students in the Emergency Distance Education Process," Journal of Advanced Academics, vol. 35, no. 2, pp. 329-356, 2024, [Online]. Available: https://doi.org/10.1177/1932202X241230589.

[11] O. Hulai, V. Shemet, I. Moroz, P. Savaryn, and V. Kabak, "Using the Labster virtual laboratory to study chemistry at a technical university," Information Technologies and Learning Tools, vol. 102, no. 4, pp. 95-107, 2024, [Online]. Available: https://doi.org/10.33407/itlt.v102i4.5737.

Hochshule Anhalt
EDITION
ICAIIT

INFORMATION TECHNOLOGY IN ECONOMICS

Section 7

APRIL 2026



Proceedings of the International Conference on Applied Innovations in IT

Digital Marketing as a Strategic Tool for Promoting Agricultural Products

Nataliia Trushkina[1], Maryna Popova[2], Rahayu Relawati[3], Bambang Yudi Ariadi[3] and Yelyzaveta Loiko[4]

[1]*Research Center for Industrial Problems of Development of the National Academy of Sciences of Ukraine, Inzhenernyi Lane 1a, 61165 Kharkiv, Ukraine*

[2]*Anhalt University of Applied Sciences, Bernburger Str. 57, 06366 Köthen, Germany*

[3]*Department of Agribusiness, Faculty of Agriculture and Animal Husbandry, University of Muhammadiyah Malang, Jalan Raya Tlogomas No. 246, 65144 Malang, Indonesia*

[4]*Department of International Economics, Borys Grinchenko Kyiv Metropolitan University, Levka Lukianenka Str. 13-B, 04212 Kyiv, Ukraine*

nata_tru@ukr.net, Maryna.Popova@hs-anhalt.de, rahayurelawati@umm.ac.id, bambang_y@umm.ac.id, y.loiko.asp@kubg.edu.ua

Keywords: Digital Marketing, Agribusiness, Digital Agriculture, Agri-food Systems Resilience, Bibliometric Analysis, Digital Transformation, Marketing Promotion, Agricultural Product, Customer Relationship Management (CRM), Marketing Strategy, Data-driven Marketing, Supply Chains, Sustainability, Sustainable Development.

Abstract: In the context of the digital transformation of the economy, the agricultural sector is undergoing significant changes in the mechanisms of production, promotion, and distribution of products. Digital marketing is evolving from a communication tool into a strategic component of agribusiness development, enabling the integration of production, logistics, and market processes within agri-food systems. The aim of the article is to identify the structural and thematic patterns in the development of scientific research on digital marketing in the agricultural sector and to substantiate its role as a strategic tool for promoting agricultural products based on a bibliometric analysis. The study is based on a bibliometric analysis of scientific publications indexed in the Scopus database, using science mapping techniques, including keyword co-occurrence analysis, overlay visualization, and country citation analysis, with the application of VOSviewer software. The results reveal key thematic research clusters, their evolution over time, and the geographical distribution of scientific activity. It is shown that the research landscape is formed at the intersection of digital agriculture, e-commerce, digital platforms, and marketing communications. At the same time, a structural imbalance is identified, with a predominance of technological studies and limited attention to marketing and behavioural aspects. The findings contribute to a deeper understanding of digital marketing as an integrative element of agri-food digital ecosystems and can be used to develop effective marketing strategies in agribusiness.

1 INTRODUCTION

The rapid development of digital technologies is transforming the global economy and reshaping the ways products are produced, distributed, and promoted. These processes are particularly significant for the agricultural sector, which plays a crucial role in ensuring food security and supporting national economic development. According to the Food and Agriculture Organization of the United Nations, agriculture accounts for approximately 4 percent of global gross domestic product, while its contribution is considerably higher in many developing countries [1].

The digitalization of agriculture contributes to the transformation of traditional agribusiness models through the integration of digital platforms, electronic commerce, customer relationship management systems, and data analytics tools. These technologies create new opportunities for promoting agricultural products, improving market access, and strengthening interaction between

producers and consumers. As noted by L. Klerkx, E. Jakku, and P. Labarthe [2], digital transformation facilitates the emergence of new agribusiness models that combine production processes, information technologies, and digital communication tools.

A key component of this transformation is the growing use of digital marketing technologies. Digital marketing includes social media, e-commerce platforms, online communication channels, and digital analytical systems that enable enterprises to improve consumer engagement and expand their presence in global markets. Research demonstrates that the implementation of digital marketing strategies contributes to higher competitiveness, sales growth, and more effective market positioning of enterprises [3].

In addition, digital technologies support the development of sustainable agriculture and improve the efficiency of agri-food value chains by enhancing production transparency, product traceability, and consumer trust [4]. They also facilitate the integration of small and medium-sized agricultural producers into international markets and expand opportunities for economic development [5].

Despite the rapid expansion of digital technologies in agriculture, the role of digital marketing as a strategic instrument for promoting agricultural products remains insufficiently explored. In particular, there is a need to identify current trends in the development of digital marketing in agribusiness and to analyse the structure and evolution of scientific research in this field.

2 LITERATURE REVIEW

2.1 Digital Transformation of Agriculture

Contemporary scientific research increasingly focuses on the digital transformation of the agricultural sector. Digital technologies are being actively integrated into agricultural production, enterprise management, and agri-food value chains, contributing to more efficient resource use, higher productivity, and the emergence of new business models.

Studies emphasize that technologies such as big data analytics, sensor systems, digital platforms, and artificial intelligence form the technological basis of smart agriculture and support data-driven decision-making processes [4]. Digital platforms and internet technologies also improve the coordination of agri-food supply chains, enhance production transparency, and strengthen interaction among market participants [6]. In addition, the application of digital analytical systems significantly improves resource management and operational efficiency in agriculture [7].

Recent research further demonstrates that the digital transformation of agri-food systems contributes to greater transparency, resilience, and sustainability of agricultural supply chains [8]. As a result, digital technologies are increasingly regarded not only as production tools, but also as strategic mechanisms for transforming agri-food systems and market interactions.

2.2 Digital Marketing and Digital Communication in Modern Business

Another important research direction concerns the role of digital marketing in the development of modern business. In the scientific literature, digital marketing is considered a set of technologies and communication tools used to promote products and establish interaction with consumers through digital channels.

Researchers note that digital marketing transforms traditional marketing approaches by expanding the use of social media, mobile technologies, online platforms, and electronic commerce [3]. The development of digital communication tools creates new opportunities for enterprises to strengthen customer engagement, improve market positioning, and build competitive advantages in the digital economy [9].

Contemporary studies also emphasize the growing importance of big data analytics and artificial intelligence in marketing activities. Digital analytical systems enable enterprises to better identify consumer needs, personalize marketing communications, and improve the effectiveness of product promotion [10]. Consequently, digital marketing is increasingly integrated with platform-based business models and data-driven decision-making processes [11].

2.3 Digital Marketing in Agribusiness

A separate strand of scientific research is devoted to the application of digital marketing in the agricultural sector. Under conditions of digital transformation, agricultural enterprises increasingly use online platforms, electronic commerce, and

social media to promote products, develop brands, and strengthen interaction with consumers.

Research demonstrates that digital platforms and e-commerce systems expand market access for agricultural producers and facilitate direct communication with consumers [12]. The use of internet marketing tools contributes to broader distribution channels, stronger brand recognition, and higher levels of consumer trust in agricultural products. Social media also plays an important role in disseminating information about product quality, sustainability, and food safety, while supporting the formation of positive brand images for agricultural enterprises [13], [14].

At the same time, existing scientific literature is still largely concentrated on technological aspects of agricultural digitalization, including precision farming, big data, and digital platforms [15]-[17]. Comparatively less attention is devoted to digital marketing as a strategic mechanism for promoting agricultural products and transforming market interactions within agri-food systems.

In addition, insufficient attention has been paid to the systematic analysis of the intellectual structure and thematic evolution of research in this field, including the identification of major research clusters, publication dynamics, and emerging scientific trends [18]-[22]. Therefore, bibliometric analysis represents an appropriate methodological approach for identifying the main directions of research development and assessing the role of digital marketing in the transformation of agribusiness.

Accordingly, the aim of this study is to identify the structural and thematic patterns in the development of scientific research on digital marketing in the agricultural sector through bibliometric analysis and to substantiate its role as a strategic tool for promoting agricultural products in the context of the digital transformation of agri-food systems.

3 METHODOLOGY

This study applies bibliometric analysis to examine the development of scientific research on digital marketing in the agricultural sector. Bibliometric methods are widely used to analyse publication dynamics, identify thematic research areas, and assess the intellectual structure of scientific fields [18]-[20].

The bibliometric dataset was obtained from the Scopus database, which provides extensive interdisciplinary coverage of peer-reviewed scientific publications and is widely used in bibliometric and science mapping studies. Scopus was selected due to its high-quality indexing standards, comprehensive metadata, and compatibility with bibliometric visualization tools such as VOSviewer (Table 1).

The search process focused on publications related to digital marketing and the agricultural sector. The search was conducted within the title, abstract, and keyword fields using the following query:

TITLE-ABS-KEY (“digital marketing” AND (agriculture OR agribusiness OR “agri-food” OR “food marketing”)).

The analysis primarily covers the period 2017-2026, which reflects the active stage of development of research on digital marketing in agribusiness. At the same time, earlier highly cited publications were also included to capture the theoretical foundations of the research field.

Table 1: Search strategy and dataset description.

Parameter	Description
Database	Scopus
Search fields	Title, Abstract, Keywords
Search query	“digital marketing” AND (agriculture OR agribusiness OR agri-food OR food marketing)
Document types	Articles, Review papers
Language	English
Data extraction	February 2026
Export format	CSV
Analysis tool	VOSviewer

3.1 Data Selection and Processing

The selection and processing of publications were conducted in several stages. Only research articles and review papers indexed in the Scopus database and published in English were included in the analysis. Duplicate and irrelevant records were excluded after examining titles, abstracts, and keywords.

After the data cleaning procedure, the final bibliometric dataset was exported in CSV format and included information on authors, publication sources, keywords, countries, and citation indicators. The processed dataset was subsequently used for bibliometric mapping and visualization.

3.2 Bibliometric Analysis and Visualization Tools

To analyse the intellectual structure and thematic development of research on digital marketing in agribusiness, bibliometric and science mapping methods were applied [20]-[22]. Visualization and network analysis were performed using VOSviewer software, which enables the construction of bibliometric maps based on citation and keyword relationships [22], [23].

The study applied the following bibliometric techniques [23]:

- keyword co-occurrence analysis to identify major thematic clusters;
- overlay visualization to examine the temporal evolution of research topics;
- country citation analysis to assess the geographical distribution of scientific activity.

The application of these methods made it possible to identify the dominant research directions, relationships between thematic areas, and the overall structure of scientific research on digital marketing in the agricultural sector.

4 RESULTS

This section presents the results of the bibliometric analysis of scientific publications on digital marketing in the agricultural sector. The analysis identifies major research trends, thematic structures, and geographical patterns of scientific activity related to the digital transformation of agribusiness.

4.1 Dynamics of Scientific Publications

The analysis of publication dynamics based on the Scopus database demonstrates the rapid expansion of research on digital marketing in agriculture. Prior to 2018, publication activity remained limited and fragmented. However, beginning in 2019, the field entered a phase of intensive growth characterized by increasing scientific interest and the gradual formation of a stable research domain.

As shown in Table 2, the number of publications increased from 10 in 2019 to 37 in 2025, reflecting a 3.7-fold increase over the period. The most significant growth was recorded in 2019 and 2022, indicating the acceleration of research activity related to digital transformation processes in agribusiness.

Table 2: Dynamics of scientific publications on digital marketing in the agricultural sector (scopus, 2017-2026).

Year	Number of publications	Absolute increase	Growth rate, %
2017	4	-	-
2018	2	-2	50.0
2019	10	+8	500.0
2020	11	+1	110.0
2021	13	+2	118.2
2022	22	+9	169.2
2023	26	+4	118.2
2024	30	+4	115.4
2025	37	+7	123.3
2026	12	-25	32.4

Note: Data for 2026 are partial and reflect the status of the Scopus database at the time of data extraction (February 2026).

The observed growth dynamics are associated with the rapid expansion of digital technologies, the increasing use of online marketing channels in agribusiness, and the acceleration of digitalization during the COVID-19 pandemic. At the same time, the decline in 2026 is explained by incomplete database coverage at the time of data collection and should not be interpreted as a negative trend.

To identify the intellectual foundations of the research field, an analysis of the most cited publications in the Scopus dataset was conducted (Table 3).

The results indicate that the intellectual core of the field extends beyond narrowly defined digital marketing and includes broader issues related to digital agriculture, supply chain transformation, electronic commerce, social media, and agri-food resilience. The most cited publications primarily focus on digitalization processes, online market interaction, and technological transformation of agricultural systems.

The analysis also demonstrates the interdisciplinary nature of the field, which integrates digital technologies, marketing, logistics, and agri-food management. Consequently, digital marketing in agriculture should be considered not only as a promotional instrument, but also as a component of broader digital ecosystems connecting production, communication, and distribution processes.

Additional insight into the disciplinary structure of the field is provided by the analysis of leading academic journals publishing research on digital marketing and digital technologies in agriculture (Table 4).

Table 3: Most cited publications in the scopus dataset on digital technologies and marketing in the agricultural sector (scopus, 2000-2024).

Authors	Year	Citations	Main scientific focus
D. J. Choruma et al. [24]	2024	107	Digitalization of agriculture, technology adoption practices, barriers and opportunities for small-scale producers
M. Alkahtani et al. [25]	2021	64	Electronic management of agricultural supply chains, blockchain technologies, cooperative strategies
L.A. Duram [26]	2000	59	Socio-economic and institutional factors of agricultural sector development
A.U. Din et al. [27]	2022	49	Impact of the pandemic on food supply chains and the role of electronic commerce
M. Prosper Bright et al. [28]	2021	36	Digital and organizational transformations in agricultural consulting and food provision systems
A. Elghannam et al. [29]	2017	36	Social media and the development of short agri-food supply chains
K.A. Nagaty [30]	2023	29	Artificial intelligence-powered Internet of Things in commercial and industrial applications
I. Grigorescu et al. [31]	2022	28	Resilience of peri-urban small-scale farming under pandemic conditions
S. Ravi & S. R. C. Rajasekaran [32]	2023	26	Digital marketing in rural areas
K.M.I. Bashir et al. [33]	2019	26	Marketing strategies and competitiveness of agri-food companies in international markets

Table 4: Leading academic journals publishing research on digital marketing and digital technologies in the agricultural sector.

Rank	Journal	Publisher	Research focus	CiteScore 2024	SJR 2024	Quartile
1	Computers and Electronics in Agriculture	Elsevier	Digital technologies and smart farming	13.5	1.85	Q1
2	Agricultural Systems	Elsevier	Agricultural innovation and sustainability	9.8	1.65	Q1
3	Technological Forecasting and Social Change	Elsevier	Digital transformation and innovation	12.2	2.30	Q1
4	International Journal of Information Management	Elsevier	Digital technologies and information systems	20.0	3.50	Q1
5	Journal of Business Research	Elsevier	Digital marketing and business transformation	11.0	2.10	Q1
6	Journal of the Academy of Marketing Science	Springer	Marketing research and strategy	25.0	4.50	Q1
7	International Journal of Production Research	Taylor & Francis	Digital supply chains	8.5	1.75	Q1
8	Journal of Retailing and Consumer Services	Elsevier	Consumer behaviour and digital retail	10.5	1.90	Q1
9	British Food Journal	Emerald	Agri-food marketing	5.2	0.85	Q2
10	Journal of Rural Studies	Elsevier	Rural development and agricultural innovation	8.0	1.60	Q1

Note: CiteScore and SJR indicators are presented according to Scopus and SCImago data for 2024.

The results show that the research field is largely shaped by interdisciplinary journals focused on digital transformation, innovation, information systems, and agricultural development. Journals such as Computers and Electronics in Agriculture, Agricultural Systems, and Technological Forecasting and Social Change occupy leading positions, indicating the close connection between digital marketing research and broader processes of technological modernization of agri-food systems.

At the same time, the presence of marketing-oriented journals, including Journal of Business Research and Journal of the Academy of Marketing Science, reflects the growing attention to strategic,

behavioural, and communication dimensions of digitalization in agribusiness.

Overall, the results confirm that digital marketing research in agriculture is characterized by a high degree of interdisciplinarity and combines technological, managerial, and market-oriented perspectives.

4.2 Keyword Co-Occurrence Analysis

To identify the thematic structure of the research field, a keyword co-occurrence analysis was conducted using VOSviewer (Fig. 1).

The analysis demonstrates that the concepts marketing and digital marketing occupy central positions within the network and are connected with several major thematic clusters. The first cluster is related to electronic commerce and digital distribution channels and includes terms such as e-commerce, online marketing, and marketplaces. The second cluster focuses on strategic management and supply chain coordination, including concepts such as marketing strategy, decision support systems, and supply chains.

Another cluster is associated with sustainability and agricultural innovation, reflecting the growing importance of transparency, traceability, and sustainable production within agri-food systems. In addition, socio-behavioural topics related to social media marketing, consumer interaction, and COVID-19 are also strongly represented in the network.

Thus, digital marketing in agribusiness functions as an integrative research domain connecting technological, managerial, and socio-economic dimensions of agricultural development.

4.3 Evolution of Research Directions Over Time

Overlay visualization analysis reveals a gradual evolution of scientific discourse in the field (Fig. 2). Earlier studies were primarily focused on information technologies, electronic commerce, and rural digitalization. Subsequently, research attention shifted toward online marketing, digital platforms, and supply chain integration.

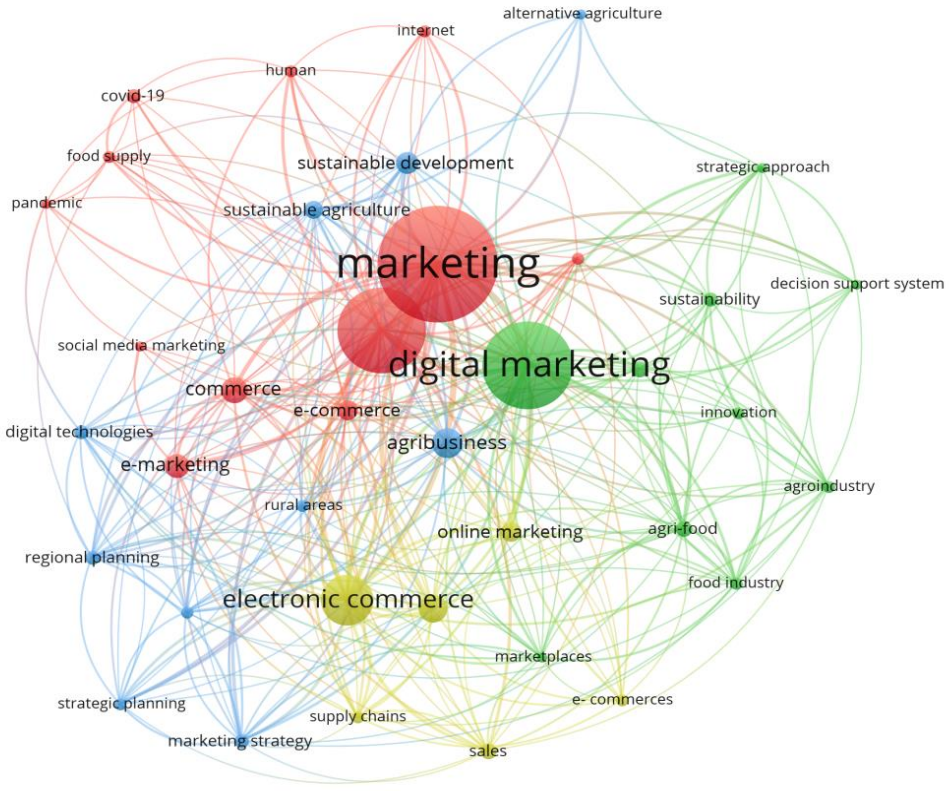


Figure 1: Co-occurrence map of keywords in research on digital marketing in the agricultural sector.

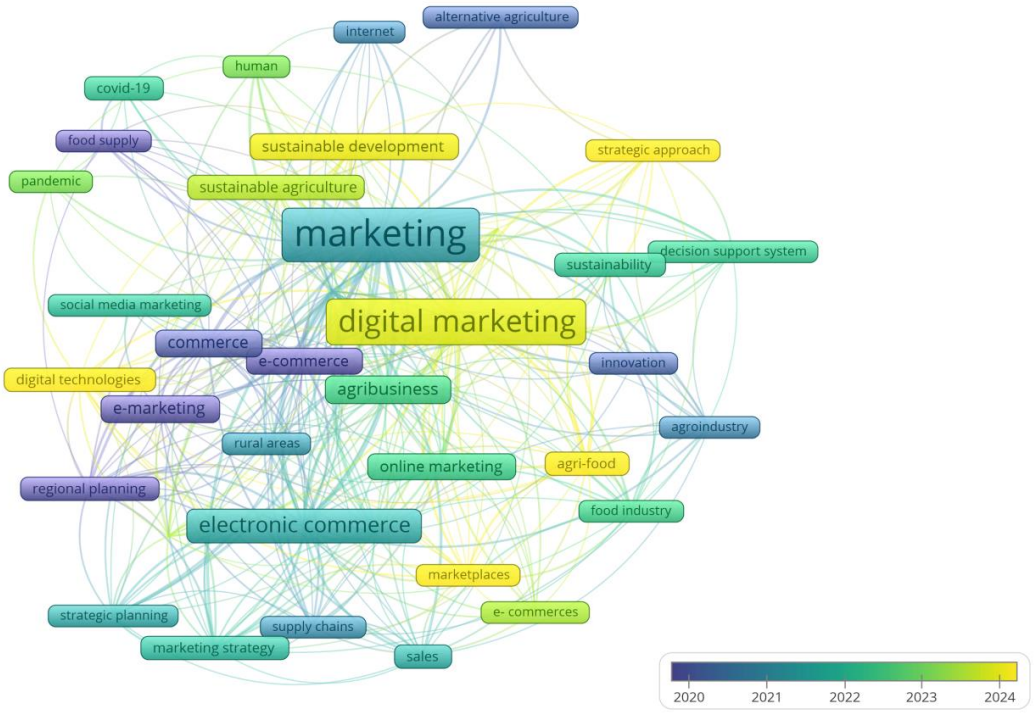


Figure 2: Overlay visualization of the evolution of research on digital marketing in the agricultural sector.

More recent studies increasingly emphasize sustainability, innovation, decision support systems, and data-driven management approaches. This transformation indicates a transition from purely technological perspectives toward more integrated approaches in which digital marketing is considered a strategic tool for value creation and market coordination within agri-food systems.

4.4 Geographical Structure of Scientific Publications

The geographical analysis demonstrates that scientific activity is concentrated primarily in countries with advanced digital economies, including the United States, China, and European Union member states (Fig. 3). These regions benefit from developed digital infrastructure, high levels of innovation investment, and active implementation of digital technologies in agriculture.

At the same time, developing countries remain comparatively underrepresented despite the growing importance of digital transformation for agricultural modernization. This indicates the need to strengthen international collaboration and expand research related to regional and institutional specificities of agribusiness digitalization.

The obtained results confirm that digital marketing research in agriculture is increasingly shifting from isolated technological applications toward integrated models combining electronic commerce, supply chain management, sustainability, and consumer interaction.

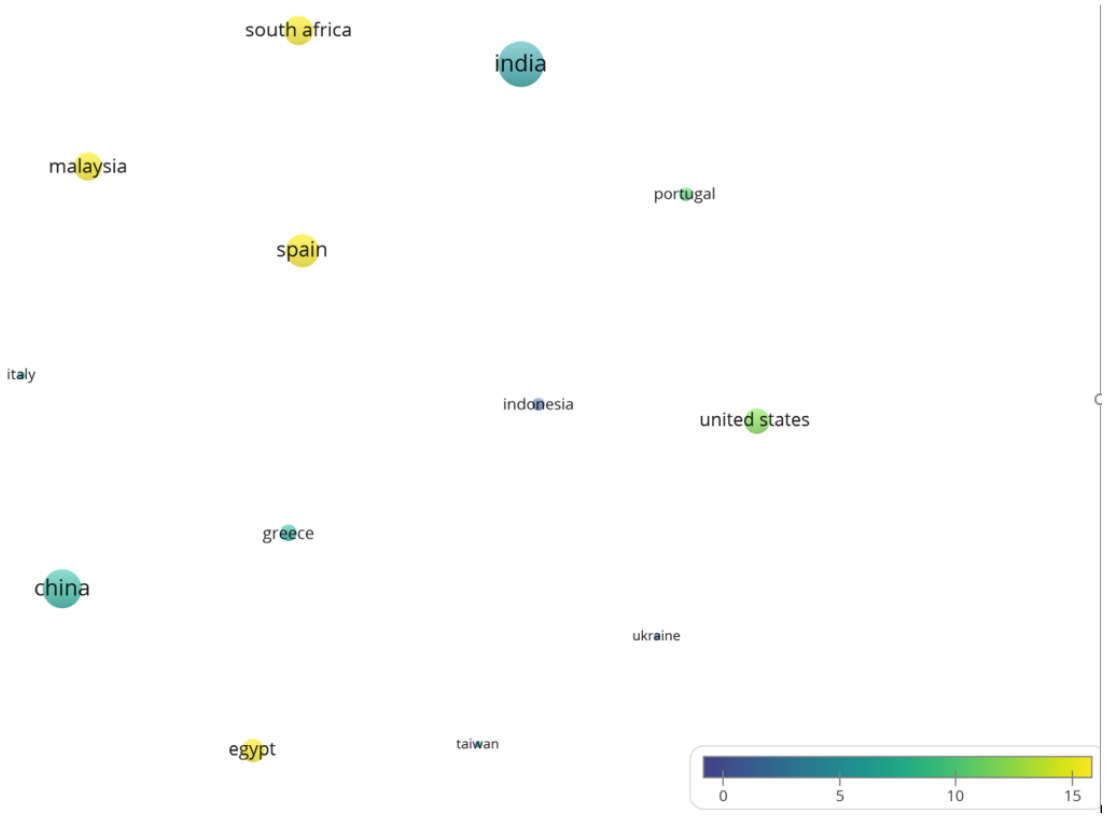


Figure 3: Geographic distribution of scientific publications on digital marketing in the agricultural sector.

4.5 Conceptual Framework of Digital Marketing in Agribusiness

The bibliometric analysis demonstrates that research on digital marketing in agriculture is formed at the intersection of digital agriculture, information technologies, electronic commerce, and marketing. Based on the obtained results, a conceptual framework of digital marketing development in agribusiness is proposed (Fig. 4) [23], [34], [35].

The framework reflects the interconnection between digital transformation drivers, agricultural technologies, digital marketing instruments, and business outcomes. At the first level, global digitalization processes and technological innovation shape the external environment of agribusiness. At the second level, technologies such as big data analytics, IoT, artificial intelligence, and digital platforms support data-driven decision-making and operational efficiency.

At the third level, digital marketing tools — including social media marketing, e-commerce platforms, content marketing, and digital communication channels — facilitate interaction with consumers and strengthen market presence. The integrated application of these technologies contributes to market expansion, stronger brand recognition, improved traceability, and increased competitiveness of agricultural enterprises.

Thus, digital marketing in agribusiness should be considered not only as a communication tool, but also as a strategic mechanism integrating technological innovation, market coordination, and consumer interaction within agri-food systems.

The proposed framework reflects a logically structured interconnection between the key elements of digital transformation and the strategic development of agribusiness in the digital economy.

At the first level, a determining role is played by global processes of digital transformation, including the development of the digital economy, the globalization of agri-food markets, the formation of sustainable food systems, and technological innovation. These factors shape the external environment in which agricultural enterprises operate and determine the conditions for the digital modernization of agri-food markets.

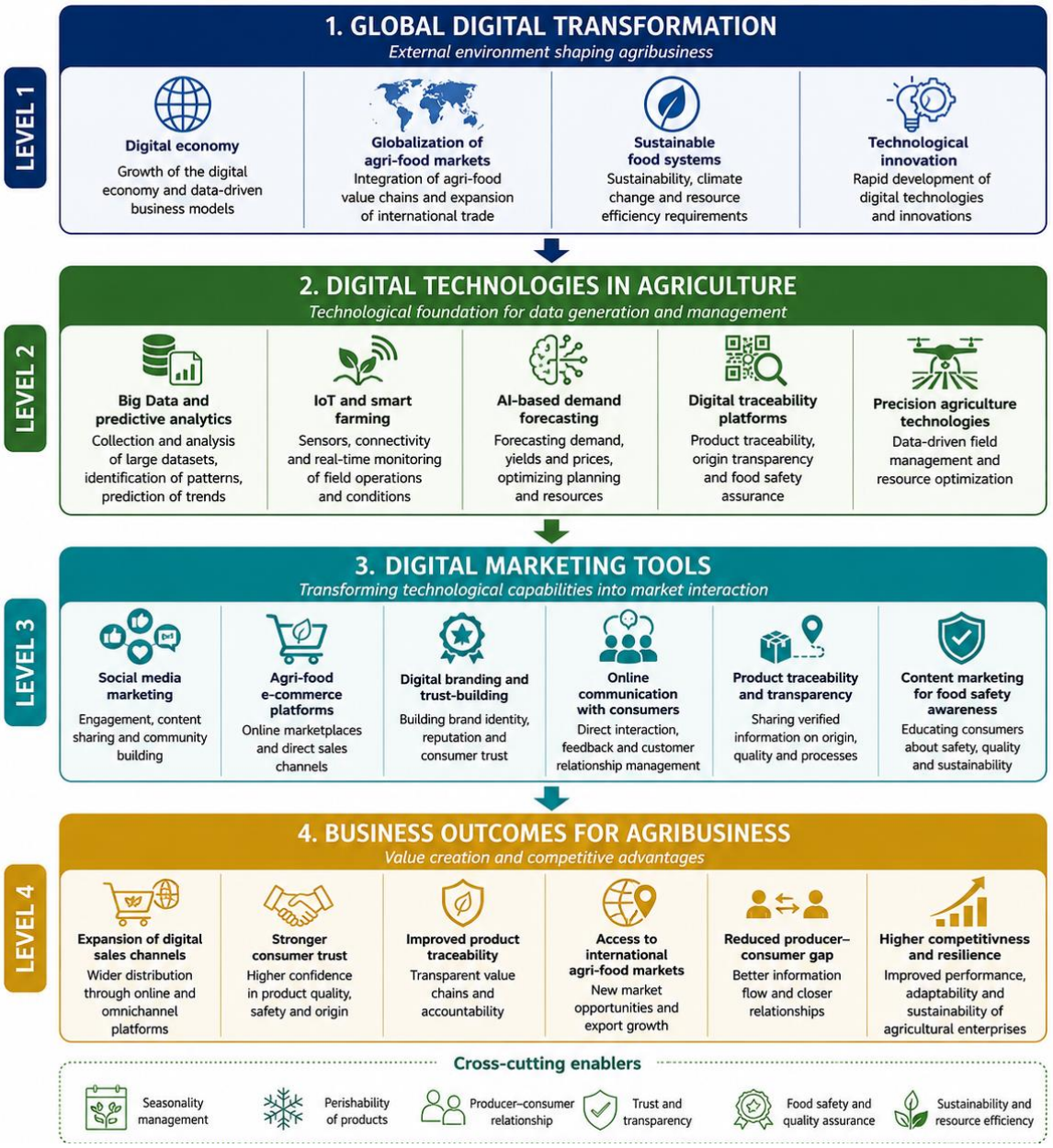


Figure 4: Conceptual framework of digital marketing development in agribusiness.

The second level is represented by digital technologies in agriculture. The use of big data and predictive analytics, the Internet of Things, artificial intelligence, precision agriculture technologies, and digital traceability platforms enables the accumulation, processing, and analysis of information. In the agricultural sector, these technologies are particularly important for managing seasonality, forecasting demand fluctuations, monitoring product quality, and ensuring traceability throughout agri-food supply chains. As a result, agricultural enterprises can improve managerial decision-making, optimize resource use, and increase operational efficiency.

At the third level, digital marketing tools transform technological capabilities into market interaction mechanisms. Social media marketing, agri-food e-commerce platforms, digital branding, online communication channels, and content marketing contribute to the development of direct relationships with consumers and the strengthening of market presence. In addition, digital marketing

tools help agricultural enterprises reduce the communication gap between producers and consumers, increase transparency regarding product origin and food safety, and strengthen consumer trust in agricultural products.

The fourth level reflects the key business outcomes for agribusiness. These include the expansion of digital sales channels, stronger brand recognition, improved product traceability, access to international agri-food markets, reduction of the producer-consumer gap, and increased competitiveness and resilience of agricultural enterprises. The integrated application of digital technologies and digital marketing tools also contributes to improving the adaptability of agri-food systems to market, logistical, and external disruptions.

Thus, digital marketing in agribusiness should be considered not only as a communication tool, but also as a strategic mechanism that integrates technological innovation, market coordination, consumer interaction, and sustainable development within agri-food systems (Table 5).

Table 5: Key components of the conceptual framework of digital marketing in agribusiness. Source: compiled by the authors based on [23], [34], [35].

Component	Key elements	Expected impact
Digital transformation drivers	Digital economy, technological innovation, globalization of agri-food markets	Transformation of the business environment
Digital technologies in agriculture	Big Data, IoT, artificial intelligence, digital platforms	Improved decision-making and operational efficiency
Digital marketing tools	Social media marketing, e-commerce, digital communication channels	Enhanced interaction with consumers
Business outcomes	Brand development, market expansion, competitiveness	Sustainable growth of agribusiness

The developed framework makes it possible to systematically represent the transformation of digital technologies into marketing value, consumer interaction, and competitive advantages within agri-food systems. Its practical significance lies in its potential to serve as a conceptual basis for the development of integrated digital growth strategies for agricultural enterprises, taking into account the specific characteristics of agri-food markets.

To deepen the understanding of interaction processes within the digital environment of agribusiness, an integrated digital marketing ecosystem model is proposed (Fig. 5). Unlike generalized digital ecosystem models, the proposed framework reflects the specific features of agri-food systems, including supply chain integration, food safety requirements, product traceability, sustainability challenges, and the growing importance of trust-based communication between producers and consumers.

The proposed ecosystem reflects a multi-level system of interaction among institutional, technological, production, logistical, and market actors within the agri-food sector.

Government institutions form the regulatory and institutional environment by supporting agricultural digitalization, food safety standards, sustainable development policies, and the expansion of digital infrastructure. Their role is particularly important in ensuring transparency, traceability, and the implementation of digital standards within agri-food markets.

Technology companies and research institutions contribute to the development of innovative digital solutions, including artificial intelligence systems, IoT technologies, predictive analytics, blockchain tools, and digital traceability platforms. These technologies enable agricultural enterprises to optimize production processes, improve demand forecasting, and strengthen data-driven decision-making.

Digital platforms act as an integrating element of the ecosystem by connecting producers, consumers, logistics operators, retailers, and other market participants. E-commerce systems, CRM platforms, social media, and online marketplaces facilitate communication, information exchange, and the coordination of agri-food supply chains.

Agri-food supply chains occupy an important place within the ecosystem because digital marketing in agriculture is closely interconnected with logistics, storage, distribution, and product delivery systems. The integration of digital technologies into supply chains contributes to improving product traceability, reducing transaction costs, and increasing the resilience of agri-food systems.

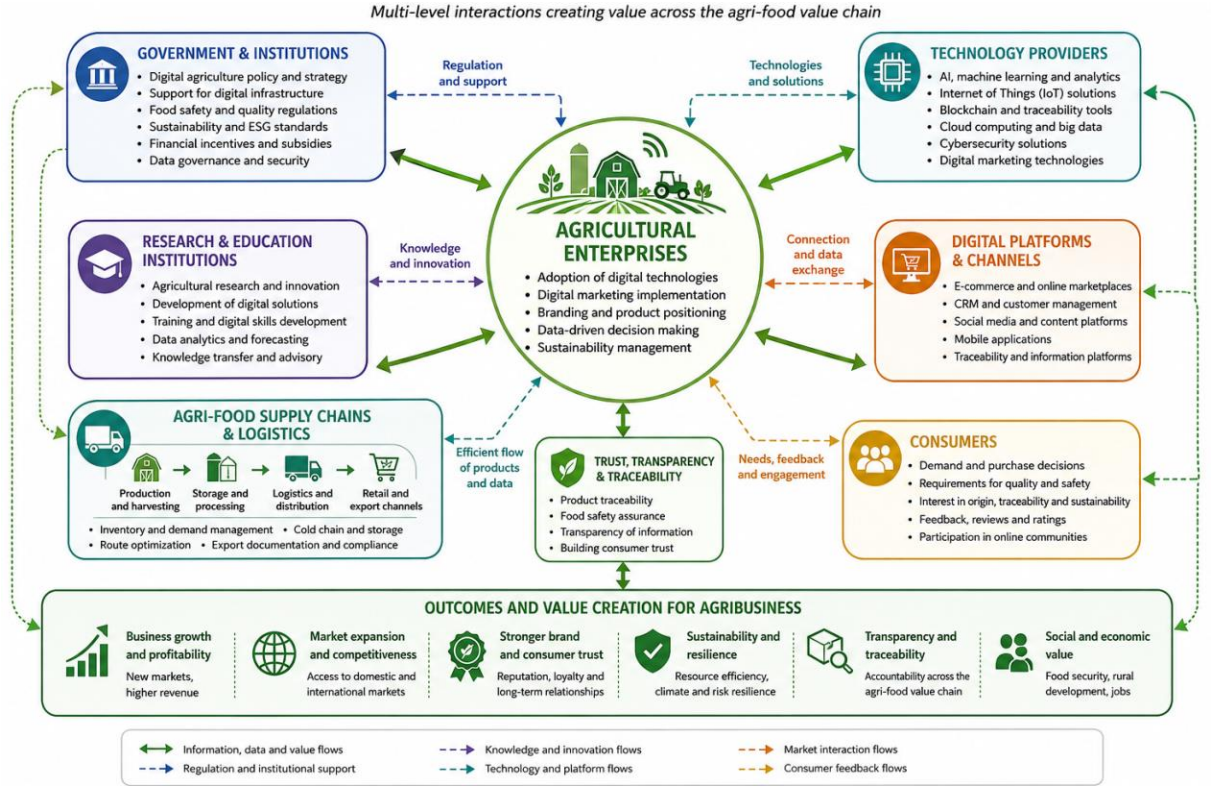


Figure 5: Integrated digital marketing ecosystem in agribusiness. Source: developed by the authors based on [23], [34], [35] using AI tools.

Table 6: Digital marketing tools used in agribusiness. Source: compiled by the authors based on [3]-[4], [7], [8], [10], [11].

Digital marketing tool	Application in agribusiness	Expected outcomes
Social media marketing	Promotion of agricultural products via social media platforms	Increased brand awareness and customer engagement
E-commerce platforms	Online sales through marketplaces and digital platforms	Expansion of sales channels and market access
Content marketing	Creation of digital content about product quality and origin	Increased consumer trust and loyalty
Data-driven marketing	Use of analytics to study consumer behavior	Personalization and higher campaign effectiveness
Digital platforms	Integration of producers, suppliers, and buyers	Improved coordination and market transparency
Mobile marketing	Use of mobile applications and communication tools	Faster interaction and broader access to information

Agricultural enterprises occupy the central position within the ecosystem, as they implement digital technologies and digital marketing tools in production, communication, branding, and sales activities. The use of digital marketing channels enables agricultural producers to establish direct interaction with consumers, reduce the producer-consumer gap, and strengthen trust in agricultural products.

Consumers, in turn, shape market demand and define requirements for product quality, environmental sustainability, transparency of origin, and food safety. In the digital economy, consumers increasingly influence the formation of marketing strategies through feedback mechanisms, digital communication, and participation in online agri-food communities.

Thus, digital marketing in agribusiness operates within a complex ecosystem of technological, institutional, and market interactions, where the strategic importance of communication, transparency, sustainability, and trust becomes increasingly significant for the development of competitive and resilient agri-food systems.

For the practical implementation of the proposed model, it is advisable to consider the key digital marketing tools (Table 6) and their applications.

As shown in Table 6, digital marketing combines communication and analytical tools. Their integrated application enables enterprises not only to promote products, but also to better understand consumer needs and adapt marketing strategies accordingly.

Overall, the implementation of digital marketing tools enhances the flexibility of agricultural enterprises, strengthens their market adaptability, and supports long-term competitiveness.

4.6 Research Gaps and Emerging Trends

Despite the rapid growth of scientific publications on digital marketing in agriculture, the conducted bibliometric analysis reveals several structural imbalances and research gaps within the existing scientific discourse.

First, the contemporary research landscape is largely dominated by technological studies focused on smart farming, artificial intelligence, IoT systems, big data analytics, and digital platforms. While these areas form an important technological basis for the digital transformation of agriculture, considerably less attention is paid to marketing, communication, and behavioural dimensions of digital interaction within agri-food systems.

Second, the analysis demonstrates the insufficient development of studies related to consumer behaviour, trust formation, customer engagement, and digital communication between producers and consumers. At the same time, in the context of increasing consumer requirements regarding food safety, sustainability, product origin, and transparency, these issues are becoming strategically important for agricultural enterprises.

Third, a significant geographical imbalance in scientific research is identified. The majority of publications are concentrated in developed countries, whereas developing economies and small-scale agricultural producers remain underrepresented in the global scientific discourse. This limits the understanding of regional specificities of digital marketing implementation in agribusiness under different institutional and technological conditions.

In addition, the results indicate the insufficient integration of digital marketing research with studies on agri-food supply chains, logistics, and sustainability management. Most existing studies examine digital technologies either from a technological or marketing perspective, while integrated interdisciplinary approaches remain relatively underdeveloped.

At the same time, the bibliometric analysis demonstrates the emergence of new promising research directions associated with digital traceability systems, AI-driven marketing analytics, sustainability-oriented branding, food transparency, and platform-based interaction models within agri-food ecosystems.

Thus, the future development of scientific research in this field requires a transition from fragmented technological approaches toward more integrated interdisciplinary models that combine digital technologies, marketing strategies, supply chain management, sustainability, and consumer interaction within agri-food systems.

5 DISCUSSION, LIMITATIONS AND FUTURE RESEARCH

5.1 Discussion of Results

The results of this study provide a more nuanced understanding of the positioning of digital marketing within contemporary agribusiness research. The bibliometric evidence indicates that the field emerges at the intersection of digital agriculture, information systems, supply chain management, e-commerce, and marketing communications. Rather than forming an isolated discipline, digital marketing operates as an integrative layer linking technological innovation with market interaction, consumer engagement, and value creation in agri-food systems.

These findings align with previous research suggesting that digital transformation in agriculture leads to the emergence of interconnected ecosystems involving producers, technology providers, platform operators, logistics actors, and research institutions [2]. The growing adoption of big data analytics, IoT solutions, artificial intelligence, and digital platforms supports a shift toward more data-driven and adaptive management models in agriculture [4], [6].

At the same time, the results refine earlier perspectives that primarily emphasized technological modernization and smart farming solutions. The current analysis shows a gradual but

clear shift toward marketing-oriented and behavioural dimensions of digital transformation. In this context, digitalization extends beyond production efficiency and increasingly shapes communication strategies, consumer engagement, digital branding, trust-building, and demand creation.

Digital marketing tools such as social media platforms, CRM systems, mobile applications, and e-commerce environments enable agricultural enterprises to develop more direct and personalized relationships with consumers [3], [9]. In parallel, data-driven technologies and artificial intelligence enhance segmentation accuracy, forecasting capabilities, and strategic decision-making processes [10].

However, the analysis also reveals a persistent fragmentation in the literature. Technological, marketing, and behavioural perspectives are still often studied in isolation. Limited attention is given to integrated digital ecosystems that combine communication, logistics coordination, sustainability, traceability, and consumer trust. This suggests a need for more holistic and interdisciplinary research frameworks.

Another important observation concerns geographical imbalance. Scientific output is concentrated mainly in developed economies, while developing regions and small-scale agricultural producers remain underrepresented. This restricts the generalizability of findings and limits understanding of how digital marketing functions under diverse institutional and infrastructural conditions.

Finally, sustainability, transparency, and traceability are becoming increasingly central to agri-food systems. As consumer expectations evolve, digital marketing is no longer limited to promotion; it also plays a strategic role in building trust, ensuring product transparency, and supporting sustainable value chains.

Overall, digital marketing is gradually shifting from a tactical communication instrument to a strategic component of agribusiness transformation.

5.2 Research Limitations

Several limitations should be acknowledged when interpreting the findings.

First, the analysis is based exclusively on the Scopus database. Although Scopus provides extensive coverage and high-quality indexing, relevant studies indexed in Web of Science, Dimensions, or Google Scholar may not be fully represented. This may lead to partial coverage of the global research landscape.

Second, bibliometric methods are inherently limited in terms of interpretative depth. While they are effective for identifying patterns, trends, and structures, they do not allow for detailed assessment of theoretical frameworks, methodological rigor, or practical effectiveness of digital marketing tools.

Third, the study presents aggregated global trends without fully capturing regional differences in digital readiness, institutional environments, and technological infrastructure. These contextual factors are critical for understanding the real-world implementation of digital marketing in agriculture.

Finally, the study does not include empirical validation at the enterprise level. As a result, conclusions remain conceptual and should be interpreted as descriptive rather than explanatory of causal relationships.

5.3 Future Research Agenda

Future research should address several key directions.

First, expanding the dataset beyond a single database would improve the comprehensiveness and robustness of bibliometric findings. Combining multiple scientometric sources would provide a more complete view of the research landscape.

Second, methodological triangulation should be encouraged. Integrating bibliometric analysis with systematic literature reviews, qualitative synthesis, and empirical case studies would allow for deeper insights into both structure and substance of the field.

Third, further studies should focus more strongly on behavioural and relational aspects of digital marketing in agribusiness, particularly consumer trust, engagement, transparency, and food perception in digital environments.

Fourth, stronger integration between digital marketing, supply chain management, logistics, sustainability frameworks, and platform ecosystems is needed. This would support the development of more realistic and applicable models for modern agrifood systems.

Overall, future research should move toward interdisciplinary approaches that combine technological, economic, and behavioural perspectives in a unified analytical framework.

6 CONCLUSIONS

This study provides a structured overview of the development of research on digital marketing in the agricultural sector based on bibliometric analysis. The results confirm that the field is highly

interdisciplinary, emerging at the intersection of digital agriculture, information technologies, supply chain management, e-commerce, and marketing communication.

A key trend identified in the study is the gradual transition from a technology-centred perspective toward a more integrated view that incorporates marketing strategies, consumer interaction, and value creation processes. While technological themes such as IoT, artificial intelligence, and big data remain dominant, increasing attention is being paid to their application in market-oriented and consumer-focused contexts.

At the same time, the study highlights several important gaps. Research on consumer behaviour, trust formation, digital branding, and value co-creation remains relatively underdeveloped compared to technological studies. In addition, geographical concentration of research in developed countries limits the inclusiveness of current scientific discourse.

The contribution of this study lies in systematizing existing knowledge and positioning digital marketing as an integrative mechanism connecting technological innovation with market and consumer dynamics in agribusiness. The proposed conceptual framework and ecosystem model provide a basis for understanding how digital technologies are translated into marketing value and competitive advantage.

From a practical perspective, the findings highlight the importance of integrated digital marketing strategies for agricultural enterprises, including e-commerce systems, CRM tools, data analytics, and omnichannel communication. For policymakers, the results emphasize the need to strengthen digital infrastructure and support digital transformation in rural areas. For researchers, the study highlights the importance of interdisciplinary approaches and empirical validation of theoretical models.

In conclusion, digital marketing in agribusiness should be understood not merely as a promotional function, but as a strategic mechanism integrating technology, communication, supply chains, and consumer trust within modern agri-food systems.

ACKNOWLEDGMENTS

We acknowledge support by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) and the Open Access Publishing Fund of Anhalt University of Applied Sciences.

REFERENCES

[1] Food and Agriculture Organization of the United Nations, The State of Food and Agriculture 2023. Revealing the True Cost of Food to Transform Agrifood Systems, Rome, 2023, [Online]. Available: https://doi.org/10.4060/cc7724en.

[2] L. Klerkx, E. Jakku, and P. Labarthe, “A review of social science on digital agriculture, smart farming and agriculture 4.0: New contributions and a future research agenda”, NJAS: Wageningen Journal of Life Sciences, vol. 90-91(1), 100315, pp. 1-16, 2019, [Online]. Available: https://doi.org/10.1016/j.njas.2019.100315.

[3] Y. K. Dwivedi, E. Ismagilova, D. L. Hughes et al., “Setting the future of digital and social media marketing research: Perspectives and research propositions”, International Journal of Information Management, vol. 59, 102168, 2021, [Online]. Available: https://doi.org/10.1016/j.ij.infomgt.2020.102168.

[4] S. Wolfert, L. Ge, C. Verdouw, and M. J. Bogaardt, “Big data in Smart Farming - a review”, Agricultural Systems, vol. 153, pp. 69-80, 2017, [Online]. Available: https://doi.org/10.1016/j.agsy.2017.01.023.

[5] K. Fuglie, M. Gautam, A. Goyal, and W. F. Maloney, Harvesting Prosperity: Technology and Productivity Growth in Agriculture, World Bank, Washington, DC, 2020, [Online]. Available: https://doi.org/10.1596/978-1-4648-1393-1.

[6] C. Verdouw, S. Wolfert, and B. Tekinerdogan, “Internet of Things in agriculture,” CABI Reviews, vol. 11(35), pp. 1-12, 2016, [Online]. Available: https://doi.org/10.1079/PAVSNNR201611035.

[7] A. Kamilaris, A. Kartakoullis, and F. X. Prenafeta-Boldú, “A review on the practice of big data analysis in agriculture”, Computers and Electronics in Agriculture, vol. 143, pp. 23-37, 2017, [Online]. Available: https://doi.org/10.1016/j.compag.2017.09.037.

[8] A. Kumar, S. Mangla, P. Kumar, and M. Song, “Mitigate risks in perishable food supply chains: Learning from COVID-19”, Technological Forecasting and Social Change, vol. 166, Art. 120643, 2021, [Online]. Available: https://doi.org/10.1016/j.techfore.2021.120643.

[9] J. R. Saura, P. Palos-Sánchez, and M. Correia, “Digital marketing strategies based on the e-business model: Literature review and future directions”, in Organizational Transformation and Managing Innovation in the Fourth Industrial Revolution, IGI Global, Hershey, Pennsylvania, 2019, [Online]. Available: https://doi.org/10.4018/978-1-5225-7074-5.ch005.

[10] T. H. Davenport, A. Guha, D. Grewal, and T. Bressgott, “How artificial intelligence will change the future of marketing”, Journal of the Academy of Marketing Science, vol. 48(7553), pp. 1-19, 2020, [Online]. Available: https://doi.org/10.1007/s11747-019-00696-0.

[11] P. C. Verhoef, T. Broekhuizen, Y. Bart et al., “Digital transformation: A multidisciplinary reflection and research agenda”, Journal of Business Research, vol. 122, pp. 889-901, 2021, [Online]. Available: https://doi.org/10.1016/j.jbusres.2019.09.022.

[12] M. Hingley and A. Lindgreen, "Marketing of agricultural products: Case findings", British Food Journal, vol. 104(10), pp. 806-827, 2002, [Online]. Available: https://doi.org/10.1108/00070700210448908.

[13] X. Y. Leung, B. Bai, and K. A. Stahura, "The marketing effectiveness of social media in the hotel industry: a comparison of Facebook and Twitter", Journal of Hospitality & Tourism Research, vol. 39(2), pp. 147-169, 2015, [Online]. Available: https://doi.org/10.1177/1096348012471381.

[14] E. Pantano, G. Pizzi, D. Scarpi, and C. Dennis, "Competing during a pandemic? Retailers' ups and downs during the COVID-19 outbreak", Journal of Business Research, vol. 116, pp. 209-213, 2020, [Online]. Available: https://doi.org/10.1016/j.jbusres.2020.05.036.

[15] M. Ben-Daya, E. Hassini, and Z. Bahroun, "Internet of things and supply chain management: A literature review", International Journal of Production Research, vol. 57(15-16), pp. 4719-4742, 2017, [Online]. Available: https://doi.org/10.1080/00207543.2017.1402140.

[16] E. D. Lioutas and C. Charatsari, "Smart farming and short food supply chains: Are they compatible?", Land Use Policy, vol. 94, Art. 104541, 2020, [Online]. Available: https://doi.org/10.1016/j.landusepol.2020.104541.

[17] C. Eastwood, M. Ayre, R. Nettle, and B. D. Rue, "Making sense in the cloud: Farm advisory services in a smart farming future", NJAS: Wageningen Journal of Life Sciences, vol. 90-91(1), 100298, 2019, [Online]. Available: https://doi.org/10.1016/j.njas.2019.04.004.

[18] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines", Journal of Business Research, vol. 133, pp. 285-296, 2021, [Online]. Available: https://doi.org/10.1016/j.jbusres.2021.04.070.

[19] M. Aria and C. Cuccurullo, "Bibliometrix: An R-tool for comprehensive science mapping analysis", Journal of Informetrics, vol. 11, iss. 4, pp. 959-975, 2017, [Online]. Available: https://doi.org/10.1016/j.joi.2017.08.007.

[20] I. Zupic and T. Čater, "Bibliometric methods in management and organization", Organizational Research Methods, vol. 18, iss. 3, pp. 429-472, 2015, [Online]. Available: https://doi.org/10.1177/1094428114562629.

[21] M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma, and F. Herrera, "Science mapping software tools: Review, analysis, and cooperative study among tools", Journal of the American Society for Information Science and Technology, vol. 67, iss. 7, pp. 1382-1402, 2011, [Online]. Available: https://doi.org/10.1002/asi.21525.

[22] N. J. van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping", Scientometrics, vol. 84, pp. 523-538, 2010, [Online]. Available: https://doi.org/10.1007/s11192-009-0146-3.

[23] A. Kwilinski, N. Trushkina, Yu. Remyha, and T. Patlachuk, "Impact of artificial intelligence on customer relationship management: A bibliometric analysis", Proceedings of International Conference on Applied Innovation in IT, vol. 13(1), pp. 295-303, [Online]. Available: http://dx.doi.org/10.25673/119246.

[24] D. J. Choruma, T. L. Dirwai, M. J. Mutenje et al., "Digitalisation in agriculture: A scoping review of technologies in practice, challenges, and opportunities for smallholder farmers in sub-saharan africa", Journal of Agriculture and Food Research, vol. 18, 101286, 2024, [Online]. Available: https://doi.org/10.1016/j.jafr.2024.101286.

[25] M. Alkahtani, Q. S. Khalid, M. Jalees et al., "E-agricultural supply chain management coupled with blockchain effect and cooperative strategies", Sustainability, vol. 13(2), 816, 2021, [Online]. Available: https://doi.org/10.3390/su13020816.

[26] L. A. Duram, "Agents' perceptions of structure: How Illinois organic farmers view political, economic, social, and ecological factors", Agriculture and Human Values, vol. 17(1), pp. 35-48, 2000, [Online]. Available: https://doi.org/10.1023/A:1007632810301.

[27] A. U. Din, H. Han, A. Ariza-Montes et al., "The Impact of COVID-19 on the Food Supply Chain and the Role of E-Commerce for Food Purchasing", Sustainability, vol. 14(5), 3074, 2022, [Online]. Available: https://doi.org/10.3390/su14053074.

[28] M. Prosper Bright, N. Terrence Kudzai, and C. Ngavaite, "The impact of COVID-19 on agricultural extension and food supply in Zimbabwe", Cogent Food and Agriculture, vol. 7(1), 1918428, 2021, [Online]. Available: https://doi.org/10.1080/23311932.2021.1918428.

[29] A. Elghannam, M. Escribano, and F. Mesias, "Can social networks contribute to the development of short supply chains in the Spanish agri-food sector?", New Medit, vol. 16(1), pp. 36-42, 2017, [Online]. Available: https://www.scopus.com/pages/publications/85028706185?origin=resultslist.

[30] K. A. Nagaty, "IoT Commercial and Industrial Applications and AI-Powered IoT", in Frontiers of Quality Electronic Design (QED): AI, IoT and Hardware Security, Cham, Springer International Publishing, pp. 465-500, 2023, [Online]. Available: https://doi.org/10.1007/978-3-031-16344-9_12.

[31] I. Grigorescu, E.-A. Popovici, N. Damian et al., "The resilience of sub-urban small farming in Bucharest Metropolitan Area in response to the COVID-19 pandemic", Land Use Policy, vol. 122, 106351, 2022, [Online]. Available: https://doi.org/10.1016/j.landusepol.2022.106351.

[32] S. Ravi and S. R. C. Rajasekaran, "A perspective of digital marketing in rural areas: A literature review", International Journal of Professional Business Review, vol. 8(4), e01388, 2023, [Online]. Available: https://doi.org/10.26668/businessreview/2023.v8i4.1388.

[33] K. M. I. Bashir, J.-S. Kim, M. Mohibbullah, J. H. Sohn, and J.-S. Choi, "Strategies for improving the competitiveness of Korean seafood companies in the overseas halal food market", Journal of Islamic Marketing, vol. 10(2), pp. 606-632, 2019, [Online]. Available: https://doi.org/10.1108/JIMA-03-2018-0056.

[34] A. Kwilinski, L. Hnatyshyn, O. Prokopyshyn, and N. Trushkina, "Managing the logistic activities of agricultural enterprises under conditions of digital economy", Virtual Economics, vol. 5(2), pp. 43-70, [Online]. Available: https://doi.org/10.34021/ve.2022.05.02(3).

[35] R. Relawati, B. Y. Ariadi, and N. Trushkina, "Digital marketing strategies for agriculture: Indonesia and Ukraine", BIO Web of Conferences, vol. 201, 03003, 2025, [Online]. Available: https://doi.org/10.1051/bioconf/202520103003.

Consequences of the Functioning of the New Computerised Transit System (NCTS) for Customs Logistics

Viktoriia Rudenko[1], Olena Moskvichova[2], Halyna Pohrishchuk[1] and Nadiia Holovai[1]
[1]*Vinnytsia Educational and Research Institute of Economics, West Ukrainian National University, Honty Str. 37, 21017 Vinnytsia, Ukraine*
[2]*Department of Marketing and International Trade, National University of Life and Environmental Sciences of Ukraine, Heroiv Oborony Str. 15, 03041 Kyiv, Ukraine*
v.rudenko@wunu.edu.ua, devenezia@nubip.edu.ua, g.pohrishchuk@wunu.edu.ua, n.golovay@wunu.edu.ua

Keywords: New Computerised Transit System (NCTS), Customs Administration, Customs Transit, Customs Logistics, Customs Authorities, Foreign Economic Activity Entities, Customs Control, International Trade.

Abstract: The growth in international trade volumes and the digitalization of customs procedures make the study of the of NCTS's impact on the development of customs logistics in the context of common transit relevant. The purpose of the article is to study the functioning of NCTS and its impact on customs logistics, taking into account information-technological, operational-procedural, economic, organizational-institutional, and international aspects. The study uses methods of analysis and synthesis, systemic, structural-functional, comparative approaches, and the generalization method. The architecture and principles of the functioning of NCTS are revealed, its advantages and challenges for customs authorities and entities of foreign economic activity are systematized. The dynamics of the implementation of NCTS in Ukraine in 2022-2025 were analyzed, which showed a significant increase in the volume of transit declarations and increased integration into the European customs space. The main consequences of the functioning of NCTS for customs logistics were determined, in particular, increasing the transparency of transit flows, accelerating customs procedures, reducing transaction costs, and strengthening coordination between customs administrations. At the same time, challenges associated with the system's technical complexity, the need to develop digital infrastructure, cybersecurity and digital competencies of users were outlined. A conceptual model of the impact of NCTS on customs logistics was proposed. The theoretical significance of the results lies in the development of scientific approaches to the study of digital transit systems, and the practical significance lies in the possibility of using the results to improve customs policy and adapt the national customs system to common transit standards.

1 INTRODUCTION

In the current context of globalization of international trade and integration of national economies, digital technologies increasingly have a decisive impact on the efficiency of customs procedures and logistics chains. One of the key initiatives of digital transformation in the field of customs administration is the New Computerised Transit System (NCTS), which was developed within the framework of the Convention on a Common Transit Procedure and implemented in the territory of the European Union and other contracting parties. NCTS is an example of an information and communication system that provides electronic processing of transit procedures, integration of customs services, coordination of data on the movement of goods and the application of a single guarantee throughout the entire transit route. NCTS is designed to increase the transparency of customs control, reduce administrative barriers, and contribute to the optimization of logistics processes in international trade.

The effectiveness of customs logistics, as a component of transport and logistics systems, directly depends on the efficiency of customs clearance of goods, the consistency of customs procedures between states, and the ability to adapt logistics models to the requirements of the modern digital economy. It is in this context that it is important to assess the consequences of the implementation of NCTS, since the system forms not only a technological platform, but also affects the organizational and procedural aspects of transit transportation. Analysis of the effects of NCTS allows us to understand how electronic transit declarations, a single guarantee and unified data

standards transform the interaction between customs authorities, foreign economic activity entities, and other participants in logistics chains.

Existing digital transit systems contribute to increasing the throughput capacity of customs checkpoints, reducing the time of customs clearance procedures, and reducing operating costs, which are critically important factors in increasing the competitiveness of national transport and logistics systems in international trade. At the same time, the experience of NCTS participating countries indicates the existence of certain challenges related to the requirements for technical integration, the level of digital competence of personnel and the adaptation of business processes of foreign economic activity entities to new standards of electronic interaction. Thus, a comprehensive assessment of the consequences of the functioning of NCTS from the standpoint of customs logistics is an urgent task of scientific and practical nature.

2 LITERATURE REVIEW

The issue of digitalization of customs procedures and implementation of electronic transit systems, in particular NCTS, is widely presented in modern scientific research in the field of customs and customs logistics. The works emphasize that information technologies provide a transition from paper procedures to electronic data exchange, increasing the transparency, speed, and efficiency of customs control, as well as contributing to the reduction of administrative costs and optimization of transit flow management [1], [2].

A significant part of the literature is devoted to the institutional and managerial aspects of the functioning of computerized transit systems. Documents of international organizations and scientific studies emphasize the system-forming role of NCTS in ensuring the continuity of international supply chains, strengthening customs security, and forming a new logic of customs logistics based on risk management and real-time information exchange [3], [4].

Some studies have focused on the impact of NCTS on transit logistics parameters. It has been proven that the automation of procedures and the standardization of electronic messages contribute to the reduction of border crossing times, the reduction of transactional and indirect logistics costs, and the increase of the competitiveness of international transport corridors [5], [6], [7].

The issues of international coordination and harmonization of transit regimes occupy a significant place in the literature. The authors emphasize that the functioning of NCTS is based on close information interaction between customs administrations of different countries, which contributes to the unification of procedures and increased predictability of transit operations [8], [9], [10]. At the same time, the importance of adapting national customs systems and legislation to the requirements of the Convention on the Common Transit Procedure and EU standards is emphasized [11]. Comparative studies of transit systems consider NCTS as a reference model of electronic transit, which can be adapted in other countries, provided that there is proper institutional, legal, and technical preparation [12].

In general, the analysis of scientific sources indicates the complex impact of NCTS on customs logistics and forms a theoretical and methodological basis for studying its consequences in the context of increasing the efficiency of customs transit and integrating national customs systems into international supply chains.

3 METHODOLOGY

The methodological basis of the study is based on systemic, institutional, and process-oriented approaches to analysing the impact of NCTS on customs logistics. The research relies on a combination of qualitative and quantitative analytical methods, which enabled a comprehensive assessment of the institutional, information-technological, operational, and logistical aspects of the functioning of the common transit system. The logic of the study involved a sequential examination of the principles underlying the operation of NCTS, an analysis of its implementation in Ukraine, an evaluation of its implications for customs logistics, and the development of a conceptual model of the system's impact on international transit processes.

The information base of the study consists of regulatory and legal documents of the European Union, the World Customs Organization, official materials of the State Customs Service of Ukraine, as well as scholarly publications indexed in Scopus and other peer-reviewed academic databases. The analysis incorporated sources addressing the digitalisation of customs procedures, the functioning of NCTS, customs logistics, transit management, and risk management in international trade.

To achieve the stated objective, the study employed methods of theoretical generalisation, structural-functional, comparative, and statistical analysis. The structural-functional method was applied to examine the architecture of NCTS and the mechanisms of information interaction between customs administrations. Comparative analysis made it possible to identify the advantages and

disadvantages of the system's functioning for customs authorities and foreign economic operators. Statistical analysis was used to assess the dynamics of transit declarations processed under the common transit procedure in Ukraine during 2022–2025 and to identify structural changes between customs offices of departure and destination.

The empirical basis of the research was formed by official statistical data of the State Customs Service of Ukraine regarding the number of transit declarations processed within the framework of NCTS during 2022-2025. The analysis covered indicators related to declarations processed by customs offices of departure, customs offices of destination, and the overall volume of transit operations. This made it possible to assess the intensity of Ukraine's integration into the European transit area, identify trends in the development of digital transit, and reveal institutional features of the adaptation of the customs system to functioning under the common transit procedure.

To generalise the research findings, the method of conceptual modelling was employed, on the basis of which a model of the impact of NCTS on customs logistics was developed. The model was constructed according to the principle of a cause-and-effect relationship between the institutional preconditions for the functioning of the system, the digital mechanisms of NCTS, the transformation of customs procedures, and the strategic implications for international customs logistics. This approach enabled the systematisation of the key effects of transit procedure digitalisation and the identification of the main factors influencing the effectiveness of NCTS functioning.

4.1 Content and Principles, Advantages and Disadvantages of the Functioning of NCTS

The New Computerised Transit System (NCTS) is a European computer system designed to implement the common customs transit procedure within the European Union, the European Free Trade Association countries, and other contracting countries, including Ukraine, Turkey, and Serbia. The system operates on the basis of the Convention on the Common Transit Procedure, ensuring the movement of goods between the customs territories of the contracting parties without changing their customs status until the moment the cargo is presented at the customs office of destination. The basis of its operation is the electronic submission and processing of transit declarations, which replaces paper documents and a number of formalities inherent in traditional transit procedures. The system covers all EU Member States, as well as countries that are parties to the Convention on the Common Transit Procedure, providing a single platform for international cooperation of customs authorities.

Architecturally, NCTS is implemented through national components that are integrated through a central domain in Brussels, which provides the connection of all European customs services and the exchange of electronic messages in real time. These messages contain data on cargoes, senders and recipients, means of transport, routes, and guarantees accompanying the movement of goods. This approach contributes to increased transparency of transit flows, optimization of logistics routes, and rapid assessment of customs risks.

The functioning of NCTS is based on the principles of digitalization of customs procedures, unification of information exchange, guaranteeing customs obligations, decentralization with central coordination, and monitoring and control in real time. The information and technical architecture of the system ensures the exchange of transit data between customs authorities, foreign economic activity operators, and central services in real time, which contributes to increasing the controllability of transit traffic and reducing administrative barriers.

The first fundamental principle of the NCTS is the digitalization of customs procedures, which involves the electronic submission and processing of transit declarations at all stages of goods transportation. The transit declaration submitted through the NCTS national interface contains a complete set of data on the goods, their sender and recipient, means of transport, route, as well as warranty obligations accompanying the movement of goods [5]. As a result of the acceptance of the transit declaration, the system generates a unique identification number, which is used to monitor the movement of cargo in the customs transit regime and further risk analysis within the framework of the accompanying security and customs control procedures.

The second key principle of the NCTS is the unification of information exchange, which consists in the integrated exchange of information between the customs authorities of the contracting parties, which is implemented through the central communication network and the national components of the NCTS. This involves not only the exchange of data on transit declarations but also the transmission of customs messages on the physical movement of goods, changes in transit status, and possible exceptions or incidents during transportation [4]. Thanks to a single information environment, administrations are able to coordinate customs operations at the international level, ensuring transparency, efficiency, and mutual trust between participating states.

Another important principle of the NCTS is the guarantee of customs obligations, which involves the provision of valid guarantees to secure possible customs payments in the event of a violation of transit rules. This guarantee can be both individual and general, including simplification in the form of a guarantee waiver for authorized operators, which is achieved on the basis of an assessment of their compliance with risk and financial reliability criteria [2]. This principle helps to balance the fiscal interests of states and the economic interests of businesses, reducing the costs of customs procedures and stimulating efficient activity in international trade.

The fourth principle of the NCTS operation is decentralization with central coordination, as it is architecturally implemented as a distributed computerized system with national portals for operations and a single central communication infrastructure, which allows integrating of local customs information systems and using of open data exchange protocols. [13]. Such a structure promotes scalability, allows national components to be adapted to common standards, and provides the ability to simultaneously process a large number of transactions in accordance with the needs of international trade.

The fifth principle of the NCTS operation is real-time monitoring and control, which consists in tracking the movement of goods within the framework of a single transit procedure. Customs authorities pay attention to ensuring the authenticity of data, the correctness of the route, and compliance of transit operations with established rules, which significantly reduces the risks of fraud, smuggling, and violation of customs obligations.

The implementation of NCTS as a single computerized transit system between the countries party to the Convention on a common transit procedure presents customs administrations and foreign economic activity entities with a new level of information integration, which has both significant advantages and certain challenges. The advantages of the operation of NCTS for both groups of users stem from the principles of electronic exchange, standardized transit control, and simplified procedures that were incorporated into the system during its design at the level of the European Union and the Contracting Parties to the Convention. On the other hand, the disadvantages reflect the complexity of technical integration, the need for significant investments in technology and user skills, as well as the dependence on the stability of the IT infrastructure (Table 1) [6], [8], [11], [14].

Thus, NCTS offers significant advantages, including optimization of procedures, reduction of transaction costs, reduction of customs clearance time, and increased security of transit operations for both customs authorities and foreign economic activity entities. However, the challenges associated with the complexity of implementation, the need for training, technical support, and adaptation to changes require a systematic approach to managing these processes and further improving the relevant IT and legal infrastructure.

Table 1: Advantages and disadvantages of NCTS functioning.

Participant	Advantages	Disadvantages
Customs authorities	▪ Ensuring electronic data exchange between customs authorities with prompt access to information on transit operations. ▪ Reduction of customs transaction time thanks to a single transit declaration and guarantee for the entire NCTS territory.	▪ The need for significant investment in the IT infrastructure and staff training, which can be challenging for customs services with limited budgets or experience with digitalization. ▪ Dependence on the availability and stability of the IT system: technical failures can lead to delays and the need to apply backup procedures.
Foreign economic activity entities	▪ Acceleration of customs procedures and reduction of transaction costs through electronic processing of transit declarations and interaction with customs authorities in real time. ▪ Reduction of administrative burden and costs of submitting paper documents, which is especially important for small and medium-sized businesses. ▪ Increased business confidence in electronic solutions, as evidenced by the growing use of NCTS and the expansion of transit operations volumes.	▪ The need to adapt business processes and the IT systems to the requirements of the new NCTS phases, which may require additional efforts and costs. ▪ Insufficient digital competence of individual enterprises may limit the effective use of the system in the early stages of implementation.

4.2 Analysis of the Implementation of NCTS in Ukraine

Ukraine officially joined NCTS on October 1, 2022, after the successful completion of the international assessment of the national IT system and the implementation of a set of regulatory, technical, and organizational requirements specified by the Convention on the Common Transit Procedure. A prerequisite for accession was Ukraine's acquisition of the status of a country-participant of the Convention on the Common Transit Procedure [15] and the International Convention on the Simplification and Harmonization of Customs Procedures [16], which ensured the integration of the national customs system into the European transit space.

The implementation of transit operations within the framework of NCTS in Ukraine for the period of the fourth quarter of 2022-2025 indicates a deep institutional transformation of the customs and logistics system, which is manifested simultaneously in an explosive growth in the volume of transit declarations and a gradual structural restructuring between the customs offices of departure and destination (Fig. 1) [17].

The total number of processed transit declarations increased from 0.7 thousand in 2022 to 142.0 thousand in 2025, i.e., more than 200 times, while the dynamics are phased: a sharp expansionary jump in 2023 (41.2 thousand), further significant growth in 2024 (94.3 thousand), and relatively more moderate, but stabilized growth in 2025 (142.0 thousand). Such a trajectory is typical for the implementation of complex digital infrastructures, when the initial institutional adaptation and mass connection of market participants are replaced by the stage of operational saturation of the system.

A structural analysis of the distribution of transit operations between customs offices of departure and destination ones demonstrates the dominance of the first category throughout the period, but with a clearly expressed tendency towards a gradual equalization of the functional load. The volumes of clearance at customs offices of departure increased from 0.4 thousand to 96.5 thousand, while at customs offices of destination − from 0.3 thousand to 45.5 thousand. Accordingly, the share of operations at the destination stage increased from approximately 18.4% in 2023 to 32.0% in 2025, which indicates an increase in the role of the final stage of transit and a gradual increase in the institutional maturity of the system for controlling the movement of goods. At the same time, the preservation of a significant gap between the two links indicates incomplete harmonization of the transit cycle and potential asymmetry in administrative procedures.

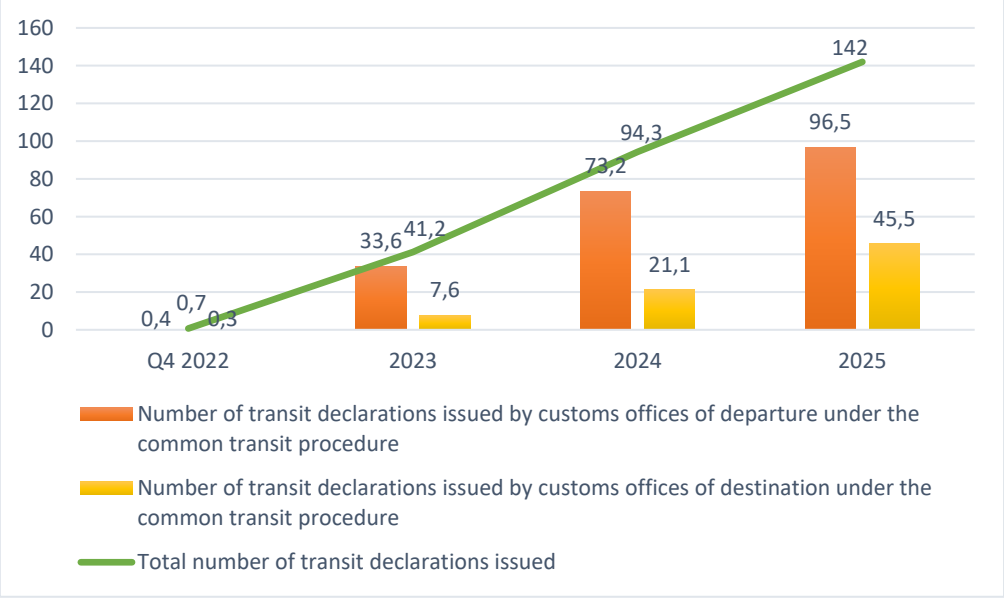


Figure 1: Dynamics of processed customs declarations within the framework of the Convention on the Common Transit Procedure in Ukraine in 2022-2025, thousands.

The identified trends form a double effect of the implementation of NCTS in Ukraine. On the one hand, the system provides a significant expansion of transit activity, which strengthens the country's positioning as an important transit corridor in European logistics chains. At the same time, the digitalization of procedures reduces transaction costs, speeds up customs clearance, increases the predictability of logistics operations, and minimizes the space for administrative discretion, which together contribute to increased transparency and reduced corruption risks. In addition, the gradual increase in the share of destination customs offices reflects the institutional strengthening of the system and increased efficiency of the final stage of customs control.

On the other hand, the empirical dynamics reveal a number of systemic limitations. The extremely rapid growth in 2023 placed a significant strain on the customs infrastructure, potentially leading to the uneven adaptation of regional customs offices and the formation of operational bottlenecks. In addition, the dominance of customs offices of departure over customs offices of destination throughout the period indicates incomplete unification of transit completion procedures and residual institutional asymmetry between the stages of the operational cycle. An additional challenge is the high dependence on the digital infrastructure of the system, which increases the importance of cybersecurity, stability of IT solutions, and continuous staff training. Finally, the volatility of growth rates in 2023–2024 reflects the adaptive nature of the system's development and complicates long-term forecasting of transit flows.

In conclusion, the functioning of NCTS in Ukraine is characterized by both a structurally positive dynamics of integration into the European customs space, and the presence of institutional and technological limitations. Its further effectiveness will be determined by the ability to ensure a balance between the scaling of transit operations, the alignment of the internal structure of customs administration, and the strengthening of the digital and organizational capacity of the customs system.

4.3 Consequences of the Functioning of NCTS

Based on Ukraine's experience in implementing NCTS, it can be argued that this system creates a new quality in the functioning of customs logistics, forming several groups of consequences that cover organizational, informational, operational, and strategic aspects of managing transit cargo flows (Table 2) [1], [3], [4], [7], [8].

Taken together, the consequences determine the efficiency of logistics chains in international trade and affect the interaction of customs authorities and foreign economic activity entities in the context of common transit. We propose to include the following consequences of the functioning of NCTS:

1) Informational and technological consequences. NCTS is an interconnected IT platform that provides real-time exchange of customs data between customs authorities of the countries participating in the Convention on a Common Transit Procedure. Such integration contributes to the unification of electronic transit declarations, automatic data processing, and reduction of manual operations, which significantly increases the speed and reliability of document processing. This eliminates information gaps between different customs systems and creates transparent conditions for monitoring transit operations, which is critically important for customs logistics, since data on cargo movements can be used to forecast flows and optimize routes [1]. In addition, new phases of NCTS (for example, Phase 5 and the planned Phase 6) provide for even greater integration with other customs services, including security and safety mechanisms, allowing customs authorities to carry out risk-based control without disrupting the logistics flow.
2) Operational and procedural consequences. The transition to NCTS means using a single transit declaration and a single guarantee for the entire route of cargo movement from the customs office of departure to the customs office of destination, eliminating the need to complete separate documents in each country at each transit point. This approach helps to reduce the time for customs procedures, reduce queues at checkpoints, and increase the throughput capacity of logistics corridors [3]. From the point of view of customs logistics, this means reducing administrative delays and increasing the efficiency of logistics processes, since the weight and quantitative parameters of cargo movement are recorded once and accepted by all participating customs offices. In addition, the system simplifies control over compliance with customs requirements, in particular due to a single data standard and guarantee management mechanisms, which reduces the likelihood of document fraud and counterfeiting. The electronic format minimizes the risk of human error, which has a positive effect on the quality of logistics procedures and the level of service.

Table 2: Consequences of the functioning of NCTS for customs logistics.

Group of consequences	Characteristics	Impact on customs logistics
Informational and technological	Ensuring the exchange of electronic transit declarations and cargo data in real time; integration of customs databases	Increases transparency of logistics flows; reduces errors associated with manual data entry; facilitates rapid decision-making regarding routes and risks
Operational and procedural	Use of a single declaration and guarantee for the entire transit route; automation of document control	Reduces customs clearance time; increases the throughput capacity of checkpoints; reduces administrative delays
Economic	Reducing transaction costs and optimizing logistics resources; the ability to forecast cargo flows	Reduces logistics costs for carriers and foreign economic activity entities; increases the competitiveness of international transportation
Organizational and institutional	Coordination of actions of customs services of participating countries; standardization of procedures and guarantees	Increases the coherence of customs authorities' actions; facilitates the integration of foreign economic activity entities into international logistics processes
International	Improving the security of transit chains; risk analytics; forecasting changes in logistics flows	Increases the resilience of logistics systems; allows routes and resources to be adapted to fluctuations in the international trade; facilitates strategic planning

3) Economic consequences. The economic effect of the implementation of NCTS is manifested in the reduction of logistics costs and increased competitiveness of international transportation. In particular, over the years of application of the NCTS system, Ukraine has significantly increased the number of transit declarations, which indicates its real impact on logistics activity and the reduction of barriers to cross-border transportation [8]. Integrated transit management allows operators to reduce costs associated with duplication of customs procedures, waiting at borders, and administrative red tape, which together contribute to the optimization of logistics chains and increased efficiency of international trade.
4) Organizational and institutional consequences. NCTS affects the structure of interaction between customs administrations and foreign economic activity entities, stimulating the development of new forms of cooperation and standardization of customs procedures. The unification of customs authorities in a common information space allows coordinating logistics processes within the framework of unified standards of the Convention on the Common Transit Procedure, which significantly enhances the consistency of actions of customs authorities and reduces uncertainty in customs and logistics operations [4]. It also launches mechanisms for unifying approaches to ensuring transit, authorizing operators, and optimizing resources at border crossing points. Such changes contribute to the establishment of unified customs logistics practices at the international level, which significantly expand the opportunities for the development of transit corridors and the integration of national customs services into global logistics networks.
5) International consequences. At the international level, the functioning of NCTS contributes to the strengthening of security, predictability, and efficiency of transit chains, which is important for the global logistics strategies of states and enterprises. Enhanced digital interaction allows customs services to carry out sophisticated risk analytics, predict changes in flows, and adapt logistics solutions in accordance with fluctuations in international markets and transport corridors [7]. This is a significant factor in building resilient logistics systems that can withstand external shocks, including geopolitical risks or changes in global trade.

Analysis of the consequences of the functioning of NCTS from the perspective of customs logistics shows that the digitalization of transit procedures creates a comprehensive positive effect on the management of goods flows, information support for customs operations, reduction of logistical barriers and integration of customs administrations. At the same time, these changes transform logistics practices aimed at increasing the efficiency, security, and integration of the international logistics chains.

4.4 Conceptual Model of the Impact of NCTS on Customs Logistics

The practice of implementing NCTS in Ukraine shows that the system generates both positive effects − acceleration of customs clearance, reduction of transaction costs, and increased transparency of transit − as well as new institutional and technological challenges related to the adaptation of IT infrastructure, the need for digital competencies, ensuring cybersecurity, and harmonization of procedures between various participants in the transit process. This indicates the need to form a conceptual model that allows for a comprehensive reflection of the mechanisms of the impact of NCTS on customs logistics and systematize the relationships between its structural elements (Fig. 2).

The proposed conceptual model of the impact of NCTS on customs logistics reflects the cause-and-effect relationship between the institutional prerequisites for the functioning of the system, its digital and organizational mechanisms, operational transformations of customs procedures, and the final results for international customs logistics. The model is built according to the logic of a sequential transition from input factors to integral logistics effects and allows for a comprehensive characterization of the impact of NCTS on the functioning of transit flows in the international environment.

The first block of the model forms the basic environment for the functioning of the NCTS. It includes institutional and legal, technological, and organizational prerequisites. The institutional component covers the Convention on the Common Transit Procedure, the harmonization of customs procedures, and the participation of the EU, EFTA, and other contracting parties in the single transit space. It is regulatory unification that provides the legal basis for the functioning of the common transit system and mutual recognition of customs procedures between states. Technological factors include the development of IT infrastructure, data standardization, ensuring interoperability and cybersecurity, without which the functioning of electronic transit is impossible. Organizational factors are related to the level of digital competence of personnel, change management, and coordination between customs administrations and entities of foreign economic activity.

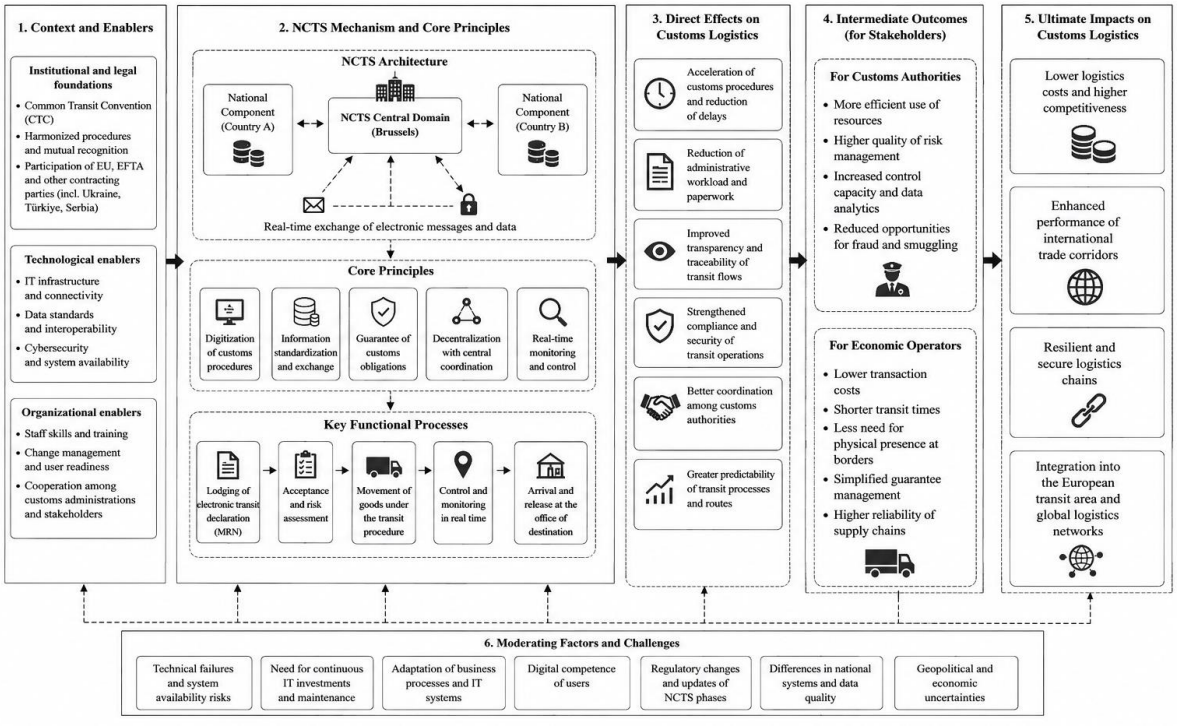


Figure 2: Conceptual model of the impact of NCTS on customs logistics.

The second block of the model characterizes NCTS as a mechanism for transforming customs logistics. The central element is the system architecture, consisting of national components and a central domain in Brussels, which provides for the exchange of electronic messages and transit data in real time. Such an architecture creates a single information environment for customs administrations of different countries, and ensures the continuity of transit operations.

Within this block, the basic principles of NCTS functioning are highlighted: digitalization of customs procedures, standardization of information exchange, guaranteeing customs obligations, decentralization with central coordination, and monitoring of transit operations in real time. Their implementation ensures the transition from traditional paper procedures to integrated digital management of transit flows. The key functional processes of the system include the electronic submission of a transit declaration, verification and risk assessment, movement of goods within the transit procedure, real-time control, and completion of transit at the customs office of destination. The combination of these processes ensures continuous digital support of cargo flows along the entire route of their movement.

The third block of the model reflects the direct consequences of the functioning of NCTS for customs logistics. First of all, the system contributes to the acceleration of customs procedures and the reduction of delays at the borders due to the automation of document flow and the use of a single transit declaration. At the same time, the administrative burden and the volume of paper procedures are reduced, which increases the efficiency of logistics operations. An important consequence is also the increase in transparency and traceability of transit flows, as information on the movement of goods becomes available to customs authorities in real time. This creates the prerequisites for strengthening control over compliance with customs rules, reducing the risks of fraud, and improving coordination between customs administrations.

The fourth block describes the intermediate results of the system for the main participants in the transit process − customs authorities and foreign economic operators. For customs administrations, NCTS provides more efficient use of resources, improved risk management, and expanded analytical capabilities based on electronic data. In addition, the system reduces opportunities for smuggling and documentary fraud through integrated control of transit operations. For businesses, the benefits are manifested in reduced transaction costs, reduced transit time, simplified guarantee procedures, and increased reliability of logistics chains. This creates a more predictable environment for international trade and stimulates the development of cross-border transportations.

The fifth block of the model reflects the final strategic consequences of the implementation of NCTS for customs logistics. As a result of the digitalization of transit procedures and the integration of customs administrations, an efficient, transparent, and secure system of international transit is formed. This contributes to reducing logistics costs, increasing the competitiveness of transport corridors and integrating states into global logistics networks. At the same time, the system increases the resilience of logistics chains to external risks, and creates conditions for the strategic development of customs infrastructure in accordance with the requirements of the digital economy.

A separate place in the model is occupied by a block of modifying factors and challenges that may affect the effectiveness of the NCTS. These include technical failures and risks of system availability, the need for constant investment in digital infrastructure, adaptation of business processes and software, an insufficient level of digital competences of users, differences between national information systems, and the impact of geopolitical and economic risks. The presence of these factors indicates that the effectiveness of the NCTS depends not only on the technological capabilities of the system, but also on the level of institutional and organizational readiness of states and businesses to operate in a digital transit environment.

Thus, the proposed conceptual model demonstrates that the NCTS is not only a tool for electronic customs clearance, but also a comprehensive digital platform for the transformation of customs logistics, which integrates information, organizational, and economic mechanisms for managing international transit flows.

5 CONCLUSIONS

The study found that NCTS is not only a tool for digitizing customs procedures, but a comprehensive mechanism for transforming customs logistics, which combines information and technology, organizational, and economic elements of transit flow management. It was determined that the key principles of the system are digitization of customs procedures, unification of information exchange,

guaranteeing customs obligations, decentralized coordination, and real-time monitoring.

The analysis of the implementation of NCTS in Ukraine in 2022–2025 showed a rapid growth in the volume of transit operations and increased integration of the national customs system into the European logistics space. It was found that the functioning of the system contributes to increasing the transparency of transit procedures, reducing administrative delays, optimizing logistics routes, and strengthening the control capabilities of customs authorities. At the same time, a structural asymmetry between the customs offices of departure and destination ones was identified, which indicates the need for further harmonization of individual stages of the transit cycle.

The scientific novelty of the study lies in the systematization of the consequences of the functioning of NCTS for customs logistics in information-technological, operational-procedural, economic, organizational-institutional, and international areas. The proposed conceptual model allowed us to reflect the relationship between the institutional prerequisites of the system's functioning, digital transformation mechanisms, and strategic logistics effects.

REFERENCES

[1] M. Antov, "The role of information technologies in the development of customs control in the Republic of Bulgaria," World Customs Journal, vol. 11, pp. 101-116, Sep. 2017, doi: 10.55596/001c.115706.

[2] S. Pashova, "Electronization of the Bulgarian customs administration," Customs Scientific Journal, vol. 7, no. 1, pp. 34-49, 2017.

[3] A. A. P. Azcárraga, T. Matsudaira, G. Montagnat-Rentier, J. Nagy, and R. J. Clark, Customs Matters: Strengthening Customs Administration in a Changing World. Washington, DC, USA: International Monetary Fund, 2022. [Online]. Available: https://doi.org/10.5089/9798400200120.071.

[4] K. Mikuriya, Compendium of Best Practices in the Area of Transit. Brussels, Belgium: World Customs Organization, Jun. 2020.

[5] S. Dalton and B. Stosić, "The importance of applying lean innovation for enhancing harmonization of customs procedures in context of digitalization of customs administration – A case study of Serbia," Lex Localis – J. Local Self-Government, vol. 19, pp. 305-327, Apr. 2021, doi: 10.4335/19.2.305-327(2021).

[6] V. Lebid, T. Anufriyeva, H. Savenko, and V. Skrypnyk, "Study of efficiency of simplification of customs formalities on the digitalization basis," Technol. Audit Prod. Reserves, vol. 1, no. 4, pp. 49-53, 2021, doi: 10.15587/2706-5448.2021.225627.

[7] A. Erceg, "The new computerized transit system and eCustoms influence on savings in transit time and costs," in Proc. 13th Int. Scientific Conf. Business Logistics in Modern Management, Osijek, Croatia, 2013, pp. 143-152.

[8] V. Fostolovych, L. Mykhalchyshyna, and Z. Simanaviciene, "International cooperation in the conditions of the NCTS joint transit system: In Ukrainian dimension," Public Policy and Accounting, vol. 1, no. 7, pp. 47-53, 2023, doi: 10.26642/ppa-2023-1(7)-47-53.

[9] A. Grainger and J. Boelts, A Practical Guide to Cross-Border Facilitation. Vienna, Austria: Organization for Security and Co-operation in Europe, Sep. 2025.

[10] R. Kril, "Common transit procedure and NCTS implementation: The case of Ukraine," Lex Portus, vol. 7, no. 3, pp. 32-59, 2021, doi: 10.26886/2524-101X.7.3.2021.2.

[11] O. Ilchenko, "Common customs transit procedure for Ukraine: Legal aspects," Visegrad J. Human Rights, no. 1, pp. 90-94, 2023, doi: 10.61345/1339-7915.2023.1.12.

[12] A. I. Dusmukhametov, E. R. Bokiev, and A. A. Saidov, "The study of the transit system of the European Union and the features of the information system 'E-Transit' of the Republic of Uzbekistan," Digital Transformation and Artificial Intelligence, vol. 1, no. 3, pp. 148-153, Oct. 2023.

[13] European Commission, "New Computerised Transit System (NCTS)," Directorate-General for Taxation and Customs Union. [Online]. Available: https://taxation-customs.ec.europa.eu/online-services/online-services-and-databases-customs/new-computerised-transit-system-ncts_en.

[14] F. Jane, NCTS: The New Computerised Transit System Made Easy. iCustoms.ai, Oct. 2023. [Online]. Available: https://www.icustoms.ai/blogs/ncts-new-computerised-transit-system.

[15] Convention between the European Economic Community, the Republic of Austria, the Republic of Finland, the Republic of Iceland, the Kingdom of Norway, the Kingdom of Sweden and the Swiss Confederation, on a Common Transit Procedure, 1987. [Online]. Available: https://eur-lex.europa.eu/eli/convention/1987/415/oj/eng.

[16] World Customs Organization, International Convention on the Simplification and Harmonization of Customs Procedures (Revised Kyoto Convention), 1973. [Online]. Available: https://www.wcoomd.org/en/Topics/Facilitation/Instrument%20and%20Tools/Conventions/pf_revised_kyoto_conv/Kyoto_New.

[17] State Customs Service of Ukraine, "NCTS." [Online]. Available: https://customs.gov.ua/news/ncts-26.

IT Standardization in the Area of Information Security with a Focus on ISO/IEC 27001

Lazar Krstić and Marija Krstić
Department of Higher Business School Leskovac, Academy of Applied Studies Southern Serbia, Partizanska Str. 7, 16000 Leskovac, Serbia
krstic.lazar@vpsle.edu.rs, krstic.marija@vpsle.edu.rs

Keywords: Standard, Information Security, ISO/IEC 27000, ISO/IEC 27001, T-test.

Abstract: Due to the intensive use of information technologies, the protection and security of information have become a key challenge and imperative in the business environment. Threats to information security are constantly growing, and more and more organizations know that poor information security can be "expensive", regardless of whether their or their customers' confidential data is at risk. In this regard, various standards and guidelines for information security have been developed. In this paper, the focus is on the ISO/IEC 27000 family of standards for information security management, more precisely, the most well-known standard for information security and the fourth most widespread ISO standard - ISO/IEC 27001. The paper aims to analyze the current state of standardization in the field of information protection and security through the analysis of available sources of knowledge and to point out the importance of applying the ISO/IEC 27000 family of standards in practice and then use statistical analysis, specifically, using adequate T-test to examine whether the price of the most popular and, in practice, the most applied standard from the ISO/IEC 27000 family, the ISO/IEC 27001 standard, is statistically significantly different from the price of other published standards from the said family.

1 INTRODUCTION

Information technology (IT) is a broad term that includes designing, developing, implementing, supporting, studying, or managing computer information systems, hardware, and software applications. At its core, IT is concerned with improving diverse human problem-solving efforts through the design, development, and use of technology-based systems and processes that improve the efficiency and effectiveness of information and relevant knowledge in a variety of strategic, tactical, and operational situations [1].

Standardization in IT contributes to the more efficient establishment of information functions, their greater stability, and easier transition. Applying international, national, and internal standards creates the conditions for developing an efficient, economical, reliable, and secure software product [2].

Nowadays, the use of information defence technology is no longer enough. It is essential to implement an effective information security management system.

Implementing the ISO/IEC 27001 standard often poses significant financial and organizational challenges, especially for small and medium-sized enterprises. Costs arise from purchasing the standard, engaging experts and consultants, training employees, adapting business processes, upgrading technical infrastructure, and the certification and recertification processes. Therefore, analyzing the economic aspects of implementing the standard is as important as analyzing its security benefits.

The aim of the research is to analyze standardization in the field of information security, with a particular focus on the ISO/IEC 27001 standard, and to examine whether its cost differs statistically significantly from the costs of other standards in the ISO/IEC 27000 family. The methods used in the paper include analysis, comparison, and statistical testing, with a single-sample Student's t-test performed in IBM SPSS Statistics.

Table 1: Literature review.

Title of work	Description
The ISO/IEC 27001 information security management standard: literature review and theory-based research agenda	This paper is based on a systematic review of the existing literature. The identified research themes and sub-themes are organized around five overarching research perspectives: integration with other standards, underlying motivations, implementation-related challenges, potential outcomes, and relevant contextual factors. Through a structured analysis of the academic literature on ISO/IEC 27001, the study provides a comprehensive overview of current knowledge and establishes a solid foundation for further investigation. In addition, the paper highlights several directions for future research, with particular emphasis on interdisciplinary studies positioned at the intersection of information security and quality management [3].
Overview of the Impact of Human Error on Cybersecurity based on ISO/IEC 27001 Information Security Management	The research presented in this paper focuses on defining practical rules that organizations should establish to support cybersecurity objectives and ensure effective data protection, particularly for incidents caused by human error. These rules are derived from the ISO/IEC 27001 standard and are intended for use in everyday organizational practices. Their adoption can increase employee awareness of individual behavior in information systems, thereby reducing both the likelihood and the potential impact of human-related security incidents on organizational data and systems [4].
A Model of an Information Security Management System Based on NTC-ISO/IEC 27001 Standard	The paper presents a model of the information security management system compliant with the NTC-ISO/IEC 27001:2013 standard. This model is valid for any organization and helps them understand their current information security status. The proposed model will enable organizations to implement systematic and appropriate controls, procedures, and policies necessary to preserve the integrity, confidentiality, and availability of information assets [5].
Comparative Study Between the Integration of ITIL and ISO/IEC 27001 with the Integration of COBIT and ISO/IEC 27001	The paper compares the integration of ITIL and ISO/IEC 27001 with the integration of COBIT and ISO/IEC 27001. Factors considered include the increased credibility of information security services, the cost of integrating the two standards, implementation time, the elimination of multiple processes, and improved understanding across parts of the organization. Various sources of literature are used as research methods. The research results can serve as a reference for organizations to determine which of the two information security standards is better to integrate [6].
Significance of ISO/IEC 27001 in the Implementation of Governance, Risk and Compliance	In this research paper, the researchers established the relationship between ISO 27001:2013 and GRC and discussed the standard and GRC objectives. In organizations, 'Governance', 'Risk', and 'Compliance' (GRC) are among the most fundamental and robust pillars that work together to ensure organizations achieve their goals by effectively using available people, processes, and technology. ISO 27001:2013 covers all GRC objectives within its Information Security Management System (ISMS) through which a practical GRC framework can be established and maintained [7].

2 LITERATURE REVIEW

Few recent freely available works deal with the ISO/IEC 27000 family of standards, precisely the ISO/IEC 27001 standard. Table 1 provides an overview of works dealing with the issue of the ISO/IEC 27001 standard from various aspects. The table contains information about the year of publication of the work, the authors of the work, the title, and a short description.

3 INFORMATION SECURITY

The contemporary business landscape relies heavily on information technologies, as information itself represents a strategic resource that significantly influences organizational continuity and development. Irrespective of the medium in which it is stored or processed, information requires adequate protection. Achieving this level of protection assumes that users understand basic security principles and the measures required to implement them. In this context, maintaining the confidentiality, integrity, and availability of information remains a fundamental objective. Nevertheless, information security extends beyond the mere deployment of technical solutions provided by modern IT systems. Information security is the ability to safeguard information throughout its lifecycle (acquisition, processing, storage, transmission, and use) against diverse threats. The relevance of this issue has grown alongside the rapid expansion of computer technologies and the increasing interconnection of systems within

networked environments. As a result, information security has evolved into a broader concern that demands continuous attention and a comprehensive approach.

4 STANDARDIZATION AND IT

A standard represents a formally defined document that sets out requirements, recommendations, or characteristics intended to ensure that materials, products, processes, and services fulfill their intended purpose. Such documents are developed through a consensus-based process and are formally adopted by recognized standardization bodies. Their application is not limited to technologically advanced economies; standards are widely used across regions worldwide, including both developed and developing countries [8].

An IT standard is a rule, principle, technique, process, or pattern designed to ensure consistency in information technology services' planning, development, operation, and management [9]. IT standardization refers to making software, systems, and infrastructure used throughout an organization fully consistent and integrated. Once this is done, it is clear that the systems will more easily "talk" to each other, helping to perform business processes more efficiently and enabling jobs to be completed faster [10].

4.1 Organizations for Standardization

Standards are of great benefit and importance to consumers and users, as well as to trade, science, and the state, because they provide innovative technologies and increased safety for society [11]. There are three levels of standardization, i.e. standardization can take place at the international, European, or national levels. All three levels are interconnected, and all complement each other. This is ensured by the structure of standardization organizations and internal regulations [12]. The three levels of standardization are:

- International level (ISO);
- European level (CEN, CENELEC, ETSI);
- National level (National Standardization Bodies - NSBs).

The International Organization for Standardization - ISO and the International Electrotechnical Commission - IEC constituted the joint technical committee JTC 1, tasked with adopting standards in information technologies.

4.2 ISO/IEC 27000 Family of Standards and ISO/IEC 27001

To address increasingly sophisticated security challenges, organizations should take an integrated approach to information protection by implementing an Information Security Management System (ISMS) to protect the confidentiality, integrity, and availability of information assets. The ISO/IEC 27000 series provides an internationally recognized framework for information security management and organizational resilience, and certification to these standards will also prevent clients from conducting repeated audits. Among these, ISO/IEC 27001 (a joint effort by ISO and IEC) is a central and popular standard that sets out a structured set of security objectives and controls in Annex A. Recent updates to this standard have provided greater flexibility in its requirements, replacing a mandate to use specific controls with an organization's ability to choose controls that best suit its risk management process [13], [14].

The ISO/IEC 27000 family of standards provides a comprehensive framework for information security management, with ISO/IEC 27001 at its core, specifying requirements for an ISMS that integrates people, processes, and IT systems. Applicable to organizations of all sizes and sectors, these standards help manage the security of critical assets, including financial data, intellectual property, employee information, and third-party data. As technology and organizational environments evolve, the family continues to expand, offering guidance to address changing information security needs worldwide. ISO/IEC 27001 is an international standard for information security (in Serbia, SRPS ISO/IEC 27001:2014) that specifies requirements for an ISMS. As part of the ISO 27000 series, it provides a framework for organizations to establish, implement, operate, monitor, review, and continuously improve their information security management system [15].

ISO 27001 helps organizations set clear responsibilities for information. Implementing information security standards ensures that security becomes part of the organization's culture and resilience against cyber threats. ISO 27001 is a standard that an organization should maintain by conducting risk assessments, which enables management and key stakeholders to maintain information security risks. The safeguards (or controls) that an organization must implement are usually policies, procedures, and technical implementation (such as software and hardware). However, in most cases, organizations already have

all the necessary security controls in place, but they are not always implemented correctly. Since such an implementation will require the management of many policies, procedures, people, and resources, ISO 27001 describes how to link all these elements together in an ISMS. Therefore, the whole concept of ISMS is not only about IT security (like firewalls and antivirus). These are process management, legal protection, human resource management, physical asset protection, and more [16].

The ISO/IEC 27001 standard is comprehensive, as it treats information security from three aspects [17]:

- IT - analyzing and defining the performance of IT equipment, access rights, encryption, passwords, protocols, and policies from the aspect of data and information security risks;
- administrative - defining clear instructions, policies, and procedures for the generation of information, its distribution, and preservation (storage);
- physical - physical access control, employee records, video surveillance, protection of work premises.

The advantages of the information security management system are [2]:

- Meeting legal requirements - there are more and more laws, regulations, and contractual requirements related to information security, most of which can be addressed by applying ISO/IEC 27001 - this standard provides the perfect methodology to comply with them all.
- Gaining a marketing advantage - if an organization gets certified and competitors don't, it gives them an advantage in the eyes of customers sensitive to their data protection.
- Lower costs - the underlying philosophy of ISO/IEC 27001 is to prevent security incidents, and every incident, big or small, costs money. Therefore, the organization will save a lot of money by preventing incidents.

Better organization - fast-growing organizations often do not have time to stop and define their processes and procedures. The consequence is that employees often do not know what, when, and who should do it. Implementing ISO/IEC 27001 helps solve such a situation because it encourages organizations to write their core processes (even those unrelated to security), which allows them to reduce losses and optimize employees' working time.

4.3 IEEE Standardization in the Field of Cybersecurity and Privacy

IEEE (Institute of Electrical and Electronics Engineers) is the world's largest technical professional organization dedicated to advancing technology through technology standard setting and educational activities [18]. IEEE standards are developed within its societies and coordination committees through a consensus-based process approved by the American National Standards Institute. This process brings together volunteers representing diverse perspectives, who contribute without compensation, to produce the final standard. While IEEE oversees the development process and ensures fairness, it does not test, evaluate, or verify the technical accuracy of the information contained in the standards [19].

The IEEE Cybersecurity and Privacy Standards Committee is responsible for standardizing cybersecurity and privacy standards in the IEEE. Cyber security focus is but not limited to cryptographic techniques, cyber incident management, identity management, IT system security assessment, information security management systems, network security, security automation, continuous monitoring, supply chain risk management, software, and system assurance safety engineering standards. On the other hand, privacy includes, but is not limited to, standardization related to privacy risk identification and mitigation methods and technology [20].

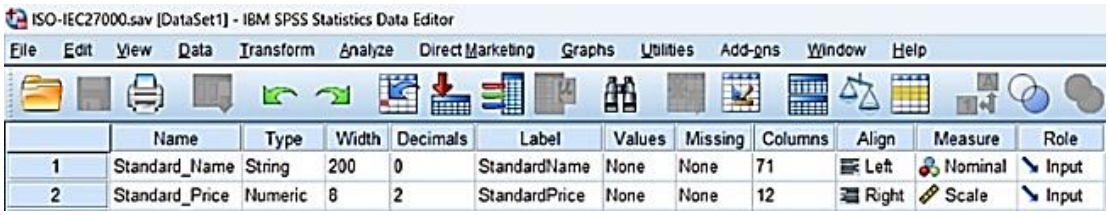
5 RESEARCH METHODOLOGY

Basic methods, including analysis, synthesis, concretization, and generalization, as well as general scientific methods, analytical-deductive and comparative, were applied in the work. Basic methods (analysis, synthesis, concretization, and generalization) were used to have a more superficial understanding of concepts from the theory of information security, standardization in that area, ISO/IEC and IEEE standards, i.e., interpretation of T-test results, and conclusion. The analytical-deductive method was used to operationalize appropriate analytical procedures of a qualitative-quantitative character, making adequate conclusions. The comparative method was used when comparing scientific and professional papers dealing with similar

issues to those with which this paper deals. Also, the paper used an appropriate statistical method for hypothesis testing, while the sample description was presented with the help of descriptive statistics indicators.

The research problem investigates whether the price of the ISO/IEC 27001 standard differs statistically from the average price of other standards in the ISO/IEC 27000 family. Since the objective is to compare the sample mean with a predefined reference value, a one-sample t-test was applied. The essence of the one-sample Student's t-test is to determine whether the sample mean differs statistically significantly from a predefined reference value, allowing assessment of whether the observed difference is due to random variation or is statistically significant. Statistical analysis was conducted using IBM SPSS Statistics, a software tool for statistical data processing. The dataset included published standards from the ISO/IEC 27000 family, obtained from the official website of the International Organization for Standardization (ISO), where each row corresponds to a single standard and its price.

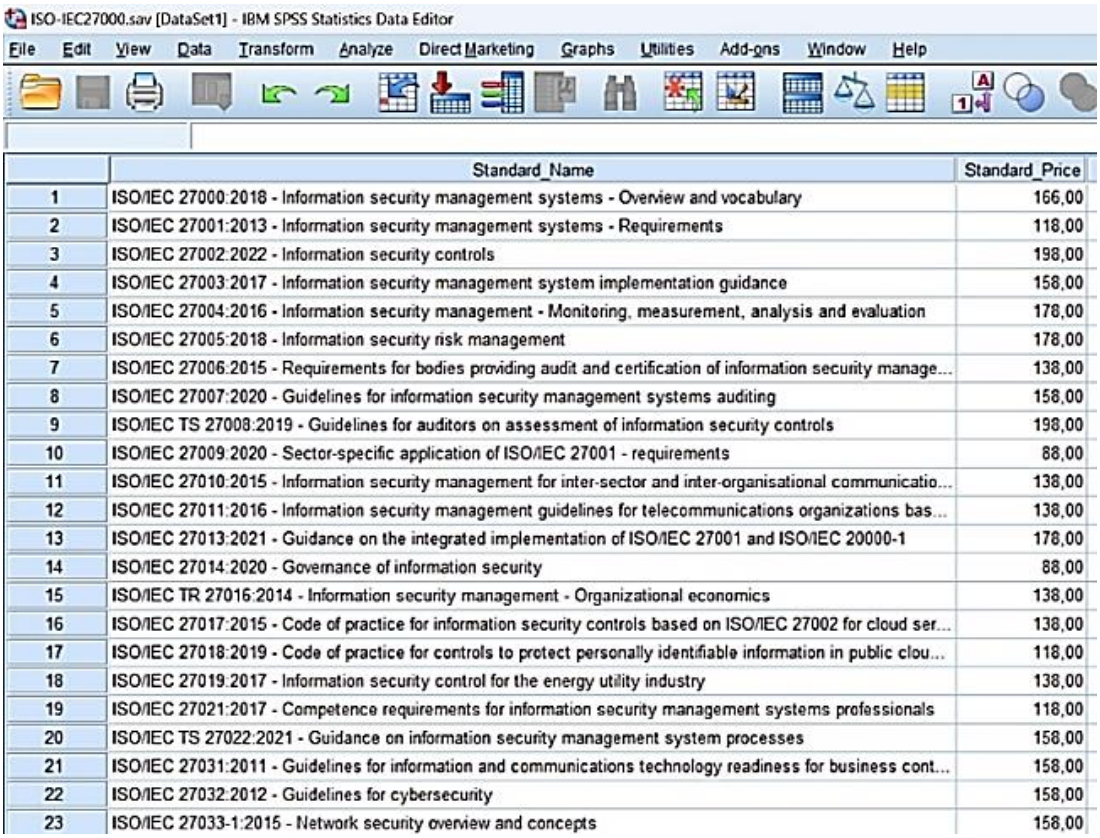
In the statistical analysis, two variables were used: Standard_Name, a nominal variable for standard identification, and Standard_Price, a continuous numerical variable representing the price of each standard. Figure 1 presents the variables used in the analysis, highlighting the nominal variable (Standard_Name) and the continuous dependent variable (Standard_Price), while Figure 2 illustrates the dataset employed for the statistical analysis.


ISO-IEC27000.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Standard_Name	String	200	0	StandardName	None	None	71	Left	Nominal	Input
2	Standard_Price	Numeric	8	2	StandardPrice	None	None	12	Right	Scale	Input

Figure 1: Presentation of the variables used in the statistical analysis.


ISO-IEC27000.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	Standard_Name	Standard_Price
1	ISO/IEC 27000:2018 - Information security management systems - Overview and vocabulary	166,00
2	ISO/IEC 27001:2013 - Information security management systems - Requirements	118,00
3	ISO/IEC 27002:2022 - Information security controls	198,00
4	ISO/IEC 27003:2017 - Information security management system implementation guidance	158,00
5	ISO/IEC 27004:2016 - Information security management - Monitoring, measurement, analysis and evaluation	178,00
6	ISO/IEC 27005:2018 - Information security risk management	178,00
7	ISO/IEC 27006:2015 - Requirements for bodies providing audit and certification of information security manage...	138,00
8	ISO/IEC 27007:2020 - Guidelines for information security management systems auditing	158,00
9	ISO/IEC TS 27008:2019 - Guidelines for auditors on assessment of information security controls	198,00
10	ISO/IEC 27009:2020 - Sector-specific application of ISO/IEC 27001 - requirements	88,00
11	ISO/IEC 27010:2015 - Information security management for inter-sector and inter-organisational communicatio...	138,00
12	ISO/IEC 27011:2016 - Information security management guidelines for telecommunications organizations bas...	138,00
13	ISO/IEC 27013:2021 - Guidance on the integrated implementation of ISO/IEC 27001 and ISO/IEC 20000-1	178,00
14	ISO/IEC 27014:2020 - Governance of information security	88,00
15	ISO/IEC TR 27016:2014 - Information security management - Organizational economics	138,00
16	ISO/IEC 27017:2015 - Code of practice for information security controls based on ISO/IEC 27002 for cloud ser...	138,00
17	ISO/IEC 27018:2019 - Code of practice for controls to protect personally identifiable information in public clou...	118,00
18	ISO/IEC 27019:2017 - Information security control for the energy utility industry	138,00
19	ISO/IEC 27021:2017 - Competence requirements for information security management systems professionals	118,00
20	ISO/IEC TS 27022:2021 - Guidance on information security management system processes	158,00
21	ISO/IEC 27031:2011 - Guidelines for information and communications technology readiness for business cont...	158,00
22	ISO/IEC 27032:2012 - Guidelines for cybersecurity	158,00
23	ISO/IEC 27033-1:2015 - Network security overview and concepts	158,00

Figure 2: Representation of a part of the data set used in the statistical analysis.

Table 2: Descriptive statistics of the variables in the analysis.

One-sample statistics				
	N	Mean	Std. Deviation	Std. Error Mean
Standard_Price	65	139.0462	35.33608	4.38290

Table 3: One sample t-test results.

One-sample test						
	Test Value = 118					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Standard_Price	4.802	64	.000	21.04615	12.2903	29.8020

6 RESULTS AND DISCUSSION

Today, the cost and efficiency of implementing the ISO/IEC 27001 standard are addressed through cost-benefit analyses, risk assessments, and return-on-investment models. Organisations increasingly adopt a risk-based approach, implementing only those security controls that are economically justified and tailored to their business needs. In addition, integration with other standards and the use of cloud and automated solutions help reduce the costs of implementing and maintaining the information security system.

In order to empirically examine the cost differences identified through these approaches, a one-sample T-test was applied, and the following hypotheses were formulated:

- H0: The price of the ISO/IEC 27001 standard is statistically not significantly different from the price of other published standards from the ISO/IEC 27000 family.
- H1: The price of the ISO/IEC 27001 standard is statistically significantly different from the price of other published standards from the ISO/IEC 27000 family.

After performing the T-test for one sample in the SPSS tool, the results shown in Tables 2 and 3 were obtained.

Table 2 shows descriptive statistics, i.e., the number of samples, mean, standard deviation, and standard error of the mean. The average value of the test result is 139.04. Table 3 shows the one-sample T-test statistics. In the first row, the mean value of the population is entered (test value = 118). It is necessary to refer to the fourth column (Sig. (2-tailed)) for the significance of the T-test statistic. If $p > .05$, the null hypothesis is not rejected. In this case, the p-value is less than 0.05 ($p = .000 < .05$), so the null hypothesis is rejected, and it is concluded that the price of the ISO/IEC 27001 standard is statistically significantly different from the price of other published standards from the ISO/IEC 27000 family.

The results indicate a significant difference between the true mean (M=118.00) and the mean test score (M=139.04; SD=35.33), [$t(64) = 4.802$, $p = .000$]. Therefore, the null hypothesis that there is no difference between the true mean and the comparison value is rejected, and it is concluded that the mean test score is significantly different from the true population mean.

This statistically significant difference in the price of the ISO/IEC 27001 standard compared with the other ISO/IEC 27000 standards suggests that cost is a potential barrier to decision-making for organizations considering certification. This result may also affect decision-making processes, especially for small and medium-sized enterprises, by highlighting the role of cost-related factors in the adoption of information security standards. This study provides empirical evidence on economic factors in information security, alongside technical and organizational factors. By quantitatively demonstrating price differences across these standards, the analysis has contributed to a more complete picture of the information security standards adoption process.

The strength of the ISO/IEC 27001 standard is its ability to help an organization establish an information security risk management framework that protects vital information. This is based on various studies indicating that organizations that adopt the ISO/IEC 27001 standard will see improved employee awareness, greater security uniformity, and better defenses against online attacks. However, achieving this standard will depend on the organization's commitment to incorporating security controls into its culture, thereby enabling the development of a customized standard to suit the organization’s needs. From this perspective, ISO/IEC 27001 certification alone is not enough; the organization’s culture plays a critical role in the standard’s success.

7 CONCLUSIONS

Implementing IT standardization is one of the fastest ways to achieve business goals. In other words, IT standardization can help organizations achieve their goals faster and continuously upgrade their growth and profits.

Awareness of the importance of information security is increasing, a positive trend that should continue. The International Organization for Standardization (ISO) remains committed to helping global organizations by developing standards based on input from subject matter experts around the world.

The ISO/IEC 27000 family of standards was created to help organizations manage the risks of cyberattacks and internal threats to data security. The ISO/IEC 27001 standard from the specified family of standards provides a framework for ISMS organizations and is certifiable and widely used, with a tendency of constant growth. It consists of processes, policies, and resources that can be used to systematize an organization's security requirements.

By statistical analysis, specifically by applying the Student's T-test for one sample over the created data set, which contains published standards from the ISO/IEC 27000 family, it was necessary to determine and give an answer to the question whether the price of the most popular and in practice the most applied standard from the ISO family /IEC 27000, the ISO/IEC 27001 standard, statistically significantly differs from the price of other published standards from the mentioned family. Based on the obtained results, it is concluded that the price of the ISO/IEC 27001 standard is statistically significantly different from the price of other published standards from the ISO/IEC 27000 family.

The importance of the ISO/IEC 27001 standard for organizations is reflected in the fact that it enables systematic information security management through risk identification and control, protection of confidentiality, integrity, and availability of data, as well as compliance with legal and contractual requirements, thereby reducing security incidents and strengthening the trust of customers and partners. On the other hand, the cost of implementing and certifying this standard plays a key role because it affects the scope, quality, and sustainability of the information security management system, so adequate investment represents a long-term investment in the stability of the business, its reputation, and the competitive advantage of the organization.

The ISO/IEC 27001 standard cannot be downloaded for free from official sources but must be purchased from ISO or national standards bodies. Cost can be a significant barrier for small businesses and other organizations, especially in lower-income countries. This limitation reduces the wider availability of knowledge, since the standard contains recommended practices and controls for information security management, which, in the interest of general security improvement, should be more accessible to the public. Difficult access to the standard further reduces the likelihood that organizations will understand its importance and consider its implementation.

REFERENCES

[1] The MBA Institute, “Definitions of IT,” 2025, [Online]. Available: https://themba.institute/information-systems-for-managers/definitions-of-it/, [Accessed: Aug. 20, 2025].

[2] PC Press, “Standardi u informacionim sistemima i tehnologijama,” 2025, [Online]. Available: https://pcpress.rs/standardi-u-informacionim-sistemima-i-tehnologijama/, [Accessed: Aug. 20, 2025].

[3] G. Culot, G. Nassimbeni, M. Podrecca, and M. Sartor, “The ISO/IEC 27001 information security management standard: literature review and theory-based research agenda,” The TQM Journal, vol. 33, no. 7, pp. 76-105, 2021, [Online]. Available: https://doi.org/10.1108/TQM-09-2020-0202.

[4] A. Y. El-Bably, “Overview of the impact of human error on cybersecurity based on ISO/IEC 27001 Information Security Management,” Journal of Information Security and Cyber Research (JISCR), vol. 4, no. 1, pp. 95-102, 2021, [Online]. Available: https://doi.org/10.26735/WLPW6121.

[5] O. Fonseca-Herrera, A. Rojas, and H. Florez, “A model of an information security management system based on NTC-ISO/IEC 27001 standard,” IAENG International Journal of Computer Science, vol. 48, no. 2, pp. 213-222, 2021, [Online]. Available: https://www.iaeng.org/IJCS/issues_v48/issue_2/IJCS_48_2_01.pdf, [Accessed: Aug. 27, 2025].

[6] N. K. Gunawan, R. B. Hadiprakoso, and H. Kabetta, “Comparative study between the integration of ITIL and ISO/IEC 27001 with the integration of COBIT and ISO/IEC 27001,” in IOP Conf. Series: Materials Science and Engineering, Jakarta, Indonesia, Nov. 21-22, 2020, pp. 1-5, [Online]. Available: https://doi.org/10.1088/1757-899X/852/1/012128.

[7] S. Choubey and A. Bhargava, “Significance of ISO/IEC 27001 in the implementation of governance, risk and compliance,” IJSRNSC Int. J. of Scientific Research in Network Security and Communication, vol. 6, no. 2, pp. 30-33, 2018, [Online]. Available: https://ijsrnsc.org/index.php/j/article/view/130/130, [Accessed: Sept. 2, 2025].

[8] Institut za standardizaciju Srbije, "Šta je standard?" 2025, [Online]. Available: https://iss.rs/sr_Latn/shta-je-standard_p13.html, [Accessed: Sept. 3, 2025].

[9] CIO Wiki, "IT Standard (Information Technology Standard)," 2025, [Online]. Available: https://cio-wiki.org/wiki/IT_Standard_(Information_Technology_Standard), [Accessed: Sept. 4, 2025].

[10] 1+1 Technology, "Achieve your business goals faster with IT standardization," 2025, [Online]. Available: https://www.1plus1tech.com/achieve-your-business-goals-faster-with-it-standardization/, [Accessed: Sept. 5, 2025].

[11] DKE Standards, "The importance of standardization – benefits and advantages," 2025, [Online]. Available: https://www.dke.de/en/standards-and-specifications/importance-of-standardization, [Accessed: Sept. 6, 2025].

[12] Standards+Innovation, "Standardization organizations," 2025, [Online]. Available: https://www.standardspluselearning.eu/b-1-1-standardization-organizations, [Accessed: Sept. 7, 2025].

[13] D. Ganji, H. Mouratidis, and S. M. Gheytassi, "Towards a modelling language for managing the requirements of ISO/IEC 27001 standard," in Proc. SOFTENG 2019 – Fifth Int. Conf. on Advances and Trends in Software Engineering, Valencia, Spain, Mar. 24-28, 2019, pp. 17-23, [Online]. Available: https://www.researchgate.net/publication/332801832.

[14] A. Renvall, Improving Cybersecurity Through ISO/IEC 27001 Information Security Standard in the Context of SMEs, M.S. thesis, Metropolia Univ. of Applied Sciences, Helsinki, Finland, 2018, [Online]. Available: https://www.theseus.fi/bitstream/handle/10024/157277/Renvall_Aleksi_final.pdf.

[15] IT Governance, "ISO 27000 series of standards," 2025, [Online]. Available: https://www.itgovernance.co.uk/iso27000-family, [Accessed: Sept. 13, 2025].

[16] Auro Standard, "ISO 27001," 2025, [Online]. Available: https://www.aurostandard.org/standardi/iso-27001/, [Accessed: Sept. 13, 2025].

[17] Jonik, "ISO 27001:2013," 2025, [Online]. Available: https://www.jonik.rs/rs/iso-27001-2013, [Accessed: Sept. 15, 2025].

[18] IEEE, "About IEEE," 2025, [Online]. Available: https://www.ieee.org/, [Accessed: Sept. 16, 2025].

[19] ISO, "ISO/IEC/IEEE 18882:2017," 2025, [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso-iec-ieee:18882:ed-1:v1:en, [Accessed: Sept. 16, 2025].

[20] IEEE Computer Society, "Cybersecurity and Privacy Standards Committee," 2025, [Online]. Available: https://www.computer.org/volunteering/boards-and-committees/standards-activities/committees/cybersecurity-privacy, [Accessed: Sept. 20, 2025].

From Cost Structure to ESG Sustainability: Structural and Cross-Country Evidence from the EU Livestock Sector

Ayagoz Orazbayeva[1,2], Maira Bauer[1,2] and Aliya Ismailova[2]
[1]*Humboldt Innovation GmbH, Humboldt University of Berlin, Unter den Linden 6, 10117 Berlin, Germany*
[2]*Institute of Business and Digital Technologies, Saken Seifullin Kazakh Agrotechnical Research University, Zhenis Avenue 62, 010011 Astana, Kazakhstan*
a.orazbaeva@kazatu.edu.kz, mairak@bk.ru, ismailova61@mail.ru

Keywords: Cost Structure, ESG Sustainability, Livestock Sector, Cross-Country Analysis, Sustainability Assessment, Energy Intensity.

Abstract: The transition toward climate-neutral development in the European Union has intensified the need for structured and data-driven approaches to sustainability assessment in agriculture. Despite extensive research on environmental performance in livestock systems, empirical models explicitly linking managerial cost structure to ESG sustainability outcomes at a cross-country level remain limited. This study develops an integrated analytical framework to examine whether cost allocation patterns and energy structure systematically determine sustainability profiles in the EU livestock sector. The empirical analysis is based on a harmonized panel dataset covering Germany, France, the Netherlands, Spain, and Poland for the period 2000-2023. A multi-stage modelling approach combining k-means and fuzzy clustering, regression analysis, and Random Forest classification was implemented in MATLAB. The clustering stage identified three stable ESG structural profiles. Regression results indicate that energy share and environmental investment allocation significantly explain variation in GHG intensity ($R^2 = 0.9173$). Machine learning classification further confirms the structural importance of energy-related cost components in differentiating sustainability classes. The findings demonstrate that sustainability performance in livestock systems is not random but structurally embedded in managerial cost architecture. The study contributes to sustainability accounting by integrating ESG metrics with sector-level cost indicators and provides a scalable analytical framework for sustainability benchmarking and policy evaluation in the EU agri-food sector.

1 INTRODUCTION

The transition of the European Union (EU) toward climate-neutral development has significantly increased the demand for quantitative tools capable of assessing sustainability performance in agriculture. Livestock production remains one of the most emission-intensive branches of the agricultural sector and is closely associated with methane emissions, energy consumption, and resource allocation efficiency [1], [2]. At the same time, EU climate policy and decarbonization strategies exert growing pressure on agricultural systems to reduce greenhouse gas (GHG) intensity while preserving economic viability [3].

Existing approaches to agricultural sustainability assessment are predominantly based on multi-criteria and composite index frameworks, including systems such as IDEA4 [4], [5]. While these methodologies provide structured evaluation tools, they often rely on predefined weighting schemes and do not fully integrate managerial cost architecture into sustainability analysis. Research in environmental and carbon management accounting demonstrates that operational cost components — such as feed expenditure, fossil energy use, veterinary services, and environmental investments — directly influence environmental performance indicators [6]. However, empirical cross-country models explicitly linking sector-level cost structure to carbon intensity and ESG differentiation remain limited.

Recent advances in machine learning (ML) enable the identification of nonlinear patterns and latent structural relationships in sustainability data. ML-based ESG analytics have been widely applied in corporate finance and environmental performance prediction [7], [8], whereas in agriculture, ML applications predominantly focus on yield forecasting

or production monitoring [9]. The integration of clustering techniques and ensemble learning methods into cross-country livestock sustainability modelling remains insufficiently explored.

Furthermore, carbon footprint and emission intensity studies confirm that energy structure and management practices play a decisive role in environmental outcomes [1], [10]. Nevertheless, these analyses rarely incorporate managerial accounting variables into predictive sustainability frameworks, thereby limiting their ability to connect cost allocation decisions with ESG performance.

This study addresses this methodological gap by developing an integrated ML-based analytical framework for assessing ESG sustainability in the livestock sector across five EU countries — Germany, France, the Netherlands, Spain, and Poland — for the period 2000-2023. The proposed framework combines clustering techniques to identify ESG structural profiles, regression modelling to quantify the influence of cost allocation patterns on GHG intensity, and Random Forest classification to determine sustainability classes and assess predictor importance.

The contribution of this research is threefold. First, it integrates ESG sustainability metrics with managerial accounting indicators at the sectoral level. Second, it applies a multi-stage analytical framework to cross-country panel data, enhancing methodological robustness. Third, it provides an interpretable structural linkage between cost architecture, energy dependence, and sustainability differentiation.

The remainder of the paper is structured as follows. Section 2 describes the dataset and methodological approach. Section 3 presents the empirical results. Section 4 discusses implications, and Section 5 concludes.

2 DATA AND METHODS

2.1 Data Description

The empirical analysis is based on a harmonized cross-country panel dataset covering five EU member states — Germany, France, the Netherlands, Spain, and Poland — for the period 2000-2023. These countries were selected to ensure structural heterogeneity within the sample, representing both highly industrialized livestock systems (Germany, the Netherlands, France) and structurally transitional systems (Spain, Poland). Together, they account for a substantial share of EU livestock production and provide variation in technological development, energy dependence, and sustainability investment patterns, enabling meaningful cross-country structural comparison. The dataset integrates production, environmental, and managerial cost indicators derived from FAOSTAT, Eurostat, and national statistical sources. After data harmonization and cleaning, the constructed panel comprises annual observations for each country over the study period.

The analytical variables include sector-level cost structure indicators (share_feed, share_energy, share_vet_medicine, share_labor_wages, share_env_taxes_fees, share_env_investments), environmental performance indicators (GHG_intensity, energy_intensity), and country dummy variables to capture structural heterogeneity across national livestock systems.

To ensure cross-country comparability and eliminate scale effects, all continuous variables were standardized prior to modelling using z-score normalization. This procedure allows consistent estimation within clustering and machine learning algorithms while preserving relative structural differences among observations.

Variables with insufficient observations or structural breaks over the observation period were excluded from subsequent modelling stages to maintain statistical consistency and avoid bias from incomplete time series. The resulting dataset provides a coherent analytical basis for examining the structural linkage between cost allocation patterns and ESG sustainability outcomes in the EU livestock sector.

2.2 Clustering Analysis

To identify latent ESG structural profiles within the EU livestock sector, clustering analysis was conducted using a multi-algorithm approach implemented in MATLAB. Prior to clustering, all continuous variables were standardized using z-score normalization to ensure comparability across indicators and prevent dominance of scale-sensitive variables.

Three complementary clustering techniques were applied:

- K-means clustering, providing compact partitioning of observations by minimizing within-cluster variance;
- Fuzzy C-means (FCM) clustering, allowing partial membership and identification of transitional sustainability trajectories;

- Hierarchical clustering (Ward linkage), enabling structural validation of grouping patterns.

The optimal number of clusters was determined as k = 3 based on within-cluster sum of squares analysis, cluster interpretability, and stability across repeated random initializations. The selection of k = 3 was guided by within-cluster variance reduction and structural interpretability of sustainability profiles. Given the moderate sample size and the panel nature of the dataset, emphasis was placed on stability across repeated initializations and cross-method consistency (k-means, fuzzy C-means, hierarchical clustering) rather than reliance on a single cluster validity metric. Multiple runs of the k-means algorithm confirmed convergence stability and consistent partitioning outcomes.

The clustering input vector included the following standardized variables:

- GHG_intensity;
- share_energy;
- share_feed;
- energy_intensity;
- share_env_investments.

This five-dimensional feature space integrates managerial cost allocation and environmental performance indicators, allowing identification of structurally differentiated sustainability configurations. The combined use of hard (k-means) and soft (FCM) clustering enhances robustness by distinguishing stable structural profiles from transitional observations.

2.3 Regression Analysis of Carbon Intensity

To evaluate the structural relationship between managerial cost allocation and environmental performance, regression analysis was conducted in MATLAB using a linear modelling framework. Greenhouse gas intensity (GHG_intensity) was specified as the dependent variable representing sector-level environmental outcomes across countries and years.

The explanatory variables included energy cost share (share_energy), environmental investment share (share_env_investments), operational energy intensity (energy_intensity), and country dummy variables. The inclusion of country dummies allows control for structural heterogeneity across national livestock systems. Given the limited cross-sectional dimension of the dataset (five countries), a parsimonious linear modelling framework was adopted to preserve interpretability while accounting for cross-country fixed structural effects.

All continuous variables were standardized prior to estimation to ensure comparability across indicators and to avoid scale-driven distortions. The regression was estimated using the full available panel sample within MATLAB's linear modelling environment. Model performance was evaluated using goodness-of-fit indicators, including R^2, root mean squared error (RMSE), and mean absolute error (MAE).

The objective of this regression stage is not to establish strict causal inference, but rather to assess the explanatory capacity of managerial cost architecture in accounting for variation in GHG intensity across countries and years within the proposed sustainability framework.

A potential limitation of the regression specification is the possibility of endogeneity between explanatory variables and GHG intensity. In particular, environmental investment allocation may partly reflect a response to previously observed emission levels rather than acting solely as an exogenous determinant. Given the scope of the study and the aggregate nature of the panel data, the model is interpreted as capturing structural associations rather than strict causal relationships. Addressing endogeneity through instrumental variable approaches or dynamic panel techniques remains an important direction for future research.

2.4 Random Forest ESG Classification

To further evaluate sustainability differentiation, a Random Forest classification model was implemented in MATLAB. The objective of this stage was not predictive forecasting, but structural validation of sustainability profiles and identification of dominant explanatory variables.

Since the constructed regression panel did not contain a predefined ESG class label, the classification target variable was derived from the k-means clustering results (k = 3). The resulting clusters were ordered into sustainability levels, where Class A represents the most sustainable structural profile and Class C represents the least sustainable configuration.

The classification input variables included:

- share_feed;
- share_energy,
- share_env_investments;
- GHG_intensity;
- energy_intensity;
- country dummy variables.

The Random Forest model was trained using MATLAB's ensemble learning framework. Model performance was evaluated using out-of-bag (OOB) validation, which provides an internal cross-validation mechanism and reduces the risk of overfitting.

Permutation-based feature importance was calculated to assess the relative contribution of each predictor to classification accuracy. This procedure enables identification of structurally dominant sustainability drivers within the integrated analytical framework.

The classification stage complements clustering and regression analysis by providing an independent validation of sustainability differentiation and highlighting the explanatory relevance of managerial cost architecture.

2.5 Integrated ML Framework

The proposed methodological architecture integrates clustering, regression analysis, and ensemble classification into a unified analytical system for ESG sustainability assessment in the livestock sector. Rather than treating these methods as isolated analytical tools, the framework applies them sequentially to construct a structurally consistent interpretation of sustainability differentiation.

The analytical sequence operates as follows. First, clustering techniques identify latent ESG structural profiles based on multidimensional cost-emission indicators. This stage reveals endogenous sustainability groupings without imposing predefined classifications. Second, regression analysis quantifies the structural relationship between managerial cost allocation and GHG intensity, providing explanatory insight into emission dynamics across countries and years. Third, Random Forest classification validates sustainability differentiation and determines the relative importance of cost and energy variables in predicting structural sustainability profiles.

The integration of these stages ensures methodological triangulation: clustering uncovers structural patterns, regression explains cost-emission relationships, and machine learning evaluates classification robustness and predictor dominance. This layered approach strengthens internal consistency and reduces reliance on any single modelling technique.

Overall, the framework enables a transition from descriptive sustainability measurement toward structurally interpretable ESG-oriented analytical control. By linking cost architecture, energy dependence, and environmental performance within a coherent modelling system, the proposed approach advances sustainability accounting methodologies at the sectoral level.

3 RESULTS

3.1 Clustering Results

The clustering analysis identified three stable ESG structural profiles across the five EU countries over the period 2000-2023. As illustrated in Figure 1, the observations form three clearly distinguishable groups within the multidimensional cost-emission space. The figure presents a three-dimensional projection of the five-dimensional clustering solution, enabling visualization of structural differentiation across sustainability profiles.

The k-means algorithm (k = 3) demonstrated robust partitioning, confirmed by repeated random initialization and convergence stability across iterations. The resulting clusters represent structurally distinct combinations of managerial cost allocation, energy dependence, and environmental performance indicators.

Cluster 1 corresponds to a resource-intensive structural configuration characterized by a high energy cost share, elevated GHG intensity, and relatively limited environmental investment allocation. This profile is associated with higher energy dependence and comparatively lower ESG sustainability levels. Similar structural contrasts between energy-intensive and environmentally modernized agricultural systems have been identified in prior sustainability assessment research [4], [5].

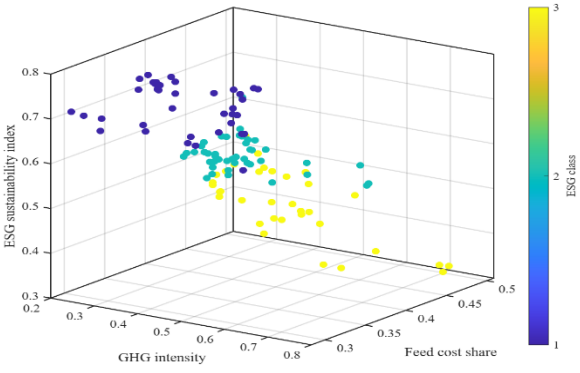


Figure 1: Three-dimensional projection of k-means clustering results (k = 3) showing ESG structural profiles in the EU livestock sector (2000-2023).

Cluster 2 represents a transitional sustainability profile, characterized by moderate indicator values and more balanced cost allocation patterns. Observations within this group demonstrate intermediate energy intensity and moderate environmental investment shares, suggesting ongoing structural adjustment.

Cluster 3 reflects an environmentally modernized structural configuration, characterized by lower GHG intensity, reduced energy intensity, and higher environmental investment allocation. Countries such as Germany and the Netherlands are more frequently associated with this profile in the later years of the observation period, indicating progressive sustainability transformation.

Fuzzy C-means clustering further confirmed the presence of transitional trajectories, particularly for Spain and Poland, suggesting gradual structural evolution rather than abrupt regime shifts. Hierarchical clustering using Ward linkage produced consistent grouping patterns, reinforcing the robustness and stability of the identified ESG structural profiles.

3.2 Regression Analysis of Carbon Intensity

The regression analysis reveals a strong structural association between managerial cost allocation and environmental performance in the EU livestock sector. The estimated model explains a substantial share of variation in GHG intensity, achieving an in-sample goodness-of-fit of $R^2 = 0.9173$.

This high explanatory power reflects the structural interdependence between energy-related cost components and emission intensity at the sectoral level. In particular, energy cost share and operational energy intensity demonstrate systematic positive associations with GHG intensity, indicating that greater dependence on energy inputs translates into higher emission outcomes. Conversely, environmental investment allocation exhibits a mitigating relationship with emission intensity, suggesting that targeted sustainability-oriented expenditures contribute to improved environmental performance.

Country dummy variables capture cross-country heterogeneity in production structures, technological development, and institutional settings. Their inclusion ensures that the identified relationships reflect structural cost-emission linkages rather than purely national differences.

The relationship between observed and predicted values, presented in Figure 2, confirms the consistency of the regression fit across the sample period. The close alignment between actual and predicted GHG intensity values indicates that cost architecture and energy dependence constitute dominant explanatory dimensions within the proposed sustainability framework.

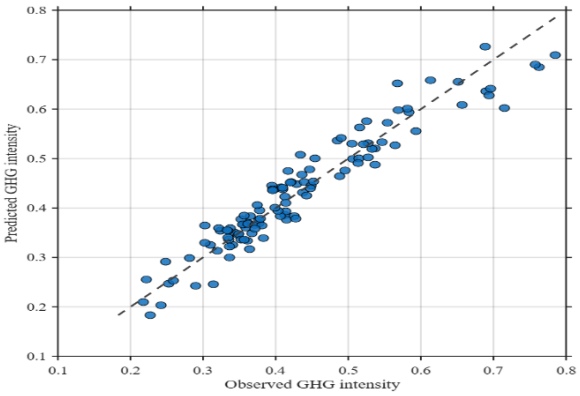


Figure 2: Actual versus predicted GHG intensity values obtained from the regression model ($R^2 = 0.9173$).

It is important to emphasize that the regression stage identifies structural associations rather than strict causal effects. The results demonstrate that variation in emission intensity across countries and years can be systematically explained by managerial allocation patterns, reinforcing the conceptual premise of the study: sustainability differentiation is embedded in cost structure dynamics.

3.3 Random Forest ESG Classification

Random Forest classification was applied to differentiate three sustainability classes (A-C) derived from the k-means clustering stage. Since the regression panel did not contain an externally predefined ESG class label, the clustering output ($k = 3$) was used to construct the target variable. The clusters were subsequently ordered according to sustainability performance, where Class A represents the most sustainable structural configuration and Class C represents the least sustainable profile.

The model was evaluated using out-of-bag (OOB) validation, ensuring internal consistency without requiring external sample partitioning. Permutation-based feature importance analysis was conducted to quantify the relative contribution of each explanatory variable to classification performance. The resulting importance ranking is presented in Figure 3.

The results indicate that energy share and environmental investment allocation consistently emerge as dominant predictors of sustainability class.

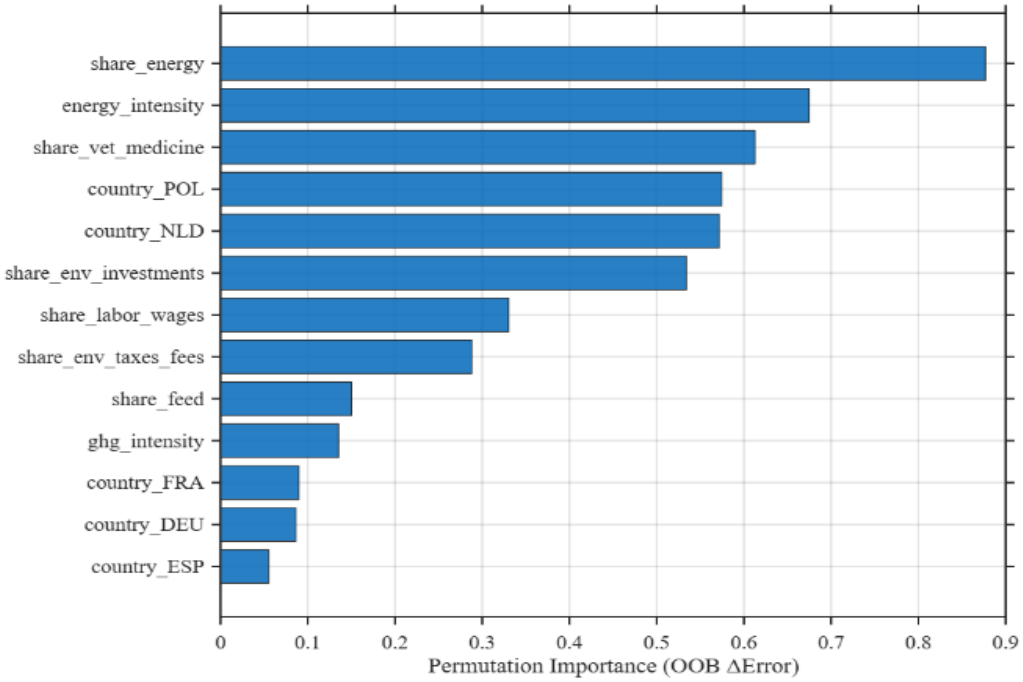


Figure 3: Permutation importance of predictors in the Random Forest ESG classification model.

This finding reinforces the regression results, where energy-related variables explain a substantial share of GHG intensity variation. The convergence between regression coefficients and permutation importance confirms that managerial cost structure is not only associated with emission intensity but also structurally determines sustainability differentiation.

Operational energy intensity further strengthens classification performance, suggesting that efficiency-related indicators act as transmission mechanisms between cost allocation and environmental outcomes. In contrast, country dummy variables demonstrate comparatively lower importance, indicating that sustainability differentiation is primarily driven by internal cost architecture rather than purely national institutional factors.

The alignment between clustering structure, regression associations, and classification importance provides methodological triangulation and strengthens the robustness of the proposed framework. Together, these results support the central argument of the study: ESG sustainability profiles in the EU livestock sector are structurally embedded in managerial allocation patterns and energy dependence dynamics.

4 DISCUSSION

The findings of this study contribute to the growing body of research on ESG-oriented managerial analytics by demonstrating that sustainability differentiation in livestock systems can be structurally derived from cost allocation patterns. Unlike conventional ESG assessments based on composite indices or externally assigned sustainability ratings, the proposed framework embeds sustainability evaluation within sector-level managerial accounting variables.

Although the analysis focuses on five EU member states, the selected sample captures substantial structural diversity within the EU livestock sector. The modelling approach prioritizes interpretability and structural consistency over methodological complexity, ensuring robustness within the constraints of sector-level panel data.

The convergence of clustering, regression, and machine learning classification results strengthens the internal consistency of the modelling approach. Across all analytical stages, energy-related cost components and environmental investment allocation emerge as structurally dominant drivers of sustainability differentiation. This consistency suggests that ESG performance is not an exogenous attribute imposed by institutional conditions, but rather an endogenous outcome of managerial resource allocation decisions.

From a theoretical perspective, the results extend sustainability accounting research by operationalizing the linkage between cost architecture and environmental performance at the sectoral level. The framework moves beyond descriptive sustainability measurement toward structurally interpretable ESG analytics, where cost monitoring functions as an instrument of environmental control. This supports the conceptual transition from traditional cost accounting toward ESG-integrated controlling systems.

From a policy perspective, the findings imply that decarbonization strategies in the livestock sector may be more effective when aligned with internal cost reallocation mechanisms. Targeted incentives for environmental investment and energy efficiency improvements could produce measurable emission reductions by reshaping cost composition rather than relying solely on regulatory constraints.

Nevertheless, several limitations should be acknowledged. The constructed panel relies on macro-level sectoral data, which may not fully capture firm-level managerial heterogeneity. Additionally, the regression framework identifies structural associations rather than causal effects. Future research may extend the model by incorporating micro-level accounting data, dynamic panel techniques, or scenario-based simulation to examine long-term sustainability transitions.

Overall, the study demonstrates that ESG sustainability profiles in the EU livestock sector are structurally embedded in managerial cost allocation and energy dependence dynamics, providing a scalable analytical foundation for sustainability-oriented controlling systems.

5 CONCLUSIONS

This study developed and empirically validated an integrated analytical framework linking managerial cost structure to ESG sustainability differentiation in the EU livestock sector. Using harmonized panel data for five EU countries (2000-2023), the research combined clustering techniques, regression analysis, and Random Forest classification to examine how cost allocation and energy dependence shape sustainability outcomes.

The results demonstrate that GHG intensity is strongly explainable within a cost-based analytical structure, with regression analysis achieving substantial explanatory power ($R^2 = 0.9173$). Machine learning classification further confirms that energy share and environmental investment allocation are dominant determinants of sustainability class differentiation. The consistency across clustering, regression, and classification stages reinforces the structural validity of the proposed framework.

The central contribution of the study lies in demonstrating that ESG sustainability performance in livestock systems is not an externally imposed attribute but a structurally embedded outcome of managerial allocation patterns. By integrating accounting indicators with environmental performance metrics, the research advances ESG-oriented sustainability analytics at the sectoral level.

The proposed framework provides a scalable methodological basis for sustainability benchmarking, strategic investment planning, and ESG-integrated controlling in the agri-food sector. Future research may expand this approach by incorporating firm-level accounting data, dynamic modelling techniques, and broader geographical comparisons to further refine structural sustainability assessment.

ACKNOWLEDGMENTS

The authors acknowledge the academic support provided during the research internship at Humboldt Innovation GmbH, Berlin. The authors also appreciate the institutional support of Saken Seifullin Kazakh Agrotechnical Research University.

REFERENCES

[1] K. R. Lucas and E. Kebreab, “Modeling livestock ecosystem services through carbon footprint,” Ecological Indicators, vol. 181, art. no. 114406, 2025, [Online]. Available: https://doi.org/10.1016/j.ecolind.2025.114406.

[2] X. Kang et al., “Livestock methane emissions in China: Spatiotemporal dynamics and mitigation strategies,” Resources, Conservation and Recycling, vol. 226, art. no. 108695, 2026, [Online]. Available: https://doi.org/10.1016/j.resconrec.2025.108695.

[3] C. Duffy, C. Doedens, R. Moriarty, H. Daly, D. Styles, and M. Meinshausen, “National temperature neutrality, agricultural methane and climate policy: Reinforcing inequality in the global food system,” Environmental Research Letters, vol. 20, no. 9, art. no. 094019, 2025, [Online]. Available: https://doi.org/10.1088/1748-9326/adf12d.

[4] F. Zahm et al., “Assessing farm sustainability: The IDEA4 method, a conceptual framework combining dimensions and properties of sustainability,” Cahiers Agricultures, vol. 33, art. no. 10, 2024, [Online]. Available: https://doi.org/10.1051/cagri/2024001.

[5] T. Deng and Z. F. Mohamad, "Context-sensitive agricultural sustainability assessment: A systematic review of frameworks and local adaptation criteria," Agricultural Systems, vol. 231, art. no. 104532, 2026, [Online]. Available: https://doi.org/10.1016/j.agsy.2025.104532.

[6] S. R. Ndemewah, K. Menges, and M. R. Hiebl, "Management accounting research on farms: What is known and what needs knowing?," Journal of Accounting & Organizational Change, vol. 15, no. 1, pp. 58-86, 2019, [Online]. Available: https://doi.org/10.1108/JAOC-05-2018-0044.

[7] G. Sariyer, M. Sarioglu, and B. Ramtiyal, "Machine learning-driven clustering based environmental, social, and governance performance prediction model," in The Unified International Conference on Emerging Technologies in Cyber-Physical Systems and Industrial AI, Cham: Springer Nature Switzerland, 2024, pp. 661-675, [Online]. Available: https://doi.org/10.1007/978-3-031-95963-9_46.

[8] J. Zhang and Z. Zhao, "Corporate ESG rating prediction based on XGBoost-SHAP interpretable machine learning model," Expert Systems with Applications, vol. 295, art. no. 128809, 2026, [Online]. Available: https://doi.org/10.1016/j.eswa.2025.128809.

[9] A. Yenkikar, V. P. Mishra, M. Bali, and T. Ara, "An explainable AI-based hybrid machine learning model for interpretability and enhanced crop yield prediction," MethodsX, art. no. 103442, 2025, [Online]. Available: https://doi.org/10.1016/j.mex.2025.103442.

[10] Y. W. Li, "Spatiotemporal investigation into water, land, energy, and carbon footprints of broiler and correlations among them in China," Discover Sustainability, vol. 6, no. 1, art. no. 794, 2025, [Online]. Available: https://doi.org/10.1007/s43621-025-01760-2.

Digital Transformation of Customer Relationship Management Systems in Logistics Systems: A Conceptual Model

Iulia Caprian[1], Viktoriia Khaustova[2], Nataliia Trushkina[2] and Kateryna Shulakova[3,4]
[1]*Moldova Cooperative-Commerce University, Bulevardul Iuri Gagarin Str. 8, 2001 Chișinău, Moldova*
[2]*Research Center for Industrial Problems of Development of the National Academy of Sciences of Ukraine, Inzhenernyi Lane 1a, 61165 Kharkiv, Ukraine*
[3]*State University of Intelligent Technologies and Telecommunications, Kuznechna Str. 1, 65023 Odesa, Ukraine*
[4]*Anhalt University of Applied Sciences, Bernburger Str. 57, 06366 Köthen, Germany*
kaprianusm@gmail.com, v.khaust@gmail.com, nata_tru@ukr.net, katejojo29@gmail.com

Keywords: Customer Relationship Management, CRM Systems, Digital Transformation, Logistics System, Digital Technologies, Artificial Intelligence, Big Data Analytics, Logistics Management, Digital Platform, Customer Experience, Digital Marketing, Supply Chain Management, Sustainable Development.

Abstract: In the context of the digital transformation of the economy, customer relationship management systems play a key role in improving business process efficiency and enhancing the competitive advantages of enterprises. Their application is especially important in logistics systems, where effective customer relationship management directly affects service quality, order processing speed, and customer satisfaction. The purpose of the article is to examine the trends in the digital transformation of customer relationship management systems in logistics systems and to substantiate a conceptual model of their integration into the digital business ecosystem of an enterprise. The methodological framework is based on general scientific and specific research methods, including analysis and synthesis, system analysis, generalization, comparative analysis, and the use of statistical and analytical data from international organizations and research agencies. The study identifies current trends in the global customer relationship management market and determines key areas of their application in logistics companies. It is shown that the implementation of such systems enhances customer data management, supports personalized communication, optimizes order processing, and improves logistics service quality. The role of digital technologies, including artificial intelligence, big data analytics, cloud services, and digital platforms, is substantiated. The scientific novelty lies in the development of a conceptual model integrating digital transformation drivers, technologies, customer relationship management systems, logistics processes, and business outcomes. The proposed model has practical value for designing digital transformation strategies in logistics enterprises.

1 INTRODUCTION

In the contemporary conditions of the formation of the digital economy, a profound transformation of enterprise business processes is taking place, accompanied by the active implementation of digital technologies [1], [2], platform solutions, and data analytics systems. One of the key directions of these changes is the digital transformation of customer relationship management systems (Customer Relationship Management, CRM) [3], [4], which are gradually evolving from tools for recording customer information into comprehensive digital platforms for managing customer experience and are becoming an important factor in enhancing the competitiveness of enterprises.

The global market for customer relationship management systems demonstrates a stable growth trend. According to the analytical company Statista [5], in 2023 its volume amounted to approximately 79 billion dollars, whereas in 2017 it was estimated at about 36 billion dollars, indicating more than a twofold increase over six years. According to forecasts, by 2030 the market volume may exceed 145 billion dollars [5], which confirms the strategic importance of customer relationship management technologies in the modern business environment.

The growing role of customer relationship management systems is largely driven by the transformation of approaches to the formation of competitive advantages, in which customer experience is becoming increasingly central. In particular, according to the results of the PwC Global Consumer Insights Survey [6], 73 percent of consumers identify customer experience as a key factor in purchasing decisions, while 65 percent are willing to abandon a brand after a single negative interaction experience. At the same time, research by Salesforce [7] shows that 88 percent of customers expect personalized interaction that takes into account their individual needs and behavioral characteristics.

An important driver of the development of customer relationship management systems is the global process of digital transformation of the economy. According to estimates by International Data Corporation (IDC) [8], global spending on digital transformation in 2023 exceeded 2.3 trillion dollars and may reach 3.4 trillion dollars by 2026. A significant share of these investments is directed toward the development of digital platforms for managing customer data, analytical systems, cloud services, and artificial intelligence technologies, which creates the preconditions for the qualitative modernization of customer relationship management systems.

At the same time, the role of digital channels of interaction with customers is increasing. According to Gartner [9], by 2025 more than 80 percent of interactions between companies and customers will be conducted through digital channels, including mobile applications, social networks, chatbots, and electronic service platforms. The use of artificial intelligence technologies and big data analytics enables enterprises not only to analyze customer behavior but also to forecast demand, personalize offerings, and improve the effectiveness of managerial decision-making.

Digital customer relationship management systems are of particular importance in enterprise logistics systems. The growth of electronic commerce and digital trading platforms, according to the United Nations Conference on Trade and Development [10], necessitates the integration of logistics processes with customer experience management systems. In particular, the global volume of electronic commerce in 2022 exceeded 26.7 trillion dollars, which is almost 60 percent higher compared to 2016 [11]. Under such conditions, delivery speed, reliability of logistics operations, and service quality become decisive factors in meeting customer needs.

The efficiency of logistics systems at the macro level is reflected in the Logistics Performance Index, which is calculated by the World Bank [12]. According to the 2023 report, the highest index values are observed in Singapore (4.3), Finland (4.2), and Denmark (4.1), which indicates a high level of digitalization and integration of logistics processes. At the same time, for Ukraine this indicator is 2.7, which indicates significant potential for improving the efficiency of logistics systems, particularly through the implementation of digital solutions and the development of customer relationship management technologies [12].

Thus, the development of digital technologies, the rapid growth of electronic commerce, the increasing importance of customer experience, and the integration of logistics and marketing processes create an objective need to reconsider approaches to customer relationship management. Under these conditions, the integration of customer relationship management systems with enterprise logistics processes becomes especially important, forming the basis for the development of new conceptual models of customer relationship management in the digital economy.

2 LITERATURE REVIEW

In scientific literature, the development of customer relationship management systems is examined in the context of the transformation of enterprise business models under the influence of digital technologies. Researchers emphasize that customer relationship management systems are gradually evolving from tools for automating marketing and sales processes into comprehensive digital platforms for managing customer experience, which integrate data analytics, digital communication channels, and business process management systems [13].

A. Payne and P. Frow [14] consider customer relationship management as a strategic management concept aimed at building long-term relationships with customers through the integration of marketing, information, and managerial processes of an enterprise. The authors emphasize that the effective functioning of customer relationship management systems requires the use of modern information technologies, customer data analysis systems, and tools for the personalization of customer interaction [14].

Further development of research in this field is associated with the study of the impact of digital transformation on customer relationship management. In modern conditions, digital technologies, including artificial intelligence, big data, cloud computing, and digital platforms, significantly expand the capabilities of customer relationship management systems in terms of analyzing customer behavior, forecasting demand, and forming personalized offers [15]. In particular, studies by international analytical agencies indicate that the use of big data analytics technologies in customer relationship management systems makes it possible to increase the accuracy of consumer demand forecasting by 30 to 40 percent and significantly improve the effectiveness of marketing communications [15].

Considerable attention in contemporary research is also paid to customer experience management, which is regarded as one of the key factors in forming competitive advantages of enterprises in the digital economy. According to Salesforce research [7], the majority of customers expect personalized interaction from companies based on the analysis of their individual needs and behavioral characteristics. In this regard, customer relationship management systems are increasingly integrated with digital communication channels, including social networks, mobile applications, and electronic customer service platforms.

A separate direction of scientific research is related to the study of the role of customer relationship management systems in logistics and transport systems. Under the conditions of digitalization of the economy, the efficiency of logistics processes largely depends on the level of integration of information systems for supply chain management with customer relationship management systems [16]-[20]. Digital platforms for managing logistics operations enable enterprises to ensure transparency of delivery processes, improve the quality of logistics services, and promptly respond to changes in customer needs.

At the same time, the results of studies by the World Bank indicate that the efficiency of logistics systems largely depends on the level of digitalization of infrastructure and the use of information technologies in the management of logistics operations [12]. Increasing the level of digital integration of logistics systems and customer relationship management platforms contributes to the optimization of interaction processes between enterprises and customers, reduction of delivery time, and improvement of customer satisfaction.

At the same time, the analysis of contemporary scientific publications shows that a significant part of research is focused mainly on individual aspects of the development of customer relationship management systems, in particular on issues of digital marketing, customer experience management, or business process automation. The issues of integrating customer relationship management systems with enterprise logistics processes, as well as the formation of comprehensive conceptual models of the digital transformation of customer relationship management systems in logistics systems, remain insufficiently studied.

3 METHODOLOGY

In the context of contemporary processes of digital transformation of the economy, there is a need to reconsider traditional approaches to customer relationship management and the organization of logistics processes of enterprises. Under conditions of rapid development of digital technologies, the expansion of electronic commerce, and the growing importance of customer experience, customer relationship management systems are becoming a key element of the digital infrastructure of enterprises. This requires the application of a comprehensive methodological approach that makes it possible to study the interconnections between technological, organizational, and managerial aspects of enterprise functioning in the digital environment.

The purpose of the study is to substantiate the theoretical foundations and to develop a conceptual model of the digital transformation of customer relationship management systems in logistics systems, taking into account current trends in the development of digital technologies, customer-oriented business models, and the digitalization of logistics processes.

The methodological basis of the study is formed by a combination of general scientific and specific methods of scientific inquiry, which allow for a comprehensive examination of the processes of digital transformation of customer relationship management systems in enterprise logistics systems. The study is based on a system approach, which involves considering customer relationship management systems as a complex, multi-level system of interrelated elements that integrates marketing, information, and logistics processes of the enterprise. The application of the system approach makes it possible to analyze the interaction

of digital technologies, customer data, communication channels, and logistics operations within a unified digital environment.

To generalize contemporary scientific approaches to the digital transformation of customer relationship management systems, methods of analysis and synthesis were used, which made it possible to systematize the results of scientific research in the fields of customer relationship management, digital marketing, customer experience management, and the digitalization of logistics systems. The use of these methods contributed to the formation of the theoretical basis of the study and the identification of key factors of the digital transformation of customer relationship management systems.

To study current trends in the development of digital technologies and customer relationship management systems, comparative and statistical analysis methods were applied, based on the generalization of international statistical data, analytical reports, and global indices of digital development. This approach makes it possible to identify the main patterns in the development of digital technologies, electronic commerce, and digital channels of customer interaction, as well as to assess their impact on the transformation of logistics processes and customer relationship management systems.

An important element of the research methodology is the method of conceptual modelling, which was used to develop a conceptual model of the digital transformation of customer relationship management systems in logistics systems. The application of this method makes it possible to integrate the results of theoretical analysis, empirical data, and contemporary approaches to digital business process management into a coherent conceptual framework that reflects the relationships between digital technologies, customer analytics, customer experience management systems, and enterprise logistics processes.

For the visual representation of the research results, a graphical method was used, which makes it possible to illustrate the structure of the conceptual model and the interconnections between the key elements of the digital transformation of customer relationship management systems.

The information base of the study consists of statistical data from international organizations and analytical agencies, including the World Bank, the United Nations Conference on Trade and Development, International Data Corporation, Gartner, Salesforce, and PwC. The use of these sources makes it possible to obtain reliable data on the development of the digital economy, the spread of digital technologies, and the transformation of logistics systems in the global economic environment.

At the same time, the conducted study has certain methodological limitations that should be taken into account when interpreting the obtained results. First, the statistical data of international organizations used in the study are aggregated in nature and reflect general trends in the development of digital technologies and logistics systems, which may limit the possibilities for a detailed analysis of the specific features of the functioning of customer relationship management systems at the level of individual enterprises. Second, the proposed conceptual model has a generalized nature and is primarily aimed at reflecting the key relationships between the main elements of customer relationship management systems in the digital environment. Third, the study is based mainly on secondary statistical data and analytical reports of international organizations, which limits the ability to take into account sectoral and regional features of the implementation of customer relationship management systems in the activities of specific enterprises.

Taking into account the outlined methodological limitations, the results of the study provide a theoretical generalization of the key patterns of the digital transformation of customer relationship management systems and form an analytical basis for the further development of conceptual and applied approaches to their integration into enterprise logistics systems.

4 RESULTS

4.1 Trends in the Digital Transformation of CRM Systems and Logistics

The digital transformation of the economy is radically changing approaches to customer relationship management and the organization of enterprise logistics processes. Under current conditions, the boundaries between marketing, logistics, and information technologies are gradually disappearing, which contributes to the formation of integrated digital management systems [21], [22]. The key drivers of these changes include the rapid growth of e-commerce, the spread of digital

platforms, the development of artificial intelligence and Big Data technologies, and the growing role of customer-oriented business models. As a result, CRM systems are evolving from customer accounting tools into comprehensive analytical platforms integrated with the enterprise's logistics and marketing processes.

According to international analytical agencies, the global CRM market demonstrates stable positive dynamics [8], [23]. This is confirmed by the data presented in *Table 1*, which show not only the steady expansion of the market but also the strengthening strategic role of CRM solutions in business digitalization.

As shown in Table 1 [8], [23], during 2017-2024 the global CRM market more than doubled, indicating a high level of demand for digital customer relationship management solutions. The projected acceleration of growth in the forecast period is associated with the active implementation of artificial intelligence, process automation, and predictive analytics. Therefore, CRM systems increasingly act as the core of enterprise digital transformation, ensuring the integration of data, analytics, and managerial decision-making.

Table 1: Dynamics of the global CRM systems market.

Year	Market volume, *USD billion*	Growth rate, *%*
2017	36	–
2019	48	33.3
2021	63	31.3
2024	73	15.9
2028*	131	79.5
2030*	163	24.4

Note: projected values of the indicators.

The deepening digitalization of business is accompanied by the wide adoption of CRM technologies across various sectors of the economy. According to international studies, CRM systems are used by more than 90% of companies with more than 10 employees [7]. This indicates that CRM platforms have already become one of the basic elements of the digital infrastructure of modern enterprises.

The scale of CRM adoption by companies of different sizes is summarized in Table 2 [7], [9]. The data show that the degree of CRM integration depends significantly on the scale of business activity, resource availability, and organizational maturity.

As can be seen from Table 2, CRM systems are most actively used by large enterprises, where they are embedded into key business processes and serve as an instrument for coordinating marketing, service, and sales activities. At the same time, adoption in the small business segment remains lower due to financial, technological, and organizational barriers. Nevertheless, the gradual diffusion of cloud-based solutions and modular CRM platforms creates conditions for expanding their use among small and medium-sized enterprises. Consequently, the further development of CRM systems is associated not only with technological sophistication but also with increasing accessibility and adaptability for businesses of different sizes.

The evolution of CRM platforms is directly determined by the introduction of modern digital technologies that expand their functional capabilities and analytical potential. In contemporary business practice, CRM systems no longer perform merely operational functions related to storing customer data. They are transforming into complex digital environments capable of collecting, processing, interpreting, and using information in real time.

The main technologies that shape the digital transformation of CRM systems are presented in Table 3 [9], [21], [22].

Table 2: Level of CRM systems adoption in companies.

Indicator	Value	Interpretation
Share of companies using CRM	91%	CRM has become a standard tool of digital management
Share of small businesses	~50%	Resource constraints hinder implementation
Share of medium-sized businesses	70–85%	Active digitalization
Share of large companies	>90%	High level of CRM integration
Average CRM market growth rate	10–14%	Stable development of the industry

As shown in Table 3, the use of digital technologies ensures a transition from reactive to proactive customer relationship management. In particular, artificial intelligence technologies make it possible to forecast customer behavior, identify hidden demand patterns, and automate communication scenarios. Big Data technologies, in turn, provide a deeper analysis of customer needs, preferences, and interactions across different channels. Cloud computing enhances system flexibility and scalability, while IoT and digital platforms contribute to integrating customer data with logistics and operational processes. Thus, CRM

systems are increasingly turning into management decision-support platforms that connect marketing, logistics, and service processes within a single digital architecture.

Table 3: Key technologies of CRM digital transformation.

Technology	Main functions	Impact on CRM
Artificial Intelligence	prediction of customer behavior	personalization and automation
Big Data	processing of large data sets	improved analytical accuracy
Cloud computing	cloud infrastructure	scalability and accessibility
Internet of Things	integration of physical objects	expansion of customer-related data
Digital platforms	omnichannel communication	integration of interaction channels

Alongside the development of CRM systems, logistics processes are also undergoing intensive digital transformation. In the modern economy, logistics efficiency is becoming one of the key determinants of enterprise competitiveness, as it affects delivery speed, service quality, cost levels, and the transparency of supply chains. For this reason, the digitalization of logistics should be considered not as a separate process but as a structural component of the overall transformation of business management systems.

To assess the effectiveness of logistics systems, the Logistics Performance Index (LPI), calculated by the World Bank, is widely used [12]. The leading countries according to this indicator are presented in Table 4 [12].

The analysis of Table 4 shows that the leaders are countries with highly developed digital infrastructure and a high level of integration of information systems into logistics activities. Their high LPI values result not only from the quality of transport networks or customs procedures but also from the use of digital platforms, data exchange systems, real-time tracking tools, and analytical instruments for supply chain coordination. This confirms that digitalization is one of the decisive factors in improving logistics efficiency and in ensuring the adaptability of enterprises to changes in customer demand and market conditions.

The parallel development of CRM systems and logistics solutions creates objective preconditions for their integration into a unified digital ecosystem of business process management. Such integration allows enterprises to synchronize customer information with operational logistics processes, thereby improving service quality, increasing forecast accuracy, and optimizing delivery management.

Table 4: Leading countries by the Logistics Performance Index (LPI, 2023).

Country	LPI value	Characteristic
Singapore	4.3	high level of logistics digitalization
Finland	4.2	integration of digital services
Denmark	4.1	efficient infrastructure
Germany	4.1	developed logistics networks
Netherlands	4.1	digital logistics platforms

The main effects of integrating CRM and logistics systems are summarized in Table 5 [1]-[4].

Table 5: Effects of CRM and logistics systems integration.

Integration direction	Effect	Business result
Customer data	personalization	increased loyalty
Delivery management	optimization	cost reduction
Demand analytics	forecasting	reduced risks
Communications	omnichannelity	improved service quality

As can be seen from Table 5, the integration of CRM and logistics provides a synergistic effect manifested in enhanced business process efficiency, improved customer experience, and strengthened enterprise competitiveness. The integration of customer data enables more accurate personalization of service, while the connection between CRM and delivery management systems helps to optimize logistics routes, reduce delays, and improve order fulfillment quality. In turn, demand analytics supports more accurate forecasting and lowers the risks associated with inventory imbalances or supply chain disruptions. Thus, the integration of CRM systems and logistics platforms should be regarded as one of the strategic directions of enterprise digital transformation.

4.2 Conceptual and Ecosystem Models of CRM Digital Transformation in Logistics Systems

The generalization of the results presented in Tables 1-5 provides grounds for developing the author's conceptual model of CRM systems digital transformation in logistics systems, shown in Figure 1. The proposed model reflects the systemic logic of CRM transformation as a multi-level process that combines the digital environment, digital technologies, CRM mechanisms, logistics processes, and business outcomes within a unified management system. Unlike traditional approaches, where CRM is mainly viewed as a tool for sales support or marketing automation, the proposed model considers CRM as the integration core of the enterprise's digital business architecture.

As shown in Figure 1, the transformation process begins with the influence of the external digital environment, including the development of the digital economy, the rapid growth of e-commerce, market globalization, and technological innovation. The next level is represented by digital technologies such as artificial intelligence, Big Data analytics, cloud computing, the Internet of Things, and digital platforms, which create the technological basis for integrated digital management.

The central place in the model is occupied by the CRM system, which integrates customer data, analytics, marketing processes, and communication channels into a unified management environment. The logistics level includes order management, supply chain coordination, delivery optimization, and logistics service management. The final level reflects business outcomes, including improved customer experience, operational efficiency, competitiveness, and business growth. Thus, Figure 1 demonstrates that the effect of digital transformation is achieved through the integration of digital technologies, CRM systems, and logistics processes within a coherent business architecture.

While Figure 1 presents the hierarchical logic of transformation, Figure 2 reflects the ecosystem dimension of digital customer relationship management in logistics systems. The ecosystem approach makes it possible to reveal the interconnections among digital platforms, CRM systems, logistics operations, marketing activities, customer experience, and the analytical data layer.

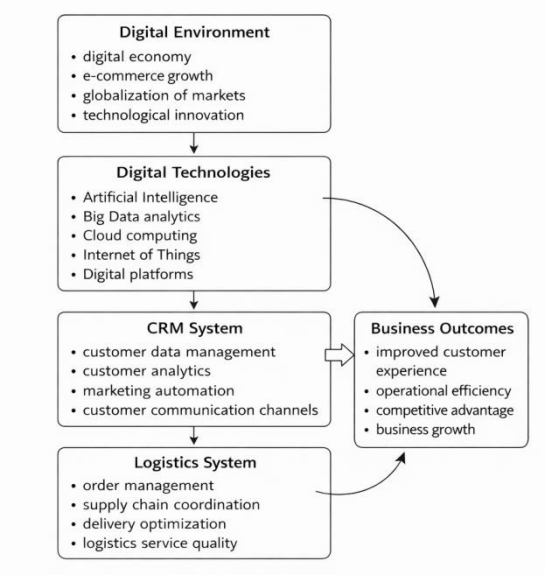


Figure 1: Conceptual model of CRM systems digital transformation in logistics systems.

As illustrated in Figure 2, digital platforms form the technological environment of the ecosystem and include cloud infrastructure, omnichannel communication tools, e-commerce solutions, and social media integration mechanisms. They ensure continuous interaction among customer data, operational information, and marketing processes.

Within this ecosystem, CRM performs the role of the central coordination hub by integrating customer information, analytics, communication channels, and decision-making processes. Through CRM, customer data are connected with logistics execution, marketing activities, and customer service processes.

The logistics subsystem covers order tracking, delivery logistics, supply chain visibility, and transport management, while the marketing subsystem includes campaign management, personalized offers, customer segmentation, and communication activities. Their integration within the ecosystem supports coordinated customer service and logistics performance.

An important component of the ecosystem is the Data and Analytics Layer, which supports data processing, customer segmentation, predictive analytics, and performance measurement. This analytical layer integrates all subsystems at the data level and provides adaptability and feedback within the digital ecosystem.

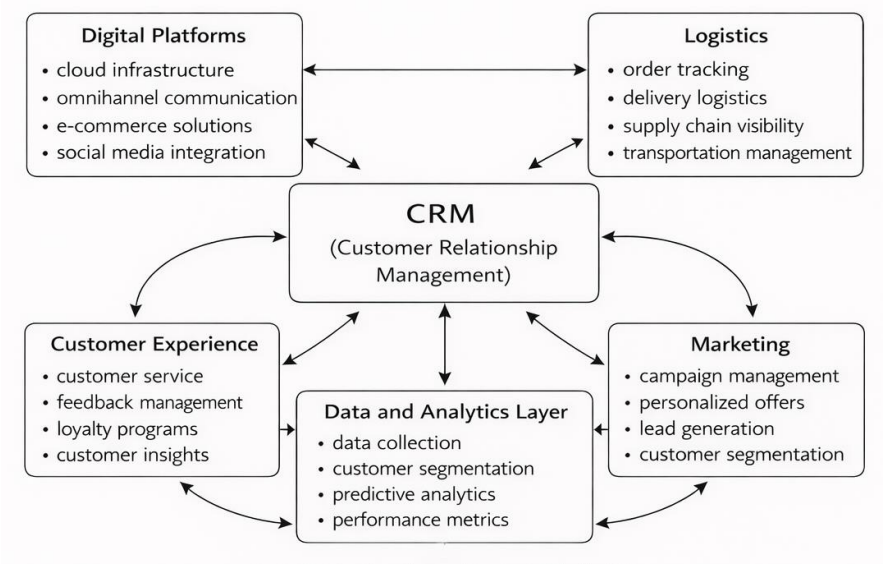


Figure 2: Ecosystem of CRM digital transformation in logistics systems.

In the proposed model, customer experience is interpreted as an integrated outcome of the coordinated interaction among digital platforms, CRM systems, logistics processes, marketing activities, and analytics. Therefore, Figure 2 supplements Figure 1 by demonstrating the internal structure of the digital ecosystem and the interconnected nature of customer-oriented logistics management.

4.3 Mini-cases of CRM Integration in Logistics Companies

4.3.1 Mini-Case 1. Amazon

Amazon's digital ecosystem shows a high level of integration between CRM technologies and logistics processes. The company actively uses artificial intelligence and predictive analytics tools to personalize customer interaction, forecast demand, and improve delivery operations. Its CRM system is connected with order tracking services, recommendation algorithms, and customer support channels, which helps coordinate customer data with logistics activities. This approach improves service speed, delivery accuracy, and customer satisfaction.

4.3.2 Mini-Case 2. DHL

DHL applies a digital approach to logistics management through the integration of CRM platforms with shipment monitoring systems and digital communication channels. The use of data analytics and real-time technologies allows the company to track delivery status, coordinate logistics operations, and provide more personalized communication with customers. The integration of CRM with logistics infrastructure increases service transparency and improves customer experience management.

4.3.2 Mini-Case 3. FedEx

FedEx uses CRM solutions as part of the digital coordination of transport and logistics operations and customer service processes. The company implements omnichannel communication tools, digital shipment tracking platforms, and analytical systems for forecasting customer needs. The combination of CRM technologies with logistics platforms helps optimize delivery routes, reduce response time to customer requests, and improve the flexibility of logistics operations.

The presented mini-cases confirm the practical importance of integrating CRM systems with logistics processes and demonstrate that the use of digital platforms, data analytics, and artificial intelligence technologies creates the basis for the development of customer-oriented logistics management models.

5 DISCUSSION, LIMITATIONS AND FUTURE RESEARCH

The results of the study confirm that the transformation of customer relationship management systems in logistics goes beyond the technological modernization of separate business processes. In today's digital environment, CRM systems are gradually becoming a coordination center that integrates customer data, communication channels, analytics, service processes, and logistics operations. This approach creates the conditions for moving from fragmented customer interaction management to a more flexible model of coordination between marketing, service, and logistics decisions.

The obtained results are consistent with the approach of A. Payne and P. Frow [14], who consider CRM as a strategic concept of long-term customer relationship management rather than only a tool for sales automation or marketing communications. At the same time, this study develops this approach further by explaining the role of CRM in connecting customer analytics with logistics processes, including order management, supply coordination, demand forecasting, and supply chain transparency.

The ideas of F. Buttle and S. Maklan [13], who define CRM as a multi-component system integrating information technologies, analytical tools, and management processes, are also supported by the results of the study. At the same time, the proposed conceptual model allows CRM to be viewed more broadly as a system-forming element of the enterprise digital business ecosystem. In this ecosystem, CRM supports interaction between digital platforms, marketing channels, analytical modules, and logistics infrastructure.

This integration becomes especially important in logistics systems, where the level of customer service depends not only on communication quality, but also on order processing speed, forecasting accuracy, delivery transparency, and the ability to respond quickly to changing customer needs. In this context, CRM systems perform a coordination function by connecting customer analytics with logistics operations.

The theoretical contribution of the study lies in the development of the conceptual understanding of CRM as a tool for integrating digital, marketing, and logistics processes within an enterprise. Unlike approaches that mainly examine CRM within marketing or information technology frameworks, this study focuses on its role in creating an integrated digital environment for customer value management. The proposed model systematizes the relationships between digital technologies, data flows, logistics operations, and business results. This expands the scientific understanding of enterprise digital transformation mechanisms.

At the same time, the study has several limitations. First, it is mainly based on international statistical data, analytical reports, and scientific sources. This made it possible to identify global trends in the development of CRM systems and digital logistics solutions, but limited the opportunities for a deeper analysis of sectoral and regional features of their practical implementation. In addition, the proposed model has a conceptual character and does not include a quantitative assessment of CRM integration with logistics platforms at the level of individual enterprises.

Another limitation is that some behavioral, organizational, and institutional aspects of digital transformation were outside the scope of the study. In particular, further attention should be paid to staff readiness for change, the level of digital skills, data compatibility, cybersecurity, personal data protection, and regulatory requirements related to customer data use. In logistics management practice, these factors may strongly influence the effectiveness of digital solutions.

Future research should focus on the empirical testing of the proposed conceptual model using examples from enterprises in different sectors of the economy. Special attention should be given to assessing the impact of CRM systems on logistics efficiency, customer service quality, demand forecasting accuracy, supply chain adaptability, and the ability of enterprises to respond quickly to market changes.

Further studies may also focus on the development of the ecosystem approach to customer relationship management. In particular, important areas include the creation of integrated analytical platforms, the assessment of interaction between CRM and logistics systems, and the development of tools for measuring the effectiveness of digital coordination of business processes. Additional attention should also be paid to the role of artificial intelligence, predictive analytics, digital leadership, and change management in the development of next-generation CRM systems.

6 CONCLUSIONS

The study confirmed the growing role of CRM systems in coordinating customer, service, and

logistics processes within digitally transformed enterprises. It was found that under the conditions of economic digitalization, CRM systems are evolving from tools for supporting marketing communications into integrated management platforms that combine customer analytics, communication channels, service functions, and logistics operations within a unified digital environment. The integration of CRM platforms with logistics systems contributes to more effective customer data management, optimization of order processing, improvement of logistics service quality, and greater adaptability of enterprises to changes in the market environment.

The analysis of current trends in the global CRM market confirmed the increasing importance of digital technologies in the development of customer-oriented management models. The study showed that the use of artificial intelligence, Big Data analytics, cloud services, IoT technologies, and digital platforms expands the functional capabilities of CRM systems and creates the technological basis for integrating marketing, analytical, and logistics processes. At the same time, the growing level of CRM adoption among enterprises demonstrates that CRM systems are becoming one of the key elements of modern digital business infrastructure.

The scientific novelty of the study lies in the development of a conceptual model of CRM systems digital transformation in logistics systems and in the substantiation of the ecosystem approach to customer relationship management. The proposed model reflects the relationships among the drivers of digital transformation, digital technologies, CRM platforms, logistics processes, and business outcomes. Unlike traditional approaches, the model is based on the integration of customer analytics, digital communications, logistics management, and data-driven decision-making within a unified digital business ecosystem.

The practical significance of the obtained results lies in the possibility of using the proposed model for developing digital transformation strategies for logistics enterprises, improving customer relationship management, enhancing the coordination of logistics operations, and developing personalized services. The implementation of integrated CRM solutions contributes to greater transparency of business processes, faster managerial decision-making, improved logistics coordination, and higher customer service quality. In addition, the proposed approach may be useful for enterprises seeking to strengthen their competitiveness and increase the flexibility of logistics systems under conditions of rapid market change.

Future research should focus on the empirical assessment of CRM integration effectiveness in logistics systems, as well as on studying the impact of artificial intelligence, predictive analytics, digital ecosystems, and data-driven management technologies on the development of customer-oriented logistics management models. Further studies may also examine organizational and behavioral aspects of CRM implementation, including digital leadership, change management, staff readiness for digital transformation, and data security issues in logistics systems.

REFERENCES

[1] V. Khaustova, M. Kyzym, N. Trushkina, and M. Khaustov, “Digital transformation of energy infrastructure in the conditions of global changes: bibliometric analysis”, Proceedings of the 12th International Conference on Applied Innovations in IT, vol. 12(1), pp. 135-142, 2024, [Online]. Available: http://dx.doi.org/10.25673/115664.

[2] A. Zaporozhets, V. Khaustova, I. Hubarieva, and N. Trushkina, “Digitization as a Modern Challenge for the Energy Systems’ Transformation in the World”, in Systems, Decision and Control in Energy VII. Studies in Systems, Decision and Control, vol. 596, pp. 3-31, Springer Science and Business Media Deutschland GmbH, Berlin, 2025, [Online]. Available: https://doi.org/10.1007/978-3-031-90462-2_1.

[3] A. Kwilinski, L. Hnatyshyn, O. Prokopyshyn, and N. Trushkina, “Managing the logistic activities of agricultural enterprises under conditions of digital economy”, Virtual Economics, vol. 5(2), pp. 43-70, [Online]. Available: https://doi.org/10.34021/ve.2022.05.02(3).

[4] A. Kwilinski, N. Trushkina, Yu. Remyha, and T. Patlachuk, “Impact of artificial intelligence on customer relationship management: A bibliometric analysis”, Proceedings of International Conference on Applied Innovation in IT, vol. 13(1), pp. 295-303, [Online]. Available: http://dx.doi.org/10.25673/119246.

[5] Statista, “Customer Relationship Management Software Market Size Worldwide from 2017 to 2030”, 2024, [Online]. Available: https://www.statista.com/statistics/1024763/crm-market-size-worldwide/.

[6] Global Consumer Insights Survey 2023, PricewaterhouseCoopers, London, 2023, [Online]. Available: https://www.pwc.com/gx/en/industries/consumer-markets/consumer-insights-survey.html.

[7] State of the Connected Customer, 5th ed., Salesforce Research, San Francisco, 2023, [Online]. Available: https://www.salesforce.com/resources/research-reports/state-of-the-connected-customer/.

[8] Worldwide Digital Transformation Spending Guide, International Data Corporation (IDC), Framingham, 2023, [Online]. Available: https://www.idc.com/getdoc.jsp?containerId=IDC_P33198.

[9] Future of Customer Service and Support: Forecast and Trends, Gartner Research, Stamford, 2023, [Online]. Available: https://www.gartner.com/en/customer-service-support.

[10] Digital Economy Report 2021: Cross-border data flows and development: For whom the data flow, United Nations Conference on Trade and Development, Geneva, 2021, [Online]. Available: https://unctad.org/system/files/official-document/der2021_en.pdf.

[11] Digital Trade for Development, The International Monetary Fund, the Organisation for Economic Co-operation and Development, the United Nations, The World Bank, and the World Trade Organization, Geneva, 2023, [Online]. Available: https://unctad.org/system/files/official-document/dtd2023_e.pdf.

[12] J.-F. Arvis, L. Ojala, B. Shepherd, D. Ulybina, and C. Wiederer, "Connecting to Compete 2023: Trade Logistics in an Uncertain Global Economy - The Logistics Performance Index and Its Indicators", World Bank, Washington, DC, 2023, [Online]. Available: https://doi.org/10.1596/39760.

[13] F. Buttle and S. Maklan, Customer Relationship Management: Concepts and Technologies, 4th ed., Routledge, London, 2019.

[14] A. Payne and P. Frow, Strategic Customer Management: Integrating Relationship Marketing and CRM, Cambridge University Press, Cambridge, 2013.

[15] S. Chatterjee et al., "Assessing Organizational Users' Intentions and Behavior to AI-Integrated CRM Systems: a Meta-UTAUT Approach", Information Systems Frontiers, vol. 25, pp. 1299-1313, 2023, [Online]. Available: https://doi.org/10.1007/s10796-021-10181-1.

[16] T. H. Davenport and J. Foutty, "AI-Driven Leadership", MIT Sloan Management Review, Aug. 10, 2018, [Online]. Available: https://sloanreview.mit.edu/article/ai-driven-leadership/.

[17] R. Abduljabbar et al., "Applications of artificial intelligence in transport: An overview", Sustainability (Switzerland), vol. 11, iss. 12, 189, 2019, [Online]. Available: https://doi.org/10.3390/su11010189.

[18] S. Chatterjee and R. Chaudhuri, "Customer Relationship Management in the Digital Era of Artificial Intelligence", in EAI/Springer Innovations in Communication and Computing, Berlin: Springer Science and Business Media Deutschland GmbH, pp. 175-190, 2023, [Online]. Available: https://doi.org/10.1007/978-3-031-19711-6_8.

[19] V. Kumar et al., "Understanding the role of artificial intelligence in personalized engagement marketing", California Management Review, vol. 61, iss. 4, pp. 135-155, 2019, [Online]. Available: https://doi.org/10.1177/0008125619859317.

[20] D. Ozay et al., "Artificial Intelligence (AI)-based Customer Relationship Management (CRM): a comprehensive bibliometric and systematic literature review with outlook on future research", Enterprise Information Systems, vol. 18(7), 2351869, 2024, [Online]. Available: https://doi.org/10.1080/17517575.2024.2351869.

[21] V. Kumar and W. Reinartz, Customer Relationship Management: Concept, Strategy, and Tools, 3rd ed., Springer, Berlin, 2018.

[22] D. Chaffey and F. Ellis-Chadwick, Digital Marketing: Strategy, Implementation and Practice, 8th ed., Pearson Education Limited, Harlow, 2022.

[23] Customer Relationship Management Market Size, Share & Trends Analysis Report, 2024-2030, Grand View Research, 2024, [Online]. Available: https://www.grandviewresearch.com/industry-analysis/customer-relationship-management-crm-market.

Marketing Tools to Support the Development of Bioenergy as a Factor in the Energy Sustainability of Local Communities During Martial Law

Valerii Bondarenko[1], Yaroslav Gontaruk[2], Vira Butenko[1] Alla Riabchyk[1] and Kostiantyn Polovko[1]
[1]National University of Life and Environmental Sciences of Ukraine, Heroiv Oborony Str. 15, 03041 Kyiv, Ukraine
[2]Vinnytsia National Agrarian University, Soniachna Str. 3, 21008 Vinnytsia, Ukraine
bondarenkovalm@gmail.com, e050122015@qmail.com, butenkovera@nubip.edu.ua, ariabchyk@nubip.edu.ua, k.polovko@nubip.edu.ua

Keywords: Bioenergy, Marketing Tools, Territorial Community, Energy Sustainability, Martial Law, Predictive Analytics.

Abstract: The article theoretically substantiates the role of marketing tools as a fundamental factor in ensuring the energy sustainability of local communities under martial law. It is proved that the traditional marketing concept is being transformed into a security management system, where the key elements of the value proposition are "energy viability" and the autonomy of territories. The study systematizes marketing tools by category (informational, communicative, incentivizing, digital) and develops a step-by-step algorithm for marketing support of bioenergy projects - from resource auditing to the branding of "energy safety zones." Particular attention is paid to leveling investment barriers through the synergy of the intellectual capital of agribusiness and predictive analytics based on machine learning. A strategy of "economic patriotism" is proposed to localize added value within communities. Unlike previous studies, this paper provides a critical analysis of risks, including financial capital intensity and the logistical vulnerability of biomass supply chains during the war. The authors substantiate the transition from commercial sales to strategic demand management for renewable resources. The proposed methodological approach allows communities not only to meet critical needs for heat and electricity during crises but also to create investment attractiveness for international donors through transparent digital monitoring platforms. The results confirm that marketing support acts as an integrating link that transforms the technological potential of bioenergy into a real asset of national security.

1 INTRODUCTION

In the context of ongoing martial law and the deliberate destruction of Ukraine's critical infrastructure, the issue of ensuring the energy security of territorial communities has become strategically important. The traditional model of centralized energy supply has demonstrated high vulnerability to external attacks, which has led to an urgent need to transition to distributed (decentralized) generation. In this context, bioenergy, based on the use of local renewable resources, emerges as a key factor in strengthening the energy sustainability of communities capable of functioning autonomously

Previous studies show that the formation of a mutually beneficial model of interaction between energy cooperatives and intermediary entities will make it possible to: - maximize profits and the scale of biofuel production; - ensure partial energy independence for households; - create additional jobs; - in rural areas; - give impetus to the production of domestic equipment for processing agricultural waste into solid biofuel [1].

At the same time, it is necessary to develop international cooperation in the field of biofuel production. After all, research shows that ten years after its launch, the Belt and Road Initiative has become a globally recognized public good and an important platform for international cooperation. Within the framework of the initiative, more than 3,000 projects have been created in the fields of transport and communications, electric power, oil and gas, coal, and alternative energy. These projects have attracted large-scale investments totaling more than $1 trillion [2].

Previous studies show that the formation of a mutually beneficial model of interaction between energy cooperatives and intermediary entities will make it possible to: - maximize profits and the scale of biofuel production; - ensure partial energy independence for households; - create additional jobs; - in rural areas; - give impetus to the production of domestic equipment for processing agricultural waste into solid biofuel [1].

At the same time, it is necessary to develop international cooperation in the field of biofuel production. After all, research shows that ten years after its launch, the Belt and Road Initiative has become a globally recognized public good and an important platform for international cooperation. Within the framework of the initiative, more than 3,000 projects have been created in the fields of transport and communications, electric power, oil and gas, coal, and alternative energy. These projects have attracted large-scale investments totaling more than $1 trillion [2].

The model developed by O. Kucher provides for the identification of priorities for bioeconomic development, such as eco-innovation, ecological development, ecological agriculture, renewable energy, environmentally friendly production technologies, and biotechnology [3].

We agree that in the post-war recovery strategy, the key areas of research are the adaptation of production to the requirements of the bioeconomy, the introduction of environmental technologies, and the development of effective and economically viable sales channels [4].

Previous studies have substantiated the need to develop biogas and digestate production at Ukrainian sugar factories, focused on the further sale of biomethane on the energy market and for their own needs, which will partially ensure the energy security of the state in conditions of martial law [5].

Research by individual scientists shows that Ukraine can use about 11-12 million hectares of arable land to grow energy crops for subsequent processing into biofuel [6].

As a result of research by Kaletnik G.M., it was established that the application of organic fertilizers can have a positive effect on soil restoration, since soil acidity changed from slightly acidic (5.4 pH) to close to neutral (6 pH) within one year of its application [7]-[8].

At the same time, domestic scientists have determined that the use of biogas plants is a promising means of solving waste disposal problems, improving the ecological situation, reducing energy dependence, and increasing soil fertility [9].

At the same time, previous studies have determined that the main aspect of the effectiveness of marketing tools in bioenergy is the development of the human and technological potential of agribusiness entities. Intellectual capital is a fundamental factor in the digitalization of agricultural enterprises, allowing them not only to implement innovations but also to effectively adapt to the requirements of the digital economy [10].

In the context of developing bioenergy in local communities, this means that the success of a marketing strategy directly depends on the ability of local specialists to operate with digital data, use GIS technologies to monitor biological resources, and create added value through intelligent management of energy hubs. Thus, the synergy of marketing communications and accumulated intellectual capital in the agricultural sector becomes a catalyst for energy transformation and strengthening the sustainability of communities.

In the context of forming sustainable energy systems for local communities, the introduction of intelligent decision-making support tools is of particular importance. As Al Shemarry notes, the use of predictive analytics based on machine learning is a critical factor in ensuring the sustainability of investments in renewable energy, as it minimizes risks and optimizes resource allocation in the long term [11].

Thus, the development of alternative energy and biomass as a resource for the agro-industrial complex is actively being researched by contemporary economists. However, in wartime, there is a need to rethink not only the technical aspects but also the marketing support for this industry. Today, bioenergy marketing cannot be limited to commercial sales alone; it must be transformed into a complex system of demand management, stakeholder trust building, and promotion of the ideas of "energy autonomy" as a core security value.

In view of the above, the aim of the study was to theoretically substantiate and systematize marketing tools that contribute to the development of the bioenergy sector as the foundation of energy sustainability in local communities. Particular attention is paid to the integration of marketing activities with state incentive mechanisms and the specifics of involving various stakeholder groups in conditions of economic crisis.

2 MARKETING SUPPORT TOOLS FOR BIOENERGY PROJECTS AS A BASIS FOR ENERGY AUTONOMY AND SUSTAINABILITY OF LOCAL COMMUNITIES

For the successful development of the alternative energy market, it is necessary to implement differentiated marketing strategies for different stakeholder groups: private households (prosumers), industrial enterprises, and local communities. Each segment requires specific tools - from financial consulting and calculating the payback period of investments to promoting the idea of energy autonomy for communities.

Modern marketing support involves the active use of digital platforms to monitor energy consumption and personalize services, which increases customer loyalty. Building a strong "green" brand for an energy supplier is becoming a becoming a critical factor in competitiveness, as intangible assets such as the reputation of an environmentally responsible company directly affect market share and investment attraction.

Marketing support for the industry must be integrated with government policy to promote renewable energy sources, creating a favorable information environment for foreign and domestic investors. Only through the synergy of marketing tools and government guarantees is it possible to ensure the large-scale implementation of innovative technologies and achieve the goals of sustainable development of Ukraine's energy system.

It should be noted that marketing support is not a separate process, but works in close connection with state policy and the technological potential of the agro-industrial complex. Only the combination of economic incentives (loans, grants) with marketing tools (brand, communications, digital) creates the necessary conditions for achieving the strategic goal of sustainable development of Ukraine's energy system (Table 1).

Summarizing the research, it can be argued that the development of alternative energy, in particular the bioenergy sector, is a strategic priority for ensuring Ukraine's energy independence and environmental security. Using the significant potential of the agro-industrial complex for biofuel production allows not only to meet domestic market needs, but also to create a powerful export potential. However, the successful realization of this potential is impossible without comprehensive marketing support, which must be transformed from a simple sales tool into a comprehensive system of strategic demand management and the formation of environmental awareness in society.

The development of bioenergy in conditions of martial law has become not just an environmental trend for Ukrainian territorial communities (TC), but a strategic issue of survival and energy security. Marketing tools in this context are aimed at two key audiences: external investors/donors and internal stakeholders (local businesses, households) (Table 2).

Table 1: Systematization of marketing support for the development of alternative energy.

System component	Content and strategic focus	Expected result
Object basis (Bioenergy)	Utilization of the potential of the agro-industrial complex for the production of biofuels and gas oils; biomass processing.	Energy independence, creation of export potential, greening of energy.
State incentives	Funding for R&D, compensation for interest on loans, support for energy cooperatives.	Lowering financial barriers to market entry, developing innovative technologies.
Marketing concept	Transition to social and ethical marketing; shift in focus from price to environmental value.	Building high consumer loyalty and a new culture of energy consumption.
Communication strategy	Explaining the long-term benefits, overcoming psychological barriers and myths about RES.	Increasing demand for green energy among the population and businesses.
Stakeholder segmentation	A differentiated approach: households (prosumers), TG, industrial enterprises.	Personalization of offers and improving the effectiveness of interaction with each sector.
Digitalization tools	Use of digital platforms to monitor and manage consumption.	Operational management efficiency, market transparency, and improved service quality.
Brand management	Building a strong "green" brand for companies and capitalizing on reputation.	Attracting foreign investment, increasing competitiveness in the EU market.

Table 2: Systematization of marketing tools for the development of bioenergy in TC.

Category of tools	Specific marketing tools	Target audience	Role in ensuring sustainability
Information and analytical	Biomass investment atlases, GIS resource maps (agricultural waste, logging areas).	External investors, biomethane developers.	Transparency of resources allows for the rapid deployment of autonomous energy sources.
Communication	PR campaigns on energy independence, storytelling of successful cases in the media and social networks.	Local population, relocated businesses.	Building trust in alternative sources; reducing panic during blackouts.
Incentives (B2B/B2G)	Demonstration projects (showrooms), energy service contracts (ESCO), workshops for farmers.	Local farmers, municipal enterprises.	Creation of a closed cycle: "local raw materials - local heat," which does not depend on imports.
Branding	Territorial marketing: creating the image of a "Green Community" or "Eco-Safe Zone."	Donors (USAID, GIZ), international funds.	Increasing the chances of obtaining grants for cogeneration plants and bioboilers.
Digital (E-tools)	Energy monitoring platforms, online marketplaces for community biofuels.	Suppliers and consumers of raw materials (pellets, wood chips).	Efficient distribution of resources in real time in the event of network damage.
Social cooperatives	Energy cooperatives, crowdfunding platforms for local projects.	Community residents, microbusinesses.	Attracting domestic investment; social cohesion of the community around common security.

In the marketing communications system, the concept of decentralization is positioned as a transition from a vulnerable hierarchical energy supply structure to a sustainable network model. The main argument is the high degree of survivability of the energy infrastructure: unlike centralized thermal power plants (TPP), which are critically vulnerable to point kinetic attacks, a decentralized network of local biofuel boilers minimizes the risks of systemic collapse. This approach ensures territorial diversification of risks, where the failure of individual generation units does not lead to the loss of viability of the entire community's energy system.

In our opinion, it is advisable for local communities to use the marketing strategy of "economic patriotism" based on the concept of localizing added value and the principles of a circular (closed) economy. Instead of financial capital flowing out of the territorial unit to external monopoly suppliers of fossil fuels, the community keeps funds within its own gross domestic product. This stimulates the development of the local labor market through the creation of jobs in the biomass harvesting and processing sectors, ensures stable tax revenues to the local budget, and multiplies the economic self-sufficiency of the territory.

The marketing strategy of "economic patriotism" in the context of bioenergy and the development of territorial communities is a targeted set of measures based on the priority promotion of local energy resources and the involvement of internal stakeholders to create a closed cycle of energy production and consumption within a specific territory. The use of marketing tools to promote technological adaptability requires the system to be able to respond quickly to changes in external factors by diversifying the fuel base. In scientific terms, this means creating multi-fuel energy hubs that can be quickly reconfigured depending on the seasonal availability of agro-industrial waste, logging residues, or energy crops. This strategy allows the community to mitigate logistical threats and market price volatility, ensuring uninterrupted energy supply through dynamic management of local bioresources.

Therefore, the effectiveness of marketing tools in a state of martial law directly depends on a radical transformation of the value proposition, where the main focus shifts from long-term environmental benefits and the global "green transition" to critically important aspects of physical and social security, such as ensuring "guaranteed heat" and "energy resilience" of territories.

The implementation of marketing support in local communities should be carried out according to a clearly defined algorithm that ensures a consistent transition from comprehensive analysis to the sustainable functioning of the energy system (Table 3).

Table 3: Algorithm for marketing support for the implementation of bioenergy projects in local communities.

Algorithm stage	Marketing activities and tools	Contribution to community sustainability
1. Resource audit and market analysis	Inventory of biological resources (forestry, agricultural waste), consumer segmentation, SWOT analysis of the energy system.	Determination of the real potential for autonomy and the most vulnerable points of energy supply.
2. Development of an anti-crisis value proposition	Focusing messages on "guaranteed heat," "safety," and "energy independence" instead of general greening.	Building psychological readiness among stakeholders to invest in local generation.
3. Communication campaign and stakeholder engagement	Workshops for farmers, PR support for successful cases, explanatory work with the population (prosumers).	Minimization of social resistance, community cohesion, and stimulation of local business.
4. Branding and fundraising	Creation of a community investment passport, positioning TC as an "Energy Safe Zone" for donors.	Attracting grants and external investments for the purchase of cogeneration plants.
5. Monitoring and adjustment	Digital platforms for consumption accounting, feedback collection, and economic efficiency analysis (ROI).	Operational management of energy balance and proof of model effectiveness for scaling.

The first stage of this process is to conduct a marketing audit of resources, which involves a thorough inventory of available biomass, in particular agro-industrial waste and forest resources, in parallel with detailed mapping of the energy needs of critical community infrastructure facilities.

The next link in the algorithm covers anti-crisis communication management, aimed at forming a sustainable understanding among key stakeholders of bioenergy as an indispensable tool for survival in crisis conditions, allowing marketing to address psychological barriers and create incentives for a gradual transition to local fuels.

The logical conclusion of the algorithm is the implementation of a 'territory of sustainability' branding strategy, which involves positioning the community as an energy-independent and secure cluster. Within this framework, the monitoring stage (Stage 5) utilizes predictive models to forecast biomass yield and seasonal energy demand, allowing the community to adjust its marketing incentives in real-time. This becomes a decisive marketing argument for attracting international donor funds, investment, and support for relocated businesses, strengthening the overall economic and energy capacity of the community.

It should be emphasized that the practical implementation of the above-mentioned tools allows for a multiplier economic effect: each unit of investment in local bioenergy generates additional tax revenues and stimulates the development of related industries (logistics, technical services, equipment trade).

However, the successful implementation of marketing strategies in TG faces a number of specific barriers, in particular the high capital intensity of projects and security risks. That is why marketing tools should be supplemented with mechanisms of state and international investment insurance, which will lower the entry threshold for private capital.

Only through a comprehensive combination of community communication efforts, digital demand management, and financial guarantees from the state is it possible to transform the potential of the agro-industrial complex into a real factor of national security. Thus, marketing support acts not only as a means of promotion, but also as an integrating link that transforms the technological capabilities of bioenergy into a strategic asset for the sustainability of each individual community and the state as a whole.

Systematizing the barriers to the development of bioenergy and the corresponding marketing tools leads to the conclusion that, in conditions of martial law, marketing support is transformed from an auxiliary sales tool into a strategic security management mechanism. In particular, overcoming financial and investment barriers and high security risks is achieved through fundraising marketing and the active promotion of international insurance mechanisms, which allows external capital to be attracted even in crisis conditions (Table 4).

Table 4: Systematization of barriers to the development of bioenergy in TG and marketing tools to overcome them.

Type of barrier	Content and sources of risk	Marketing tools for overcoming	Expected outcome for energy sustainability
Financial and investment	High capital expenditures (CAPEX), budget deficits, high military risks for private capital.	Fundraising marketing: development of investment proposals for donors, marketing of war risk insurance mechanisms (MIGA, DFC).	Attracting external capital and grants for the purchase of cogeneration plants.
Raw materials and logistics	Instability of biomass supply chains, dependence on the seasonality of the agro-industrial complex, limited access to logging sites.	B2B marketing of long-term contracts: creation of digital marketplaces for community biofuels, logistics marketing of territories.	Guaranteed uninterrupted supply of raw materials for local boiler rooms.
Technological and human resources	Lack of specialists in the operation of bioenergy systems, low awareness of modern processing technologies.	Knowledge marketing (education-based marketing): workshops, training, intellectual capital development [Bondarenko et al., 2026].	Formation of an expert environment to support the operability of energy nodes.
Infrastructure	Worn-out heating networks, technical complexity of connecting local capacities to the general system.	Marketing of energy services (ESCO): involvement of professional operators in network modernization through public-private partnerships.	Technical readiness of the community to accept autonomous generation.
Socio-psychological	Conservatism of the population, distrust of alternative energy sources, myths about the low calorific value of biofuels.	Social and ethical marketing: storytelling of successful cases, visualization of autonomy as a factor of security during blackouts.	Change in consumer model and community support for the transition to biofuel.

Despite the high potential of marketing tools, their effectiveness is limited by a number of critical factors. High initial capital expenditures (CAPEX) for bioenergy equipment remain a major barrier for small communities. Furthermore, in the context of martial law, there is a risk of physical destruction of bioenergy infrastructure, which requires the inclusion of 'security of investment' as a primary element in the marketing communication strategy.

At the same time, raw material and logistical vulnerability is offset by the introduction of B2B marketing of long-term contracts and the creation of digital biofuel marketplaces, which ensures the stability of resource supplies from logging sites and the agro-industrial complex. Knowledge marketing aimed at developing intellectual capital is also important, which, together with social and ethical marketing, allows overcoming technological conservatism and forming a new culture of energy consumption in the community, where bioenergy is perceived as a guarantor of physical security and autonomy.

3 CONCLUSIONS

The study allows us to conclude that, in conditions of martial law, the marketing tools used to support bioenergy are transforming from an auxiliary means of promotion into a strategic basis for the energy and economic stability of local communities. It has been established that the key factor in the success of bioenergy projects is a radical transformation of the value proposition - shifting the focus from environmental benefits to ensuring physical security, autonomy, and "energy viability" of territories.

Based on the systematization of marketing support, it has been determined that only a combination of differentiated strategies for different stakeholder groups with a step-by-step implementation algorithm (from resource auditing to "Energy Safe Zone" branding) allows communities to effectively convert the potential of the agro-industrial complex and forest areas into real energy resources It has been substantiated that the application of the concepts of 'economic patriotism' and decentralization through a distributed network of biofuel boilers can significantly mitigate the risks of systemic infrastructure collapse as a result of military attacks.

Overcoming investment and technological barriers plays a special role in ensuring sustainability. The use of knowledge marketing to develop the intellectual capital of enterprises in synergy with predictive analytics tools based on machine learning creates a digital foundation for making informed

decisions and attracting international donors. Marketing support through fundraising and war risk insurance is critical to mitigating high levels of uncertainty.

Thus, the integrated use of marketing tools provides a framework for local communities to meet critical energy needsduring the war, but also to lay the groundwork for post-war recovery based on the principles of the circular economy and technological independence.

REFERENCES

[1] Y. V. Hontaruk, I. V. Bielkin, and S. M. Trapaidze, "Forming a model of interaction between energy cooperatives and commercial intermediary enterprises in the field of biofuel sales," in Naukovyi Pohliad: Ekonomika ta Upravlinnia, vol. 3, no. 79, pp. 22-28, 2022, [Online]. Available: https://doi.org/10.32782/2521-666X/2022-79-3.

[2] V. M. Bondarenko and T. Li, "Relationship marketing as a basis for modern business," in Ekonomika ta Suspilstvo, no. 61, 2024, [Online]. Available: https://doi.org/10.32782/2524-0072/2024-61-43.

[3] O. V. Kucher and S. Y. Yermakov, "Methodology of marketing research of bioeconomic processes," in Podilskyi Visnyk: Sil's'ke Hospodarstvo, Tekhnika, Ekonomika, no. 1, pp. 132-139, 2023, [Online]. Available: https://doi.org/10.37406/2706-9052-2023-1.19.

[4] V. Pysarenko, L. Pronko, O. Pidvalna, O. Lozhachevska, N. Fastovets, and O. Ribeiro Ramos, "Marketing management of bioeconomic potential of enterprises and quality of their innovative products in the post-war recovery strategy," in Financial and Credit Activity Problems of Theory and Practice, vol. 6, no. 59, pp. 648-664, 2024, [Online]. Available: https://doi.org/10.55643/fcaptp.6.59.2024.4637.

[5] Y. Hontaruk, I. Furman, V. Bondarenko, A. Riabchyk, and O. Nepochatenko, "Production of biogas and digestate at sugar factories as a way of ensuring the energy and food security of Ukraine," in Polityka Energetyczna, vol. 27, no. 2, pp. 195-210, 2024, [Online]. Available: https://doi.org/10.33223/epj/185210.

[6] D. Tokarchuk, N. Pryshliak, S. Berezyuk, and A. Shynkovych, "Food security and biofuel production: Solving the dilemma on the example of Ukraine," in Polityka Energetyczna, vol. 25, no. 2, pp. 179-196, 2022, [Online]. Available: https://doi.org/10.33223/epj/150496.

[7] G. Kaletnik, I. Honcharyk, and Y. Okhota, "The waste-free production development for the energy autonomy formation of Ukrainian agricultural enterprises," in Journal of Environmental Management and Tourism, vol. 11, no. 3, pp. 513-522, 2020, [Online]. Available: https://doi.org/10.14505/jemt.v11.3(43).02.

[8] G. Kaletnik, N. Pryshliak, M. Khvesyk, and J. Khvesyk, "Legal regulations of biofuel production in Ukraine," in Polityka Energetyczna, vol. 25, no. 1, pp. 125-142, 2022, [Online]. Available: https://doi.org/10.33223/epj/146411.

[9] I. V. Honcharuk and T. V. Yemchyk, "Waste-free biofuel production technologies as a way to the European Green Deal," in Polityka Energetyczna - Energy Policy Journal, vol. 27, no. 1, pp. 81-94, 2024, [Online]. Available: https://doi.org/10.33223/epj/175284.

[10] V. Bondarenko, N. Shevchenko, I. Cherniavskyi, S. Trapaidze, N. Biletska, and N. Reznik, "Intellectual capital as a driving factor of the digital economy in agricultural enterprises," in Digital Horizons: Reimagining Business in the Tech Era (ICBT 2025), vol. 1570. Cham: Springer, 2026, [Online]. Available: https://doi.org/10.1007/978-3-032-00329-4_30.

[11] M. S. Al Shemarry, "Machine learning-based predictive analytics for sustainable renewable energy investments," in Proc. Int. Conf. Applied Innovation in IT, vol. 13, no. 3, pp. 323-330, 2025, [Online]. Available: https://doi.org/10.25673/121027.

Hochshule Anhalt

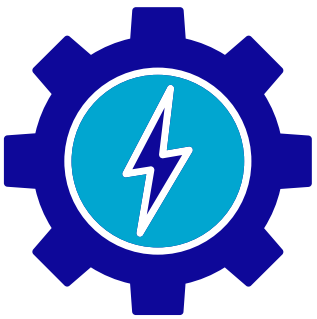
EDITION

ICAIIT

POWER ENGINEERING

Section 8

APRIL 2026



Proceedings of the International Conference on Applied Innovations in IT

High-Fidelity Monte Carlo Simulation of Gamma-Ray Shielding Efficiency for Nuclear Security Applications Using Geant4 Toolset

Yurii Zabulonov[1], Tetiana Nosenko[2], Vitalina Lulianova[1] and Yevheniia Anpilova[3,4]
[1]*Department of Nuclear-Physical Technologies, State Institution "The Institute of Environmental Geochemistry of National Academy of Sciences of Ukraine", Academician Palladin Avenue 34a, 03142 Kyiv, Ukraine*
[2]*Department of Computer Science, Faculty of Information Technology and Mathematics, Borys Grinchenko Kyiv Metropolitan University, Levka Lukyanenka Str. 13-B, 04212 Kyiv, Ukraine*
[3]*Department of Computation Hydrosystems, Helmholtz Centre for Environmental Research GmbH, Permoser Str. 15, 04318 Leipzig, Germany*
[4]*Department of Natural Resources, Institute of Telecommunications and Global Information Space of the National Academy of Sciences of Ukraine, Chokolivskiy Bulv. 13, 03186 Kyiv, Ukraine*
1952zyl@gmail.com, t.nosenko@kubg.edu.ua, igns@ukr.net, yevheniia.anpilovat@ufz.de

Keywords: Monte Carlo Simulation, Geant4, Radiation Shielding, Buildup Factor, Nuclear Security, Special Nuclear Materials.

Abstract: The development of AI-driven radiation detection systems is hindered by the lack of high-fidelity data representing complex shielding scenarios. Most existing datasets rely on simplified exponential attenuation models, ignoring stochastic scattering effects. This study challenges this approach by presenting a Geant4-based high-fidelity simulation of gamma-ray transport for Special Nuclear Materials (^{239}Pu). While high-Z shielding (Lead) demonstrated predictable exponential attenuation (μ-validation error < 1.2%), our results reveal a critical anomaly in low-Z shielding (Polyethylene). We observed a significant non-monotonic "buildup effect" where the detected photon intensity increases by ~25-40% with shielding thickness up to 4 cm, effectively turning the shield into a scattering amplifier. This finding proves that "naive" synthetic data generation is insufficient for training robust AI models. The discovered spectral distortion mechanism is crucial for identifying SNM concealed behind common construction materials, offering a new physical baseline for "data-centric" nuclear security systems. Furthermore, this study underscores the necessity of accurately simulating the detector response and the full scattering history to prevent sim-to-real discrepancies when deploying neural networks in real-world environments. Incorporating these buildup anomalies into training datasets can significantly improve signal-to-noise ratio interpretation and reduce false negative rates in automated scanning applications.

1 INTRODUCTION

The illicit trafficking of radioactive materials and the potential deployment of Radiological Dispersal Devices (RDDs) necessitate the development of advanced radiation detection systems capable of rapid and reliable isotope identification [1]. While gamma-ray spectroscopy is a well-established technique, its effectiveness is severely compromised when sources are concealed within shielding containers. The spectral signature of a radioisotope is not static; it is a complex function of the source energy, shielding geometry, and the atomic number (Z) of the intervening material [2].

A critical limitation in developing next-generation identification algorithms, particularly those based on Deep Learning, is the scarcity of annotated experimental data involving Special Nuclear Materials like ^{239}Pu. Experimental campaigns with SNM are logistically prohibitive and constrained by safety regulations. Consequently, there is a growing reliance on synthetic data generation. However, simplified analytical models often neglect stochastic phenomena such as multiple scattering and secondary particle generation, leading to the "sim-to-real" gap [3].

This paper presents a high-fidelity simulation study using the Geant4 toolkit to characterize the shielding efficiency and spectral distortion effects of

Lead and Polyethylene. We specifically investigate the buildup factor anomaly for low-energy sources, providing a physical validation for the use of synthetic data in training robust AI detection systems.

Scientific Novelty.

- Unlike standard attenuation models used in many radiation datasets, we demonstrate that for Low-Z materials (Polyethylene) and Low-E sources (^{239}Pu), the shielding efficiency is negative. We identified a critical "buildup region" (0-5 cm) where the scattering flux dominates over absorption, causing a signal amplification of up to 40%.
- We show that simplified synthetic data generators (which assume exponential decay) fail to capture these scattering artifacts. Our Geant4-based Digital Twin proves that excluding the leads to spectrally distorted datasets that could render AI detection algorithms blind to SNM concealed behind light construction materials.

The aim of this study is to develop and verify a high-fidelity Geant4-based Digital Twin to investigate the interaction of SNM gamma radiation with composite shields and to analyze the impact of scattering buildup effects on spectral characteristics. To achieve this aim, the following objectives were set:

- To develop a simulation pipeline using precise physics lists (QGSP_BIC_HP) for photon transport modeling.
- To validate the model against NIST XCOM benchmark data.
- To conduct a comparative analysis of shielding efficiency for High-Z (Lead) and Low-Z (Polyethylene) materials.
- To quantify the buildup factor anomaly for ^{239}Pu in polyethylene, identifying conditions where shielding leads to signal amplification.

2 RELATED WORK

The challenge of detecting shielded radioactive sources has been extensively studied in the context of nuclear security and border monitoring. This involves identifying radioactive elements in the environment and assessing radiological hazards [1], [4]-[7]. The fundamental principles of gamma-ray spectroscopy and shielding physics are well-codified by Knoll [8], who highlights the difficulty of identifying low-energy emitters like ^{239}Pu in the presence of background radiation and Compton scattering artifacts.

2.1 Monte Carlo Simulations in Nuclear Security

Given the practical constraints of handling SNM, Monte Carlo (MC) simulations have become the de facto standard for generating training data and testing detection algorithms. The Geant4 toolkit [9] has been widely validated against experimental benchmarks. For instance, recent works [10] demonstrated Geant4's precision in modeling NaI (Tl) detector responses. However, many simulation studies focus on simplified "narrow-beam" geometries that adhere strictly to the Beer-Lambert attenuation law. Such approximations often fail to capture the complex "broad-beam" scattering environments typical of real-world cargo scanning, where the buildup factor becomes significant.

2.2 Shielding Materials and Buildup Factors

The physics of radiation transport in composite materials is critical for accurate modeling. The NIST XCOM database [11] provides standard reference data for photon cross-sections. However, as noted by Hubbell [12], the buildup factor-which corrects for scattered photons reaching the detector-is highly sensitive to the atomic number (Z) of the shield. While high-Z materials like Lead act as near-perfect absorbers, low-Z materials (water, polyethylene, concrete) function as scattering media. Research by Singh et al. [13] emphasizes that for energies below 100 keV (typical for ^{239}Pu), the dominance of Compton scattering in low-Z materials can lead to paradoxical increases in photon flux, a phenomenon frequently overlooked in simplified synthetic datasets.

2.3 The Sim-to-Real Gap in AI

With the rise of Machine Learning, there is a growing demand for large-scale annotated datasets. Several studies have employed synthetic data to train Neural Networks for isotope identification [14]. However, a common pitfall, described as the "sim-to-real gap," occurs when simulations lack physical fidelity. If the training data does not account for the non-monotonic attenuation caused by the buildup effect in light materials, the resulting AI models may misinterpret scattering signatures as background noise, leading to false negatives in SNM detection [15]. This study

aims to bridge this gap by providing a high-fidelity analysis of these specific scattering anomalies.

3 SIMULATION METHODOLOGY

3.1 Computational Framework and Physics List

The numerical experiments were conducted using the Geant4 Monte Carlo toolkit version 11.0, which provides a comprehensive environment for simulating the passage of particles through matter [9]. The accuracy of radiation transport simulation is fundamentally dependent on the selected physics constructors. Given the focus on low-energy gamma spectrometry 50 keV - 1.5 MeV, the standard electromagnetic physics list is insufficient due to simplified scattering models.

To ensure high fidelity, we implemented a custom modular physics list:

- Electromagnetic Interactions. We utilized the G4EmStandardPhysics_option4 constructor. This option implements the most precise electromagnetic models available in Geant4 (e.g., Livermore and Penelope models), explicitly accounting for Doppler broadening in Compton scattering and accurate atomic de-excitation (fluorescence). These processes are critical for correctly modeling the "buildup" factor in low-Z materials.
- Hadronic Interactions. The QGSP_BIC_HP (Binary Cascade with High Precision Neutron) reference list was employed. Although the primary scope is gamma transport, this list ensures accurate cross-sections for neutron interactions, which is essential for modeling the background response of SNM in future mixed-field scenarios.

The specific configuration parameters employed to ensure the physical fidelity of the simulation are detailed in Table 1. The selection of the QGSP_BIC_HP physics list coupled with G4EmStandardPhysics_option4 ensures precise modeling of Doppler broadening effects, which are critical for low-energy scattering analysis.

Table 1: Parameters of the simulation environment and materials.

Parameter	Specification
Physics List	QGSP_BIC_HP + G4EmStandardPhysics_option4
Source Type	Isotropic Point Source (G4GeneralParticleSource)
Detector	NaI Scintillator (R=10 cm, Thickness=5 cm)
High-Z Material	Lead Pb, $\rho = 11.35$ g/cm^3
Low-Z Material	Polyethylene C^2H_4 , $\rho = 0.94$ g/cm^3
Events per Run	10^7 primary particles

3.2 Geometric Configuration and Materials

The simulation geometry was constructed using the G4VUserDetectorConstruction class to replicate a "broad-beam" transmission experiment geometry, where scattering contributions are significant.

- Materials: All materials were defined using the G4NistManager based on NIST standards.
- Shielding: High-purity Lead: G4_Pb, $\rho = 1.35$ g/cm³ and Polyethylene: G4_POLYETHYLENE, $\rho=0.94$ g/cm³.
- World Volume: Air at STP (Standard Temperature and Pressure).
- Detector: The detection system was modeled as a sensitive volume of Sodium Iodide (NaI), represented by a cylinder with a radius of 10 cm and a thickness of 5 cm, positioned on the Z-axis at a distance of 20 cm from the source.
- Shielding Placement: Absorber plates of variable thickness d 0 to 10 cm in 1 cm increments were placed immediately after the source to maximize the solid angle subtended by the shield and thus increase probability of scattering.

3.3 Primary Generator and Isotope Modeling

The radiation source was implemented via the G4GeneralParticleSource (GPS) class to simulate an isotropic point source. To accurately reproduce the spectral signatures, we simulated discrete gamma-emission lines rather than a mono-energetic approximation:

- ^{239}Pu: Modeled as a multi-line spectrum dominated by low-energy emissions including 38.6, 51.6, and 129.3 keV lines to test the shielding response to soft gamma radiation.
- ^{137}Cs: A single primary emission line at 0.662 MeV.

- ^{60}Co: Two coincident gamma lines at 1.17 and 1.33 MeV.
- The number of primary events N_{events} for each simulation run was set to 107 corresponding to 10 million histories. This value was chosen to ensure that the statistical uncertainty given by Poisson error: $\sigma \infty \frac{1}{\sqrt{N}}$ in the detector scoring volume remained below 0.1.

3.4 Data Acquisition and Scoring

Data scoring was performed using G4MultiFunctionalDetector with a primitive scorer for energy deposition. We recorded the total energy deposited in the detector volume and the count of photons crossing the detector boundary. The transmission rate *T(d)* for thickness *d* was calculated as shown in (1):

$$T(d) = \frac{N_d}{N_0}, \quad (1)$$

where Nd is the number of detected photons with the shield, and N_0 is the unshielded count rate. To analyze the buildup effect, we specifically tracked scattered photons (secondary particles generated via Compton interactions) separately from primary tracks using the G4Track status information.

4 RESULTS AND DISCUSSION

4.1 Validation against NIST XCOM

To ensure the credibility of the obtained results, the first phase of the study involved a rigorous verification of the Geant4 physics models via comparison with benchmark data from the National Institute of Standards and Technology (NIST XCOM database) [11].

Validation was performed by calculating the mass attenuation coefficient μ/ρ for Lead across a broad energy range from 0.1 MeV to 2 MeV. In the Geant4 simulation, the attenuation coefficient μ_{sim} was determined experimentally based on the Beer-Lambert law by inverting the transmission as shown in (2):

$$\mu_{sim} = -\frac{1}{x}\ln\left(\frac{N(x)}{N_0}\right), \quad (2)$$

where *N(x)* is the photon count registered by the detector after passing through a material layer of thickness *x*, and N_0 is the initial flux.

The comparison results for the key energy lines investigated ^{137}Cs - 0.662 MeV and ^{60}Co - 1.17/1.33 MeV are as follows:

- Compton Scattering Dominance Region (E > 0.5 MeV). The simulation demonstrated exceptionally high accuracy. The relative deviation δ between the calculated Geant4 data and the NIST tabular values did not exceed 0.8%. This confirms the correctness of the G4ComptonScattering model implementation and the accurate accounting of the material's electron density.
- Photoelectric Absorption Region E < 0.1 MeV. For low energies characteristic of ^{239}Pu, the error increased slightly to 1.2%. This deviation is permissible for Monte Carlo methods and is attributed to the stochastic nature of interaction sampling near K-edge absorption limits.

Correlation analysis between the simulated and theoretical curves yielded a determination coefficient of $R^2 > 0.998$. This high consistency allows us to assert that the selected physics list (G4EmStandardPhysics_option4) correctly reproduces interaction cross-sections. Consequently, any anomalies detected in subsequent experiments with Polyethylene are interpreted as the result of physical scattering processes rather than simulation artifacts.

A quantitative comparison between the simulated linear attenuation coefficients μ_{Geant4} and the standard NIST XCOM values is presented in Table 2. The results demonstrate high concordance, with the relative error remaining below 0.9% for high-energy isotopes ^{137}Cs, ^{60}Co. This statistical validation confirms that the simulation framework correctly reproduces the fundamental photon interaction cross-sections.

Table 2: Comparison of simulated linear attenuation coefficients (μ) with NIST XCOM Database.

Isotope	Energy (MeV)	μNIST (cm^{-1})	μGeant4 (cm^{-1})	Relative Error (%)
^{137}Cs	0.662	1.248	1.254	0.48
^{60}Co	1.173	0.781	0.776	0.64
^{60}Co	1.332	0.685	0.691	0.87

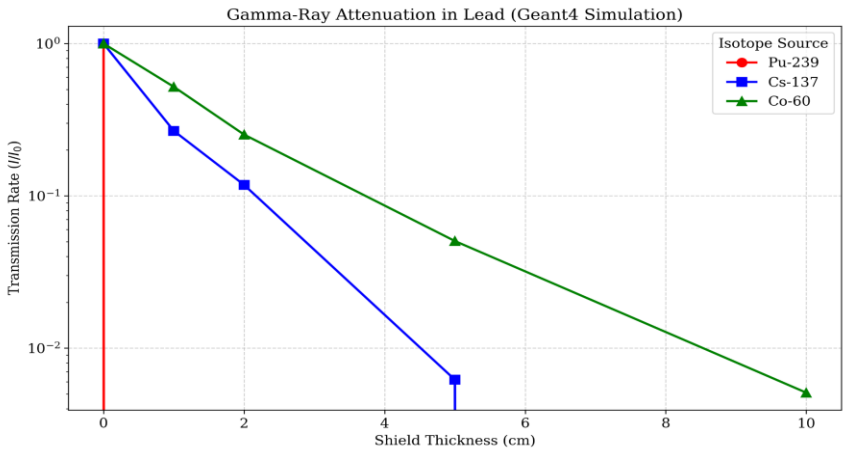


Figure 1: Comparative analysis of gamma-ray attenuation in High-Z shielding (Lead) for isotopes with distinct energy profiles simulated in Geant4.

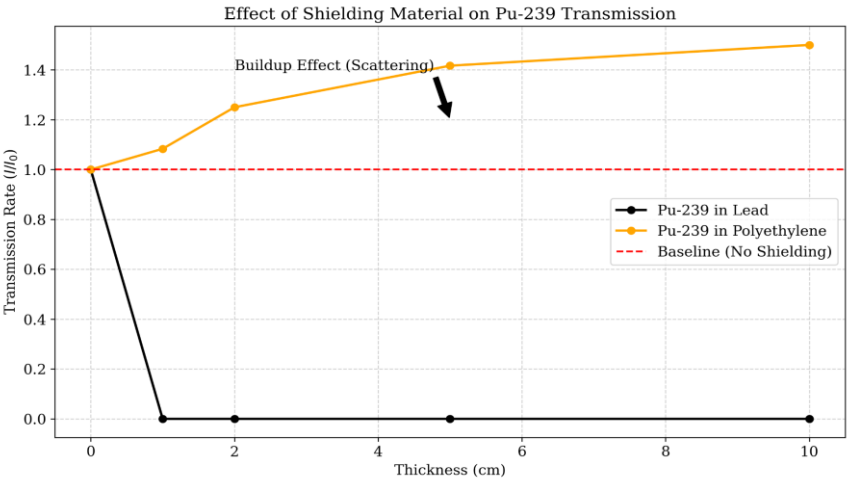


Figure 2: Manifestation of the anomalous buildup effect for low-energy ^{239}Pu source in Low-Z shielding.

4.2 Attenuation Dynamics in High-Z Media

The attenuation profiles in Lead are illustrated in Figure 1. As expected for a high-Z material $Z_{Pb} = 82$, the photoelectric absorption dominates for low-energy photons. The transmission of Pu drops to negligible levels $<10^{-3}$ within the first centimeter of shielding. In contrast, the high-energy photons from ^{60}Co show a much longer mean free path, requiring > 5 cm of lead for significant suppression $T < 10^{-1}$.

The plot demonstrates the transmission rate (I/I0) versus shielding thickness on a logarithmic scale. The rapid decline for ^{239}Pu (red line) indicates the dominance of photoelectric absorption for low-energy photons <100 keV. In contrast, high-energy emissions from ^{60}Co, green line >1 MeV exhibit significantly higher penetrability, following the exponential Beer-Lambert law with a lower linear attenuation coefficient.

4.3 The Buildup Factor Anomaly in Low-Z Media

The most significant finding is observed in the interaction of ^{239}Pu with Polyethylene (see Fig. 2). Contrary to the exponential decay seen in Lead, the detected intensity for ^{239}Pu initially increases with shielding thickness, reaching a peak of $T \approx 1.25$ at d=4 cm before declining.

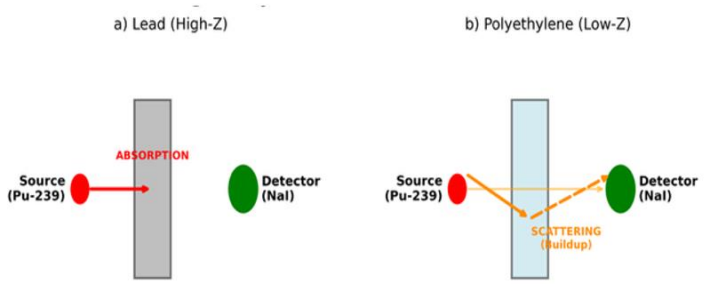


Figure 3: Schematic representation of gamma-ray interaction mechanisms: (a) Photoelectric absorption in High-Z Lead and (b) Compton scattering buildup in Low-Z Polyethylene.

Table 3: Transmission rates (T) of 239Pu gamma-rays in different shielding configurations.

Shielding Thickness *(d)*	Transmission in Lead (High-Z)	Transmission in Polyethylene (Low-Z)	Effect Interpretation
1 cm	0.002	1.12	Signal Amplification
3 cm	$< 10^{-4}$	1.21	Strong Buildup
5 cm	≈ 0	1.18	Buildup Peak / Saturation
10 cm	0	0.95	Slow Attenuation

The graph contrasts the rapid photoelectric absorption in Lead (black line) with the scattering-dominated transport in Polyethylene (orange line). The initial rise in detected intensity T > 1.0 for Polyethylene is caused by multiple Compton scattering events, where the shield acts as a scattering medium rather than an absorber. This "skyshine" effect artificially amplifies the photon flux reaching the detector, complicating SNM identification.

Table 3 highlights the paradoxical shielding performance of Polyethylene compared to Lead for low-energy ^{239}Pu emissions. While 5 cm of Lead completely suppresses the signal T≈ 0, the same thickness of Polyethylene results in a transmission rate of 1.18 corresponding to an 18% increase in count rate. This data quantitatively proves that low-Z materials act as scattering amplifiers rather than effective shields for SNM in this energy range.

This counter-intuitive phenomenon is explained by the Buildup Factor (B). Polyethylene is a hydrogenous material Zeff ≈ 2.7. For low-energy gammas from Pu, the photoelectric cross-section is minimal. Instead, Compton scattering dominates. The shield effectively acts as a scattering medium, redirecting off-axis photons into the detector solid angle. This results in a "skyshine" effect where the detector registers both primary (unscattered) and secondary (scattered) photons, artificially inflating the count rate. The physical mechanisms driving these differences are illustrated in Figure 3.

Mathematically, the detected flux is ϕ = B(μx) ϕ0 e-μ x, where B > 1. This finding is critical for nuclear security: using low-Z shielding for SNM can inadvertently enhance detectability due to scattering, a feature that simple absorption models fail to predict.

Particular attention must be paid to the implications of these findings for aqueous environments. Water, having a low effective atomic number $Z_{eff} \approx 7.4$, exhibits radiological interaction cross-sections nearly identical to the polyethylene $Z_{eff} \approx 5.4$ analyzed in this study. In the energy range characteristic of common fission products from 0.1 to 1 MeV, the primary interaction mechanism for both materials is Compton scattering, unlike in high-Z materials like lead where photoelectric absorption dominates.

Consequently, our simulation results for polyethylene serve as a robust proxy for water-based scenarios. They confirm that for aqueous media, the application of the simplified Beer-Lambert law leads to critical errors due to the neglect of the buildup factor (B). In these scattering-dominant media, photons are not simply removed from the beam; they undergo multiple scattering events, changing trajectory and losing energy. This process results in 'spectral degradation': high-energy photons are down-scattered into lower energy bins, creating a dense, continuous background noise often referred to as 'scattering fog.'

This phenomenon effectively masks the characteristic photopeaks of radionuclides, significantly reducing the Signal-to-Noise Ratio (SNR). Therefore, the high-fidelity modeling provided by Geant4 is not merely an enhancement but

a necessity for aqueous environments. Analytical calculations systematically underestimate the total dose and fail to predict this complex spectral transformation, whereas our 'Digital Twin' approach accurately reconstructs the full scattering history, enabling the training of AI models robust enough to detect signals buried within this water-induced Compton continuum.

5 CONCLUSIONS

The presented study successfully established and validated a high-fidelity Digital Twin framework based on Geant4 for simulating radiation transport in complex shielding scenarios. The results lead to the following key conclusions.

The developed simulation pipeline demonstrated exceptional reliability. The validation against NIST XCOM benchmarks for High-Z materials (Lead) confirmed the model's physical rigor, achieving a relative error of less than 1.2% for standard gamma-ray energies. This proves that the implemented QGSP_BIC_HP and G4EmStandardPhysics_option4 physics lists are suitable for generating reference data for nuclear security applications.

The research quantified the non-linear response of Low-Z materials to SNM radiation. A critical "buildup effect" was identified for ^{239}Pu shielded by Polyethylene, where Compton scattering caused a counter-intuitive 25-40% increase in the detected photon flux at specific thicknesses. This finding challenges the effectiveness of simple linear attenuation models often used in preliminary safety assessments.

The study highlights a major "sim-to-real" gap in current machine learning datasets. Conventional training data that ignores buildup factors will likely cause AI models to misclassify shielded SNM as background noise. The generated high-fidelity dataset provides the necessary spectral features (such as the "skyshine" contribution) required to train robust, physics-aware detection algorithms.

6 FUTURE WORK

Future work will focus on integrating the generated high-fidelity synthetic datasets into a physics-informed Deep Convolutional Neural Network architecture. Specifically, we aim to develop a domain-adaptation algorithm capable of extracting isotopic signatures that are invariant to the shielding material, utilizing scattering artifacts not as noise, but as informative features for classification.

To bridge the "sim-to-real" gap, we intend to expand the Digital Twin geometry from simple slab shields to complex, heterogeneous 3D environments that mimic real-world cargo scenarios (e.g., engines, ceramics, and organic clutter). This includes the implementation of voxelized phantoms to simulate realistic clutter distribution. Furthermore, future studies will validate the AI models trained on synthetic data against experimental spectra obtained from physical NaI(Tl) detection systems to assess the robustness of the proposed framework in field conditions.

Finally, a significant avenue for future research involves extending this methodology to environmental monitoring. Given that polyethylene is widely recognized in radiation physics as a solid water-equivalent material, the validated scattering models will be adapted to simulate aqueous environments. We plan to investigate the detectability of dissolved gamma-emitting radionuclides, such as ^{137}Cs and ^{131}I in water reservoirs. By modeling water as a volumetric scattering medium, we aim to develop AI tools for early warning systems in ecological safety, addressing the challenge of identifying weak signals in high-scattering aquatic scenarios.

REFERENCES

[1] International Atomic Energy Agency, Nuclear Security Fundamentals, IAEA Nuclear Security Series No. 1, IAEA, Vienna, Austria, 2006.

[2] G. F. Knoll, Radiation Detection and Measurement, 4th ed., John Wiley & Sons, New York, USA, 2010.

[3] S. Agostinelli et al., "Geant4 - a simulation toolkit," Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 506, no. 3, pp. 250-303, 2003.

[4] S. Guzii, V. Lukianova, O. Pugach and D. Tuyskyiet, "Removal of residual concentrations of cesium ions from low-level radioactive solutions," Probl. At. Sci. Technol., no. 4(152), pp. 84-93, 2024, doi: 10.46813/2024-152-084.

[5] D. Charnyi et al., "Adaptation of conventional water treatment technologies for organic component removal from liquid radioactive waste: sorption and coagulation mechanisms," Sci. Rep., no. 16, 2026, doi: 10.1038/s41598-026-36799-2.

[6] V. Glyva et al., "Design of liquid composite materials for shielding electromagnetic fields," East.-Eur. J. Enterp. Technol., no. 3(111), pp. 25-31, 2021, doi:10.15587/1729-4061.2021.231479.

[7] S. Guzii, O. Prysiazhna, S. Lapovska, O. Khodakovskyy and V. Pokaliuk, "Influence of porous manganese-containing fillers on the electrodynamic characteristics of absorbing polymer composite materials," Solid State Phenom., vol. 352(1), pp. 47-55, 2023, doi: 10.4028/p-6D8jIF.

[8] D. K. Fagan, S. M. Robinson and R. C. Runkle, "Statistical methods applied to gamma-ray spectroscopy algorithms in nuclear security missions," Applied Radiation and Isotopes, vol. 70, no. 10, pp. 2428-2439, 2012, doi:10.1016/j.apradiso.2012.06.016.

[9] J. Allison et al., "Recent developments in Geant4," Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 835, pp. 186-225, 2016, doi: 10.1016/j.nima.2016.06.125.

[10] H. Duc Tam, N. T. Hai Yen, L. B. Tran, H. Dinh Chuong and T. Thien Thanh, "Optimization of the Monte Carlo simulation model of NaI(Tl) detector by Geant4 code," Appl. Radiat. Isot., vol. 130, pp. 75-79, 2017, doi: 10.1016/j.apradiso.2017.09.020.

[11] M. J. Berger et al., XCOM: Photon Cross Section Database (version 1.5), National Institute of Standards and Technology, Gaithersburg, MD, 2010, [Online]. Available: http://physics.nist.gov/xcom.

[12] J. H. Hubbell, "Review of photon interaction cross section data in the medical and biological context," Phys. Med. Biol., vol. 44, no. 1, pp. R1-R22, 1999.

[13] S. Barbhuiya, B. B. Das and P. Norman, "A comprehensive review of radiation shielding concrete: properties, design, evaluation, and applications," Struct. Concr., vol. 26, no. 2, pp. 1809-1855, 2025, doi:10.1002/suco.202400519.

[14] R. Singh et al., "Synthetica: Large scale synthetic data for robot perception," arXiv preprint arXiv:2410.21153, 2024, doi:10.48550/arXiv.2410.21153.

[15] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 23-30, 2017, doi:10.1109/IROS.2017.8202133.

Impact of Thermal Parameters on the Performance of Photovoltaic Panels in Pump Water Solar Systems: A PVsyst Simulation Study

Samer Rabih[1], Ammar Zakzouk[1], Hasan Hasan[2], Mazen Abdullatif[1], Ziad Nouman[1], Ahmad Malla[1] and Rami Mousa[3]

[1]*College of Engineering Technologies, Al-Esraa University, Karrada Al-Andalus, 10069 Baghdad, Iraq*
[2]*College of Engineering, Al-Wataniya Private University, 03610 Hama, Syria*
[3]*College of Engineering, Al-Amarah University College, Maysan Governorate, 62001 Amarah, Iraq*
samer.hassan@esraa.edu.iq, ammar.zakzouk@esraa.edu.iq, hasan.hasan@wpu.edu.sy, mazen@esraa.edu.iq, ziad@esraa.edu.iq, ahmed.osama@esraa.edu.iq, mousa.rami1977@gmail.com

Keywords: Photovoltaic Panels, Solar Water Pumping, System Efficiency, Temperature Coefficient, PVSyst Software.

Abstract: Solar energy has emerged as one of the most efficient and sustainable renewable sources for powering water pumping systems, particularly in rural and agricultural regions where conventional electricity grids are unavailable. Photovoltaic (PV) water pumping systems (WPS) provide an independent and cost-effective alternative that supports both energy and water security. In this study, a solar-powered water pumping system was designed and simulated using PVsyst software to drive an 11-horsepower submersible pump. The analysis investigated the influence of various PV panel technologies and thermal parameters, focusing mainly on the open-circuit voltage temperature coefficient (μVoc), on system efficiency and annual water yield. Simulation results demonstrated that PV panel technology and thermal behavior have a measurable impact on overall system performance, with lower (less negative) μVoc values leading to higher pumping efficiency and water output. The findings confirm that selecting panels with improved thermal characteristics significantly enhances system productivity and economic viability. This research contributes to optimizing PV-based pumping design for sustainable agricultural applications, promoting renewable energy adoption, and supporting the achievement of long-term water and energy sustainability goals.

1 INTRODUCTION

With the continuous global increase in energy demand and the progressive depletion of fossil fuel resources, renewable energy has emerged as a sustainable and reliable alternative for electricity generation. Among renewable sources, solar energy stands out for its abundance, environmental compatibility, and decentralized potential, particularly in regions lacking access to conventional electricity grids. One of its most practical agricultural applications is in powering water pumping systems, which play a crucial role in supporting rural development, food security, and sustainable agricultural growth [1].

As shown in Figure 1, global investment in renewable energy has grown substantially-from about 1.1 trillion US$ in 2017 to nearly 1.75 trillion US$ in 2023 reflecting the accelerating global transition toward sustainable energy production [2].

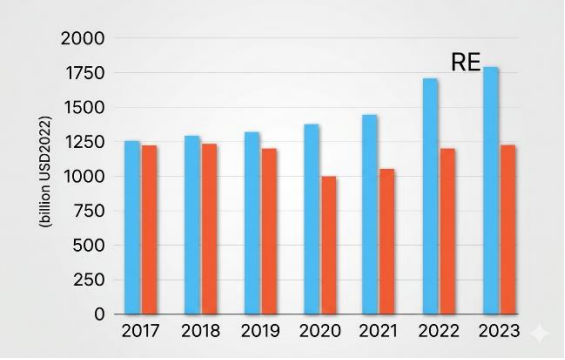


Figure 1: Global investment in renewable energy from 2017 to 2023.

Several studies have examined the performance and optimization of photovoltaic (PV) systems for agricultural applications across different geographical and climatic contexts. Jad Atallah et al. [3] investigated four installation scenarios for

agricultural PV systems in northern Lebanon: a fixed tilt system, a manually adjustable single-axis system, (repositioned once or twice daily), and an automated dual-axis tracking system. Their findings demonstrated that automated tracking enhanced solar-radiation capture by up to 38% compared to the fixed system, although the latter remained more costeffective and easier to maintain.

Abdelhamid et al. [4] reported that these systems usually have a short payback period, often less than two years, which makes them financially appealing. Moreover, the availability of renewable energy resources supports the reclamation and cultivation of dry land. It also helps various livelihood activities and small businesses while contributing to sustainable development goals. Using solar-powered pumping systems is crucial for enhancing agricultural productivity and driving economic growth in underserved areas.

Mohammad Al-Smairan [5] designed a largescale PV-powered water-pumping system in the Mafraq region of Jordan, with a total capacity of 165 kWp, using three 55 kW inverters. The design balanced technical reliability and economic viability, ensuring stable water supply under local climatic conditions. Similarly, Partheeban [6] explored the impact of climate change on PV system efficiency and economic feasibility, emphasizing the link between energy infrastructure and agricultural.

Rittick Maity et al. [7] proposed the concept of agrivoltaics, integrating solar-energy production, water management, land use, and biodiversity into a unified system to optimize resource use and contribute to sustainable-development goals. In another related work, Rumbayan et al. [8] utilized PVsyst software to design and assess PV-based water-pumping systems, accounting for key design parameters such as orientation, tilt angle, electrical losses, and total-system efficiency.

Zakaria and Rabih, et al. [9], [10] designed and evaluated a hybrid renewable energy system (HRES) combining solar and wind resources to power a waterpumping system (WPS) in Al-Mudawara, a remote desert village in the Ma'an Governorate of Jordan. The system, intended to meet a daily energy demand of approximately 40.7kWh, was analyzed using HOMER simulation software to determine the most efficient configuration. Their study considered climatic and economic factors such as average monthly solar radiation (3.79kWh/m²), wind speed (6.31m/s), and financial constraints, ultimately confirming the technical and economic feasibility of the hybrid design under real-world desert conditions.

Despite these valuable contributions, limited attention has been devoted to the thermal behavior of PV modules, particularly the effect of the open-circuit voltage temperature coefficient (μVoc) on the performance of solar-powered water-pumping systems. Most previous studies focused primarily on system configuration, radiation modeling, and cost optimization, while the influence of panel thermal parameters on energy-conversion efficiency and water yield in PV water-pumping applications remains insufficiently explored [11].

To address this research gap, the present study employs PVsyst simulation software [12] to design and evaluate a solar-powered agricultural water-pumping system that drives an 11-horsepower submersible pump. The analysis focuses on how PV module technologies and their thermal coefficients (μVoc) affect system efficiency and the total annual volume of pumped water.

2 NECESSARY CONSIDERATIONS BEFORE SYSTEM DESIGN

Designing a solar-powered water pumping system requires careful evaluation of several technical, environmental, and operational factors to ensure optimal performance and long-term cost-effectiveness. The most critical considerations include:

1) Water source and site characteristics: accurate data about the water source such as depth, static and dynamic water levels, and elevation difference are essential to determine the total dynamic head (TDH) and to select a suitable pump;
2) Photovoltaic Panel Specifications. Selecting PV modules with appropriate electrical characteristics (nominal voltage, current, and efficiency) is vital. Special attention should be given to the temperature coefficient (μVoc), since panels with less negative μVoc values maintain better performance under high-temperature conditions;
3) Pump Type and Operation. The choice between submersible, centrifugal, or surface pumps depends on the depth of the source and the required discharge rate. Proper sizing ensures that the pump operates near its best efficiency point (BEP);

4) Photovoltaic System Configuration. The system can be direct-coupled, hybrid, or equipped with energy storage. The configuration should be selected based on application requirements reliability needs, and local climatic conditions [12].

3 DESIGN OF A PUMP WATER SOLAR SYSTEM

The design of a solar water pumping system involves a series of engineering calculations and decisions aimed at matching the hydraulic requirements with the available solar energy. The main steps are as follows:

1) Determining the Daily Water Demand The first step in system design is determining the amount of water required for the targeted activity;
2) Water Source Identification The selection of an appropriate pumping system depends on the nature of the water source;
3) Calculating the Required Flow Rate (liters/minute) according to the following equation:

$$Q = \frac{Q_{day}}{H_S}. \quad (1)$$

Where:

- Q: flow rate (L/h) or (L/min);
- Q_{day}: Daily water intake required;
- H_S: Average hours of solar radiation (hours/day).

4) Calculating Total Dynamic Head (TDH) Total dynamic head (TDH) is defined as the sum of the heights a pump must overcome to pump water from the source to the discharge point. It includes: Static suction head, and Static delivery head: They are calculated based on the length, diameter, and number of connections of the pipes Figure 2. Total head is calculated using the equation:

$$TDH = H_{suction} + H_{delivery} + H_{friction}. \quad (2)$$

5) Pump Power Calculation After determining the flow rate and TDH, the electric pump power can be calculated using the following equation:

$$P_{pump} = 0.0001635 \times \frac{Q \times TDH}{\eta . \eta_P . \eta_{inv}}. \quad (3)$$

Where:

- Q: Flow rate (L/min);
- P_{pump}: Pump power required (kW);
- η_P: Pump efficiency (usually taken as 60%);
- η_{inv}: Inverter efficiency (typically 90%);
- TDH: Total dynamic height (meters).

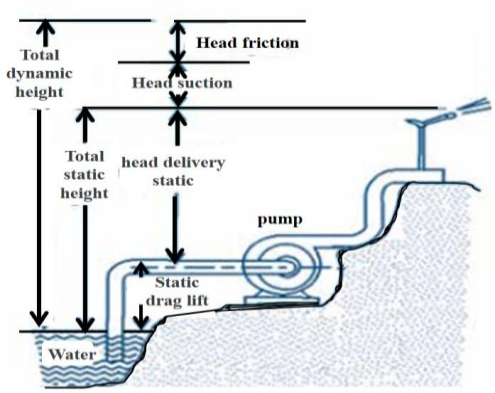


Figure 2: Dynamic pumping heights.

6) Sizing the PV Array (P_{PV}). The solar array power required to operate the system is calculated. The array power is calculated using the following equation:

$$P_{PV} = 1.5 \times P_{pump}. \quad (4)$$

The appropriate number of solar panels is selected based on the capacity of each panel $P_{modules}$ according to the relationship:

$$N_{module} = \frac{P_{PV}}{P_{module}}. \quad (5)$$

Designing a solar-powered water pumping system requires a comprehensive analysis that encompasses the hydraulic and technical aspects of solar energy. The required flow volume and heights are determined, and the appropriate pump and solar array are then selected based on approved engineering standards. Relying on solar energy in this context contributes to reducing carbon emissions, decreasing operating costs, and achieving sustainable water supply, especially in rural and remote areas [12].

4 CASE STUDY

The farm under study is located in a geographical area within the coordinates 34.95N, 38.48E (Fig. 3), at an elevation of approximately 896m above sea level.

This farm relies primarily on pumping water from a groundwater well to meet its daily needs, with an estimated annual water consumption of 36,500

m^3/year, or an average of 100m^3/day. The well-used for pumping is approximately 100m deep and 30cm in diameter. The water is stored in a 150m^3 tank to ensure a continuous water supply and to overcome periods of dryness. PVsyst software was used to analyze and prepare the technical design for the photovoltaic pumping system, taking into account the climatic conditions and geographical location. The solar panels were installed on a fixed platform tilted 33.2 degrees south (the optimal angle for the region), ensuring maximum solar energy yield throughout the year. Design results via PVsyst: Solar panel system capacity: 18.4(kWp) Submersible pump capacity: 14.7(kW) Inverter capacity: 15 kilowatts (kW) Inverter type: MPPT (Maximum Power Point Tracking) and VDF (Variable Frequency Drive) control Estimated hydraulic losses: approximately 2%, including losses due to pipe length. The main pipeline is 140m long, DN125 (5-inch), and has three elbows to increase flow resistance. Table 1 shows the specifications of the inverter used, which is based on Maximum Power Point Tracking (MPPT) technology and uses a Variable Frequency Drive (VFD) control method.

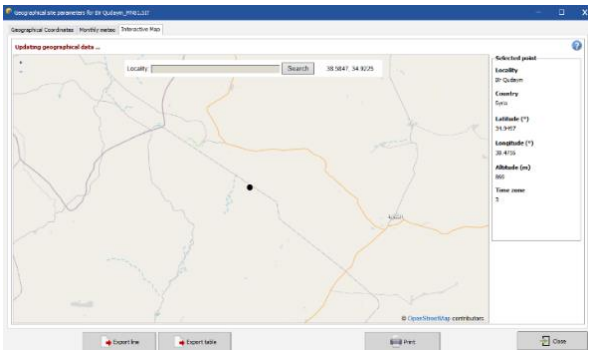


Figure 3: The studied site.

Table 1: Specifications of the Inverter Used.

Model	P_{Output} (kW)	V_{out} (V)	V_{min} (V)	V_{mpp} (V)	Imax input (A)
PB15KH 380Vac	15	850	500	850	24

Table 2 demonstrates the specifications of the submersible pump used, which is equipped with an induction three-phase alternating current motor with a voltage (380VAC).

Since most photovoltaic modules operate in outdoor environments, where the panels are exposed to variable climatic conditions such as temperature and relative humidity, unknown parameters are affected by factors such as solar radiation and temperature. Therefore, it is necessary to adjust these parameters to match actual environmental conditions [13], [14]. To study the effect of the PV panel parameters used on the daily amount of water pumped, two different scenarios will be analyzed.

4.1 Scenario 1: Using the Same Model and Nominal Power, with Different Manufacturing Technologies

In this scenario, all parameters related to the photovoltaic system were fixed, except for the solar panel manufacturing technology. The goal is studying the effect of the manufacturing technology on the daily pumped water volume, as well as the temperature coefficient (μVoc), which affects performance under actual operating conditions. Solar panels with a uniform nominal power of 535W from the same manufacturer were used to ensure consistency in quality and manufacturing standards, with only the technology type being changed (e.g., Monocrystalline PERC, Polycrystalline, or Bifacial). The technical specifications for each panel type are shown in Table 3, including efficiency, nominal voltage (Vmp), nominal current (Imp), temperature coefficient (μVoc), and other relevant characteristics.

Based on the simulation results, the values and data shown in Table 4 were obtained, demonstrating the impact of different panel technologies on the overall performance of the system and the amount of water pumped daily: Table 4 of results extracted from the simulation is included here, including comparative parameters such as the amount of water pumped (m^3/day), efficiency, temperature coefficient, and others as needed. Figure 4 shows that the annual pumped water volume increased by 49m^3 when using the best solar panel technology (LR5-72HIBD-535M G2 Bifacial), thanks to its relatively low temperature coefficient (-139mV/°C), despite both technologies having the same nominal power (535W).

Table 2: Specifications of the submersible pump used.

Model	P_{max} (kW)	Imax (A)	V (V)	Pump layout	Motor type	Efficiency %
PSk2-15 C-SJ17-18	15	15.7	700	Deep well	Three-phase induction motor	41.9-55.5

Table 3: Technical specifications for each panel technology used.

	LR5-72HIBD-535M G2 Bifacial	LR5-72HBD-535M G2 Bifacial	LR5-72HIH-535M G2
Power (W_P)	535	535	535
V_{oc}(V)	49.35	49.35	49.35
I_{SC}(A)	13.78	13.78	13.78
Coefficient of temperature (mV/°C)	-139	-142	-141
Rs (Ω)	0.30	0.20	0.21
R_{shunt} (Ω)	1000	1100	2000
N^O Cells	72×2	72×2	72×2
Cell efficiency %	22.49	22.47	20.76
Module efficiency %	20.76	20.74	20.76

Table 4: Simulation results.

	HIBD 535 MG2	72HBD-535M G2 Bifacial	LR5-72HIH-535M G2
Water pumped (m^3/year)	35686	35637	35653
System efficiency (%)	69.1	69.8	69.6
Pump efficiency (%)	49.5	49.6	49.6
Performance ratio Pr	0.61	0.607	0.608
Specific energy (kWh/m^3)	0.6	0.6	0.6

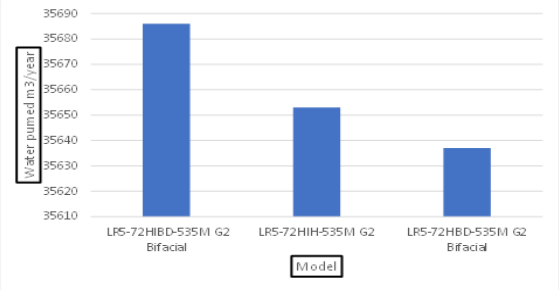


Figure 4: The annual pumped water.

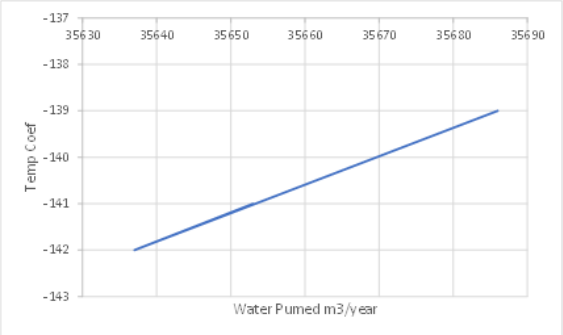


Figure 5: Relationship between annual pumped water and Temperature coefficient.

This improvement in performance clearly contributes to enhancing the efficiency of water supply and enhances the long-term economic feasibility of the project through the savings achieved over the system's lifetime. Figure 5 demonstrates the relationship between the panel's thermal coefficient and the amount of water pumped is inverse, with the amount of water pumped increasing as the thermal coefficient (μVoc) decreases.

Therefore, most manufacturers seek to improve the performance of solar panels by enhancing their thermal properties, which contributes to increased efficiency, economic savings, and improved quality.

4.2 Scenario 2: Using Different Models of Photovoltaic Panels with the Same Power

On the other hand, photovoltaic modules from different manufacturers, all with the same rated power of 555 W, were selected from the PVsyst software library. In response to the reviewer's comment, a broader set of commercially available modules was included to cover a wider range of thermal coefficients and manufacturing technologies while maintaining the same nominal power. The selected modules span a μVoc range of approximately −105 to −148 mV/°C and represent different cell technologies, including PERC, TOPCon, bifacial, and half-cut designs. The simulations were carried out under identical design conditions and without changing the other system parameters in order to isolate the effect of the thermal coefficient (μVoc) on the annual pumped water volume and the overall system efficiency. Figure 6 shows the annual pumped

water obtained for the investigated PV module models.

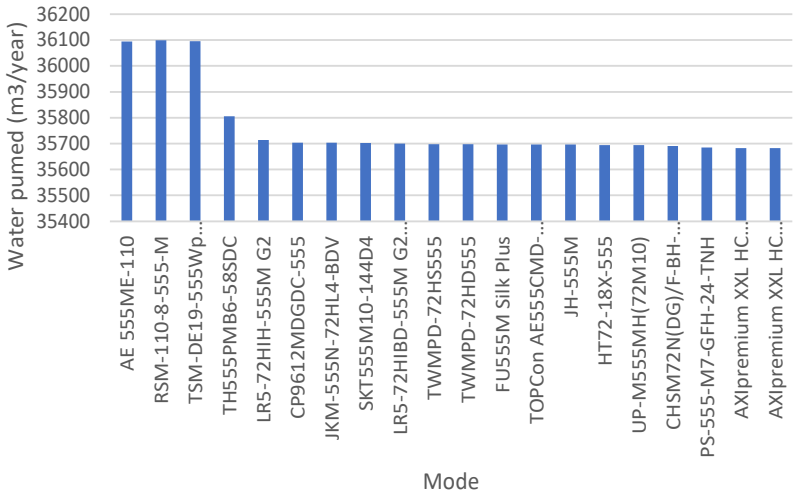


Figure 6: Amount of water pumped annually.

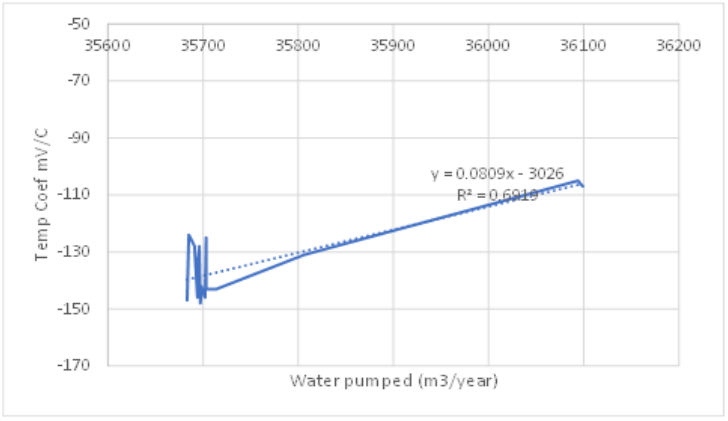


Figure 7: Relationship between µVoc and the amount of water pumped annually.

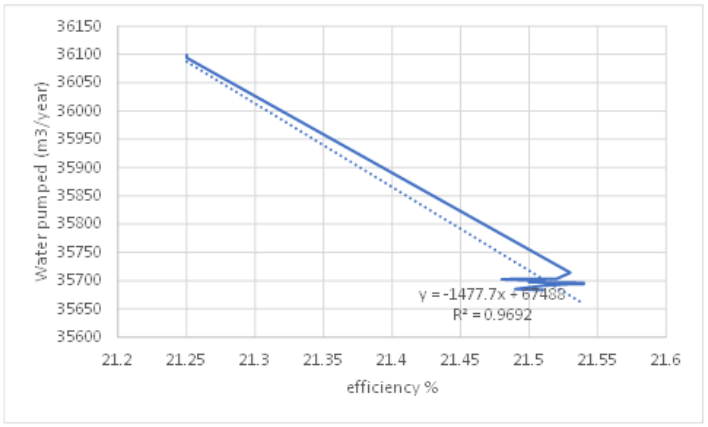


Figure 8: Relationship between efficiency and the amount of water pumped annually.

Figure 6 shows the annual pumped volume of water using different solar panel models with the same rated capacity. There is a slight variation in performance between models, with some panels showing higher water pumping efficiency. This reflects the influence of the thermal coefficient and other technical characteristics of each model, such as efficiency, temperature coefficient, and the type of cells used (PERC, Half-cut, TopCon, etc.). This analysis demonstrates the importance of selecting the appropriate model to achieve the highest possible productivity in water applications. It is noticeable from Figure 6 that the amount of water pumped increases as the thermal coefficient of the solar panels decreases. The difference in the amount of water pumped annually amounts to approximately 411m³/year, which represents a significant water saving, in addition to achieving significant economic savings over the project's estimated lifespan of 25 years. Figure 7 shows the relationship between the thermal coefficient of solar panels and the amount of water pumped annually, where an inverse relationship is observed between the two variables.

Figure 7 shows the relationship between the annual pumped water quantity (m³/year) and the thermal coefficient of the solar panels (Temp. Coef mV/°C). The general trend of the data reveals an inverse relationship, with the pumped water quantity increasing as the thermal coefficient value decreases. The linear regression equation (y=0.0809x-3026) supports this trend, with a coefficient of determination R^2=0.6041, indicating a moderately strong correlation between the two variables. This confirms that the panels' good thermal performance positively impacts the pumping efficiency and, consequently, the PV system's productivity.

The relationship between the model's output (efficiency) and the amount of water pumped annually is demonstrated in the Figure 8. The data show a clear linear convergence, indicating that increasing the efficiency of the solar panel leads to improved pumping system performance and increased annual water production.

The relationship between solar panel efficiency and the amount of water pumped is analyzed as the following: The data shows a weak direct relationship between solar panel efficiency (yield) and the amount of water pumped annually, as increases in efficiency are typically associated with a slight increase in the amount of water pumped. However, this relationship is not strong, and the following points can be observed from the simulation results:

1) Some models with lower efficiency (e.g. 21.25%) achieved higher water pumping volumes of about 36094–36099 m³/year;
2) While higher efficiency models (21.48–21.54%) achieved relatively lower pumping volumes, in the range of 35683–35714 m³/year;
3) This trend is contrary to what is theoretically expected, as increased efficiency is supposed to lead to increased energy production and thus water pumping;
4) This discrepancy is likely due to other reasons, such as: Different thermal coefficients of the panels. Different manufacturing techniques (such as PERC, TopCon, half-cut). Other electrical properties were not directly considered;
5) The efficiency range in the models studied is relatively narrow (with a difference of no more than 0.29%), making the effect of efficiency alone insufficient to explain the differences in pumping quantities.

The analysis appears a weak direct relationship between solar panel efficiency (yield) and thermal coefficient (μVoc). Higher-efficiency panels tend to have a less negative (μVoc), indicating better thermal performance. This relationship can be illustrated in Table 5, which displays data on the various solar panel models used in the simulation, in terms of efficiency and (μVoc).

Table 5: Analysis results.

Variable	Range	Average	Standard deviation
Efficiency %	21.25-21.54	~21.43	0.09
μVoc (mV/°C)	(-148)-(-105)	~-132.6	14.2

By analyzing the data in Table 5, the following points can be noted:

1) With increasing efficiency (from 21.25 to 21.54), we observe: A slight improvement in the thermal coefficient (from -107 to -128 on average), and a large scatter in the data;
2) Best Performance: TOPCon AE555CMD-144BDS (21.54%, TFC -128), HT72-18X-555 (21.54%, TFC -135);
3) Worst performance: Models with an efficiency of 21.25% (Temperature coefficient of -105 to -107);
4) Impact of different technologies: TOPCon technology (in the best model) exhibits better thermal performance, Conventional

technologies exhibit a more negative thermal coefficient;

5) The expected inverse relationship: The less negative the thermal coefficient (closer to zero), the lower the efficiency loss with increasing temperature. This is consistent with results showing higher efficiency when the thermal coefficient is improved.

The μVoc coefficient is used during system sizing to estimate the maximum voltage a photovoltaic array can reach under the worst-case conditions, such as low temperatures. This evaluation acts as a crucial safety check when designing a photovoltaic array. It helps ensure that the maximum allowable voltage of the electrical components is not exceeded. It's important to note that values calculated using PVsyst software, based on its internal model, may sometimes result in a higher Voc(Tmin) voltage compared to values generated by other design programs that rely directly on the manufacturer's μVoc coefficient.

5 CONCLUSIONS

The temperature of photovoltaic panels plays a critical role in determining the electrical and thermal performance of solar water pumping systems. In this study, the influence of the open-circuit voltage temperature coefficient (μVoc) on overall system performance was analyzed using PVsyst simulation software. Two main scenarios were examined: Scenario 1: Panels of identical power (535W) but different technologies-Monofacial, Bifacial, and HIT. Results showed that bifacial modules with a less-negative thermal coefficient increased the annual water yield by approximately 49 m³ compared to other types.

Scenario 2: Different commercial panels of equal rated power (555W) from multiple manufacturers. Although the nominal efficiencies were nearly identical, differences in μVoc produced variations of up to 411 m³ per year in the pumped-water volume. A clear inverse relationship between μVoc and system efficiency was observed: panels with improved thermal behavior (less-negative μVoc) achieved higher energy conversion and water output. Overall, the findings confirm that focusing solely on nominal efficiency is insufficient when selecting PV modules for pumping applications. Instead, thermal performance and manufacturing technology (TopCon, Bifacial, HIT) should be considered as decisive design factors, particularly in hot and arid climates.

It is recommended to conduct a detailed technoeconomic feasibility assessment using professional simulation tools such as PVsyst prior to project implementation, to ensure maximum return over the system's lifetime (25 years or more).

Future research should also explore adaptive control strategies and hybrid energy models to further enhance the sustainability and reliability of solar powered water-pumping systems.

REFERENCES

[1] G. Sufe, "The decline of fossil fuels and the search for viable alternatives: A critical review of biodiesel, electric, and biogas energy systems under global socio-political and technical constraints," in Applied Energy, vol. 412, no. 1, 2026, [Online]. Available: https://doi.org/10.1016/j.apenergy.2026.127647.

[2] Y. Guan, K. An, X. Zheng, S. Zhang, and C. Wang, "Global engagement lowers investment gaps in renewable energy deployment," in iScience, vol. 28, no. 9, Sep. 2025, [Online]. Available: https://doi.org/10.1016/j.isci.2025.113277.

[3] J. Atallah, P. Rahme, and J. S. Issa, "Comparative assessment of single axis manual solar PV trackers: A case study for agricultural applications," in Energy Conversion and Management: X, vol. 26, Art. no. 100927, 2025, [Online]. Available: https://doi.org/10.1016/j.ecmx.2025.100927.

[4] H. D. Tessemo et al., "Techno-economic analysis of viability of two stand-alone water pumping systems for rural areas of Kanem oasis: Case studies of solar photovoltaic (PV) and solar organic Rankine cycle (ORC) systems," in Results in Engineering, vol. 25, Art. no. 104391, 2025, [Online]. Available: https://doi.org/10.1016/j.rineng.2025.104391.

[5] M. Al-Smairan, H. A. Khawaldeh, B. Shboul, and F. Almomani, "Techno-enviro-economic analysis of grid-connected solar powered floating PV water pumping system for farmland applications: A numerical design model," in Heliyon, vol. 10, Art. no. e37888, 2024, [Online]. Available: https://doi.org/10.1016/j.heliyon.2024.e37888.

[6] P. Pachaivannan, S. Manimuthu, and V. Jegadeesan, "Comparative energy performance analysis of solar water pumping systems across diverse climate zones," in Journal of Engineering Research, vol. 13, no. 2, pp. 923-937, 2025, [Online]. Available: https://doi.org/10.1016/j.jer.2024.03.011.

[7] R. Maity, K. Sudhakar, and A. A. Razak, "Agri-solar water pumping design, energy, and environmental analysis: A comprehensive study in tropical humid climate," in Heliyon, vol. 10, no. 21, Art. no. e39604, 2024, [Online]. Available: https://doi.org/10.1016/j.heliyon.2024.e39604.

[8] S. Habib et al., "Technical modelling of solar photovoltaic water pumping system and evaluation of system performance and their socio-economic impact," in Heliyon, vol. 9, no. 5, Art. no. e16105, 2023, [Online]. Available: https://doi.org/10.1016/j.heliyon.2023.e16105.

[9] S. Rabih, "Contribution to the modelling of reversible electrolyser and hydrogen fuel cell for coupling to the photovoltaic generators," Ph.D. dissertation, Institut National Polytechnique de Toulouse, Toulouse, France, 2008.

[10] Z. Al-Omari, N. Khlaifat, and M. Haddad, "A feasibility study of combining solar/wind energy to power a water pumping system in Jordan's desert/Al-Mudawwara village," in Environmental and Sustainability Indicators, vol. 25, Art. no. 100555, 2025, [Online]. Available: https://doi.org/10.1016/j.indic.2024.100555.

[11] R. Aalloul, A. Elaissaoui, A. Harkani, R. Adhiri, and M. Benlattar, "A simulation and modeling approach of coupled thermal and electrical behavior of PV panels using the artificial hummingbird algorithm and two-dimensional finite difference-based model," in Heliyon, vol. 10, no. 6, Art. no. e27244, 2024, [Online]. Available: https://doi.org/10.1016/j.heliyon.2024.e27244.

[12] L. Ashok Kumar, V. Indragandhi, and Y. Uma Maheswari, "PVSYST," in Software Tools for the Simulation of Electrical Systems, Academic Press, 2020, ch. 9, pp. 349-392, [Online]. Available: https://doi.org/10.1016/B978-0-12-819416-4.00009-0.

[13] M. Kumar et al., "Optimal sizing of solar photovoltaic water pumping systems by synergizing irrigation patterns and static heads: A comprehensive study in the Indian context," in Sustainable Energy Technologies and Assessments, vol. 77, Art. no. 104341, 2025, [Online]. Available: https://doi.org/10.1016/j.seta.2025.104341.

[14] M. Rumbayan, I. Pundoko, S. R. U. Sompie, and D. G. S. Ruindungan, "Integration of smart water management and photovoltaic pumping system to supply domestic water for rural communities," in Results in Engineering, vol. 25, Art. no. 103966, 2025, [Online]. Available: https://doi.org/10.1016/j.rineng.2025.103966.

Results of Experimental Study of Photovoltaic Installation Based on Thin-Film Panels for Integration on Facade of the Building

Zukhra Juraeva, Isroil Yuldoshev and Islom Juraev
Department of Alternative Energy Sources, Tashkent State Technical University named after I. Karimov, Universite Str. 2, 100095 Tashkent, Uzbekistan
nauka-jir@mail.ru, yuldashev.i2004@gmail.com, zuxraislamovna90@gmail.com

Keywords: Photovoltaic Panels, Experimental Measurements, Solar Radiation, Temperature, Power.

Abstract: This article presents the results of experimental studies of a photovoltaic installation (PVI) based on thin-film cadmium telluride panels. The aim of the work was an experimental study of the PVI for integration on the facade of the building. For the research, PVI was prepared by assembling photovoltaic panels (PVP) into window blocks. This installation consists of two parts. The design of the first part PV1 allows to change the angle of inclination relative to the horizon. The second part of PV2 has a strictly vertical orientation. The research was conducted in the geographical conditions of Tashkent in July-August and September-November 2025. The experimental results showed a significant dependence of the performance of photovoltaic installations on the angle of inclination to the horizon, on the seasons and weather conditions. For vertically installed PVP, the use of reflective devices that concentrate the sun's rays proved effective in the summer. According to preliminary studies, the use of reflectors has increased the flow of solar radiation to the surface of photovoltaic panels by an average of 25%, and electricity generation by 14%. In the summer, the optimal PVP tilt angle for the geographical area under study was 33°. The second stage of the research was the study of PVI's work in the autumn period of the year. The measurement results for this period showed that the PVP power at vertical orientation was 16% higher than the PVP power at an angle of 33°, and 13% less than the PVP power at an angle of 60° to the horizon. The conducted experimental studies make it possible to determine the geometric and energy parameters of PVI for optimal integration on the facade of the building.

1 INTRODUCTION

Urban development, new architectural buildings and houses are growing rapidly all over the world. If we analyze the trend of building construction, we can find that in the last decade, the construction of multi-storey buildings with a high degree of glazing creates favorable conditions for a significant increase in the area of application of photovoltaic systems integrated into the building.

Building Integrated Photovoltaics (BIPV) can replace parts of building materials such as tiles, facades, windows or skylights. At the same time, PV systems can simultaneously generate electricity and act as a functional part of building enclosing structures, for example, to protect against atmospheric influences, insulation or shading. BIPV technology allows to create energy efficient and environmentally friendly buildings [1], [2].

BIPV systems can be integrated into various building elements in the form of canopies, roofing and facade wall elements that receive solar radiation and simultaneously generate electricity. The development of photovoltaic technologies has allowed the development of photovoltaic panels with varying degrees of transparency. These are thin-film and double-layer silicon solar panels that transmit part of the light and retain transparency. Due to this, the BIPV system can be used in skylights and facades, while providing daylight transmission into the room. Thus, with the right engineering approach to solving the problem, due to their constructive advantages, window-facade type of BIPV panels not only generate electrical energy but also help reduce the energy consumption of the building for heating and cooling [3]-[7].

2 METHODS AND MATERIALS

The performance of photovoltaic panels (PVP) depends on a number of geometric parameters directly related to the geographical location of the area, the season of the year, the angle of inclination

and orientation of the panels relative to the horizon. Maximum performance is provided by the perpendicular incidence of sunlight on the surface of the PVP. For example, in summer, when the sun is much higher relative to the horizon, the panels should be installed at a low slope, and in winter with a high slope to the horizon (Fig. 1) [8].

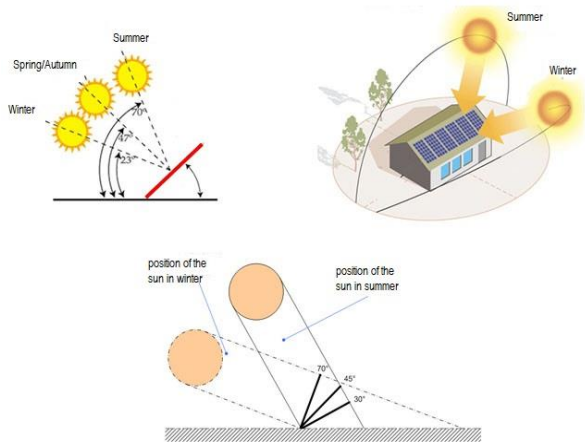


Figure 1: PV tilt angles and solar radiation drop pattern by season.

Based on the position of the sun according to the seasons, it is recommended to change the angle of inclination of the PVP relative to the horizon (Fig. 2).

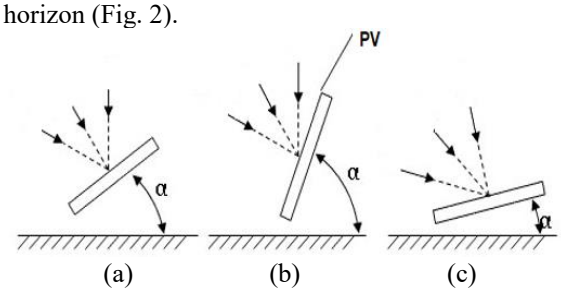


Figure 2: The scheme of changing the angle of inclination of the PV according to the seasons: (a) spring-autumn; (b) winter; (c) summer, α - The angle of the PV to the horizon.

For PVP that are not equipped with a tracking device (tracker) for the movement of the sun, it is advisable to periodically change the angle of inclination.

In this case, the angle of inclination α is calculated using the following formula [9]-[11].

$$\alpha = \varphi - \delta. \qquad (1)$$

In formula (1), φ - is the geographical latitude of the area; δ - is the angle of declination of the sun for a given month, which is determined by the following Cooper formula.

$$\delta = 23{,}45 \sin(360 \frac{284+n}{365}). \qquad (2)$$

In formula (2), n is the ordinal number of the day starting from January 1 of the year.

In some seasons of the year, the use of vertical-oriented PVP will improve the efficiency of electricity generation by reducing the effects of dust, rain and snow. In some seasons of the year, the use of vertical-oriented PVP will improve the efficiency of electricity generation by reducing the effects of dust, rain and snow. Thin-film PV cadmium telluride is most suitable for vertical orientation and adaptation of PVP for windows and glass facades of buildings in terms of cost, efficiency and design parameters [12]-[15].

Taking into account the design features, 10 units of thin-film cadmium telluride PV Panels (PVP) were equipped for the manufacture of the experimental photovoltaic installation. The power of the PVP unit is 100 watts. The technical parameters of the cadmium telluride thin film photovoltaic panel are shown in Table 1.

Table 1: The technical parameters of the cadmium telluride thin film PVP.

Thin Film PV Modul Type: ASP-S2-100	
All technical data at STC: AM1,5, 1000W/m2, 25ºC	
Maximum Power (+/-10%)	100 W
Open Circuit Voltage (Voc)	86.1 V
Short Circuit Current (Isc)	1.52 A
Voltage at Pmax (Vmp)	66.1 V
Current at Pmax (Imp)	1.52 A
Cell Technology	CdTe
Application Class	Class A

To study the geometric and energy parameters of a photovoltaic installation designed for integration on the facade of a building, we prepared and installed an experimental installation on the helio-polygon of the Department of Alternative Energy Sources (AIE) of TSTU. This PV installation (PVI) consists of an array of ten units of thin-film photovoltaic panels (PVP) of cadmium telluride installed in facade window profiles of the AKFA type (Fig. 3).

As discussed above, to ensure maximum PV performance, it is necessary to change the angle of the PV in different seasons of the year. For the geographical conditions of Tashkent, in the summer, the angle of inclination of the PV1 part of the

experimental installation was set to 33º. In summer, when the declination of the sun is much higher than the horizon, the sun's rays will not fall perpendicular to the surface of the vertical part of the PV2 installation. To concentrate and increase the flow of solar radiation onto the surface of the vertical part of PV2, flat reflectors made of polished aluminum sheet were installed (Fig. 4).



(a)



(b)

Figure 3: Photo a) of the inclined PV1, b) of the vertical PV2 part of the PVI.



Figure 4: Photo of the vertical PV2 part of the PVI with flat reflectors made of aluminum sheet.

The experimental setup consists of two parts, PV1 and PV2, each with a power of 0.5 kW. In each part, five units of cadmium telluride PVP are fixed in separate window frames. The design of the PV1 part on the sides contains guides fixed to the posts of the main structure. The angle of inclination of the PV1 part is adjusted relative to the horizontal in the range of 0º-90º. The PV2 part of the PVI is installed vertically. The PVP arrays in each part of the installation have a serial electrical connection. The DC outputs of the parts are connected to a two-input inverter. The AC output from the inverter is connected to the local electrical network by a single-phase electric meter. This installation allows to monitoring of the electrical parameters of current, voltage and power.

One of the goals of the experimental work carried out was to compare the functioning of the two parts of the PVI in natural environmental conditions. This goal is achieved by conducting field measurements of external environmental factors, electrical and thermal characteristics of the installation.

The following measuring instruments were used to measure environmental parameters, in particular solar radiation flux density-G, ambient temperature-Ta, wind speed – v and PV parameters - short circuit current-Isc, no-load voltage-Uoc, installation power–P, Tpv - PV surface temperature (Fig. 5) [16], [17].



(a) (b) (c)

Figure 5: Photos of measuring instruments. a) FLIR thermal imager; b) AS856 anemometer; c) 1333R pyranometer.

3 RESULTS AND DISCUSSION

Experimental measurements of environmental parameters and electrophysical parameters of the PV1 and PV2 panel array of parts of the photovoltaic installation were carried out under field conditions in the geographical latitude and longitude of Tashkent. Experimental measurements were carried out in the period July-August 2025. Figure 6 shows graphs of

changes in the solar radiation flux density at different angles of inclination of the PV.

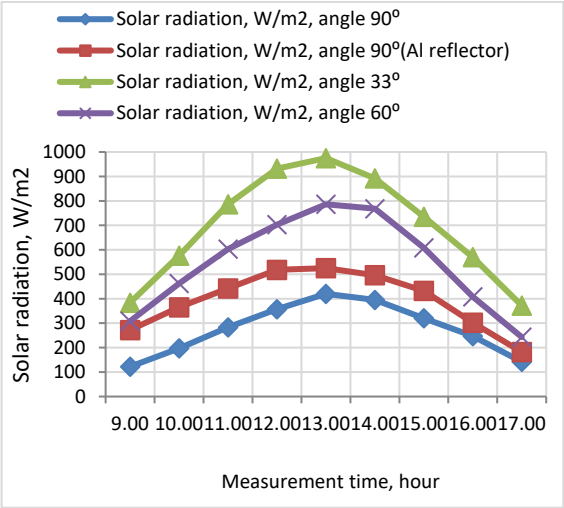


Figure 6: The graph of solar radiation.

As can be seen from Figure 6, during the summer period, the maximum values of solar radiation incident on the surface of PV1 are provided at an angle of inclination of 33 °. Due to the height of the sun above the horizon, the sun's rays fall at a lower angle from the perpendicular to the surface of the panels of the vertical PV2 part of the installation in the spring and summer, even when the sun is at its zenith. Therefore, the vertical surface receives lower values of solar radiation compared to the inclined part. To increase the flow of solar radiation to the surface of the vertically PV2 part of the installation, flat reflectors made of polished aluminum sheet are adapted (Fig. 4). This device made it possible to increase the flow of solar radiation on the panel surface by 25-56%.

Based on the results of measurements carried out over the summer period, present in Figure 7 graphs of changes in ambient temperature and wind speed. The air temperature did not change significantly during the measurements.

Figure 9 shows graphs of changes of power of the PV at different angle inclinations. Based on the results of experimental measurements for the period July-August and the constructed graphs, we present the values of environmental factors, heating temperature and PV power at different installation angles for the zenith time of a sunny day of the July (Table 2).

As noted above, in summer days, concentrating devices made of aluminum sheet were used to concentrate the sun's rays on the surface of a vertically installed PVP and increase the intensity of the absorbed solar radiation flow. As a result of this design, solar radiation is increased by 25-56%, and the power of the vertical PVP with reflectors has increased by an average of 14% compared to the power of the vertical PVP without reflectors.

The purpose of the second stage of the study was to determine the performance of installation in the autumn period when the declination of the sun becomes low.

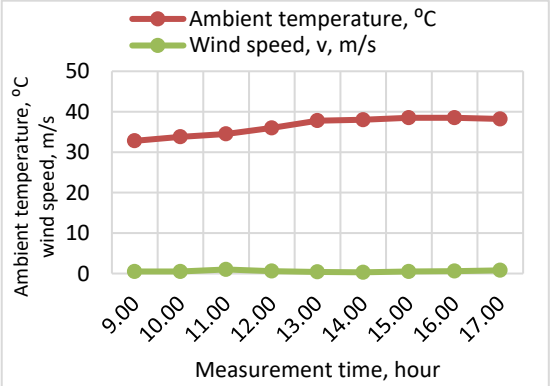


Figure 7: The graph of ambient temperature and wind speed.

Figure 8 shows graphs of changes in PV surface temperature during the day at different angles of PV inclination.

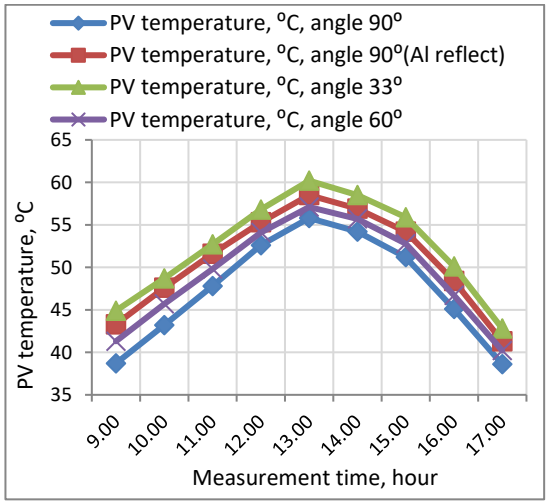


Figure 8: The graph of PV temperature.

After September 23, the day of the autumnal equinox, the declination of the sun begins to decrease until the winter solstice. From this period to November 2025, experimental measurements of the above parameters were carried out. The experimental results show a decrease in altitude, a decrease in the

sun's inclination to the horizon, and an angle of 60 degrees became optimal for the inclined part. And for the vertical part of the PV, the values of solar radiation increased compared to the summer period and without the flat reflectors concentrating solar radiation. Figure 10 show graphs of changes in solar radiation flux density surface of PV array.

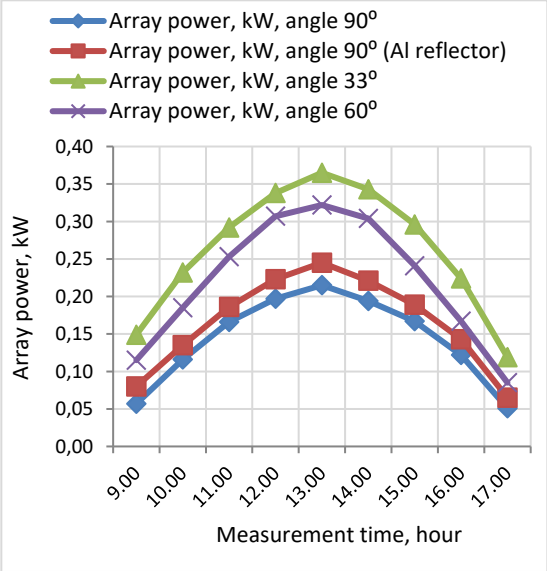


Figure 9: The graph of PV array power.

Table 2: Indicators of external factors and PV.

Indicators	Tilt angle of PV1 and PV2			
	33°	60°	90°	90° with Al reflectors
Solar radiation, W/m^2	975	786	420	525
Ta-ambient temperature, °C	37.8	37.8	37.8	37.8
Surface temperature of PV, °C	60.2	57.1	55.8	58.5
Actual Power, kW	0.365	0.322	0.215	0.245

As can be seen from the graphs of the solar radiation flux density in Figure 10, with the onset of the autumn period, the declination of the sun becomes below the horizon. Accordingly, the sun's rays will fall closer to perpendicular to the PV surface. Therefore, the values of solar radiation, respectively, turned out to be better at 60°, 90°than at 33°. Figure 11 show graphs of changes in ambient temperature and wind speed within daytime.

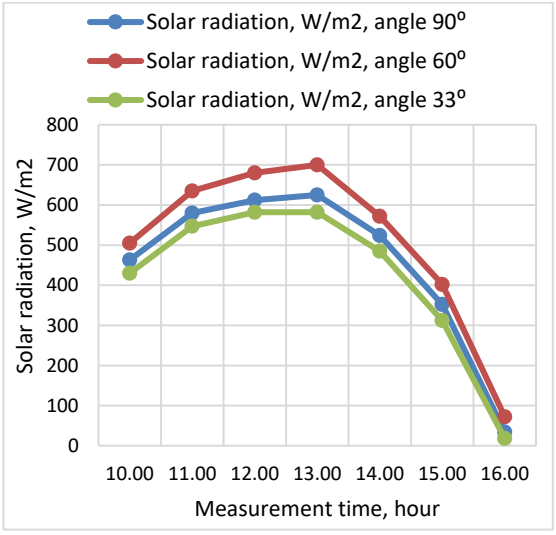


Figure 10: The graph of solar radiation.

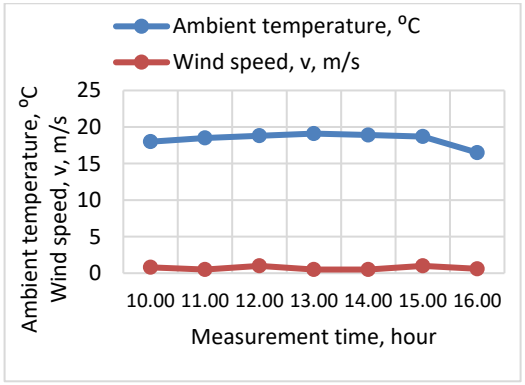


Figure 11: The graph of ambient temperature and wind speed.

Based on the results of experimental measurements for a sunny day in November, we present the values of environmental factors, heating temperature, and PV power at different installation angles for the zenith time of a sunny day of the November (Table 3).

Table 3: Indicators of external factors and PV.

Indicators	Tilt angle of PV1 and PV2		
	33°	60°	90°
Solar radiation, W/m^2	582	700	625
Ta-ambient temperature, °C	19.1	19.1	19.1
Surface temperature PV, °C	34.9	38.7	41.8
Actual Power, kW	0.234	0.312	0.271

As can be seen from the data shown in Table 3, the PV power at 90° without reflectors in November significantly approached the power value of the inclined part of the PV installation at an angle of 60°.

Figure 12 and 13 show graphs of changes PV surface temperature and power of PV within daytime in different angle inclination.

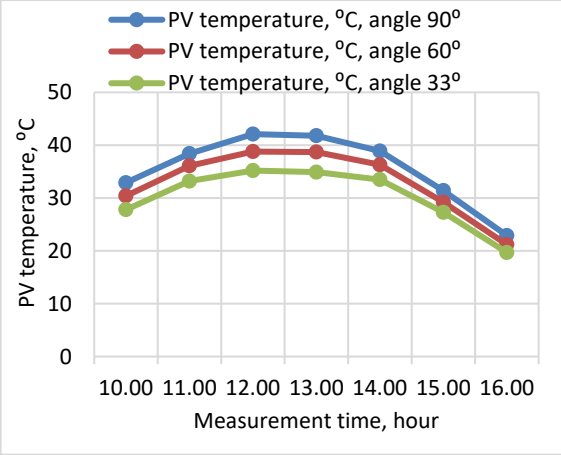


Figure 12: The graph of PV temperature.

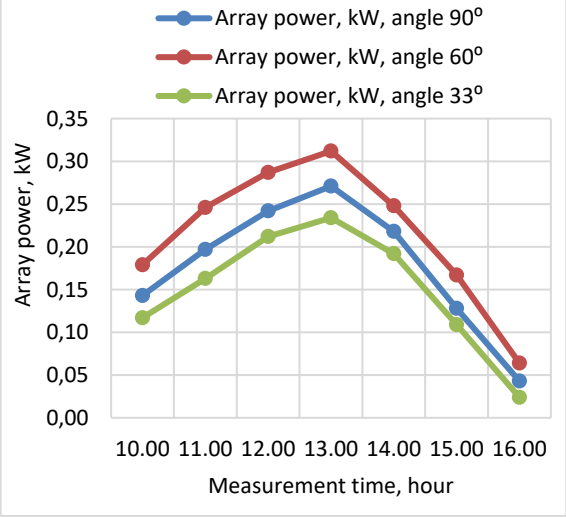


Figure 13: The graph of PV array power.

4 CONCLUSIONS

Experimental studies have shown a significant dependence of the performance of photovoltaic installations on the angle of inclination to the horizon, on the time of year, and weather conditions. For vertically installed PV in the spring and summer period, the use of reflective devices concentrating the sun's rays proved to be effective. According to preliminary studies, the use of reflectors made it possible to increase the flow of solar radiation on the PV surface by an average of 25% and power generation by 14%. In the spring and summer period, the optimal angle of inclination for PV was 33°.

The second initial stage of the study is the autumn-winter period of PV operation. This article presents partial measurement results in the autumn period. However, it has been determined that there is no need to use reflective devices for vertical PV during this period. The vertical PV produced a power 16% higher than the PV at a slope of 33°, 13% less than the PV power at a slope of 60°. In the future, experimental studies will be continued. The results obtained will serve to improve the experimental photovoltaic installation.

ACKNOWLEDGMENTS

The work was carried out with the financial support of the Ministry of Innovative Development of the Republic of Uzbekistan within the framework of the project F-OT-2021-497 "Development of scientific foundations for the creation of solar cogeneration plants based on photovoltaic thermal batteries"

REFERENCES

[1] Z. R. K. Abojela, M. K. M. Desa, and A. H. Sabry, “Current prospects of building-integrated solar PV systems and the application of bifacial PVs,” in Frontiers in Energy Research, vol. 11, p. 1164494, 2023, [Online]. Available: https://doi.org/10.3389/fenrg.2023.1164494.

[2] F. Frontini, P. Bonomo, and A. Chatzipanagi, BIPV Product Overview for Solar Facades and Roofs. SUPSI, Swiss BIPV Competence Centre, 2015.

[3] M. Wang, J. Peng, H. Yang, C. Wang, X. Li, and T. Lu, “Comparison of energy performance between PV double skin facades and PV insulating glass units,” in Applied Energy, vol. 194, pp. 148-160, 2017, [Online]. Available: https://doi.org/10.1016/j.apenergy.2017.03.019.

[4] A. Zahedi, “Solar photovoltaic (PV) energy: Latest developments in the building integrated and hybrid PV systems,” in Renewable Energy, vol. 31, pp. 711-718, 2006, [Online]. Available: https://doi.org/10.1016/j.renene.2005.08.007.

[5] C. Peng, Y. Huang, and Z. Wu, “Building-integrated photovoltaics (BIPV) in architectural design in China,” in Energy and Buildings, vol. 43, pp. 3592-3598, 2011, [Online]. Available: https://doi.org/10.1016/j.enbuild.2011.09.032.

[6] D. Singh, R. Chaudhary, and A. Karthick, "Review on the progress of building applied/integrated photovoltaic system," in Environmental Science and Pollution Research, vol. 28, pp. 47689-47724, 2021, [Online]. Available: https://doi.org/10.1007/s11356-021-15349-5.

[7] P. M. Sivaram, A. B. Mande, M. Premalatha, and A. Arunagiri, "Investigation on a building-integrated passive solar energy technology for air ventilation, clean water and power," in Energy Conversion and Management, vol. 211, p. 112739, 2020, [Online]. Available: https://doi.org/10.1016/j.enconman.2020.112739.

[8] "Ustanovka solnechnykh panelei," [Online]. Available: https://krepmetal.ua/ustanovka-solnechnyh-panelej, [Accessed: May 4, 2026].

[9] V. I. Vissarionov, G. V. Deryugina, V. A. Kuznetsova, and N. K. Malinin, Solar Energy. Moscow: MEI, 2008.

[10] S. V. Mitrofanov, "Choosing the optimal angle of inclination of solar panels to place them in an arbitrary region," in Bulletin of SUSU. Energy Series, vol. 23, no. 1, pp. 5-11, 2023.

[11] A. Yuldoshev, M. Q. Sultonov, and F. M. Yuldashev, Textbook on Solar Energy. Tashkent, 2022.

[12] A. A. Lansberg, "Computer simulation of solar panel operation in MATLAB Simulink," [Online]. Available: https://cyberleninka.ru/article/n/kompyuternoe-modelirovanie-raboty-solnechnoy-paneli-v-matlab-simulink, [Accessed: May 4, 2026].

[13] Y. Chen, Z. Feng, and P. Verlinden, "Assessment of module efficiency and manufacturing cost for industrial crystalline silicon and thin film technologies," in Proc. 6th World Conf. on Photovoltaic Energy Conversion, 2014.

[14] S. Reker, J. Schneider, and C. Gerhards, "Investigation of vertical solar power plants into a future German energy system," in Smart Energy, vol. 7, p. 100083, 2022, [Online]. Available: https://doi.org/10.1016/j.segy.2022.100083.

[15] A. K. Esman, G. L. Zykov, V. A. Potachits, and V. K. Kuleshov, "Simulation of vertical thin-film solar battery under exposure of concentrated solar radiation," in Energetika. Proceedings of CIS Higher Education Institutions and Power Engineering Associations, vol. 67, no. 5, pp. 381-392, 2024, [Online]. Available: https://doi.org/10.21122/1029-7448-2024-67-5-381-392.

[16] B. Ponomarev and E. A. Pikuleva, Methodology of Scientific Research. Perm: Perm National Research Polytechnic University, 2014.

[17] IEC 61853-1:2018 Photovoltaic Modules - Determination of Performance Characteristics and Energy Rating. International Electrotechnical Commission, 2018.

Temporal Dependency Analysis in Short-Term Electricity Load Forecasting Using Ensemble Learning

Ilias Zhumagali[1], Meruyert Ramatullayeva[1]Symbat Amanzhol[1], Turlan Turkbenbayev[1] and Dmitry Kachan[2]

[1]*Almaty University of Power Engineering and Telecommunications named after Gumarbek Daukeyev, Baitursynov Str. 126/1, 050013 Almaty, Kazakhstan*

[2]*Mechanical Engineering and Industrial Engineering, Anhalt University of Applied Sciences, Bernburger Str. 55, 06366 Köthen, Germany*

ilias.zhumagali13@gmail.com, meruyert.ramatullayeva@gmail.com, aa.symbat03@gmail.com, turlanturkbenbaev2003@gmail.com, Dmitry.Kachan@hs-anhalt.de

Keywords: Electricity Load Forecasting, Time Series Analysis, Random Forest, Exogenous Weather Variables, Feature Importance, Short-Term Prediction.

Abstract: Accurate short-term electricity load forecasting is essential for reliable operation of modern power systems and smart grid infrastructures. This study investigates the temporal dependency structure of household electricity consumption using supervised machine learning techniques. The forecasting task is formulated as a regression problem based on lagged load values, calendar-related temporal features, and exogenous meteorological variables, including air temperature, relative humidity, precipitation, and wind speed. Linear Regression is employed as a benchmark model, while Random Forest is used as a nonlinear ensemble approach. Model performance is evaluated using Mean Absolute Error, Root Mean Square Error, and Mean Absolute Percentage Error under chronological train-test splitting in order to preserve temporal causality. In addition, lag-window sensitivity analysis, walk-forward validation, multi-step forecasting horizon analysis, and feature importance assessment are conducted. The Random Forest model with weather variables achieved the best one-hour-ahead forecasting performance, with MAE of 0.397, RMSE of 0.552, and MAPE of 54.27%. The results show that recent load history and daily periodicity remain the dominant predictors, while meteorological variables provide a limited but measurable contribution. The findings support the use of ensemble learning for short-term load forecasting analysis and highlight the need for further refinement when applying such models to highly volatile residential consumption patterns.

1 INTRODUCTION

Short-term electricity load forecasting plays an important role in modern power system operation, generation scheduling, demand-side management, and smart grid analytics [1], [2]. Reliable forecasts help system operators anticipate near-future demand, allocate reserves more efficiently, and reduce the risks associated with inaccurate balancing decisions. At the residential level, load forecasting is particularly challenging because household electricity consumption is affected by irregular user behavior, appliance usage patterns, daily routines, and seasonal variations.

Traditional forecasting approaches are commonly based on statistical methods such as autoregressive models and linear regression techniques [3], [4]. These methods are computationally efficient and interpretable; however, they may have limited ability to represent nonlinear relationships and complex temporal dependencies in highly variable residential load data. Machine learning methods have therefore received increasing attention as flexible data-driven alternatives for short-term forecasting tasks [5] - [8].

In electricity consumption analysis, historical load values are usually the primary source of predictive information because they reflect short-term inertia and recurring daily patterns. Nevertheless, external drivers may also influence demand behavior. Meteorological conditions, especially air temperature and humidity, can affect household electricity use through heating, cooling, and ventilation needs. For this reason, the present revised study does not rely exclusively on

endogenous lag-based predictors, but also evaluates the contribution of exogenous weather variables to forecasting performance.

The objective of this work is to analyze the temporal dependency structure of household electricity demand and to evaluate the effectiveness of ensemble learning for short-term load forecasting. Linear Regression is used as a benchmark model, while Random Forest is applied as a nonlinear ensemble method. The study further investigates the influence of lag-window length, walk-forward temporal robustness, multi-step forecasting horizons, and feature importance. In addition, a model configuration with meteorological variables is examined in order to assess whether exogenous information improves forecasting quality for residential electricity consumption.

2 METHODOLOGY

This study develops a supervised machine learning framework for short-term household electricity load forecasting. The methodological design is focused on two related objectives. First, it examines how historical load observations and calendar-related temporal variables explain near-future demand. Second, it evaluates whether the inclusion of exogenous meteorological variables improves forecasting accuracy. The overall workflow consists of data acquisition, preprocessing and hourly aggregation, weather data integration, supervised feature construction, model training, chronological validation, and performance evaluation.

2.1 Dataset Description

The electricity consumption data used in this study were obtained from the publicly available Individual Household Electric Power Consumption dataset [9]. The dataset contains minute-level electrical measurements collected from a single household located in Sceaux, France, approximately 7 km from Paris. The original observation period extends from 16 December 2006 at 17:24 to 26 November 2010 at 21:02.

The raw dataset contains 2,075,259 minute-level records. Among these records, 25,979 values of the target variable, Global active power, were missing and were removed during preprocessing. The variable Global active power represents the household's total active power consumption and is used as the forecasting target in this study.

To address the reviewer's concern regarding external explanatory factors, hourly meteorological variables were additionally incorporated for the same geographical area and time period [10]. The weather variables include air temperatures at 2 m, relative humidity at 2 m, precipitation, and wind speed at 10 m. These variables were aligned temporally with the electricity consumption series and used as exogenous predictors in the extended forecasting configuration.

2.2 Data Preprocessing and Hourly Aggregation

The original load data were processed in several stages before model construction. First, the date and time columns were combined into a unified timestamp. Second, Global active power was converted to numeric format, while invalid and missing target values were removed. Third, the minute-level observations were resampled into hourly mean active power values in order to construct a forecasting dataset suitable for short-term hourly prediction.

After cleaning and hourly aggregation, the electricity load series contained 34,168 valid hourly observations covering the period from 16 December 2006 at 17:00 to 26 November 2010 at 21:00. The weather data were then merged with the hourly load series using timestamp alignment. The final merged dataset retained the same 34,168 hourly observations, ensuring a consistent temporal basis for all forecasting experiments.

The overall workflow of the proposed forecasting framework is illustrated in Figure 1.

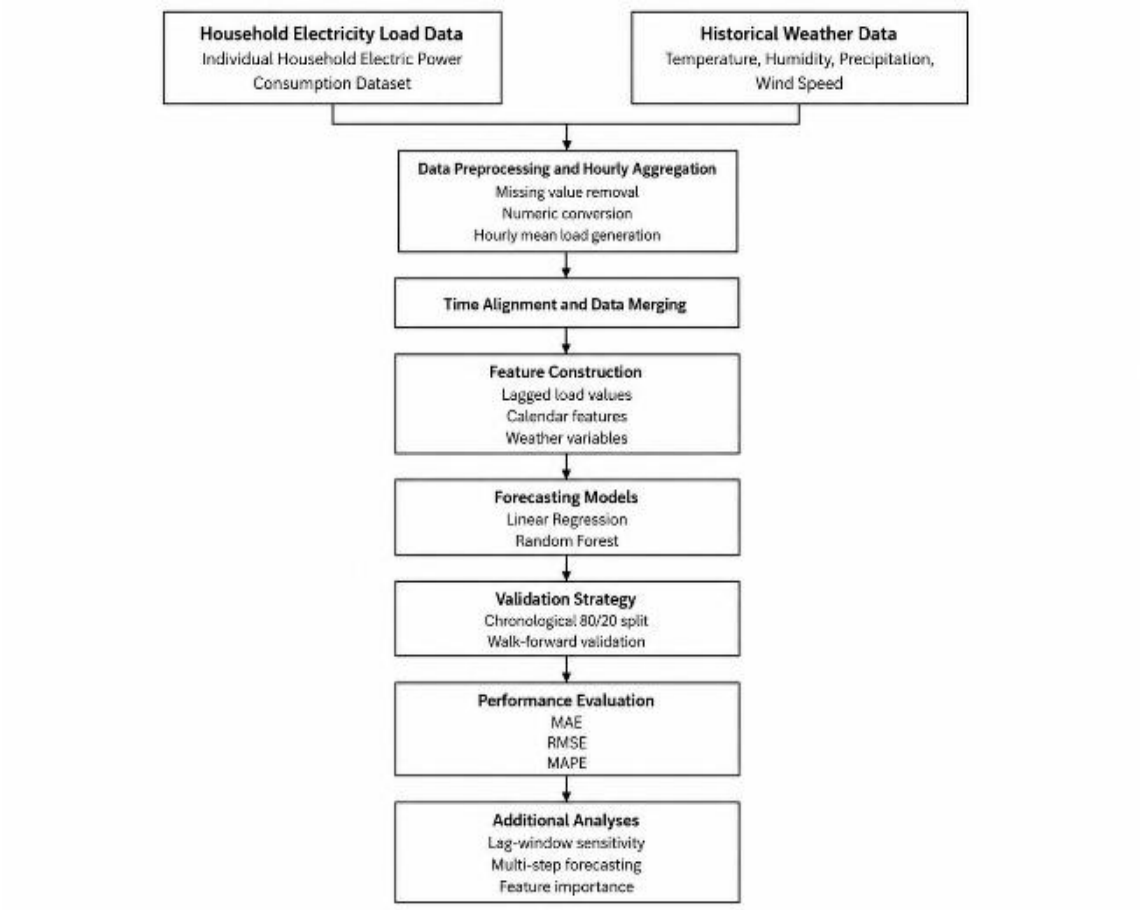


Figure 1: Overall workflow of the proposed short-term electricity load forecasting framework.

2.3 Time-Series Formulation of the Forecasting Problem

Short-term load forecasting is formulated as a supervised regression problem. Let L(t) denote the hourly mean household active power consumption at time t. The objective is to estimate the future electricity load at forecasting horizon h:

$$\hat{L}(t+h) = f\big(x(t)\big), \qquad (1)$$

where $\hat{L}(t+h)$ is the predicted load value h hours ahead, f is the forecasting model, and $x(t)$ is the feature vector available at time t.

In this study, the forecasting horizon h is varied in order to analyze how prediction accuracy changes as the forecast extends further into the future. The main one-hour-ahead prediction task is used for baseline model comparison, while additional experiments consider horizons of 3, 6, 12, and 24 hours.

The feature vector $x(t)$ is constructed from three groups of predictors:

1) Lagged historical load values.
2) Calendar-related temporal attributes.
3) Exogenous meteorological variables.

This formulation makes it possible to investigate both the internal temporal dependency structure of the electricity load series and the additional contribution of external weather conditions.

2.4 Feature Construction

The forecasting features are designed to capture short-term inertia, cyclic demand behavior, and weather-related influences on household electricity consumption.

Lag-based features are constructed from previous load observations. For a selected lag-window length n, the model receives the set of historical values:

$$L(t-1), L(t-2), \ldots, L(t-n). \qquad (2)$$

Four lag-window configurations are examined in this study: n = 6, n = 12, n = 24, and n = 48 hours.

This allows assessment of how much historical memory is required for accurate short-term prediction.

Calendar-related temporal features include hour of day, day of week, and month. These attributes are used to represent recurring daily, weekly, and seasonal consumption patterns that are common in residential electricity demand.

In the extended exogenous configuration, four meteorological variables are additionally included:

1) Air temperature at 2 m.
2) Relative humidity at 2 m.
3) Precipitation.
4) Wind speed at 10 m.

These weather variables are temporally aligned with the forecasted hour and are used to evaluate whether external environmental conditions improve forecasting accuracy compared with models relying only on internal load history and calendar information.

2.5 Forecasting Models

Two regression models are applied in this study: Linear Regression and Random Forest. Linear Regression is used as an interpretable benchmark model, while Random Forest is selected as a nonlinear ensemble method capable of capturing more complex relationships among lagged load values, temporal attributes, and weather variables.

The Linear Regression model estimates the future load as a weighted linear combination of input features:

$$\widehat{L(t+h)} = \beta_0 + \beta_1 x_1(t) + \beta_2 x_2(t) + \cdots + \beta_p x_p(t), \quad (3)$$

where β_0 is the intercept term, β_1, β_2, …, β_p are regression coefficients, and $x_1(t)$, $x_2(t)$ …, $x_p(t)$ are the predictors included in the feature vector $x(t)$.

Random Forest is an ensemble learning method that combines the outputs of multiple regression trees.

Each individual tree produces its own load forecast, and the final prediction is calculated as the average of all tree-based outputs:

$$\hat{L}(t+h) = \frac{1}{M}\sum_{m=1}^{m} Tm(X(t)), \quad (4)$$

where M is the total number of regression trees and $Tm(X(t))$ denotes the prediction generated by the m-th tree.

To evaluate the effect of external weather information, both forecasting algorithms are tested under two feature configurations:

1) Endogenous configuration: lagged load values and calendar-related temporal attributes.
2) Exogenous configuration: lagged load values, calendar-related temporal attributes, and meteorological variables.

This design enables a controlled comparison of linear and nonlinear forecasting models, as well as an assessment of whether weather-related predictors improve short-term household load prediction.

2.6 Validation Strategy

A chronological validation strategy is applied in order to preserve the temporal structure of the forecasting problem and avoid information leakage. Since electricity load forecasting is inherently time dependent, random shuffling of observations is not used. Instead, the observations are ordered by timestamp before training and testing.

For the main one-hour-ahead model comparison, lag-window analysis, and forecasting horizon experiments, the first 80% of the available observations are used for model training, while the remaining 20% are reserved for out-of-sample testing. This approach ensures that future load information is never used to predict earlier periods.

In addition to the fixed chronological split, walk-forward validation is conducted to assess the temporal robustness of the final Random Forest model with weather variables. Three sequential expanding-window splits are considered:

1) The first 50% of the data are used for training, and the following 15% are used for testing.
2) The first 65% of the data are used for training, and the following 15% are used for testing.
3) The first 80% of the data are used for training, and the remaining 20% are used for testing.

This validation procedure provides additional insight into how forecasting accuracy changes across different temporal segments of the dataset.

2.7 Evaluation Metrics

Forecasting performance is evaluated using three complementary error metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). These indicators provide different perspectives on prediction quality and are widely used in short-term load forecasting studies. MAE measures the average absolute difference between the actual and predicted load values:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |yi - \hat{y}i|, \quad (5)$$

where:

- yi is the actual load value;
- $\hat{y}i$ is the predicted load value;
- N is the number of test observations.

RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}i)^2}, \quad (6)$$

and assigns stronger penalties to larger deviations. Therefore, RMSE is particularly informative when evaluating errors during local demand peaks or abrupt changes in electricity consumption.

MAPE is calculated as:

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right|. \quad (7)$$

MAPE expresses the forecasting error in percentage terms and enables relative interpretation of model accuracy. However, its magnitude may become large when actual load values are comparatively low. For this reason, MAPE is interpreted jointly with MAE and RMSE rather than as a standalone measure.

3 RESULTS AND DISCUSSION

This section presents the experimental results obtained from the revised forecasting framework. The analysis begins with a comparison of linear and nonlinear models under endogenous and exogenous feature configurations. It then examines the influence of lag-window length, temporal robustness under walk-forward validation, forecasting horizon sensitivity, and the relative importance of individual predictors.

3.1 Baseline Model Comparison and Effect of Weather Variables

The first experiment evaluates one-hour-ahead forecasting performance using a 24-hour lag window. Four model configurations are compared: Linear Regression with endogenous temporal features, Linear Regression with additional weather variables, Random Forest with endogenous temporal features, and Random Forest with additional weather variables. The results are summarized in Table 1.

Table 1 shows that Random Forest consistently outperforms Linear Regression in both feature configurations. For the endogenous models, the nonlinear ensemble approach reduces MAE from 0.452 to 0.399 and RMSE from 0.606 to 0.555. This confirms that short-term household electricity demand contains nonlinear temporal relationships that are not fully captured by a purely linear baseline.

The inclusion of meteorological variables produces a modest improvement in both models. For Linear Regression, MAPE decreases from 64.19% to 63.11%. For Random Forest, MAE decreases from 0.399 to 0.397, RMSE decreases from 0.555 to 0.552, and MAPE decreases from 54.47% to 54.27%. Although the gain is limited, the result indicates that weather conditions provide additional explanatory information beyond historical load and calendar-related temporal features.

The Random Forest model with weather variables achieves the best overall one-hour-ahead performance and is therefore used as the principal configuration for the subsequent lag-window, walk-forward, forecasting horizon, and feature importance analyses. At the same time, the remaining MAPE above 50% indicates that residential household load remains highly volatile and difficult to predict accurately using a compact feature set. Consequently, MAPE is interpreted together with MAE and RMSE rather than in isolation.

The visual behavior of the best-performing configuration is illustrated in Figure 2, which compares actual hourly household load values with the predictions generated by the Random Forest model with weather variables for a representative segment of the test period.

Table 1: One-hour-ahead forecasting performance with and without weather variables.

Model	Feature Configuration	MAE	RMSE	MAPE (%)
Linear Regression	Endogenous	0.452	0.606	64.19
Linear Regression	Endogenous and Weather	0.450	0.605	63.11
Random Forest	Endogenous	0.399	0.555	54.47
Random Forest	Endogenous + Weather	0.397	0.552	54.27

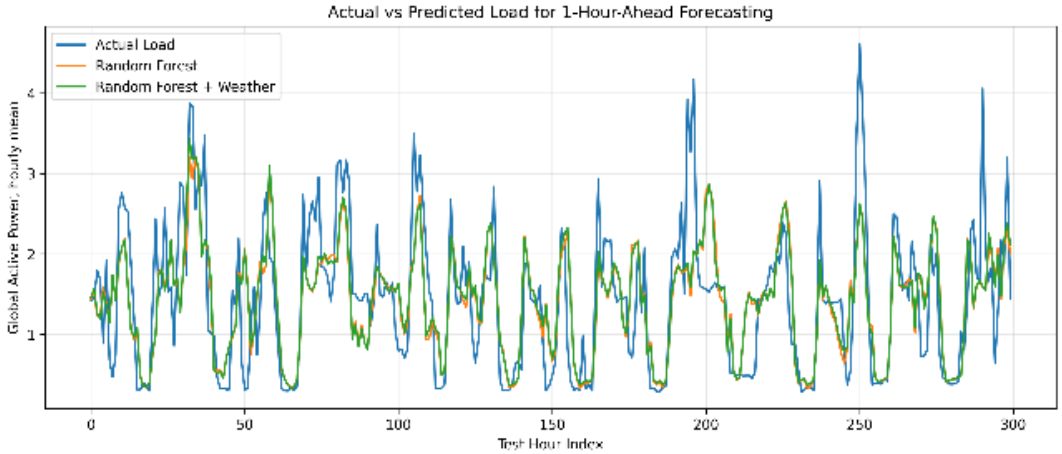


Figure 2: Actual and predicted household electricity load for one-hour-ahead forecasting using Random Forest with weather variables.

As shown in Figure 2, the model follows the general temporal evolution of household electricity demand and reproduces the main fluctuations in the test sequence. The predicted curve responds to changing demand levels and captures the broader short-term pattern of the observed load.

Nevertheless, local deviations remain visible during sharper changes in consumption. Abrupt peaks and rapid transitions are smoothed by the model, which is consistent with the RMSE value reported in Table 1. This behavior is typical for residential load forecasting, where individual appliance usage and irregular household activity can generate sudden variations that are difficult to infer from historical load, calendar attributes, and weather information alone.

The figure therefore supports the numerical results: the Random Forest model with weather variables provides a reasonable short-term representation of load dynamics, but it does not eliminate the intrinsic volatility of single-household electricity consumption.

3.2 Lag-Window Sensitivity Analysis

To evaluate the influence of historical memory length on forecasting performance, the Random Forest model with weather variables was tested using lag windows of 6, 12, 24, and 48 hours. This experiment assesses whether the inclusion of a longer sequence of previous load values provides additional predictive benefit for one-hour-ahead household electricity forecasting. The results are presented in Table 2.

Table 2: Lag-window sensitivity analysis for Random Forest with weather variables.

Lag Window (h)	MAE	RMSE	MAPE (%)
6	0.410	0.565	58.26
12	0.399	0.554	54.96
24	0.397	0.552	54.27
48	0.397	0.553	54.00

Table 2 shows that expanding the lag window from 6 to 12 hours produces a clear improvement in forecasting accuracy. MAE decreases from 0.410 to 0.399, while MAPE decreases from 58.26% to 54.96%. This indicates that very short historical windows may not provide sufficient temporal context for household load prediction.

A further increase from 12 to 24 hours leads to a smaller additional improvement. The 24-hour window achieves MAE of 0.397, RMSE of 0.552, and MAPE of 54.27%, indicating that one full daily cycle of historical information is useful for representing recurring residential demand patterns.

The 48-hour configuration yields the lowest MAPE value of 54.00% and a marginally lower MAE than the 24-hour configuration. However, the difference is very small, and RMSE slightly increases from 0.552 to 0.553. Therefore, the results suggest that most of the useful historical information is already captured within a 24-hour window, while extending the memory to 48 hours provides only limited additional benefit.

The behavior of MAE across the tested lag-window configurations is illustrated in Figure 3.

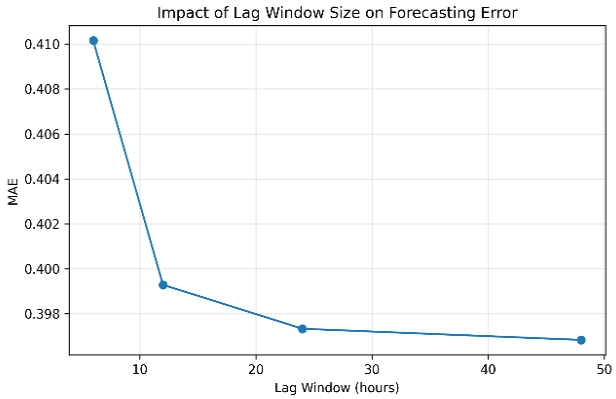


Figure 3: Impact of lag-window size on MAE for Random Forest with weather variables.

Figure 3 confirms that the largest improvement occurs when the historical window is increased from 6 to 12 hours. Beyond 24 hours, the MAE curve becomes nearly flat, which reinforces the conclusion that extending the lag structure further does not substantially improve short-term forecasting performance. This finding supports the use of a compact 24-hour lag window as a balanced configuration for the subsequent analysis.

3.3 Walk-Forward Validation

To assess the temporal robustness and generalization capability of the selected forecasting configuration, walk-forward validation was conducted using three sequential expanding-window splits. In each split, the Random Forest model with weather variables was trained only on earlier observations and evaluated on a later unseen period. The results are summarized in Table 3.

Table 3: Walk-forward validation results for Random Forest with weather variables.

Split	MAE	RMSE	MAPE (%)
1	0.460	0.645	57.60
2	0.417	0.591	57.68
3	0.397	0.552	54.27

Table 3 demonstrates that the forecasting model maintains a comparable level of performance across different temporal segments of the dataset, although the magnitude of the error varies over time. The first split produces MAE of 0.460 and RMSE of 0.645, indicating relatively higher prediction difficulty in the corresponding testing period. In the second split, MAE decreases to 0.417 and RMSE to 0.591. The third split provides the best performance, with MAE of 0.397 and RMSE of 0.552.

The MAPE values remain within a relatively narrow range from 54.27% to 57.68%. This indicates that the model does not experience an extreme collapse in generalization capability when evaluated on later unseen intervals. At the same time, the differences among splits confirm that residential household load is temporally heterogeneous and that forecasting accuracy depends on the characteristics of the specific evaluation period.

Overall, the walk-forward validation results support the temporal consistency of the proposed forecasting framework while also highlighting the persistent challenge of accurately modeling irregular household electricity consumption.

3.4 Multi-Step Forecasting Behavior

The effect of forecasting horizon length was evaluated for 1-, 3-, 6-, 12-, and 24-hour-ahead prediction tasks. For all horizons, the Random Forest model with weather variables and a 24-hour lag window was used. This experiment examines how forecasting accuracy changes as the target prediction point moves further into the future. The results are presented in Table 4.

Table 4: Multi-step forecasting performance of Random Forest with weather variables.

Forecast Horizon (h)	MAE	RMSE	MAPE (%)
1	0.397	0.552	54.27
3	0.423	0.580	58.90
6	0.428	0.587	60.12
12	0.435	0.597	61.58
24	0.444	0.606	64.07

Table 4 shows a gradual deterioration in forecasting performance as the prediction horizon

increases. The one-hour-ahead configuration achieves the best accuracy, with MAE of 0.397, RMSE of 0.552, and MAPE of 54.27%. When the horizon is extended to 3 hours, MAE increases to 0.423 and MAPE rises to 58.90%.

The degradation continues for longer forecasting horizons. At 24 hours ahead, MAE reaches 0.444, RMSE increases to 0.606, and MAPE rises to 64.07%. This pattern indicates that the available lagged load history, calendar attributes, and weather variables become progressively less informative as the forecasted time point moves further away from the latest known observations.

Although the error increase is not abrupt, the results clearly demonstrate that the proposed framework is most suitable for ultra-short-term forecasting, particularly within the 1- to 3-hour-ahead range. For longer prediction horizons, more advanced temporal modeling strategies or richer explanatory information may be required.

The increase in forecasting error across prediction horizons is illustrated in Figure 4.

Figure 4 visually confirms the progressive growth of MAE as the forecasting horizon extends. The curve rises most noticeably between the one-hour and three-hour horizons and continues to increase more gradually thereafter. This behavior reflects the accumulation of prediction uncertainty in multi-step residential load forecasting and supports the conclusion that short-term temporal dependencies are most informative for near-future demand estimation.

3.5 Feature Importance Analysis

To identify the predictors that contribute most strongly to one-hour-ahead load forecasting, feature importance analysis was performed for the Random Forest model with weather variables. This analysis provides additional interpretation of the forecasting framework by showing which historical, temporal, and meteorological variables have the greatest influence on the model output.

The relative importance of the most influential predictors is shown in Figure 5.

Figure 5 shows that the most influential predictor is the immediately preceding load value, load_lag_1, with an importance score of approximately 0.260. This confirms the strong short-term inertia of household electricity demand, where the current consumption level is highly informative for the next-hour forecast.

The second most important predictor is load_lag_23, followed by the cyclic hour-related variables. This result indicates that daily periodicity plays a substantial role in the forecasting process. The prominence of load values observed approximately one day earlier, together with hour-of-day indicators, supports the interpretation that residential demand contains recurring daily behavioral patterns.

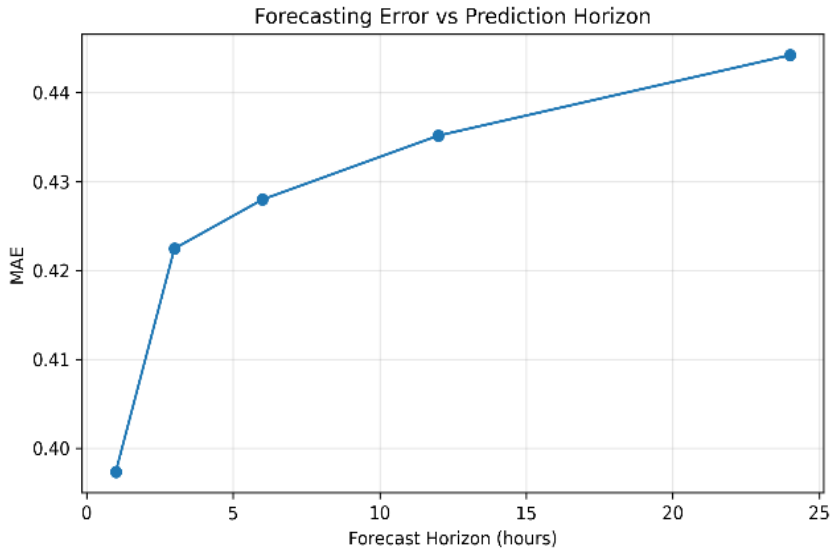


Figure 4: Forecasting error versus prediction horizon for Random Forest with weather variables

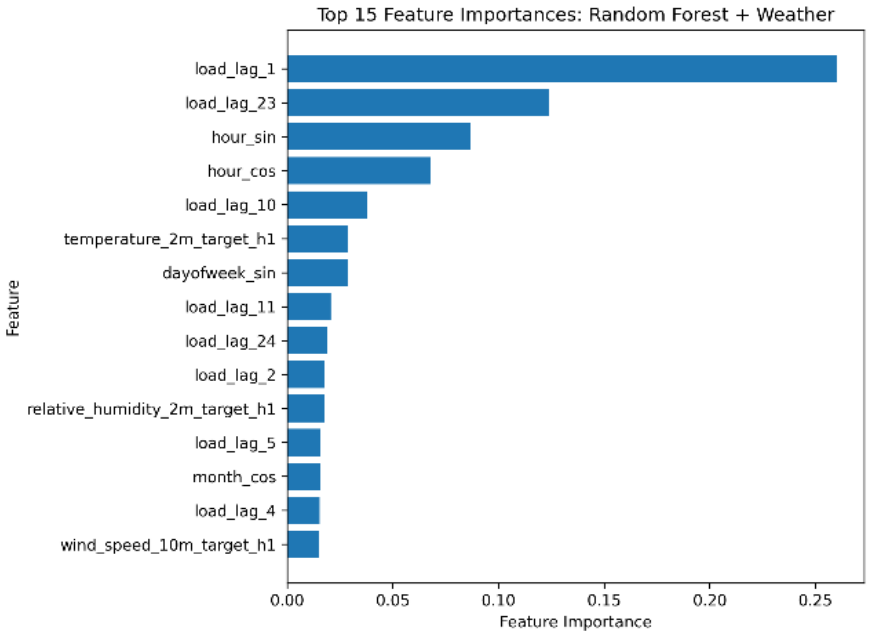


Figure 5: Feature importance ranking for the Random Forest model with weather variables.

Among the meteorological variables, air temperature appears within the ten most influential predictors. Although its contribution is smaller than that of the dominant lag-based features, its presence confirms that weather conditions provide measurable additional information for short-term household load prediction. This finding is consistent with the model comparison results in Table 1, where the inclusion of weather variables produced a modest improvement in forecasting accuracy.

Overall, the feature importance analysis demonstrates that the proposed model is primarily driven by recent load history and daily cyclic behavior, while exogenous weather information acts as a secondary but relevant explanatory factor.

4 CONCLUSIONS

This study investigated short-term household electricity load forecasting through a supervised machine learning framework focused on temporal dependency analysis. In response to the need for a more comprehensive forecasting design, the revised methodology incorporated not only lagged load observations and calendar-related temporal features, but also exogenous meteorological variables, including air temperature, relative humidity, precipitation, and wind speed.

The experimental results showed that Random Forest consistently outperformed Linear Regression in one-hour-ahead forecasting. The best-performing configuration, Random Forest with weather variables, achieved MAE of 0.397, RMSE of 0.552, and MAPE of 54.27%. The inclusion of weather information produced a modest but measurable improvement compared with the endogenous-only Random Forest model, indicating that meteorological conditions contribute additional explanatory value, although they do not dominate the forecasting process.

Lag-window sensitivity analysis demonstrated that extending historical memory from 6 to 24 hours improves model accuracy, while further expansion to 48 hours provides only marginal additional benefit. This finding suggests that a one-day historical window is sufficient to capture most of the relevant short-term temporal structure in the analyzed residential load series.

Walk-forward validation confirmed that the forecasting framework maintains broadly consistent performance across sequential temporal segments, while still reflecting the intrinsic heterogeneity of household electricity consumption. Multi-step horizon analysis further revealed that forecasting accuracy gradually deteriorates as the prediction horizon increases, supporting the conclusion that the proposed approach is most appropriate for ultra-

short-term applications, particularly within the 1- to 3-hour-ahead range.

Feature importance analysis showed that the immediately preceding load value, daily lag-related information, and hour-of-day variables are the dominant predictors. Air temperature also appeared among the most influential variables, confirming that weather effects are present but secondary relative to internal load dynamics.

Overall, the study demonstrates that ensemble learning can effectively capture nonlinear temporal dependencies in household electricity demand and that the inclusion of meteorological variables provides a limited but meaningful enhancement. At the same time, the remaining MAPE values indicate that single-household load forecasting remains a challenging problem due to irregular consumer behavior and abrupt demand fluctuations. Future research should therefore consider richer behavioral indicators, appliance-level information, probabilistic forecasting strategies, and advanced temporal models to further improve predictive robustness.

REFERENCES

[1] M. G. Pinheiro, S. C. Madeira, and A. P. Francisco, "Short-term electricity load forecasting-A systematic approach from system level to secondary substations," Applied Energy, vol. 332, p. 120493, 2023.

[2] A. A. Mamun, M. Sohel, N. Mohammad, M. S. H. Sunny, D. R. Dipta, and E. Hossain, "A comprehensive review of the load forecasting techniques using single and hybrid predictive models," IEEE Access, vol. 8, pp. 134911-134939, 2020.

[3] J. W. Taylor and P. E. McSharry, "Short-term load forecasting methods: An evaluation based on European data," IEEE Transactions on Power Systems, vol. 22, no. 4, pp. 2213-2219, 2007.

[4] T. Hong, P. Wang, and H. L. Willis, "A naive multiple linear regression benchmark for short-term load forecasting," in Proc. IEEE Power and Energy Society General Meeting, Detroit, MI, USA, pp. 1-6, 2011.

[5] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," IEEE Transactions on Smart Grid, vol. 10, no. 3, pp. 3125-3148, 2019.

[6] M. Imani, "Electrical load-temperature CNN for residential load forecasting," Energy, vol. 227, p. 120480, 2021.

[7] P. Ran, K. Dong, X. Liu, and J. Wang, "Short-term load forecasting based on CEEMDAN and Transformer," Electric Power Systems Research, vol. 214, p. 108885, 2023.

[8] X. Wen, J. Liao, Q. Niu, N. Shen, and Y. Bao, "Deep learning-driven hybrid model for short-term load forecasting and smart grid information management," Scientific Reports, vol. 14, p. 13720, 2024.

[9] G. Hebrail and A. Berard, "Individual Household Electric Power Consumption," UCI Machine Learning Repository, 2006.

[10] H. Hersbach et al., "The ERA5 global reanalysis," Quarterly Journal of the Royal Meteorological Society, vol. 146, no. 730, pp. 1999-2049, 2020.