BY ANHALT UNIVERSITY OF APPLIED SCIENCES

PROCEEDINGS OF THE 13TH INTERNATIONAL CONFERENCE ON APPLIED INNOVATIONS IN IT

VOLUME 13
ISSUE 1

EDITION
Hochshule Anhalt

Proceedings of the 13th International Conference on Applied Innovations in IT

Volume 13 | Issue 1

Koethen, Germany
26 April 2025

Proceedings of the International Conference on Applied Innovations in IT (ICAIIT), Koethen, Germany, March 13th, 2025. This volume contains peer-reviewed papers presented at the conference along with selected papers accepted for publication in the general proceedings on April 26th. The publication focuses on applied research in automation, communication, and information technologies. The presented research results are particularly relevant for both academic researchers and development engineers working with theoretical foundations and practical applications in these technical domains.

ISBN: 978-3-96057-182-7 (Online)
ISSN: 2199-8876

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

Postfach 1458
D-06354 Koethen, Germany
Phone: +49 3496 67 2327
Email: eduard.siemens@hs-anhalt.de
Web: http://icaiit.org

Content

Section 1 Communication and Data Transport Technologies

Subashkumar Rajanayagam, Max Andreas Ingrisch, Pascal Müller, Patrick Jahn and Stefan Twieg
Enhancing Voice Activity Detection for an Elderly-Centric Self-Learning Conversational Robot Partner in Noisy Environments………………………………………………………………… 1

Paul Kobina Arhin Jnr, James Ben Hayfron-Acquah, Frimpong Twum and Alfred Kumah
Designing Advanced Cryptographic Solutions for Cloud Storage Security Through Dual-Layer Encryption Protocols…………………………………………………………………………… 9

Oleksandra Yaroshevska and Veronika Kirova
Potential of Using the Ant Colony Optimization Algorithm for Optimal Network Path Selection……… 17

Serhii Siden, Roman Tsarov, Mohammed Jamal Salim, Kateryna Shulakova, Saad Malik Talha and Liliia Bodnar
Optimizing the Location of 5G Network Base Stations Taking into Account Intra-System Interference………………………………………………………………………… 27

Svitlana Skvortsova, Tetiana Symonenko and Kira Hnezdilova
Optimizing Routine Educational Tasks through Prompt Engineering: A Comparative Study of AI Chatbots…………………………………………………………………………… 35

Caroline Schon, Rong Huang, Hanna Hessenmüller, Simon Przybyl and Johannes Tümler
Classification of the Topicality and Relevance of Evaluation Tools for VR Applications……….…….. 43

Iryna Haidai, Liudmyla Suvorova, Nataliia Pankovyk, Yuliia Herasymchuk and Olha Zadoienko
Challenges and Strategies of Remote Foreign Language Learning for Students with Autism Spectrum Disorder……………………………………………………………………………….. 51

Olha Dienichieva, Iryna Biliak, Natalia Plakhotniuk, Tetiana Verhun and Kateryna Parnus
The Significance of Personality-Oriented Technology in Higher Education in Ukraine……………….. 63

Mariana Kovtoniuk, Olena Soia, Olena Kosovets and Liubov Tiutiun
Peculiarities of Forming the Professional Culture of a Mathematics Teacher by Means of Digital Technologies in Higher Education Institutions of Ukraine……………………………………….. 71

Section 2 Data Analysis and Processing

Victor Tikhonov, Yevhen Vasiliu, Eduard Siemens, Oksana Vasylenko, Olena Tykhonova, Kateryna Shulakova and Olexandr Demchenko
Quantum Field Tensor Model of Telecommunication Network Objects Interaction Based on Lie Groups………………………………………………………………………………… 79

Sofiia Drozd, Nataliia Kussul and Hanna Yailymova
A Novel Approach for Rapid Detection of Forest Degradation and Diseases Through Anomaly Analysis of Sentinel-2 Spectral Data………………………………………………………………. 87

Adil Khalil, Adel Abed Al Wahaab and Ahmed Abbas
A new Hybrid Metaheuristic Model for Image Edge Detection………………………………… 95

Andrii Liashenko and Larysa Globa
A Comprehensive Method for Anomaly Detection in Complex Dynamic IoT Systems………..… 101

Tetiana Hromko and Liudmyla Panchuk
Multidimensional Model of Sebastian Unger's Idiostyle in Poetic Creativity: Corpus Analysis and NLP Methods…………………………………………………………………,… 109

Igor Melnyk, Mykhailo Skrypka, Alina Pochynok and Olga Demyanchenko
Advanced Software Package for Estimatiom Boundary Trajectories of Electron Beams and Other Practical Applications………………………………………………………...… 119

Sergiy Prykhodko, Lidiia Makarova and Andrii Pukhalevych
Statistical Analysis of the Three-Dimensional Data of Software Metrics RFC, CBO, and WMC that are not Normally Distributed………………………………………………………… 127

Volodymyr Rusinov and Nikita Basenko
Exploration of the Efficiency of SLM-Enabled Platforms for Everyday Tasks………………… 133

Stepan Davydovych, Maryna Davydovych, Kateryna Parnus, Iryna Haidai and Liudmyla Suvorova
Engaging Students in the Digital Age: A Communicative-Pragmatic and AI-Driven Analysis of Online Prospectuses……………………………………………………. 139

Yuliia Prokop
Automated Assessment of Student Queries in Redis…………………………………………… 147

Section 3 Control and Automation

Max Andreas Ingrisch, Subashkumar Rajanayagam, Ingo Chmielewski and Stefan Twieg
Calibration of the Open-Vocabulary Model YOLO-World by Using Temperature Scaling…………….. 155

Andrej Sarjaš and Dušan Gleich
VSC Adaptive Nonuniform Sampling Approach for Resource Optimization in
Networked Controlled System…………………………………………………………………… 161

Olga Ivashchuk, Vyacheslav Fedorov, Zhanat Kenzhebayeva, Oleg Ivashchuk, Moldir Allaniyazova, Kaiyrbek Makulov and Gulnar Kaziyeva
Methods and Algorithms for Decision-Making in Agro-Industrial Environmental Management……….. 169

Svitlana Antoshchuk, Oleksii Danchuk and Tetiana Kunup
Scalable Video Processing and Frame Analysis System for Automated Monitoring of
Chicken Behavior Based on Artificial Intelligence Technologies……………………………………… 177

Yehor Zheliazkov
System for Simultaneous Control of Multiple Unmanned Aerial Vehicles……………………………… 185

Nataliia Zaiets, Nataliia Lutska and Lidiia Vlasenko
Integrated Machine Learning Models for Bakery Product Defect Prediction…………………………… 193

Aleksy Kwilinski, Maciej Kardas and Nataliia Trushkina
Application of X-bar R Control Charts for Process Efficiency Monitoring:
A Data-Driven Approach in Quality Management………………………………………………… 203

Ivan Omeko, Stefan Twieg and Martin Obert
Sensor-Based Gait Analysis: A Comparative Study of Ultrasonic and Laser Sensors for
Gait Monitoring in Rollator-Assisted Walking…………………………………………………… 213

Eugene Karashevych, Svitlana Sulima and Mariia Skulysh
Intelligent Learning Support System……………………………………………………………… 223

Nataliia Trushkina
The Relationship Between Artificial Intelligence and Critical Infrastructure Development:
Bibliometric Analysis…………………………………………………………………………… 231

Section 4 Power Engineering

Boysori Yuldoshov, Azamat Narbayev, Tulkin Buzrukov, Gayrat Raimov, Jasur Khaliyarov, Sojida Raupova and Sirojiddin Toshpulatov
Analysis of Electrical Parameters of a Metal Collector Photovoltaic-Thermal Device……………….… 243

Farkhodjon Turaev, Isroil Yuldoshev, Bozorbek Botirov, Fazliddinov Sadulla
Experimental Comparative Study on the Use of Photovoltaic Converters for Cooling Photovoltaic Modules………………………………………………………………………….… 249

Isroil Yuldoshev, Sanjar Shoguchkarov, Shakhnoza Rustamova, Bozorbek Botirov, Tulkin Jamolov and Zukhra Jurayeva
Additional Protection Device for Grid Inverter of Photovoltaic Station……………………………… 255

Section 5 Information Technology in Economics

Olha Kovalchuk, Oksana Tulai, Olga Zavytii, Ludmila Babala, Roman Ivanytskyy and Pavlo Chopyk
Global Energy Systems Modeling: Structure and Environmental Impacts………………………… 261

Oleksandr Sharyhin, Oksana Klochko and Ihor Tverdokhlib
Automation of Plagiarism and AI Detection in IT Students Papers with a Software Tool Using API……………………………………………………………………………………… 269

Oksana Vasylenko, Kateryna Shulakova, Liliia Bodnar, Eduard Siemens, Roman Tsarov and Oleksandr Vudvud
Global DevOps Market Dynamics and Workforce Analysis………………………………………… 279

Iryna Polishchuk, Yuriy Dovhan, Oksana Yakushevska, Larysa Dovhan, Oksana Kravchenko and Viktoriia Petrova
Digital Marketing: Tools, Trends and Perspectives…………………………………………………… 287

Aleksy Kwilinski, Nataliia Trushkina, Yuliia Remyha and Tamila Patlachuk
Impact of Artificial Intelligence on Customer Relationship Management: a Bibliometric Analysis…………………………………………………………………………… 295

Mirzaeva Azizakhon and Mirzamakhmudova Madina
Priority Directions for Life Insurance Development in Uzbekistan……………………………… 305

Tatyana Oklander, Anastasiia Pandas, Mykhailo Oklander, Oksana Yashkina and Hlib Chaban
Digital Marketing Tools for Preserving Cultural Heritage in the Context of Sustainable Urban Development………………………………………………………………… 313

Drishtanti Saha, Pratik Sarkar, Markus Holz, Cornelia Scott and Halina Falfushynska
Multivariate Statistical Analysis of the Drivers of Women's Empowerment in Labor Market………… 321

SECTION 1

COMMUNICATION AND DATA TRANSPORT TECHNOLOGIES

Proceedings of the 13th International Conference on Applied Innovations in IT (ICAIIT)

APRIL 2025

Enhancing Voice Activity Detection for an Elderly-Centric Self-Learning Conversational Robot Partner in Noisy Environments

Subashkumar Rajanayagam[1], Max Andreas Ingrisch[1], Pascal Müller[2], Patrick Jahn[2] and Stefan Twieg[1]
[1]*Department of Electrical, Mechanical and Industrial Engineering, Anhalt University of Applied Sciences, Bernburger Str. 57, 06366 Koethen, Germany*
[2]*Health Service Research Working Group|Acute Care, Department of Internal Medicine, Faculty of Medicine, Martin-Luther-University Halle-Wittenberg, Magdeburger Str. 12, 06112 Halle (Saale), Germany*
{subashkumar.rajanayagam, max.ingrisch, stefan.twieg}@hs-anhalt.de, {pascal.mueller, patrick.jahn}@uk-halle.de

Keywords: Voice Activity Detection, Human Robot Interaction, Conversational Robot Partner, Elderly-Centric.

Abstract: Voice Activity Detection (VAD) is a root component in Human-Robot Interaction (HRI), especially for use cases such as a self-learning personalized conversational robot partner designed to support elderly users with high acceptance. While state-of-the-art, lightweight deep-learning–based VAD models achieve high precision, they often struggle with low recall in environments with significant background noise or music. In contrast, traditional lightweight rule-based VAD methods tend to yield higher recall but at the expense of precision. These limitations can negatively affect user experience, particularly among elderly individuals, by causing frustration from missed spoken inputs and reducing overall usability and acceptance of the conversational robot partners. This study investigates noise-suppressing preprocessing techniques to enhance both the recall and precision of existing VAD systems. Experimental results demonstrate that effective noise suppression prior to VAD processing substantially improves voice detection accuracy in noisy settings, ultimately promoting better interaction quality in elderly-centric robotic applications. Moreover, optimal sample rate, frame duration, thresholds and voice activity modes were identified for the robot Double3—the conversational robot partner platform for seniors in a care home, co-creatively developed by reflecting with the nursing staff. An open-source dataset and a dataset collected and annotated in-house with the Double3 robot were evaluated for robustness in benchmarks.

1 INTRODUCTION

Voice Activity Detection (VAD) is a fundamental component in Human-Robot Interaction (HRI), particularly for a self-learning Conversational Robot Partner (CRP) designed for elderly users operating in indoor noisy environments. Robust VAD is essential for enabling these systems to effectively separate spoken content from ambient noise, thereby supporting self-learning personalized processes and maintaining contextual awareness through speaker diarization, speech recognition, and large language model-based text generation. Achieving high recall ensures no relevant speech is missed, while high precision minimizes the downstream tasks' processing of non-speech segments. This balance enhances interaction quality and optimizes resource utilization, energy consumption, and bandwidth efficiency. [1]

State-of-the-art lightweight deep learning VAD Silero [2] and rule-based VAD WebRTC [3] excel in high precision and recall respectively in environments with background noise or music. However, Silero often suffers from low recall, while WebRTC struggles with low precision. This can frustrate the elderly users and ultimately reduce the usability and acceptance of the system. A noise suppression methodology as a preprocessing step to the VADs is investigated to address these shortcomings without sacrificing their strengths. This approach effectively mitigates the limitations inherent to each system, preserving their advantages while enhancing overall performance.

The CRP born from the project EduXBot[1,2] is developed for German senior residents at a nursing home facility in indoor noisy environments by evaluating and reflecting with the nursing staff. Most modern VADs are mainly developed for English and

[1] https://www.hs-anhalt.de/eduxbot/uebersicht.html
[2] https://drks.de/search/de/trial/DRKS00034195

their range of dominant frequencies lies from 1000 Hz to 2000 Hz, whereas the German language is from 125 Hz to 3000 Hz. [4] However, language agnosticism has been studied and results show no significant difference between the accuracy of German and English languages in various SOTA VAD benchmarks. [5] Therefore, this is not considered in this study.

Furthermore, since supervised learning based VADs such as Silero are susceptible to acoustic mismatch in different environments, experiments with the open-source dataset AVA Speech and a custom-recorded and annotated dataset based on the Double3 robot were conducted. [6][7] This is to prove that no separate training would be necessary, which is normally inherent in unsupervised VADs and noise grouping. [8][9][10][11] Finally, the optimal sampling rate, frame duration and VAD Modes/thresholding value were identified for the CRP development platform Double3 in which the VAD was deployed. The Double3 platform demonstrated consistent acceptance among nursing staff, as evidenced by increased System Usability Scale (SUS) scores across three iterations, rising progressively from below 70 to above 75. [12]

2 BACKGROUND

2.1 Base Models

Two lightweight voice activity detection (VAD) models in our experiments are employed: one based on deep learning and the other rule-based. The deep-learning-based model utilized is Silero, a lightweight VAD trained on datasets covering over 6,000 languages. Silero employs Short-Time Fourier Transform (STFT) as features and operates efficiently at a sampling rate of 16 kHz with a 30ms frame duration. Processing each frame requires less than 1ms on a single CPU thread, making it highly suitable for real-time applications [2]. Additionally, the model allows adjustable parameters, including threshold levels and minimum durations for speech and silence, enabling effective customization for various use cases.

In contrast, WebRTC is a rule-based VAD model utilizing Gaussian Mixture Models (GMM). It processes six frequency bands ranging from 80 Hz to 4000 Hz, represented as log-energy values. Optimized for real-time web communication through fixed-point arithmetic operations, WebRTC is highly suitable for deployment on edge devices due to its lightweight design. It supports multiple sampling rates of 8, 16, and 32 kHz and accepts frame durations of 10, 20, and 30ms. The presence of voice activity is determined by applying predefined rule-based criteria, enhancing its efficiency in real-time scenarios [3].

2.2 Datasets

Our experiments were conducted using two distinct datasets. The first dataset is the open-source AVA Speech dataset, which consists of densely annotated audio clips recorded at sampling rates of 44.1 kHz and 48 kHz. The dataset provides approximately 40 hours of audio segmented into four categories: no speech, clean speech, speech with music, and speech with noise. Specifically, clean speech segments account for 14.55%, speech with music 13.46%, speech with noise 24.32%, and no speech 47.68% of the total duration. The signal-to-noise ratios (SNR) for clean speech, speech with music, and speech with noise segments are 40.8 dB, 11.7 dB, and 16.2 dB, respectively [7].

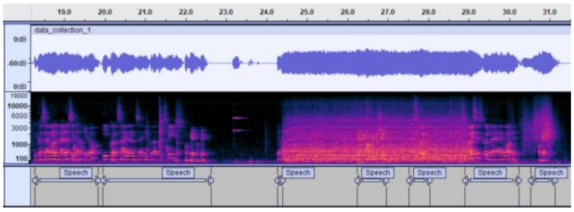


Figure 1: Annotation procedure with Audacity for the in-house Double3 dataset for the first 31.5 seconds.

The second dataset was collected in-house using a Double3 robot's microphone, incorporating typical environmental background noises such as coffee machines, doors opening and closing, keyboard typing, mouse clicks, and background music in an indoor environment similar to the nursing home's communal space. This verification dataset, recorded at a sampling rate of 44.1 kHz, comprises approximately 20 minutes of densely annotated audio. Annotations were conducted manually using the Audacity software [13], by carefully listening, analyzing waveforms and spectrograms (see Figure 1). The speech segments were directly labeled using Audacity, while the remaining non-speech segments were identified and annotated using a Python script based on gaps between speech segments. The speech segment accounts for 58.02% and non-speech 41.98% of the total duration.

2.3 Noise Suppression

Recent advancements in lightweight, real-time deep learning-based noise suppression and speech enhancement methods, such as DeepFilterNet, have significantly improved the feasibility of deploying efficient models on edge devices, including robotic platforms equipped with embedded systems [14]. In our study, the DeepFilterNet2 (DFN2) model was explored as a preprocessing step prior to the VAD stage. [14] DFN2 leverages the harmonic structure inherent in speech signals, achieving efficient speech enhancement with a real-time factor of 0.04, thus ensuring suitability for real-time applications. [14]

3 METHODOLOGY

The first objective of this study is to evaluate the performance and limitations of the lightweight VAD models Silero (deep-learning-based) and WebRTC (GMM-based) by benchmarking them against the AVA Speech and in-house Double3 datasets. Initially, WebRTC was benchmarked against both datasets across various sampling rates (8, 16, and 32 kHz), frame durations (10, 20, and 30ms), and VAD operational modes (0, 1, 2, and 3). The VAD mode "0" corresponds to the lowest detection threshold, while mode "3" represents the highest detection threshold, with incremental steps (1 and 2) between each mode. This resulted in 36 benchmark combinations for each dataset, measuring recall and precision. The benchmarks were calculated based on the total duration of speech detected rather than the number of segments, ensuring better accuracy.

The second objective is to investigate the impact of employing the DeepFilterNet2 (DFN2) noise suppression preprocessing technique to mitigate the limitations identified in the lightweight VAD models while preserving their performance. The final objective is to determine the optimal sampling rate, frame duration, threshold values, and operational modes for the VAD models when applied to the Double3-based CRP.

For Silero, only the threshold parameter was varied, evaluated at equal intervals from 0 to 1, while maintaining fixed sampling rate and frame duration (16 kHz and 30ms, respectively), as these are the only supported settings relevant for this study. The evaluation process was subsequently repeated after applying the DFN2 preprocessing step to assess its impact on Silero's performance.

4 RESULTS

This section follows the structure outlined in the previous methodology section. First, the baseline results obtained using WebRTC and Silero are presented. Subsequently, the results achieved after incorporating the DNF2 noise suppression technique as a preprocessing step are discussed.

4.1 Base Results

WebRTC was benchmarked using the AVA Speech dataset and our in-house Double3 dataset across various sampling rates, frame durations, and VAD modes.

Silero, on the other hand, was benchmarked exclusively at a sampling rate of 16 kHz and a frame duration of 30ms. However, various thresholds ranging from 0 to 1 were explored to assess their impact on performance.

4.1.1 WebRTC without Preprocessing

When looking across the sampling rate axis (x-axis) in Figure 2, no significant changes can be seen in the color and size representing the precision and recall for AVA Speech. In Figure 3, this is seen clearly, where the three colors representing the different sample rates overlay each other in most cases.

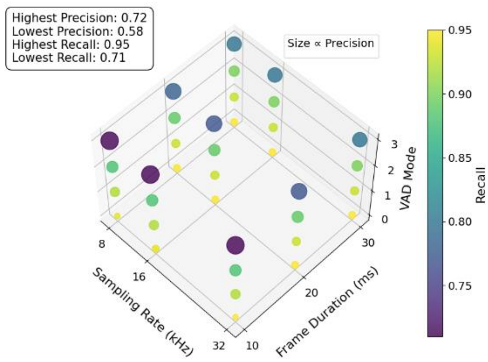


Figure 2: 5D plot of precision vs recall for WebRTC on AVA Speech dataset with various sampling and frame rates without noise suppression preprocessing.

A similar behavior is observed concerning frame duration, although predominantly at lower VAD modes (0 to 2), where threshold values are lower and less aggressive. At higher thresholds—specifically, VAD mode = 3 – the precision peaks at 72% with a frame duration of 10ms; however, this configuration also results in the lowest recall of only 70.8%.

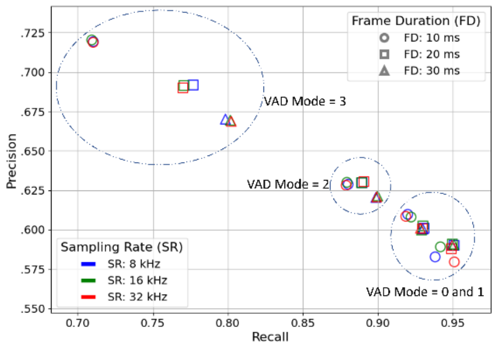


Figure 3: 4D plot of precision vs recall for WebRTC on AVA Speech dataset with various sampling and frame rates without noise suppression preprocessing. VAD modes are bounded by circles.

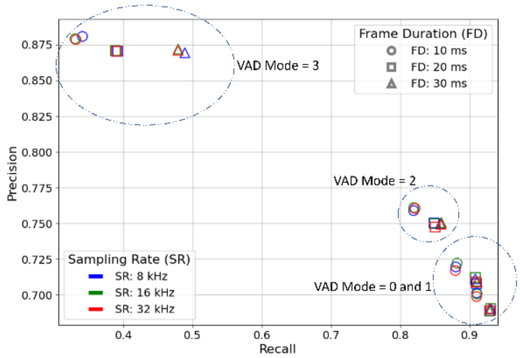


Figure 4: 4D plot of precision vs recall for WebRTC on Double3 in-house dataset for various sampling and frame rates without noise suppression preprocessing. VAD modes bounded by circles.

However, the VAD mode demonstrates a clear relationship between precision and recall for the AVA Speech dataset. As the VAD mode increases, precision improves, whereas recall decreases. At the highest VAD mode (mode 3), precision peaks at 72%, but recall drops to its lowest at 70.8%. Conversely, at the lowest VAD mode (mode 0), precision decreases to the lowest 58%, while recall reaches its highest value of 95%.

Similar observations were made when benchmarking WebRTC on the in-house Double3 dataset, as illustrated in Figure 4. Specifically, the highest precision achieved was 88%, corresponding to the lowest recall value of 33%. Conversely, the highest recall of 93% was associated with the lowest precision of 69%.

Although WebRTC demonstrates high recall for both datasets, peaking above 90%, the precision remains comparatively lower, peaking at 72% for the AVA Speech dataset. These trade-off highlights an inherent limitation of WebRTC VAD, wherein precision and recall cannot be simultaneously optimized, necessitating a compromise between the two metrics.

4.1.2 Silero without Preprocessing

The benchmark results for Silero, illustrating precision versus recall for both the AVA Speech and Double3 datasets, are shown in Figure 5. While Silero demonstrates high precision, exceeding 90% in certain thresholds, its recall remains comparatively lower, peaking at 73% for the AVA Speech dataset and 62% for the Double3 dataset when precision is 90%. Consequently, an optimal balance between precision and recall is not observed in either dataset.

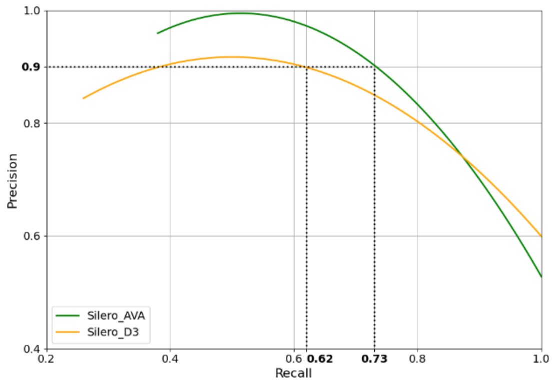


Figure 5: 2D plot of precision vs recall for Silero on AVA Speech (AVA) and Double3 (D3) in-house datasets for various thresholds without noise suppression preprocessing.

In contrast to WebRTC, which demonstrates higher recall but lower precision, Silero exhibits lower recall yet higher precision. However, both systems inherently involve a trade-off between these metrics, which constrains the ability to optimize them simultaneously.

4.2 After Noise Suppression

In this section, the results obtained after applying noise suppression (DNF2) as a preprocessing step to the VAD models WebRTC and Silero are summarized.

4.2.1 WebRTC after DNF2

Benchmarking WebRTC with DNF2 as a noise suppression preprocessing step yielded major improvements in precision. As illustrated in Figures 6 and 7, incorporating the noise suppression model (DNF2) as a preprocessing step significantly improved the precision of the WebRTC VAD model on both the AVA Speech and Double3 datasets. Specifically, for the AVA Speech dataset, the highest precision increased by 15.5%, rising from 72% to 87.5%. Similarly, the Double3 dataset exhibited a notable improvement, with the highest precision increasing by 5% from 76% to 81%, at VAD mode 2.

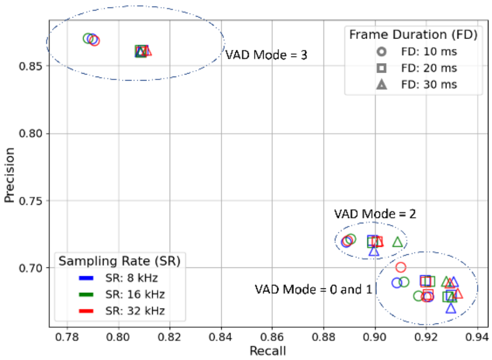


Figure 6: 4D plot of precision vs recall for WebRTC on AVA Speech dataset for various sampling and frame rates after noise suppression. VAD modes bounded by circles.

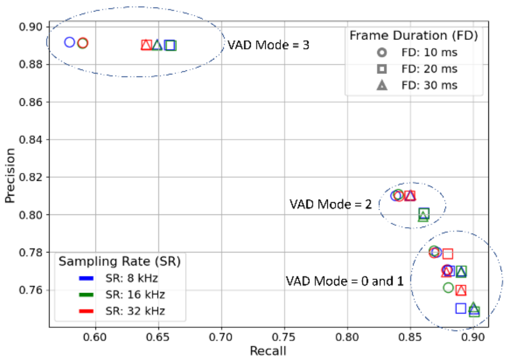


Figure 7: 4D plot of precision vs recall for WebRTC on Double3 dataset for various sampling and frame rates after noise suppression. VAD modes bounded by circles.

The sampling rate showed minimal influence on performance, consistent with observations from the baseline model without noise suppression. However, the frame duration of 10ms in the AVA Speech dataset at VAD Mode 3 achieves recall less than 80%, whereas at frame durations 20ms and 30ms are above 80%. Consequently, an 8 kHz sampling rate combined with a 20ms frame duration was selected to minimize sample size, inference time, and resource usage and maintain a precision and recall above 80% for both datasets.

With this configuration, WebRTC combined with the noise suppression preprocessing step achieved the desired balance between precision and recall. Specifically, at VAD mode 3 for the AVA Speech dataset and VAD mode 2 for the Double3 dataset, both precision and recall exceeded 80%, as seen in Figure 8. Specifically, at this configuration, the AVA Speech dataset precision increased by 17% (from 69% to 86%) and the Double3 dataset precision increased by 5% (from 75% to 80%). In both instances, recall was maintained (Double3 at 85%) or increased (AVA Speech by 3%).

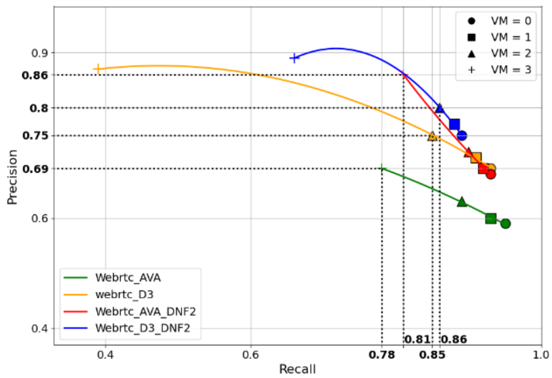


Figure 8: 2D plot of precision vs recall for WebRTC on AVA Speech and Double3 in-house datasets after noise suppression for the selected 8 kHz sampling rate and 20ms frame duration with various thresholds.

4.2.2 Silero after DNF2

Benchmarking Silero with DNF2 as a noise suppression preprocessing step yielded notable improvements in recall.

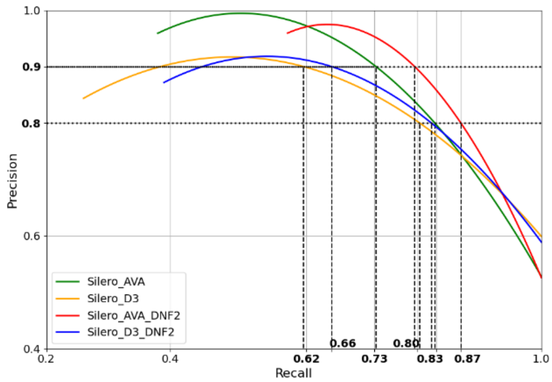


Figure 9: 2D plot of precision vs recall for Silero on AVA Speech and Double3 in-house datasets with various thresholds after noise suppression.

For the AVA Speech dataset, recall increased by 7%, rising from 73% to 80%, at a precision of 90% with a threshold of 0.20, as shown in Figure 9. Similarly, for the Double3 dataset, recall improved by 4% (from 62% to 66%) at the same precision level at a threshold of 0.25.

Although these enhancements are promising, recall remains below 80% in most instances. To explore a more optimal operating point, configuring the system to achieve 80% precision resulted in a recall increase of 4% (from 83% to 87%) for AVA Speech, while for Double3, it rose by 3% (from 80% to 83%) at the thresholds of 0.20 and 0.25 respectively.

5 CONCLUSIONS

In this study, two lightweight VAD models enhanced by incorporating noise suppression (DNF2) as a preprocessing step are evaluated. The addition of this preprocessing significantly improved WebRTC VAD performance, increasing precision by 17% on the AVA Speech dataset and by 5% on the Double3 dataset. This enhancement enabled WebRTC to achieve an optimal combination of precision and recall, with both metrics surpassing 80% at VAD mode 3 for the AVA Speech dataset and mode 2 for the Double3 dataset. The optimal sampling rate of 8 kHz, 20ms frame duration and VAD mode 2 is identified for WebRTC, aligning to the Double3 dataset. Additionally, the deep learning-based Silero model demonstrated improved recall, increasing by 4% for the AVA Speech dataset and 3% for the Double3 dataset, all when precision is at 80%.

Since WebRTC achieved better recall (85%) as compared to Silero (83%) when precision is above 80% for the Double3 dataset, WebRTC with DNF2 noise suppression is selected for the CRP at VAD mode 2.

During the analysis of our results, it was observed that speech characterized by whispering, shouting, and high-pitched tones was frequently missed by both VAD models. This was done by analyzing the before and after spectrograms and listening to the cropped regions of the false negative regions of the audio datasets. Consequently, further optimization and additional benchmarking efforts are required to enhance the detection capabilities and overall performance of these lightweight VAD models.

ACKNOWLEDGMENTS

We gratefully acknowledge the support provided by the Research, Transfer, and Start-Up Center (Forschungs-, Transfer- und Gründerzentrum), Anhalt University of Applied Sciences, and the state of Saxony-Anhalt. We also extend our sincere gratitude to the Federal Ministry of Education and Research (BMBF) for their generous support under grant number 03WIR3118A, which significantly contributed to the successful completion of this study.

REFERENCES

[1] S. Yadav, P. A. D. Legaspi, M. S. O. Alink, A. B. J. Kokkeler and B. Nauta, "Hardware Implementations for Voice Activity Detection: Trends, Challenges and Outlook," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 70, no. 3, pp. 1083-1096, March 2023, doi: 10.1109/TCSI.2022.3225717.

[2] Silero Team, "Silero Models: State-of-the-Art Speech Processing Models," GitHub repository, 2024 [Online]. Available: https://github.com/snakers4/ silero-models. [Accessed: 09-Feb-2025].

Google. (2011) WebRTC [Online]. Available:https://webrtc.org/ [accessed Feb 2025].

[3] Atlantis-vzw.com, ""Learning a foreign language with more ease"." Accessed: Feb. 02, 2025. [Online]. Available: https://www.atlantis-vzw.com/vreemde-talen?lang=en.

[4] M. Sharma, S. Joshi, T. Chatterjee, and R. Hamid, "A comprehensive empirical review of modern voice activity detection approaches for movies and TV shows," Neurocomputing, vol. 494, pp. 116–131, Jul. 2022, doi: 10.1016/J.NEUCOM.2022.04.084.

[5] R. M. Patil and C. M. Patil, "Unveiling the State-of-the-Art: A Comprehensive Survey on Voice Activity Detection Techniques", doi: 10.1109/APCIT62007.2024.10673721.

[6] S. Chaudhuri, J. Roth, D. P. Ellis, A. C. Gallagher, L. Kaver, R. Marvin, C. Pantofaru, N. Reale, L. G. Reid, K. W. Wilson, and Z. Xi, "AVA-Speech: A densely labeled dataset of speech activity in movies," Aug. 2018. [Online]. Available: https://arxiv.org/abs/1808.00606.

[7] X. L. Zhang and M. Xu, "AUC optimization for deep learning-based voice activity detection," Eurasip J. Audio, Speech, Music Process., vol. 2022, no. 1, pp. 1–12, Dec. 2022, doi: 10.1186/S13636-022-00260-9/TABLES/7.

[8] T. Yoshimura, T. Hayashi, K. Takeda, and S. Watanabe, "End-to-End Automatic Speech Recognition Integrated with CTC-Based Voice Activity Detection," ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc., vol. 2020-May, pp. 6999–7003, May 2020, doi: 10.1109/ICASSP40776.2020.9054358.

[9] F. Gu, M.-H. Chung, M. Chignell, S. Valaee, B. Zhou, and X. Liu, “A Survey on Deep Learning for Human Activity Recognition,” 2021. A Surv. Deep Learn. Hum. Act. Recognition. ACM Comput. Surv, vol. 54, no. 8, p. 177, 2021, doi: 10.1145/3472290.

[10] S. Twieg and B. Zimmermann. Acoustic clustering for vehicle based sounds, 2010.

[11] P. Müller, H. K. Gali, S. Rajanayagam, S. Twieg, P. Jahn, and S. Hofstetter. Making Complex Technologies Accessible Through Simple Controllability: Initial Results of a Feasibility Study. Applied Sciences, 15(2), 1002, 2024 [Online]. Available: https://doi.org/10.3390/app15021002.

[12] Audacity(R) software is copyright (c) 1999-2014 Audacity Team. [Website: http://audacity.sourceforge.net/. It is free software distributed under the terms of the GNU General Public License.] The name Audacity(R) is a registered trademark of Dominic Mazzoni.

[13] H. Schröter, T. Rosenkranz, A. Escalante-B., and A. Maier, "DeepFilterNet: Perceptually motivated real-time speech enhancement," in Proc. 17th Int. Workshop Acoustic Signal Enhancement (IWAENC), 2022.

Designing Advanced Cryptographic Solutions for Cloud Storage Security Through Dual-Layer Encryption Protocols

Paul Kobina Arhin Jnr[1], James Ben Hayfron-Acquah[2], Frimpong Twum[2] and Alfred Kumah[3]
[1]*Department of Computer Science and I.T, University of Cape Coast, CC-123-1749, Ghana*
[2]*Department of Computer Science, KNUST, AK-385-1973, Ghana*
[3]*Department of IT Infrastructure, University of Health and Allied Sciences, PMB 31, Ghana*
parhin@ucc.edu.gh , jbhayfron-acquah@knust.edu.gh, ftwum.cos@knust.edu.gh , akumah@uhas.edu.gh

Keywords: Cloud Storage, Data Integrity, Hybrid Cryptosystem, ChaCha20-Poly1305, ElGamal Encryption.

Abstract: This paper presents a novel hybrid cryptosystem combining ElGamal and ChaCha20-Poly1305 for securing data in cloud storage environments. Elgamal, an asymmetric encryption algorithm is utilized for secure key exchange to ensure the safe transmission of keys. ChaCha20-Poly1305, a high-performance symmetric encryption algorithm is used for efficient data encryption and authentication, thereby addressing the performance limitations of using traditional asymmetric algorithms alone. Using ChaCha20-Poly1305 for bulk data encryption and transmission significantly improves speed and reduces computational overhead compared to using only asymmetric algorithm. This makes it ideal for modern cloud storage applications. This novel system is made to provide robust defence against common cloud-based attacks such as eavesdropping, replay attacks, brute force and man-in-the-middle attacks. By making use of the strengths of both asymmetric and symmetric encryption algorithms, the system will ensure high security with better performance. The proposed hybrid Algorithm was tested and compared with the traditional Elgamal alone, and it achieved 83% faster encryption, 79.5% faster decryption and 69.5% lower memory usage. This shows its efficiency for secure cloud data storage and transmission.

1 INTRODUCTION

Cloud storage has in our modern world, become a vital key component of information technology infrastructure [1]. Cloud storage provides a large range of computational resources and storage capabilities [2]. This technology has completely transformed data-centric centres, ranging from healthcare and banking to e-commerce and artificial intelligence by offering flexible and scalable services [3][4]. However, the proliferation of cloud-based solutions has also increased concerns about data security and privacy [5]. As companies and individuals continually entrust sensitive information to cloud platform services, there is an urgent need for robust encryption techniques to protect against unauthorized access, data breaches, and other cyber threats [6]. Most classical cloud computing security models use symmetric algorithms for fast data encryption and asymmetric key algorithms for safe key management and secure key exchange [7]. Although most of these traditional models use well known algorithms like AES for bulk data encryption and have withstood many cyber-attacks, they are still not without challenges [8]. These challenges include performance overhead, implementation complexity, and the constantly changing cryptographic environment, which demands continuous scrutiny to maintain confidentiality and integrity in the face of emerging cyber-attacks [9].

ElGamal cryptosystem is a highly secured asymmetric algorithm whose strength lies in the difficulty of the discrete logarithm problem over large finite fields [10]. This makes it very difficult for attackers to solve or breakthrough. Its resilience against known cryptanalytic attacks makes it a reliable choice for key exchange and encryption [11]. This makes the Elgamal cryptosystem secure for encrypting data, exchanging keys, and creating digital signatures [12]. Until today, Elgamal still remains a trusted algorithm in public-key cryptography [13]. Even though Elgamal offers strong security through its complex mathematical foundations, this same complexity can highly affect its performance by making it relatively slow when used alone, particularly in high-volume, cloud-based data [14].

On the other hand, ChaCha20-Poly1305 has emerged as a state-of-the-art Authenticated Encryption with Associated Data (AEAD) algorithm known for its speed, resistance to timing attacks, and robust authentication guarantees [15]. Originally, it was developed as a more efficient alternative to AES in software implementations but has received increasing attention for its consistent performance across a range of platforms and architectures [16].

To address the challenge of the Elgamal cryptosystem performance limitations, this paper proposes a novel hybrid cryptographic scheme that combines the strengths and key exchange capabilities of ElGamal with the efficiency and authenticity of ChaCha20-Poly1305. By creating a hybrid of asymmetric key exchange technique with a high-performance AEAD cipher, the work seeks to introduce an exceptional end-to-end secure system for cloud storage and data transmission that addresses both confidentiality and integrity requirements.

By leveraging ElGamal for secure session key establishment and ChaCha20-Poly1305 for fast bulk encryption and authentication, this hybrid approach ensures strong and robust security for cloud storage and transmission while maintaining high speed. Even if some parts of the cloud infrastructure are compromised, it would be very hard for an attacker to figure out the encryption keys or alter the data without being detected. This makes the system highly secure and efficient for protecting sensitive information stored and transmitted in the cloud.

2 REVIEW OF RELATED LITERATURE

Cloud computing has indeed changed and transformed how data is stored and processed, but it also poses serious security risks. In order to defend and protect cloud storage systems against these risks like data breaches and cyberattacks, various cryptographic algorithms and security mechanisms have been proposed. This literature review examines different cryptographic techniques and security implementations in cloud computing storage.

Peng et al. [17] conducted a thorough review of secure cloud storage based on cryptographic algorithms. The findings from their work proved the importance of encryption algorithms such as AES (Advanced Encryption Standard), RSA (Rivest-Shamir-Adleman), and DES (Data Encryption Standard) in securing cloud storage. According to their research, classical cryptography may result in performance overhead even though it offers robust security.

Salman and Sulaiman [18] also conducted a thorough analysis into multiple cryptographic algorithms, including Blowfish, RC6, and Feistel, and compared their efficiency in cloud environments. The findings from the research showed that, lightweight encryption algorithms such as Blowfish are more suitable for cloud applications due to their lower computational costs.

Ahmad and Garko [19] also conducted a thorough review into hybrid cryptographic algorithms in cloud computing. They analyzed the effectiveness of these algorithms in securing large-scale cloud environments. Their study identified gaps in user authentication and key management in hybrid encryption techniques.

Sharma et al. [20] also conducted an investigation into hybrid cryptographic algorithms in cloud storage, analyzing their encryption and decryption speeds. Their findings also suggested that, even though hybrid cryptographic techniques provide enhanced security, they often introduce additional computational overhead.

A hybrid cryptographic scheme that integrates Elliptic Curve Cryptography (ECC) and Triple Data Encryption Standard (TDES) was proposed by Kaur and Jain [21]. Their experimental results showed that ECC-TDES enhances data security while maintaining accuracy, but it slightly increases encryption time.

Mendez [22] also introduced DNA cryptography as a novel method for enhancing cloud security. The study proposed a Bi-Directional DNA Encryption Algorithm (BDEA), which combines bio-molecular principles to enhance encryption strength. The research discovered that, the novel DNA cryptography offers a promising direction for unbreakable encryption, but practical implementations require further exploration.

Kavya and Acharva [23] conducted a comparative analysis on different homomorphic encryption schemes. They concluded that, while homomorphic encryption is highly secure, it remains computationally expensive.

Madhavi and Sivareddy [24] as well examined using public key encryption algorithms for secure cloud storage. Their work introduced the Dual-Server Public Key Encryption with Keyword Search (DS-PEKS) to mitigate keyword guessing attacks. The findings suggested that DS-PEKS enhances privacy in cloud-based search operations but requires further optimization.

Gadad and Anbusezhiyan [25] also came up with a novel approach for securing cloud storage. This was done by encoding plaintext into an intermediate compressed format before encryption. Their study proved that, integrating compression with encryption enhances both security and transmission efficiency.

Sharma et al. [26] also conducted a comparative analysis into symmetric encryption algorithms for securing cloud storage. This was done by comparing AES, Blowfish, and DES based on encryption time and throughput. Their findings suggested that optimizing symmetric algorithms can improve both security and transmission speed.

Sasikumar and Nagarajan [27] also conducted a comprehensive meta-analysis on various cryptographic security algorithms for cloud computing. Their study highlighted the integration of machine learning with cryptographic algorithms to enhance security. The research proposed the use of AI-driven anomaly detection to identify security threats in real time.

2.1 Cryptographic Algorithms Used in Leading Cloud Storage Systems

This section discusses some leading cloud storage systems and the cryptographic algorithms they use to enhance security.

Amazon Web Services (AWS) Cloud Storage: This is one of the most commonly used cloud storage system**.** Amazon S3 and AWS Key Management Service (KMS) use Advanced Encryption Standard (AES-256) encryption to secure data at rest. For data in transit, AWS employs Transport Layer Security (TLS) to ensure confidentiality [28].

Google Cloud Storage: Another well-known cloud system is the Google cloud storage. Google Cloud Storage provides multiple encryption options, including AES-256 for server-side encryption and RSA for key management. Google also supports customer-managed encryption keys (CMEK) and customer-supplied encryption keys (CSEK) to enhance data protection [29].

Microsoft Azure Storage: Microsoft Azure Storage using AES-256 encryption for data in storage and uses TLS for data in transit. Azure also implements a hybrid cryptographic approach combining RSA and AES for optimal security [30].

IBM Cloud Object Storage: IBM Cloud also makes use of AES-256 encryption for data in storage and TLS for secure data transfer. It also supports envelope encryption, where an AES data key is encrypted using an RSA key for additional security [31].

Most cloud storage systems use various cryptographic algorithms, with TLS, AES-256 and RSA being widely implemented for strong security. These hybrid encryption approaches are used to enhance data protection, ensuring confidentiality and integrity in cloud environments.

3 METHODOLOGY

At this stage, we implement the hybrid cryptosystem algorithm combining ElGamal and ChaCha20-Poly1305 for cloud storage. we will focus on the four (4) stages of the algorithm, namely key generation, encryption, transmission of data, and decryption. The goal is to combine the asymmetric ElGamal encryption (for securely exchanging a shared key) with the symmetric ChaCha20-Poly1305 encryption (for efficiently encrypting the actual data).

3.1 Key Generation Stage

The first stage in the hybrid cryptosystem involves key generation for both the ElGamal and ChaCha20-Poly1305 algorithms:

1) Elgamal Key Generation (Asymmetric):
   - First, we select a large prime number, p and a generator, g such that g is a primitive root modulo p. These values must be publicly known.
   - Then, Choose a private key x (a random number between 1 and p - 2)
   - Calculate the corresponding public key

   $$y = g^x \bmod p \qquad (1)$$

   Public key: (p, g, y)
   Private key: (x)
2) ChaCha20-Poly1305 Key Generation (Symmetric encryption):
   - First, Choose a random 256-bit key, ksym for ChaCha20
   - Next, Select a random nonce (12-byte) n_{nonce} for the ChaCha20 encryption.
3) The key, ksym and the nonce, nnonce are kept secret between the sender and receiver, but they need to be shared securely.
   Symmetric key: k_{sym}
   Nonce: n_{nonce}

OUTPUT: Public key: **(**p, g, y**),** Private key: (x **)**
ChaCha20 **:** k_{sym}, n_{nonce}

3.2 Encryption Stage

Once the keys are generated, the encryption process follows. Here, the sender will have to securely transmit the symmetric key, $\boldsymbol{k_{sym}}$ to the receiver. This is done by using the ElGamal algorithm, where a random integer, $\boldsymbol{k}$ is chosen by the sender. The sender will then go ahead and compute two values: $C_1 = g^k$ mod p and $C_2 = (y^k \cdot m)$ mod p. where m is the symmetric key $\boldsymbol{k_{sym}}$, which has been converted to an integer. The result will become an ElGamal ciphertext (C_1, C_2), which contains the encrypted symmetric key.

The symmetric key is then securely transmitted. The actual $\boldsymbol{D}$ (This refers to the data to be transmitted to the cloud and could be files or information to be uploaded to the cloud storage) is encrypted using the ChaCha20-Poly1305 algorithm. The encryption of the data will then yield two outputs, which are the ciphertext ($\boldsymbol{C_{data}}$), which is encrypted form of the data as well as the authentication tag (t***ag***), which is also used to verify the integrity and authenticity of the data during decryption. Here, the ChaCha20-Poly1305 algorithm offers both confidentiality and integrity assurance, ensuring that the encrypted data cannot be tampered with during transmission.

ElGamal encrypts k_{sym} to produce

$$C_{elgamal} = (C_1, C_2), \qquad (2)$$

ChaCha20-Poly1305 encrypts D to produce C_{data}, $_{tag}$.

3.3 Transmission of Data

Once the encryption process is completed, the next stage involves the transmission of the encrypted data over a channel which may not be secured to a cloud storage platform. The data transmitted here includes: $C_{elgamal} = (C_1, C_2)$, the ChaCha20-Poly1305 ciphertext C_{data}, and the authentication tag, *tag*. The nonce n_{nonce} which is used for ChaCha20 encryption is also sent to ensure that, the same nonce is used for decryption by the receiver. This process allows for secure storage and retrieval of the data from the cloud, ensuring that both the key and data are encrypted.
($C_{elgamal}$, C_{data}, *tag*, n_{nonce}) is sent over the cloud storage or secure channel.

3.4 Decryption

The final stage in the hybrid algorithm is the decryption of the data by the receiver. Here, in order for the receiver to recover the symmetric key, $\boldsymbol{k_{sym}}$, the ElGamal ciphertext (C_1, C_2) will have to be decrypted. The receiver will use their private key, $\boldsymbol{x}$ to compute the shared secret $s = C^x_1$ mod p and its inverse modulo $\boldsymbol{p}$, $\boldsymbol{s^{-1}}$. The symmetric key k_{sym}, is then recovered by computing $m = (C_2 \cdot s^{-1})$, where m is the encrypted symmetric key.
Once the symmetric key $\boldsymbol{k_{sym}}$ is successfully recovered, the receiver can decrypt the actual data which is encrypted with ChaCha20-Poly1305. Using the symmetric key k_{sym} and the nonce n_{nonce} , the receiver then applies the ChaCha20-Poly1305 decryption algorithm to retrieve the original plaintext D. Using the authentication tag, *tag*, if the integrity check passes, the decrypted data is successfully recovered and can now be used by the receiver.

Elgamal Decryption:

$$\text{Compute } s = C^x_1 \bmod p \qquad (3)$$

$$\text{Compute } s^{-1} \bmod p$$

$$\text{Recover } k_{sym} = (C_2 \cdot s^{-1}) \bmod p \qquad (4)$$

ChaCha20-Poly1305 Decryption:

$$\text{Decrypt data } D = \text{ChaCha20-Poly1305\_Decrypt}(k_{sym}, n_{nonce}, C_{data}, {}_{tag})$$

Verification:

Tag from the decrypted ciphertext is computed:

$$tag' = \text{ChaCha20-Poly1305\_GenerateTag}(k_{sym}, n_{nonce}, \text{D}) \qquad (5)$$

The recomputed *tag'* is compared with the received authentication *tag:*

```
if (tag' == tag) {
Data integrity verified
}
Else {
Data integrity verification failed
}
```

If the tags match, data integrity is verified. If they don't match, the data is considered compromised.

This novel hybrid cryptosystem combines ElGamal and ChaCha20-Poly1305. It offers a great solution for cloud storage security. For cloud storage services, the hybrid encryption system, it ensures that encryption is done efficiently on the client system before uploading. It will ensure rapid encryption of large volumes of information that will enable quick and secure transmissions even in the midst of systems where bandwidth and device capabilities are limited. The fast encryption algorithm, ChaCha20-Poly1305 minimizes device resource usage while Elgamal secures the session keys to defend the cloud against any breaches. Security is high compared to traditional RSA or Elgamal. This hybrid will also offer good performance on modern processors even without any hardware accelerators like AES-NI.

4 EXPERIMENTAL RESULTS

The Hybrid cryptosystem which combines Elgamal and ChaCha20-Poly1305 for cloud storage, was implemented on an i5 HP desktop running windows 10 with 3.2GHz of speed and RAM of 8GB. The results are shown in the tables and graphs. The data size used was 5MB. Table 1 shows the Encryption and decryption speed comparison between Elgamal and the hybrid of Elgamal and Elgachat across different key sizes, whiles Table 2 shows the memory usage comparison between Elgamal and Elgachat across different key sizes.

Table 1: Encryption and decryption speed comparison between Elgamal and the hybrid of Elgamal.

Key Size (in Bits)	Cryptographic Algorithms	Encryption Speed (ms)	Decryption Speed (ms)
1024	ElGamal	45.2	85.9
	ElgaChat	14.2	22.2
2048	ElGamal	93.1	132.8
	ElgaChat	17.9	34.4
3072	ElGamal	138.7	193.2
	ElgaChat	22.7	46.6
4096	ElGamal	188.5	261.4
	ElgaChat	34.9	53.6

Figure 1 shows a significant improvement in the proposed hybrid system as compared to Elgamal alone used in encryption. The hybrid system outperforms the Elgamal system in both Encryption and Decryption speeds across all key sizes, having encryption being up to 83.6% faster and with decryption being up to about 79.5% faster.

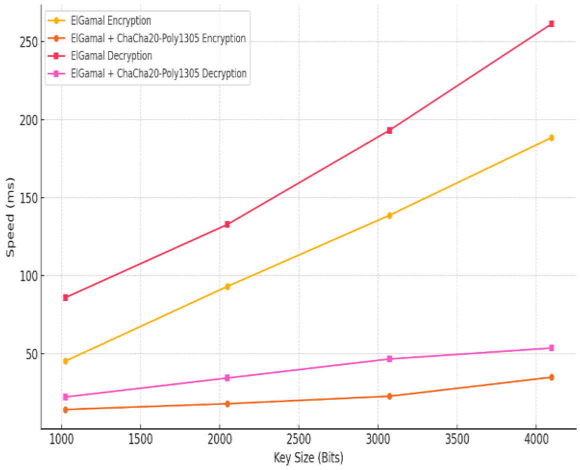


Figure 1: Encryption and decryption speed comparison between Elgamal and the hybrid of Elgamal and ChaCha20-Poly1305

Table 2: Memory Consumption comparison between Elgamal and the hybrid of Elgamal.

Key Size (in Bits)	Cryptographic Algorithms	Encryption (MB)	Decryption (MB)
1024	ElGamal	1.05	1.28
	ElgaChat	0.32	0.63
2048	ElGamal	1.55	1.92
	ElgaChat	0.53	0.95
3072	ElGamal	2.17	2.32
	ElgaChat	1.12	1.35
4096	ElGamal	3.29	4.01
	ElgaChat	1.53	2.01

Comparing the hybrid algorithm with the standalone Elgamal in terms of memory consumption, it is observed that, the hybrid of Elgamal and ChaCha20-Poly1305 uses less memory of up to 69.5% for encryption and about 51% for decryption compared to Elgamal alone. This makes it more memory efficient across all key sizes. This is shown in Figure 2.

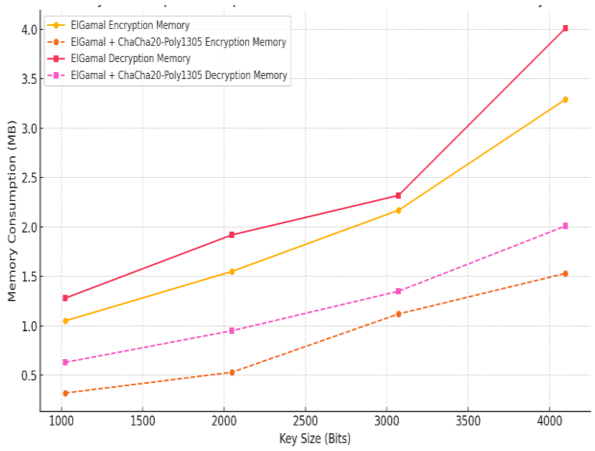


Figure 2: Memory Consumption comparison between Elgamal and the hybrid of Elgamal and ChaCha20-Poly1305.

5 CONCLUSIONS

The Hybrid cryptosystem which combines Elgamal and ChaCha20-Poly1305 for cloud storage, is designed with the aim of providing a highly secure and efficient solution for data encryption, transmission, and decryption. Elgamal is used for secure key exchange and ensures that the symmetric key required for ChaCha20-Poly1305 encryption remains securely transmitted. This addresses the challenge of securely distributing symmetric keys over insecure channels. ChaCha20-Poly1305 is a lightweight and fast encryption algorithm, which provides confidentiality, authenticity, and integrity for the data while maintaining high performance.

Furthermore, ElGamal cryptographic algorithm is relatively slow and inefficient for large datasets and this is where ChaCha20-Poly1305 excels. ChaCha20-Poly1305 is highly efficient and can handle large amounts of data with less computational overhead compared to traditional asymmetric encryption algorithms like RSA or ElGamal. The experiment conducted clearly shows that the hybrid of Elgamal and ChaCha20-Poly1305 outperforms Elgamal algorithm alone in both speed and memory efficiency across all key sizes. This makes it an efficient and scalable choice for cloud storage transactions. ChaCha20-Poly1305 also provides authenticated encryption with additional layer of security against tampering. Even though this hybrid system provides the best of security and efficiency, it may be limited in terms of implementation. Implementing a symmetric and an asymmetric cryptographic algorithm combined may be complex and may be computationally expensive and affect the overall system. The key management complexity increases as the system scales to accommodate more users. Future work may focus on optimizing key exchange process by adopting more efficient key exchange protocols.

Some of the key types of attacks that this system can mitigate include Eavesdropping attack, Replay attacks, side channel attacks, key injection attacks and ciphertext manipulation.

REFERENCES

[1] I. Odun-Ayo, O. O. Ajayi, B. Akanle, and R. Ahuja, “An Overview of Data Storage in Cloud Computing,” 2017 Int. Conf. Next Gener. Comput. Inf. Syst. (ICNGCIS), pp. 29-34, 2017, [Online]. Available: https://doi.org/10.1109/ICNGCIS.2017.9.

[2] F. Safar and R. Al King, “Data Security in Cloud Computing,” Int. J. Wireless Ad Hoc Commun., vol. 7, no. 1, 2023, [Online]. Available: https://doi.org/10.54216/ijwac.070105.

[3] W. E. Kedi, C. Ejimuda, and M. D. Ajegbile, “Cloud computing in healthcare: A comprehensive review of data storage and analysis solutions,” World Journal of Advanced Engineering Technology and Sciences, 2024, [Online]. Available: https://doi.org/10.30574/wjaets.2024.12.2.0291.

[4] R. Sivan and Z. Zukarnain, “Security and Privacy in Cloud-Based E-Health System,” Symmetry, vol. 13, p. 742, 2021, [Online]. Available: https://doi.org/10.3390/sym13050742.

[5] S. Reema, “Cloud Computing as a Solution for Security and Privacy Concerns,” Int. J. Res. Appl. Sci. Eng. Technol., 2023, [Online]. Available: https://doi.org/10.22214/ijraset.2023.49375.

[6] M. Joshi, R. Priya, and M. M. Joshi, “A Review: Analysis of Various Encryption Techniques for Securing Cloud Data,” 2022 4th Int. Conf. Adv. Comput. Commun. Control Netw. (ICAC3N), pp. 1945-1948, 2022, [Online]. Available: https://doi.org/10.1109/ICAC3N56670.2022.10074465.

[7] M. A. Althamir, A. Alabdulhay, and M. M. Yasin, “A Systematic Literature Review on Symmetric and Asymmetric Encryption Comparison Key Size,” 2023 3rd International Conference on Smart Data Intelligence (ICSMDI), pp. 110-117, 2023, [Online]. Available: https://doi.org/10.1109/ICSMDI57622.2023.00027.

[8] J. Khudair, K. A. Ghan, and M. R. B. Baharon, “Comparative Study in Enhancing AES Algorithm: Data Encryption,” Wasit Journal for Pure Sciences, 2023, [Online]. Available: https://doi.org/10.31185/wjps.100.

[9] A. V., A. P. Nirmala, B. K., Aldred Christi, and N. A., “A Review on Cloud Cryptography Techniques to Improve Security in E-health Systems,” 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), pp. 100-104, 2022, [Online]. Available: https://doi.org/10.1109/ICCMC53470.2022.9753999.

[10] A. Emmanuel, O. Aderemi, A. O., and A. O., “A Note on Time and Space Complexity of RSA and ElGamal Cryptographic Algorithms,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 12, no. 7, 2021, [Online]. Available: https://doi.org/10.14569/ijacsa.2021.0120716.

[11] T. Rezk, R. Bhargavan, and M. Blanchet, “Security Analysis of ElGamal Implementations,” Inria, 2018.

[12] M. Kumar, R. Sharma, and P. Bhushan, “An Optimal and Efficient Data Security Technique Through Crypto-Stegano for E-Commerce,” Computational Intelligence and Neuroscience, vol. 2023, no. 9906588, pp. 1-10, 2023, [Online]. Available: https://doi.org/10.1155/2023/9906588.

[13] H. Guo, X. Li, and J. Zhang, “A Novel Image Encryption Algorithm Based on ElGamal Cryptography and Chaotic System,” Physica Scripta, vol. 98, no. 11, pp. 1-12, 2023, [Online]. Available: https://doi.org/10.1088/1402-4896/98/11/115251.

[14] V. O. Waziri, J. Ojeniyi, H. Danladi, A. Isah, A. S. Magaji, and M. Abdullahi, “Network Security in Cloud Computing with Elliptic Curve Cryptography,” Network and Communication Technologies, vol. 2, no. 2, pp. 43-58, 2013, [Online]. Available: https://doi.org/10.5539/nct.v2n2p43.

[15] V. Satheesh and D. Shanmugam, “Implementation Vulnerability Analysis: A case study on ChaCha of SPHINCS,” 2020 IEEE International Symposium on Smart Electronic Systems (iSES), pp. 97-102, 2020, [Online]. Available: https://doi.org/10.1109/iSES50453.2020.00032.

[16] S. H. Tariq, R. P. Singh, and J. Gill, “Preserving Privacy in Industrial IoT: A Machine Learning Framework Enhanced by Poly1305 Encryption,” International Journal for Research in Applied Science and Engineering Technology, 2023, [Online]. Available: https://doi.org/10.22214/ijraset.2023.55767.

[17] Y. Peng, W. Zhao, F. Xie, Z. Dai, Y. Gao, and D. Chen, “Secure cloud storage based on cryptographic techniques,” J. China Univ. Posts Telecommun., vol. 19, pp. 182-189, 2012, [Online]. Available: https://doi.org/10.1016/S1005-8885(11)60424-X.

[18] D. Salman and N. Sulaiman, “A Review of Encryption Algorithms for Enhancing Data Security in Cloud Computing,” AlKadhim J. Comput. Sci., 2024, [Online]. Available: https://doi.org/10.61710/kjcs.v2i1.68.

[19] S. Ahmad and A. Garko, “Hybrid Cryptography Algorithms in Cloud Computing: A Review,” 2019 15th Int. Conf. Electron. Comput. Computat. (ICECCO), pp. 1-6, 2019, [Online]. Available: https://doi.org/10.1109/ICECCO48375.2019.9043254

[20] H. Sharma, R. Kumar, and M. Gupta, “A Review Paper on Hybrid Cryptographic Algorithms in Cloud Network,” 2023 2nd Int. Conf. Innov. Technol. (INOCON), pp. 1-5, 2023, [Online]. Available: https://doi.org/10.1109/INOCON57975.2023.10101044.

[21] S. Kaur and L. Jain, “A Hybrid Cryptographic Scheme for Improving Cloud Security Using ECC and TDES Algorithms,” Curr. J. Appl. Sci. Technol., vol. 39, no. 4, pp. 73-84, 2020, [Online]. Available: https://doi.org/10.9734/CJAST/2020/V39I4731184.

[22] R. Mendez, “Enhance Data Storage Security DNA Cryptography in Cloud,” 2020.

[23] A. Kavya and S. Acharva, “A Comparative Study on Homomorphic Encryption Schemes in Cloud Computing,” 2018 3rd IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol. (RTEICT), pp. 112-116, 2018, [Online]. Available: https://doi.org/10.1109/RTEICT42901.2018.9012261

[24] M. Madhavi and D. Sivareddy, “A Review of Research on Dual-Server Public-Key Encryption with Keyword Search for Secure Cloud Storage,” Int. J. Res., vol. 4, pp. 630-638, 2017.

[25] A. Gadad and D. Anbusezhiyan, “Cloud security: literature survey,” Int. J. Electr. Comput. Eng. (IJECE), vol. 13, no. 4, pp. 4734-4742, 2023, [Online]. Available: https://doi.org/10.11591/ijece.v13i4.pp4734-4742.

[26] A. Sharma, R. S. Thakur, and S. Jaloree, “Analyzing the Behavior of Symmetric Algorithms Usage in Securing and Storing Data in Cloud,” Springer, pp. 381-389, 2018, [Online]. Available: https://doi.org/10.1007/978-981-10-5523-2_35.

[27] K. Sasikumar and S. Nagarajan, “Comprehensive Review and Analysis of Cryptography Techniques in Cloud Computing,” IEEE Access, vol. 12, pp. 52325-52351, 2024, [Online]. Available: https://doi.org/10.1109/ACCESS.2024.3385449.

[28] S. Njuki, J. Zhang, E. C. Too, and R. Richard, “An Evaluation on Securing Cloud Systems based on Cryptographic Key Algorithms,” Proceedings of the 2nd International Conference on Algorithms, Computing and Systems, 2018, [Online]. Available: https://doi.org/10.1145/3242840.3242853.

[29] N. L. Kodumru and M. Supriya, “Secure Data Storage in Cloud Using Cryptographic Algorithms,” 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, [Online]. Available: https://doi.org/10.1109/ICCUBEA.2018.8697550.

[30] D. S. S. Marar, J. M. Joyson, and L. Ashish, “Secure Cloud Storage Using Different Algorithms in Cryptography,” International Journal for Research in Applied Science and Engineering Technology, 2022, [Online]. Available: https://doi.org/10.22214/ijraset.2022.43500.

[31] A. Poduval, A. Doke, H. Nemade, and R. Nikam, “Secure File Storage on Cloud using Hybrid Cryptography,” International Journal of Computer Sciences and Engineering, vol. 7, no. 1, 2019, [Online]. Available: https://doi.org/10.26438/ijcse/v7i1.587591.

Potential of Using the Ant Colony Optimization Algorithm for Optimal Network Path Selection

Oleksandra Yaroshevska and Veronika Kirova
Anhalt University of Applied Sciences, Bernburger Str. 57, 06366 Köthen, Germany
yaroshevska.ob@gmail.com, veronika.kirova@hs-anhalt.de

Keywords: Network Optimal Path, Ant Colony Optimization (ACO) Algorithm, Dijkstra's Algorithm, Lowest Common Ancestor (LCA), Pheromone, RTT, Python.

Abstract: The article considers the possibilities of using the Ant Colony Optimization algorithm to find the shortest path in the network based on the selected criteria. Its performance is compared to Dijkstra's algorithm and LCA algorithm, which is widely used in different network routing protocols. An overview of the ACO algorithm, including its two primary components, the "ant" and "pheromone," is provided, highlighting its efficiency for the optimal network path selection. Detailed schemes, parameters and formulas of the ACO algorithm implementation in terms of networking are shown. A comparative analysis of the performance and execution time of the ACO and two compared algorithms for the optimal network path based on Round Trip Time criteria in networks of varying scale, ranging from small to highly branched networks with thousands of nodes, is discussed. Finally, the results are analysed, and the potential for ACO to serve as a complementary algorithm to Dijkstra's and LCA in future network applications is explored.

1 INTRODUCTION

1.1 Optimal Network Path Selection Problem

As modern technologies continue to evolve and the demand for network infrastructure grows, the challenge of building extensive local and global networks becomes increasingly urgent. A key factor in maintaining network efficiency is identifying optimal routes between nodes.

Using optimal routes between nodes in a network is critical for several reasons (Figure 1).

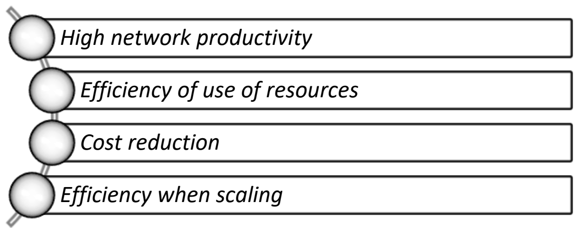


Figure 1: Reasons for the importance of route optimization in the network.

Each reason is explained detailed below [1]:

1) With optimal routing, data takes the shortest or most efficient path between nodes, reducing data transmitter delays. This facilitates faster communication and better user experience.
2) Optimal routing helps balancing the traffic by preventing certain paths from becoming congested while others are not used enough. Thus, the network works more efficiently without overloading the infrastructure.
3) There is a reduction in the need for additional infrastructure, energy consumption and maintenance that may arise from inefficient routing.
4) Smoother scaling is provided so that new nodes or connections do not decrease network efficiency.

Channel reservation is also an additional option for route optimization. This means that in case if one route fails, data can be quickly rerouted through alternate paths, increasing network reliability.

1.2 Methods to Solve the Problem and the Main Research Goal Description

Among the possible solutions to the problem of optimizing paths in a network, two groups of methods are distinguished: traditional and metaheuristic [2]. Currently, it is common to use traditional methods in networks. One of the most widespread methods for finding the best path is Dijkstra's algorithm [3]. Its advantage consists in the existence of a stable scheme using a path matrix, which helps to find the shortest or the most efficient path depending on the specific needs of the network. However, such a scheme is quite rigid, which is a drawback and may lead to difficulties in cases where more precise adjustment of the algorithm's parameters is required.

In addition to traditional approaches, modern algorithmic techniques such as the Lowest Common Ancestor (LCA) are increasingly being considered for network path optimization, especially in hierarchical or tree-structured networks. LCA algorithms are particularly effective when the network can be represented as a rooted tree and rapid queries between node pairs are required. By preprocessing the network structure, LCA allows for efficient determination of the closest shared ancestor of two nodes, which can significantly improve routing decisions in applications such as multicast routing, hierarchical clustering, and certain types of peer-to-peer networks. The use of LCA methods enhances the adaptability of routing protocols and supports real-time decision making with low computational overhead.

The other approach to solving the problem of finding the optimal path is metaheuristic methods. They are based on heuristic principles and use search in a large solution space to find approximate but high-quality answers, often operating with the processes that are inherent in nature. One of these methods is the Ant Colony Optimization (ACO) algorithm [4-6]. The basic concept of this algorithm was created based on observations of how ants find the shortest path to a food source.

The decision-making process about the optimal path includes elements of probability theory and is based on the concept of pheromones, which ants leave in nature depending on the quality of the path they have traversed. In a real network, the quality of a route could be considered, for example, as the level of delay between nodes, bandwidth or other important parameters. This article examines the potential use of ACO compared to Dijkstra's and LCA algorithms and provides a description of the problems for which ACO can be used.

2 EXPERIMENTAL TOPOLOGY

2.1 Dijkstra’s Algorithm Realization

To determine the efficiency of ACO, Dijkstra’s algorithm was implemented using various methods of data storage and processing to achieve the most valid results. Python was chosen as the programming language due to its convenience and the wide range of tools available for data analysis.

The first method involves using a standard list with element sorting to determine the smallest Round Trip Time (RTT) value, which is used in this experiment as the main parameter for path efficiency [7]. RTT is the time required to send a data packet from the sender to the receiver and return a response back. This parameter is important for evaluating network latency, as it reflects the speed of data transmission between nodes. A smaller RTT indicates faster transmission and better network performance.

The second method involves using a binary tree, where the smallest RTT value is the root of the tree. Further comparison based on processing speed determines the best method for data handling to be used as a standard when evaluating the efficiency of ACO (see section 2.3.1).

The implementation of both variations of Dijkstra’s algorithm was based on existing algorithmic frameworks [3]. This custom implementation is important to ensure that the input data in the comparison of Dijkstra’s and ACO algorithms are identical and can be adjusted during the experiment, as this is one of the factors ensuring the validity of the experiment.

The main stages of Dijkstra’s algorithm work are described below:

1) A priority queue and RTT matrices from the key node to each of the other nodes in the network are created.
   The priority queue is a list or binary tree according to one of the two methods described above. It includes "node - RTT" value pairs to identify the best paths. In this case, nodes are considered as vertices of an undirected graph, and the paths with RTT values are considered as the edges of the graph [8].
2) Each neighbor of the current node is visited.
   In the process of moving to neighboring nodes, each one is added to the priority queue along with its RTT. It is important to use nodes that have not been visited before to prevent cycles and incorrect algorithm behavior.
3) The previous matrix data is compared with the current paths to each of the neighboring nodes.

If the current path is better, the matrix is updated with the new data.

4) The path with the smallest RTT is selected. The node to which this path leads becomes the next current node.
   In the case of a list, further data sorting occurs. In the case of a binary tree, the best option is found at the root of the tree.

The algorithm repeats until every vertex of the graph is visited. As a result, the final matrix contains the best paths from the key node to every other node in the network.

In this implementation, the dictionaries are used as an analogy to matrices to improve the algorithm's productivity. Using dictionaries in Python is efficient due to their speed in handling elements (search, addition, deletion, etc.). Dictionaries are based on hash tables, which allow data to be accessed by key in O(1) time in most cases. They are also convenient for storing large amounts of data with "key-value" pairs, ensuring high performance when analyzing extensive networks with many nodes.

Two main dictionaries were created:

- Shortest path dictionary.
- Last neighbor dictionary.

The first dictionary contains information about the shortest path to each node. The second dictionary stores information about the second last node that must be visited before reaching the destination. For example, the second last node for the fifth node it could be the third, for the tenth node – the twelfth, and so on. This creates a chain effect when traversing from the starting (key) node to others.

In this way not only can the RTT of the shortest path to each node be determined, but the entire sequence of nodes along the path can be traced. This ensures accurate comparison with the ACO algorithm, where the sequence may vary depending on the specified parameters. Both dictionaries are output by the algorithm upon completion, allowing each to be used as needed.

2.2 LCA algorithm realization

To complement the comparison with the ACO algorithm, the LCA (Lowest Common Ancestor) algorithm was implemented to evaluate scenarios where hierarchical relationships between nodes are critical, such as in tree-based or partially hierarchical network topologies. Python was selected for the implementation due to its rich set of tools and the ability to prototype algorithmic logic efficiently.

In this realization, the network is modeled as a rooted tree, where each node may have multiple children, and all connections are unidirectional from parent to child. This reflects situations where pathfinding is restricted by hierarchical constraints or parent-child dependencies. The goal of the LCA algorithm is to find the common ancestor node that is lowest (i.e., deepest) in the tree for any two given nodes [9]. This approach is particularly useful in applications involving tree traversal, organizational hierarchies, or data clustering [10].

The LCA implementation is based on the binary lifting technique, which allows for efficient querying of the lowest common ancestor in logarithmic time. The preprocessing phase is designed to prepare necessary lookup tables that speed up each individual LCA query, optimizing the performance for networks with frequent ancestor-related queries.

The main stages of the algorithm are described below:

1) Tree Construction and Initialization. Each node is represented by an instance of a TreeNode class containing its value and a list of children. During initialization of the LCA class, the root node and the total number of nodes in the tree (n) are passed as parameters. Arrays are created to store the depth and parent of each node, as well as a binary lifting table (up) which enables ancestor lookup at various powers of two.
2) Depth-First Search (DFS) Traversal for Preprocessing. A depth-first traversal of the tree is performed starting from the root node. During traversal, each node's depth and parent are recorded. The up table is filled such that up[i][j] stores the 2^j-th ancestor of node i. This preprocessing allows any node to be moved upward by any power-of-two number of levels in constant time, which is essential for efficient LCA computation.
3) LCA Query Execution. To determine the lowest common ancestor of two nodes u and v, the algorithm first equalizes their depths by moving the deeper node upward. Then, starting from the highest level of the binary lifting table, both nodes are lifted together until they converge. The final result is the parent of the converging point, which represents the lowest common ancestor.
4) Performance and Application. This realization of the LCA algorithm ensures an O(n log n) preprocessing time and O(log n) query time, making it highly efficient for repeated ancestor queries in large hierarchical trees. While not directly focused on shortest paths or RTT as in

Dijkstra's or ACO, the LCA algorithm provides a foundational utility in hierarchical routing or clustering scenarios, offering a complementary perspective in network analysis.

The LCA implementation outputs the ancestor node shared by any two nodes in the shortest hierarchical path, enabling the tracing of common routes and structural relationships within tree-based network representations. This makes it a valuable comparative model alongside Dijkstra's and ACO algorithms in evaluating different types of network structures and their respective search efficiencies.

2.3 Ant Colony Optimization Algorithm Realization

ACO is based on two main terms: "ant" and "pheromone" [4]. The algorithm consists of several iterations, after each of which pheromone levels are updated. Pheromone is represented by a numerical value $0 < ph < 1$. The higher this value, the greater the probability of choosing a particular path. One iteration consists of several steps, each of which is a complete path from the start node to the final node. Such a step is analogous to an ant traveling the distance from its home to a food source in nature. For simplicity, further in the article, these steps will be referred to as ants.

The flowchart of the algorithm is shown in Figure 2.

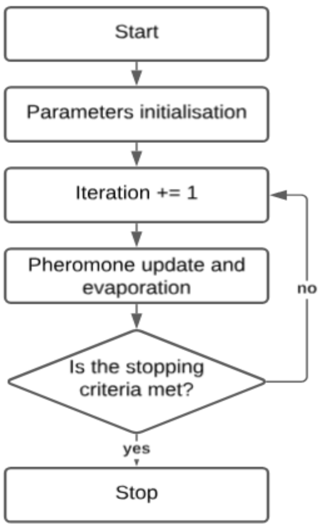


Figure 2: Base ACO block-scheme.

The main parameters of ACO are as follows:

- number of iterations;
- number of ants;
- initial pheromone level;
- pheromone decay rate;
- parameters of importance of the pheromone and RTT when calculating the probability of path selection;
- coefficients required for the mathematical calculations of probability.

After each iteration, the pheromone levels are updated, which influences subsequent iterations. Thus, the pheromone amount on paths that are traveled more frequently increases, raising the likelihood of those paths being used again while reducing the likelihood of using less efficient paths.

The stopping criterion for the algorithm in its standard implementation is the achievement of the set number of iterations. However, additional criteria may be added to increase efficiency.

The flowchart of the algorithm within each iteration is shown in Figure 3.

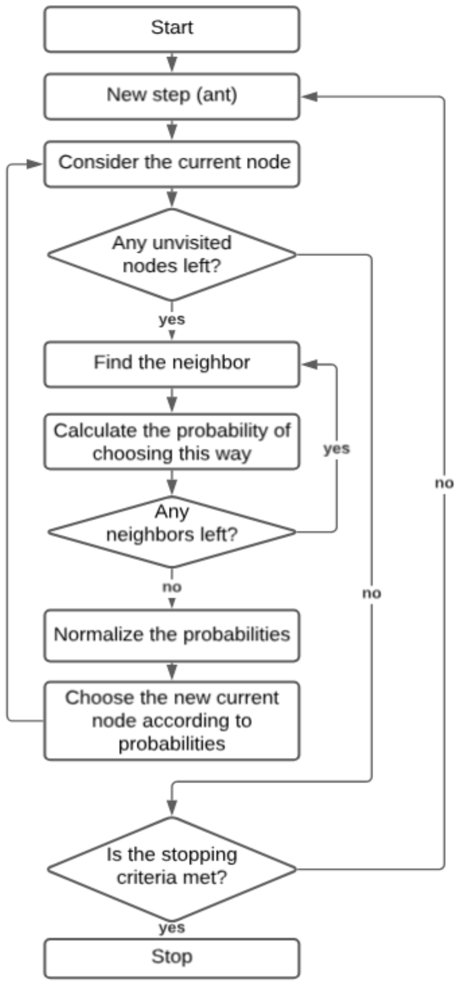


Figure 3: ACO block-scheme inside the iteration.

As shown in Figure 3, each iteration consists of a certain number of steps (ants). Each ant traverses the complete path from the starting node to the final node, after which it stores information about the path as a sequence of nodes.

After the iteration is complete, the pheromone levels on all the paths that were traversed are updated. The number of iterations and ants is determined empirically, depending on the network's scalability.

It is important to note that the probability of any path should not reach zero as long as it remains accessible, since the most optimal path may include sections with suboptimal RTT values at certain stages. The overall value across the entire path from the start to the final node will be the most optimal. Therefore, it's important to maintain a pheromone level that allows for a small probability of selecting alternative paths to those that were previously chosen.

The elements that make up the path of one ant are described in more detail below:

1) Finding the neighbors of the current node. At the first stage, the starting node is considered the current one. Subsequently, the current node will be the one the ant moves to on the path toward the final node.
2) Calculating the probability of transitioning to each neighboring node.
   This calculation is based on two main formulas.

$$P(i) = ph^{\alpha} * \frac{k}{RTT^{\beta}} . \qquad (1)$$

   Formula (1) reflects the transition strength from the current node to another specific node. This strength, also referred to as the "desire" to transition, depends on the pheromone level and the inverse value of RTT. Thus, the lower the RTT value, the greater the transition strength.

   The coefficient k is selected based on the RTT values for the specific network and is determined empirically. Coefficients α and β are used to increase the influence of pheromone or RTT. In this experiment, both are set to one, which means that the influence of pheromone and RTT is equally weighted. The initial pheromone value is set to 0.3 and is either increased or decreased depending on which paths are traversed.

$$P(norm) = \frac{P(i)}{\sum_{i=1}^{n} P(i)} \qquad (2)$$

   Formula (2) is required for normalizing the transition strength and calculating the actual probability within the range 0 < P < 1. Normalization occurs after evaluating the transition strengths for all neighboring nodes, as it requires the total sum of these strengths.
3) Selecting the next node.A scale from 0 to 1 is used for node selection. A randomly chosen number falls within a range that corresponds to a specific neighboring node based on the transition probability previously calculated. For example, with a 50% probability, half of the scale is covered. The transition to the next node then occurs.

   The condition for exiting the cycle is reaching the final node. The cycle is repeated according to the number of ants, which is one of the parameters of the algorithm.

At the end of the algorithm's execution, information is provided regarding the best path found and its total RTT value.

2.3 Results of Comparison by the Execution time Parameter

2.3.1 Dijkstra’s Algorithm Realizations Comparison

To study the performance of the algorithms, an emulation of RTT data was carried out, obtained from networks of various scalability. The data was automatically generated according to the specified number of nodes and saved in a document for convenient access and the ability to rerun the experiment with different algorithms.

For this study, the number of connections between network nodes was set to 70% of all possible connections. An essential parameter is maintaining connectivity between all nodes, meaning that each node must have at least one connection to every other node.

The comparison results based on execution time performance are shown in Table 1.

Table 1: Execution time comparison of the Dijkstra’s algorithm realizations.

Network branching (nodes / connections)		Dijkstra (binary tree)	Dijkstra (list)
10	32	0,00s	0,00s
500	87325	0,18s	0,32s
1000	349650	0,69s	1,41s
1500	786975	1,76s	3,76s
2000	1399300	3,22s	6,88s
2500	2186625	5,07s	12,47s
3000	3148950	7,12s	18,93s

Diagram of execution time comparison of Dijkstra's algorithm realizations is shown in Figure 4.

In Figure 4, the vertical axis represents the execution time of the algorithm, while the horizontal axis indicates the number of nodes in the network. It is important to note that the algorithm's execution time is stable and fluctuates within the hundredths and thousandths of a second due to a consistent pathfinding system, which allows for a reduction in the number of trials during the research to ten for networks with identical parameters.

From the data obtained in the first phase of the study, the advantage of Dijkstra's algorithm based on binary trees can be observed. In networks with a small number of nodes, this advantage is minimal and amounts to less than 1 second; however, in more complex networks, the difference becomes pronounced, highlighting the importance of using more efficient data storage and processing methods compared to standard lists.

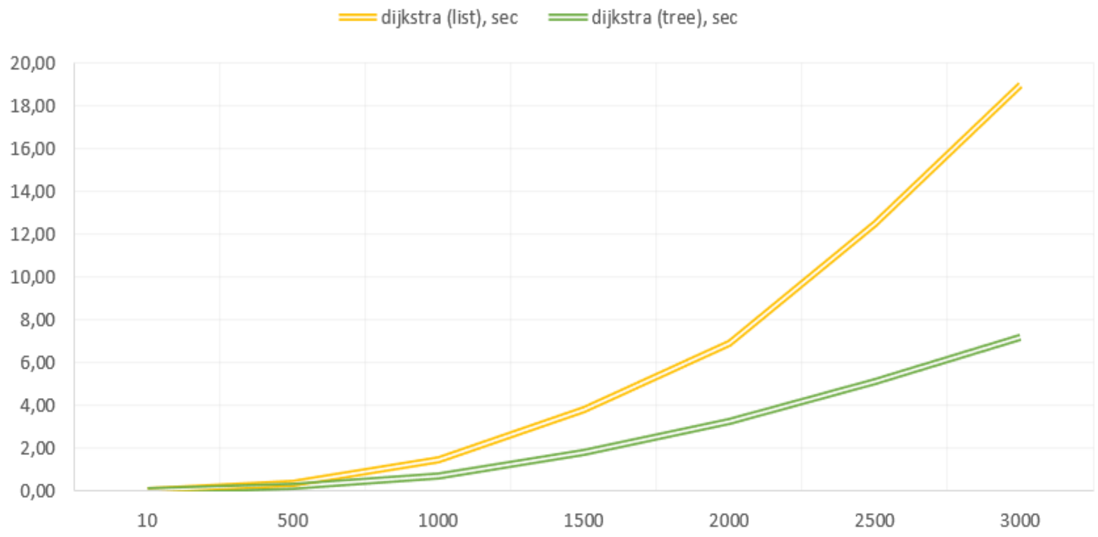


Figure 4: Diagram of execution time comparison of Dijkstra's algorithm realizations.

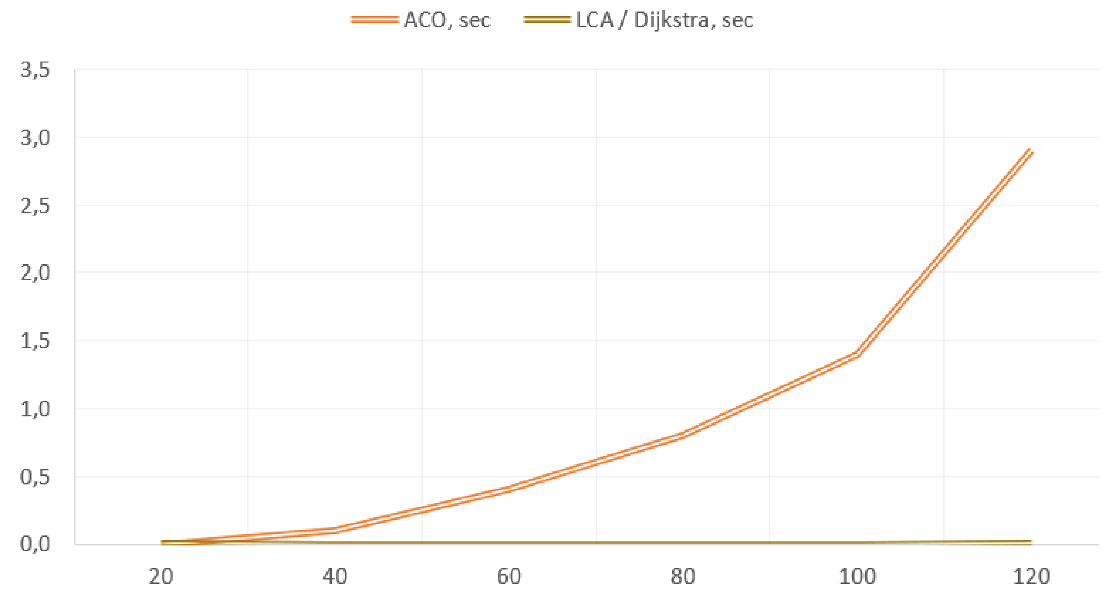


Figure 5: Diagram of execution time comparison of ACO, Dijkstra's and LCA algorithms.

2.3.2 ACO, Dijkstra's and LCA Algorithms Comparison

Based on the results of the first phase of the study, Dijkstra's algorithm was implemented using a binary tree. The next phase involves comparing ACO, Dijkstra's and LCA algorithms based on execution time to determine the potential for fully utilizing ACO in networks.

The comparison results are shown in Table 2 and the diagram of execution time comparison is shown in Figure 5.

The execution time analysis clearly demonstrates that ACO exhibits significantly lower performance in terms of speed when compared to the other two algorithms. Furthermore, it is important to note that the difference in execution times becomes more pronounced as the network size increases, as evidenced by the cases involving 20-node and 120-node topologies.

While Dijkstra's algorithm and LCA demonstrate comparable performance in small-scale networks, their efficiency diverges in larger topologies. For instance, in a simulated network consisting of 1,000 nodes, the average Round Trip Time (RTT) for queries using the LCA algorithm was measured at approximately 0.02 seconds, whereas Dijkstra's algorithm required an average of 0.71 seconds to complete equivalent path computations. This substantial difference highlights the superior scalability of the LCA approach in hierarchical network structures, where its logarithmic query complexity enables faster execution compared to the graph-based traversal required by Dijkstra's algorithm.

Table 2: Execution time comparison of ACO, Dijkstra's and LCA algorithms.

Network branching (nodes / connections)		Dijkstra (binary tree)	LCA	ACO
20	133	0,01s	0,01s	0,01s
40	546	0,01s	0,01s	0,1s
60	1239	0,01s	0,01s	0,4s
80	2212	0,01s	0,01s	0,8s
100	3465	0,01s	0,01s	1,4s
120	4998	0,01s	0,01s	2,9s

Additionally, it should be emphasized that using ACO in complex networks with more than 500 nodes is not effective at this stage of the algorithm's implementation and requires additional parameters to improve execution time.

3 PROPOSED SOLUTION

3.1 ACO Potential

Considering the execution time of Dijkstra's algorithm in networks with a large number of nodes (see section 2.3.1), as well as its efficient use in modern networks, it is proposed to utilize a different potential of ACO instead of merely competing on speed and subsequently replacing the basic path optimization algorithm.

ACO has certain application features that are absent in traditional algorithms due to their rigid working structure. These features involve dynamic parameter tuning and the ability to find alternative paths. This is achieved through the use of probabilistic elements and random selection. However, it is important to clarify that "random selection" in this context does not mean purely arbitrary choices. Rather, the decisions are guided by specific criteria such as Round Trip Time (RTT) values and the internal characteristics of the algorithm, which work together to improve overall efficiency and performance.

The search for alternative paths is a localized task within the network's operation, meaning that it can be performed without the need to gather new RTT measurements or to completely reconstruct all the paths within the network. This makes ACO particularly suitable for real-time optimizations where it is impractical to recalculate the entire network.

Such flexibility in path selection can prove useful in addressing various network management tasks, including the following:

1) Temporarily reducing the amount of traffic on a heavily congested path, allowing for better load balancing and preventing bottlenecks in data flow.
2) Identifying an alternative path for the transfer of large volumes of data, especially when the primary route is suboptimal for such specialized tasks.
3) Discovering an optimal path that bypasses a specific node, which may be temporarily unavailable or malfunctioning, thereby maintaining network connectivity and minimizing disruptions during such outages.

These capabilities make ACO a valuable tool for addressing network issues that require dynamic and responsive solutions.

3.2 Alternative Path Concept Realization

To efficiently store and manage data regarding the shortest paths discovered by the algorithm, several key data structures are utilized within the code:

- A temporary list of tuples "path – RTT value" that each ant passed in a given iteration;
- a temporary list for storing the best path during the current iteration;
- a dictionary for storing all the best paths found in each iteration, along with the corresponding RTT values, for easy output and future use.

The lists are called "temporary" because they are updated at the beginning of each new iteration, helping to collect new information about the paths traveled.

The flowchart for collecting and processing the information obtained during the algorithm's execution is shown in Figure 6.

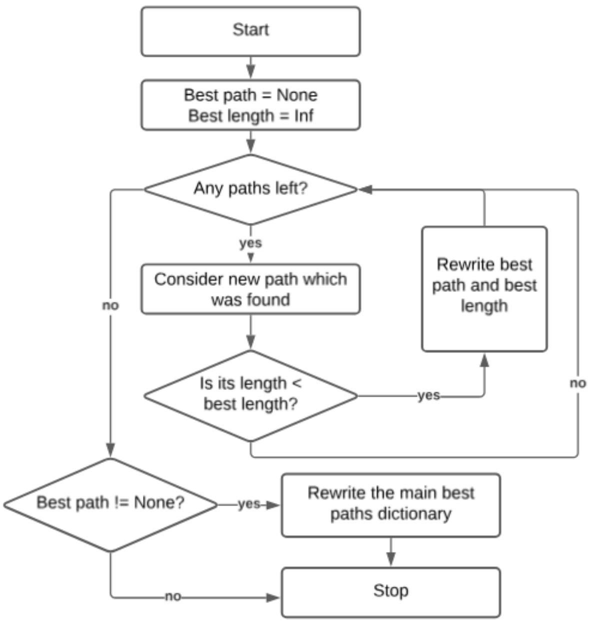


Figure 6: Block-scheme of alternative path concept realization.

This part of the algorithm begins operation after a list of all paths within the iteration has been formed. As the algorithm progresses, a dictionary is gradually populated with the best results, which is then passed on for further use at the end of the algorithm's execution.

After sorting process according to user preferences, any number of paths close to the optimal can be selected. This allows for local route adjustments where necessary within the network, without consuming excessive network resources.

4 CONCLUSIONS

This article explores the performance and potential of the Ant Colony Optimization (ACO) algorithm compared to Dijkstra’s and LCA algorithms in network pathfinding tasks.

Dijkstra’s algorithm was selected as a benchmark due to its status as a classical and widely accepted method for finding the shortest path in graphs, offering a deterministic and well-understood approach to pathfinding. In contrast, the LCA algorithm was chosen to represent more modern, structurally optimized methods, particularly suited for hierarchical or tree-based networks, where rapid ancestor queries can significantly reduce computation time.

The comparative analysis demonstrates that while ACO offers adaptability and robustness in dynamic or uncertain environments, it lags behind in terms of computational speed. Dijkstra’s and LCA algorithms show comparable performance in small networks; however, LCA outperforms Dijkstra in large-scale networks due to its logarithmic complexity in query processing. These findings suggest that although ACO holds promise in flexible and heuristic-driven scenarios, classical and hierarchical algorithms remain superior in deterministic and high-performance environments.

The main potential of ACO lies in creating alternative path, whereby its probabilistic and pheromone-based path selection allows it to adaptively find additional optimal paths in cases of congestion, large data transfers, or node failures.

The primary areas for improving ACO’s efficiency can be divided into two main groups:

1) Time control task. One effective way to improve time efficiency is by integrating time control mechanisms into the algorithm’s execution. For example, time-limiting conditions can be introduced within each iteration or across the entire algorithm run. These may involve stopping the search early if no improvement is detected over a defined number of iterations, or imposing a maximum runtime per execution

cycle. Such constraints help manage resource consumption more effectively and make the algorithm more viable for time-sensitive applications. Additionally, they enhance scalability, allowing ACO to handle larger network graphs without exponential increases in computation time.

2) Parameters research. The quality of the ACO algorithm heavily depends on the choice of key parameters, such as the number of ants, iterations, pheromone evaporation rate, and the influence of heuristic information (commonly referred to as alpha and beta). Parameter research involves systematic experimentation and analysis to identify optimal settings for various scenarios. This process may lead to the development of parameter templates tailored to specific network types (e.g., sparse, dense, hierarchical) [11], enabling faster deployment and better results without manual tuning.

Additionally, advanced approaches such as adaptive parameter tuning or machine learning-based optimization can further enhance ACO’s performance by dynamically adjusting parameters during runtime based on observed performance metrics.

The article concludes that instead of positioning ACO as a direct competitor to Dijkstra’s and LCA algorithms, it should be seen as a complementary tool for specific network challenges. Combining the efficiency of Dijkstra’s or LCA algorithm for global pathfinding with ACO’s flexibility for local adjustments can yield optimal results in network productivity.

ACKNOWLEDGMENTS

This work has been funded by DAAD, BMBF Foundation for partnership between scholars and scientists from Ukraine and Germany within the project Fit4Ukraine. The authors thank Hochschule Anhalt and Future Internet Lab Anhalt for support and the opportunity to use the equipment for research.

Also, 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] X. Liu, Y.-L. Chen, L. Y. Por, and C. Ku, "A Systematic Literature Review of Vehicle Routing Problems with Time Windows," Sustainability, vol. 15, 08 2023, [Online]. Available: https://doi.org/10.3390/su151511999.

[2] S. Rezk and K. Selim, "Metaheuristic-based ensemble learning: an extensive review of methods and applications," Neural Computing and Applications, vol. 36, pp. 17931-17959, 08 2024, [Online]. Available: https://doi.org/10.1007/s00521-024-09773-2.

[3] L. Chek, "Low Latency Extended Dijkstra Algorithm with Multiple Linear Regression for Optimal Path Planning of Multiple AGVs Network," Engineering Innovations, vol. 6, pp. 31-36, 06 2023, [Online]. Available: https://doi.org/10.4028/p-3z5h9x.

[4] Y. Razooqi, M. Al-Asfoor, and M. Abed, "Optimise Energy Consumption of Wireless Sensor Networks by using modified Ant Colony Optimization," Acta Technica Jaurinensis, vol. 17, pp. 111-117, 08 2024, [Online]. Available: https://doi.org/10.14513/actatechjaur.00744.

[5] I. Chakraborty and P. Das, "An Efficient ACO-based Routing and Data Fusion Approach for IoT Networks," SN Computer Science, vol. 4, 10 2023, [Online]. Available: https://doi.org/10.1007/s42979-023-02344-5.

[6] M. Liu, Q. Song, Q. Zhao, L. Li, Z. Yang, and Y. Zhang, "A Hybrid BSO-ACO for Dynamic Vehicle Routing Problem on Real-World Road Networks," vol. PP, pp. 1-11, 01 2022, [Online]. Available: https://doi.org/10.1109/TITS.2022.3146318.

[7] D. Stolpmann and A. Timm-Giel, "In-Network Round-Trip Time Estimation for TCP Flows," 09 2023, [Online]. Available: https://doi.org/10.48550/arXiv.2309.12345.

[8] A. Kusuma, R. Prihandini, and A. Agatha, "Graph Theory: Applications of Graphs in Map Coloring," 06 2024, [Online]. Available: https://doi.org/10.13140/RG.2.2.12345.67890.

[9] H. L. Bodlaender, et al., "Listing, Verifying and Counting Lowest Common Ancestors in DAGs," Proceedings of the 49th International Colloquium on Automata, Languages, and Programming (ICALP 2022), June 2022, [Online]. Available: https://doi.org/10.4230/LIPIcs.ICALP.2022.123.

[10] S.-B. Scholz, "A Scalable Approach to Computing Representative Lowest Common Ancestor in Directed Acyclic Graphs," Theoretical Computer Science, Elsevier BV, 2013, [Online]. Available: https://doi.org/10.1016/j.tcs.2013.01.012.

[11] K. Karpov, et al., "Available Bandwidth Metrics for Application-Layer Reliable Multicast in Global Multi-Gigabit Networks," Proceedings of International Conference on Applied Innovation in IT, vol. 8, issue 1, pp. 1-6, 2020, [Online]. Available: https://doi.org/10.25673/32742.

Optimizing the Location of 5G Network Base Stations Taking into Account Intra-System Interference

Serhii Siden[1], Roman Tsarov[1], Mohammed Jamal Salim[2], Kateryna Shulakova[1,3], Saad Malik Talha[1] and Liliia Bodnar[4]

[1]*State University of Intelligent Technologies and Telecommunications, Kuznechna Str. 1, 65023 Odesa, Ukraine*

[2]*Career Development Center, University of Information Technology and Communications, Al-Nidhal Str., 00964 Baghdad, Iraq*

[3]*Anhalt University of Applied Sciences, Bernburger Str. 57, 06366 Köthen, Germany*

[4]*The State Institution "South Ukrainian National Pedagogical University named after K. Ushynsky", Staroportofrankyvska Str. 26, 65020 Odesa, Ukraine*

ssiden@suitt.edu.ua, rcarev@gmail.com, mohamed.jamal8686@gmail.com, katejojo29@gmail.com, bodnarl79@pdpu.edu.ua

Keywords: 5G, MmWave, Structure Optimization, Genetic Algorithm, Propagation Models, SINR.

Abstract: This work is devoted to the structural optimization of 5G networks, specifically addressing the problem of base station (BS) placement optimization in indoor network deployment. A method is proposed for determining the number and optimal spatial coordinates of BSs in indoor environments, such as shopping malls or telemedicine centers, under random user distribution to ensure maximum coverage and network throughput while explicitly accounting for intra-system interference. The problem is characterized by dynamic environmental conditions, high user density, heterogeneous service demands, and the requirement for guaranteed network quality indicators, as well as the need to ensure reliable coverage in complex indoor layouts. As a result, the BS placement task is formulated as a nonlinear NP-complete integer programming problem. A genetic algorithm was employed to solve it, incorporating adaptive selection, crossover, and mutation operators. The fitness function was mathematically formulated to maximize the average user data rate while including penalty terms for BS overload, excessive BS proximity, and violations of minimum quality of service (QoS) thresholds. Numerical simulations demonstrate the effectiveness of the proposed approach, confirming that the developed method allows for structural optimization of 5G networks through intelligent base station placement under the influence of intra-system interference.

1 INTRODUCTION

In the context of rapid digital transformation and the exponential growth of data volumes, 5G technology has acquired strategic importance as a foundational component of modern information infrastructure. The efficient deployment of 5G networks is critical for ensuring sustainable economic development, strengthening national security, and enhancing the quality of life for citizens.

Networks based on 5G serve as a fundamental service platform enabling a wide range of applications spanning various societal activity sectors. In particular, in telemedicine, 5G facilitates remote medical consultations and real-time patient monitoring [1]. Furthermore, 5G contributes to the development of intelligent transportation systems, providing seamless connectivity between vehicles and infrastructure, which enables traffic optimization and improves road safety.

In the industrial domain, 5G unlocks the potential for automation of production processes and implementation of remote control and equipment monitoring systems, thereby increasing operational efficiency and productivity. In the entertainment sector, 5G supports high-resolution video streaming and virtual and augmented reality technologies, creating new opportunities for interactive content delivery.

In this context, the problem of optimizing the network design process arises it is necessary to develop a network architecture capable of supporting the efficient operation of modern services, including telemedicine systems, intelligent transport infrastructure, industrial automation, and other innovative applications.

It is evident that 5G networks operate within a constantly evolving and dynamically changing environment, where the number of users, their mobility patterns, and service demands such as data rate and latency vary continuously. These ongoing environmental changes significantly impact on the characteristics, architecture, and operational trajectory of 5G networks. Given that such influences are persistent in nature, the structure and behavior of a 5G network transform every stage of its life cycle. Under these conditions, optimizing a 5G network as a complex and adaptive system becomes feasible only through continuous adaptation. One of the most practical and effective forms of such adaptation is structural optimization, particularly the optimization of BS placement. Accurate BS positioning is essential for ensuring adequate coverage, minimizing inter-cell interference, and maintaining high network performance under dynamic conditions. To account for the impact of electromagnetic interference on bandwidth allocation, an accelerated genetic algorithm can be used [2].

An important and highly relevant task is optimizing 5G BS placement in indoor environments, such as shopping malls, medical institutions, and educational facilities. In medical institutions, where services like telemedicine and remote health monitoring are in use, the optimization of BS placement becomes a matter of vital importance.

In shopping malls, where many visitors simultaneously use mobile devices for navigation, shopping, entertainment, and communication, optimal BS deployment is essential to ensure high-quality and reliable connectivity.

The high user density, the rapid fluctuation in user presence, the diversity of services, and the requirement for robust coverage in complex indoor layouts all indicate that the problem of 5G BS placement is a complex, multi-criteria optimization challenge. It demands a comprehensive analysis that accounts for numerous factors, including coverage, throughput capacity, deployment cost, and energy efficiency. Solving this problem would enable the network's structural optimization, provide uniform coverage, avoid network congestion, and ensure high-quality service delivery to end users.

The problem of optimizing the placement of BSs has been studied by numerous researchers who have proposed various approaches and methodologies. In [3], a method for planning the optimal deployment of 5G BSs is introduced, combining conventional techniques with differential evolution algorithms while considering parameters such as transmission speed, planning accuracy, and planning depth.

In [4], under the conditions of a heterogeneous network structure, a heuristic solution based on a genetic algorithm is proposed to solve the BS placement problem. In [5], the time complexity of solving such classes of problems is addressed through optimization based on a genetic algorithm, aiming to reduce computation time.

In [6], the optimal placement of BSs in open terrain is investigated. The solution is also based on a genetic algorithm and considers factors such as installation costs, Euclidean distances between BSs, maximization of the coverage area per BS, and guaranteed throughput per user.

In summary, it can be stated that the problem of structural optimization specifically, the placement of base stations in 5G networks remains a highly relevant research challenge that has attracted significant attention from the scientific community. The proposed solutions in the literature are largely based on the use of genetic algorithms, which enable adaptation to the dynamic nature of the problem.

However, it is important to note that many existing studies have been conducted under "ideal conditions" where critical factors such as intra-system interference were not considered. This simplification limits the applicability of such models in realistic deployment scenarios, particularly in dense indoor environments.

Intra-system interference, arising from signal collisions between neighboring base stations and between user terminals, is a critical factor that limits the performance and reliability of 5G networks. As the density of BS deployments and user terminals increases, the network becomes increasingly susceptible to significant degradation in communication quality.

Existing models often fail to account for intra-system interference, which leads to reduced throughput, increased latency, deterioration in user QoS, and inefficient utilization of the radio spectrum.

Therefore, studying base station placement optimization in 5G networks with explicit consideration of intra-system interference is a pressing research and engineering challenge. The development of advanced mathematical models and optimization algorithms aimed at minimizing the impact of interference is essential to enhancing network performance and reliability, ensuring high-quality service delivery, and improving the efficiency of spectrum usage. When developing a genetic algorithm, it is necessary to consider numerical analysis methods that can be used to optimize the algorithm's parameters [7]. For effective quality of service provisioning in wireless mesh networks, packet-level resource allocation methods can be used [8].

2 PROBLEM FORMULATION

This study aims to develop a method (algorithm) for determining the spatial coordinates of base stations (BSs) in the context of deploying a 5G network in indoor environments - such as shopping centers or telemedicine facilities - under conditions of random user distribution. The proposed solution aims to ensure maximum network coverage and throughput while explicitly considering the impact of intra-system interference arising from signal collisions between users and adjacent base stations.

It is evident that during the techno-economic justification of 5G network topology design, it is necessary to consider key performance indicators such as throughput, QoS, reliability, and deployment cost.

The described problem belongs to the class of structural optimization problems, and in terms of optimization theory, it can be formulated as follows:

There exists a bounded indoor environment (e.g., a business center or a medical facility) within which a 5G network must be deployed. The network consists of several BSs, each assigned a unique identifier and associated with spatial coordinates.

Within the boundaries of this object, a fixed number of user terminals (UEs) are randomly distributed across the area (see Fig. 1).

The objective is to determine the optimal number of base stations and their optimal locations in such a way as to provide effective signal coverage over the entire area, maximize the average data rate per user and account for indoor propagation conditions, where radio signals are affected by multipath, obstructions, and intra-system interference.

This interference arises from adjacent base stations and other nearby users operating within the same frequency bands. The resulting problem requires multi-criteria optimization, where spatial arrangement directly impacts performance indicators such as coverage, throughput, and QoS.

The formulated problem of 5G base station placement optimization is a nonlinear, NP-complete, integer programming task. Exact solutions to such problems typically require exhaustive search over the solution space; however, this approach becomes impractical for high-dimensional instances due to the exponential growth of computational complexity.

As a result, a wide range of heuristic methods have been developed and are commonly employed to obtain near-optimal solutions within acceptable computational time. Among the well-known heuristic approaches for nonlinear optimization are the penalty function method, the projected gradient method, the interior-point method, and the branch and bound technique. These methods are generally based on pseudo-random search strategies and systematically evaluating the explored solution space.

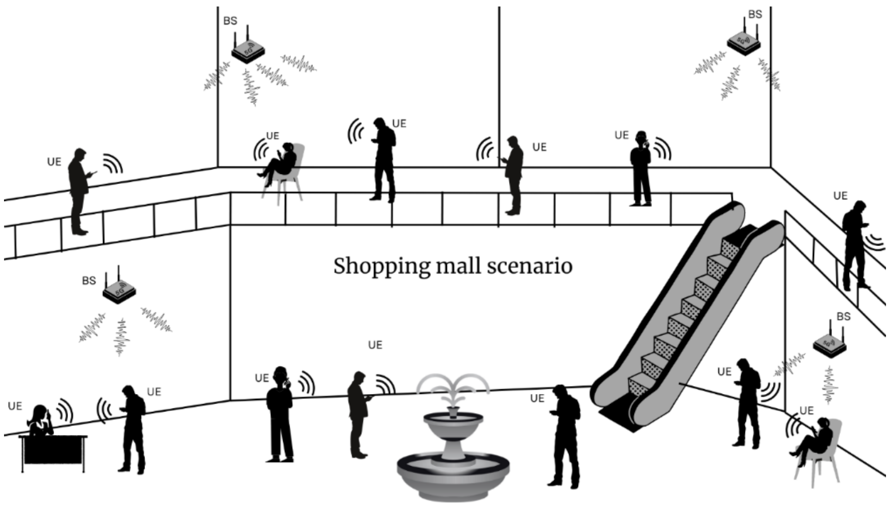


Figure 1: A typical shopping mall scenario.

In this context, nature-inspired algorithms, particularly genetic algorithms (GAs), have demonstrated significant potential. Thanks to their mutation mechanism, which introduces random perturbations into candidate solutions, GAs can escape local minima and explore broader regions of the search space.
Moreover, genetic algorithms are well suited to solving large-scale problems due to their:

- ability to handle a large number of variables;
- lack of dependence on gradient information and inherent parallelism allows for computational acceleration.

As noted in [9], GAs have been successfully applied to various optimization problems in infocommunications. For these reasons, a genetic algorithm was selected as the core method for solving the optimization problem described in this study.

3 PROPOSED METHOD

As input data for the model, we consider:

- a matrix of spatial coordinates for the base stations (BSs);
- the technical characteristics of the BSs;
- the geometric parameters of the deployment environment.

The environment is modeled as a square area of size $A \times A$ meters, containing N mobile users $n_i = (x_i, y_i)$ and M base stations $m_j = (x_j, y_j)$, which are initially randomly positioned using the Monte Carlo method.

The objective is to determine the coordinates of the base stations that maximize the average data rate for all users, while explicitly considering intra-system interference, including interference from neighboring BSs and nearby users (1).

$$M = \{m_1, m_2 ... m_K\}, \quad m_j = (x_j, y_j) \tag{1}$$

Accordingly, the objective function of the optimization problem can be formulated as follows (2):

$$\begin{aligned} \max F(M) &= \frac{1}{N}\sum_{i=1}^{N} R_i - \alpha \sum_{i=1}^{N} \delta(R_i < R_{\min}) - \\ &- \beta \sum_{i=1}^{N} \delta(n_j > n_{\max}) - \gamma \sum_{i=1}^{N}\sum_{i=1}^{N} \delta(\|b_j - b_k\| < d_{\min}) \end{aligned} \tag{2}$$

R_i is the data rate for user i, $R_{\min}$ is the minimum acceptable user rate, $n_{\max}$ is the maximum capacity per BS, $\delta(\cdot)$ is the indicator function, α, β and γ are penalty coefficients.

The achievable data rate for a given user i can be estimated (3) using the Shannon–Hartley theorem, which accounts for the signal-to-interference-plus-noise ratio (SINR):

$$R_i = B \cdot \log_2\left(1 + \frac{P_{i,j^*}}{\sum_{k \neq j} P_{i,k} + N_0}\right) \tag{3}$$

where B is the available channel bandwidth (Hz),, P_{i,j^*} is the power received by user i from its serving base station j^*, $P_{i,k}$ is the interfering power received from base station, N_0 is the thermal noise power.

The power received by user i from its serving base station j^* calculated as follows (4):

$$P_{i,j^*} = \frac{P_{BS} G_{BS}}{PL_{i,j}}, \tag{4}$$

where P_{BS} is the transmit power of the base station, G_{BS} is the antenna gain of the base station, $PL_{i,j}$ is the path loss between BS j and user i.

For the selected environment (a shopping mall), we apply the Indoor Path Loss Model [10], which captures signal propagation characteristics in complex indoor scenarios with obstacles and dense user presence (5).

$$PL_{i,j} = PL(d_0) + 10v\log_{10}\left(\frac{d_{i,j}}{d_0}\right) + \sum_{k=1}^{K} L_k + X_\sigma, \tag{5}$$

where $PL(d_0)$ is the free-space loss at the reference distance., $d_{i,j} = \|n_i - m_j\|$ is the distance between the BS and user, v is the path loss exponent (environment-dependent, e.g., 2–4), L_k is the loss introduced by the k-th obstacle (e.g., wall, glass), K is the number of obstacles, $X_\sigma \sim N(0, \sigma^2)$ is a Gaussian random variable accounting for shadow fading,.

In terms of the genetic algorithm (GA), the solution space is represented as a population consisting of multiple individuals, where each individual encodes a candidate deployment configuration of M base stations.

Formally, an individual (chromosome) can be represented as (6):

$$X = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \dots & \dots \\ x_M & y_M \end{bmatrix} \in \quad ^{M\times 2}, \qquad (6)$$

At the next step, a total of K individuals (candidate solutions) are generated to form the initial population of the genetic algorithm. $\left\{X^{(1)},...,X^{(K)}\right\}$. The coordinates of each base station $m_j = \left(x_j, y_j\right)$ are randomly initialized within the boundaries of the selected area, ensuring that all initial placements lie within the defined indoor environment $A\times A$.

For each individual, the objective (fitness) function is evaluated based on the resulting network performance. Specifically, the data rate is calculated for every user, and the average user throughput is used as the primary fitness metric.

Individuals that achieve higher average user rates are selected to advance to the next generation. Among a randomly chosen subset of the population, the best-performing individual is identified as an elite solution and retained.

Next, a crossover operation is applied—this involves combining the coordinates of two parent individuals to generate an offspring. Formally, given two parent solutions (7), (8):

$$X_{child} = \lambda X_1 + \left(1-\lambda\right) X_2,\ \ \lambda \in [0,1], \qquad (7)$$

This operation blends the base station positions from two selected parents to produce a new candidate solution, promoting diversity and the exploration of promising regions in the solution space.

Another essential step of the genetic algorithm is the mutation operation, which involves a random displacement of a single base station within the boundaries of the deployment area.

$$m_j \leftarrow m_j + \delta,\ \delta \sim N\left(0,\sigma^2\right). \qquad (8)$$

4 EXPERIMENTAL RESULTS

Figure 2a illustrates the initial random placement of 10 base stations (red triangles) within a 100×100 meter indoor area containing 200 active users (blue dots), uniformly distributed using the Monte Carlo method. This initial configuration demonstrates non-uniform coverage, with evident clustering of base stations in certain regions and coverage holes or "dead zones".

Figure 2b shows the optimized deployment of base stations (green triangles) obtained through the genetic algorithm-based optimization. The base stations are now more evenly distributed, significantly reducing poor coverage areas and more effectively adapting to user density across the environment.

As shown in Figure 2, after optimization, the base stations were placed more uniformly across the service area, leading to improved coverage and reduced dead zones.

To provide a clear visual representation of how this optimized placement affects signal quality, SINR heatmaps are presented in Figures 3a and 3b, illustrating the spatial distribution of the signal-to-interference-plus-noise ratio (SINR) before and after optimization.

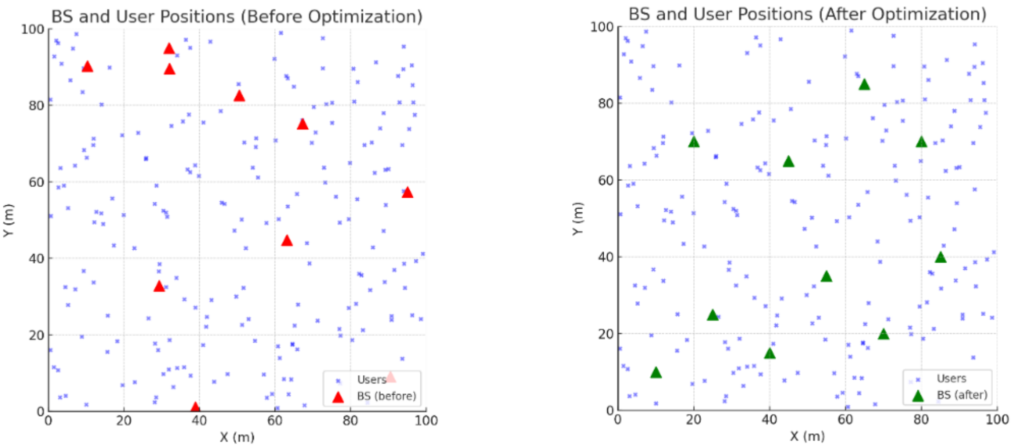
The improvement in SINR directly impacts user service quality, particularly on the achievable data rate. Figures 4a and 4b present the heatmaps of the estimated user data rates (in Mbps) before and after optimization, respectively.

It is evident that after optimization, a significant portion of the area achieves higher data rates, while low-throughput zones have been substantially minimized.

5 CONCLUSIONS

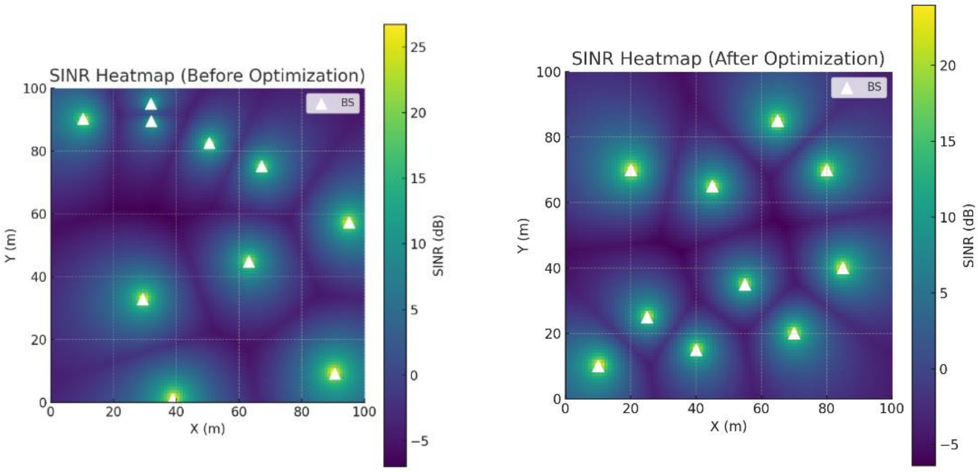
In this study, a comprehensive mathematical model of a fifth-generation (5G) mobile communication network was developed, considering the spatial distribution of base stations (BSs), intra-system interference, signal level, SINR, and data rate for each user. Particular attention was given to the problem of optimal BS placement in indoor environments, such as shopping malls and medical centers, which is especially relevant due to high user density and complex propagation conditions.

To solve the optimization problem, a genetic algorithm was implemented with adaptive operators for selection, crossover, and mutation. The mathematical formulation of the fitness function aimed to maximize the average user data rate, while incorporating penalties for base station overload, excessive deployment density, and violations of minimum quality of service (QoS) requirements.



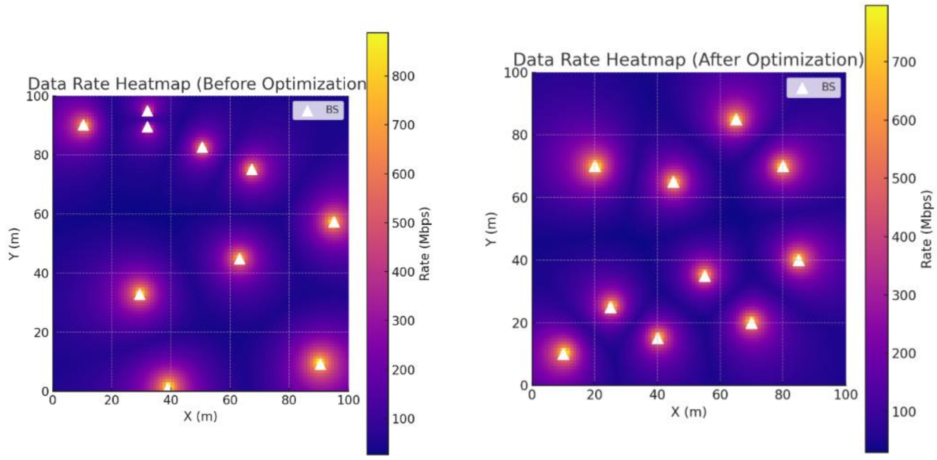
a) b)

Figure 2: Distribution of users (a) and base stations (b) within the considered area.



a) b)

Figure 3: SINR heatmap before (a) and after optimization (b).



a) b)

Figure 4: Rate heatmap before (a) and after optimization (b).

As a result of numerical simulations conducted for a 100×100 meter area with 200 users and 10 base stations, a significant improvement in key network performance metrics was observed. In particular, the average user data rate increased from approximately 123.25 Mbps to 133.76 Mbps after optimization.

Furthermore, the number of areas with poor coverage (i.e., low SINR) was significantly reduced, and users experiencing data rates below 5 Mbps were effectively eliminated. The heatmaps of SINR and data rate confirmed that the optimized BS placement provides a more uniform distribution of resources across the area and ensures effective coverage throughout the environment.

The research confirms the effectiveness of evolutionary approaches for solving the problem of radio network planning, particularly in high-density indoor environments with limited physical space.

Further development of the proposed model may include:

- support for multi-floor environments,
- path loss modeling those accounts for wall materials and obstructions,
- adaptive frequency resource management, and
- the application of deep learning methods for predictive base station placement based on user traffic patterns.

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] S.M. Talha, S. Siden, R. Tsaryov, and L. Nikityuk, "Assessment of the Possibility of Using 5G to Build Telemedicine Networks in Various Environment," 2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Dortmund, Germany, 2023, pp. 1125-1129, doi: 10.1109/IDAACS58523.2023.10348929.

[2] D. Lin and F. Labeau, "Accelerated genetic algorithm for bandwidth allocation in view of EMI for wireless healthcare," in Wireless Communications and Networking Conference (WCNC), IEEE, 2012, pp. 3312-3317.

[3] R. Sachan, S. Dash, B. Sahu, J.R., et al., "Energy efficient base station location optimization for green B5G networks," in Proc. Int. Conf., 2022, pp. 441-448.

[4] H. Ren, "Vehicle routing optimization of logistics distribution based on genetic algorithm (VRPTW)," Front. Econ. Manag., vol. 3, no. 1, pp. 294-299, 2022.

[5] P. Yu, Y. Shi, L. Wang, et al., "A method for optimizing communication network topology based on genetic algorithm," in Proc. Int. Conf., 2022, pp. 30-40.

[6] G. Chen, X. Wang, and G. Yang, "Research and Implementation of 5G Base Station Location Optimization Problem Based on Genetic Algorithm," unpublished, 2023.

[7] L.P. Feldman, Numerical Methods in Computer Science: A Study Guide, Kyiv, Ukraine: BHV Publishing Group, 2009, 479 p.

[8] H.T. Cheng and W. Zhuang, "Novel packet-level resource allocation with effective QoS provisioning for wireless mesh networks," IEEE Trans. Wireless Commun., vol. 8, no. 2, pp. 694-700, 2009.

[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. Int. Conf. Appl. Innov. IT, vol. 12, no. 1, pp. 19-24, 2023, doi: 10.25673/115637.

[10] H. A. Obeidat, R. Asif, N.T. Ali, Y.A. Dama, O.A. Obeidat, S.M.R. Jones, et al., "An indoor path loss prediction model using wall correction factors for wireless local area network and 5G indoor networks," Radio Sci., vol. 53, pp. 544-564, 2018.

Optimizing Routine Educational Tasks through Prompt Engineering: A Comparative Study of AI Chatbots

Svitlana Skvortsova[1], Tetiana Symonenko[2] and Kira Hnezdilova[3]
[1]*Department of Mathematics and Methods of its Teaching, The State Institution "South Ukrainian National Pedagogical University named after K. Ushynsky", Staroportofrankyvska Str. 26, 65020 Odesa, Ukraine*
[2]*Department of Ukrainian Philology and Social Communication, The Bohdan Khmelnytsky National University of Cherkasy, Shevchenko Avenue 81, 18031 Cherkasy, Ukraine*
[3]*Department of Primary Education, The Bohdan Khmelnytsky National University of Cherkasy, Shevchenko Avenue 81, 18031 Cherkasy, Ukraine*
skvo08@i.ua, irinka200888@i.ua, kiragnez@gmail.com

Keywords: AI in Education, Prompt Engineering, Chatbot Evaluation, Lesson Planning Automation, AI-Assisted Teaching.

Abstract: The rapid integration of artificial intelligence (AI) in education necessitates the development of effective strategies for optimizing routine teaching tasks. This study explores the relevance of prompt engineering as a tool for enhancing AI-generated responses, reducing educators' workload, and improving the efficiency of lesson planning, content creation, and assessment design. The primary objective of this research is to develop and evaluate a structured methodology for designing prompts that maximize the relevance, completeness, and applicability of AI-generated outputs. To achieve this goal, a three-phase methodology was employed: (1) a preparatory phase involving a literature review and the development of standardized educational prompts, (2) an experimental phase testing these prompts across multiple AI chatbot models (Claude, GPT, and Copilot), and (3) an analytical phase assessing chatbot responses based on predefined criteria, including relevance, accuracy, completeness, practicality, and structuredness. The results indicate significant differences in chatbot performance. Claude demonstrated superior contextual understanding, GPT provided well-balanced and structured responses, while Copilot exhibited high factual accuracy but required improvements in contextual adaptation. Statistical analysis using the Kruskal-Wallis H test confirmed these variations, highlighting the necessity of model-specific prompt optimization. The study's findings have both practical and theoretical significance. Practically, they provide educators with a structured approach to prompt engineering, enabling more effective use of AI tools in teaching. Theoretically, the research contributes to the growing field of AI-assisted education by offering insights into optimizing human-AI interaction. The conclusions emphasize the need for continued refinement of AI models and further exploration of prompt engineering techniques. Future research should focus on expanding testing across various disciplines and integrating AI-driven tools into digital learning environments to enhance personalized education.

1 INTRODUCTION

The rapid development of digital technologies has significantly impacted the cognitive processes of modern students, particularly Generation Alpha, who are accustomed to constant digital exposure. This has led to "clip thinking," where attention shifts quickly between content, posing challenges for deep focus and long-term retention [8]. Traditional educational methods may not align with their needs, necessitating innovative approaches. Integrating generative language models and prompt-based technologies into education can optimize routine tasks, enhance teaching efficiency, and create customized materials that resonate with digital learners [1], [5].

This study explores how prompt-writing technologies can optimize educational tasks and create engaging learning materials for Generation Alpha, leveraging AI models like GPT, Claude, and Copilot. The research demonstrates AI's potential in optimizing pedagogical activities, creating educational content, personalizing learning experiences, and increasing student engagement. Effective prompt engineering is crucial for successful

AI interaction, emphasizing the need for precise and structured prompts [2]. The study aims to develop and evaluate a structured methodology for creating prompts that optimize routine educational tasks.

Research Questions:

1) How do different AI chatbots (Claude, GPT, Copilot) respond to structured educational prompts?
2) What factors influence the accuracy, completeness, and applicability of AI-generated responses?
3) How can prompt engineering be refined to maximize its effectiveness in an educational setting?

2 PROMPT ENGINEERING IN EDUCATION

The process of designing, optimizing, and refining prompts to enhance the efficiency of generative AI models is known as prompt engineering. At the current stage of technological development, the prompt has evolved from a simple “query” to a crucial tool for interacting with LLMs (large language models). Prompt engineering now encompasses not only textual data but also other modalities, such as image, audio, and video processing. The prompt is input text or other forms of data provided to a language model (or another generative system) to initiate the generation of an appropriate response. The structure of a prompt may include instructions, questions, examples, context, or a combination of these elements. The functional purpose of a prompt is to set the parameters and context for the model’s operation, guiding it toward generating the desired response or behavior. There are two primary types of prompts: textual (instructions or queries) and multimodal (combinations of text with images, audio, etc.).

According to Schulhoff et al. [7], the prompt engineering process for maximizing the potential of LLMs and obtaining relevant, accurate, and useful results with minimal effort involves the following stages:

1) Design. Formulating a prompt for a specific task considering the goal, target audience, and context.
2) Optimization. Applying various techniques such as few-shot prompting (adding a few examples of desired responses to the main query), chain-of-thought prompting (encouraging the model to generate a sequential logical thought process before forming the final answer), or specifying response formats to improve accuracy.
3) Iteration. Experimental testing of different prompt variations and analyzing the obtained results.
4) Adaptation. Adjusting prompts based on the specific characteristics of the model and changes in task requirements.

The publication "The Prompt Report: A Systematic Survey of Prompting Techniques" [7] provides recommendations for creating prompts for LLMs like ChatGPT and discusses issues related to safety and reliability. Based on these recommendations, the following requirements for the prompt creation process can be formulated:

1) Clarity and specificity.
2) Contextualization.
3) Use of examples (few-shot prompting).
4) Task decomposition into steps (chain-of-thought prompting).
5) Specification of response format.
6) Avoiding ambiguous terms.
7) Testing and iteration.
8) Consideration of ethical aspects.

According to Wei et al. [9], prompt engineering is a tool for adapting language models to specific contexts. For example, in solving educational tasks, prompts should consider subject specificity and the students’ level of knowledge, among other factors.

Moreover, we agree with authors who view prompt engineering as an interdisciplinary skill that combines knowledge about query structure, AI tools, and task specifics. This position deserves attention since writing high-quality prompts in a particular field context requires subject knowledge, certain linguistic skills, logical thinking, and more. This is where the complexity and multidimensionality of prompt engineering lie. When developing prompts, it is essential to understand how linguistic nuances impact the capabilities of generative AI, ensuring the creation of authentic and well-adapted content for effective teaching and learning interactions [4].

Bozkurt and Sharma [3] emphasize the importance of developing prompt engineering skills among educators to effectively harness the full potential of generative AI in educational contexts.

They argue that co-creation involving generative AI represents a powerful approach in education, highlighting the significance of human-machine interaction facilitated by carefully crafted prompts.

Approaches to writing prompts often include using clear and specific instructions, keywords, tone, and response style settings. However, essential elements for creating a quality prompt must be highlighted:

1) Instruction. A specific task the user wants the AI to perform, such as "write a motivational speech for a lesson" or "create a story."
2) Context. Information that can guide the model to provide more accurate answers, such as the lesson topic, objectives, or story theme and style.
3) Input data. Detailed contextual information, such as structural elements of the lesson or specific problem types, or detailed descriptions of characters in a story. Providing examples of desired outcomes is also appropriate.
4) Output type or format. For example, "a fantasy story of 400 words." Currently, no definitive guidelines exist for creating quality prompts for educational purposes. However, it is essential to avoid unclear and ambiguous formulations since AI may misinterpret the request and produce unintended results. The absence of context or grammatical errors can also lead to misunderstandings by the AI. Experimenting with different formulations to find the best approach for your task is crucial.

Bozkurt and Sharma [3] suggest strategies for creating effective prompts:

1) Clearly define the goal. Specify the purpose, desired response type, or result.
2) Understand the AI model's capabilities. Leverage the model's strengths and limitations, creating prompts aligned with its expertise. Role prompting (assigning specific roles) can sometimes "break" the model's default behavior.
3) Use concise and clear formulations. Avoid confusing or irrelevant prompts.
4) Provide sufficient context. Enable the AI model to better understand the task or subject.
5) Use examples of desired outcomes. Show the AI what kind of output is expected.
6) Fine-tune and debug prompts. Make adjustments for improved results.
7) Specify output format or structure. Provide clarity on how the answer should look.
8) Include key details. Ensure the AI receives all necessary information.
9) Test different prompt variations. Identify the most effective formulation.
10) Consider safety and ethical aspects. Maintain responsible AI usage practices.

According to the authors, adhering to these strategies allows educators, researchers, and users to optimize prompt engineering for meaningful and accurate responses from language models, aligning with their specific goals and requirements. They underscore that effective prompt engineering is not merely a technical skill but an art of communication that requires understanding AI's technical capabilities and the nuances of human language and interaction [3].

2.1 Technology for Creating Prompts for Optimizing Routine Tasks in Education

Based on the analysis of scientific literature on the issues of prompt engineering, a technology for creating prompts has been developed, aimed at optimizing routine tasks in the educational sphere. Prompts, as instructions for artificial intelligence (AI) systems, allow automating such pedagogical processes as developing lesson plans, creating learning tasks, preparing didactic materials, checking work, etc. The prompt creation process is considered according to the main stages: design, optimization, iteration, and adaptation.

Design Stage. Forming a Basic Prompt The prompt creation stage involves designing an initial request, which includes formulating a clear requirement (determining the desired result), specifying the topic and goal, determining the response format and target audience. Additional details can increase the effectiveness of the prompt.

For example, "Create 5 mathematical problems for 2nd grade students on the topic 'Addition and subtraction of two-digit numbers.' The level of complexity is simple problems that are solved by arithmetic operations of addition and subtraction. The format is plot problems with practical content."

Optimization Stage. Correction and Clarification of the Prompt The next stage is the evaluation of the result – the generated AI answer. The evaluation includes checking the correspondence of the response to the query, its correctness (truthfulness) and the logic of the information presented. In the case of partial or complete non-compliance of the result with the user's expectations, the prompt is optimized. Optimization may include: adding clarifications, breaking the task into subtasks, providing examples, indicating specific requirements. For example, the optimization of the previous prompt may look like this: "Additional requirements: each task must contain a context that describes situations from the school life of students; the formulation of the task must correspond to the canonical structure (first the condition, then the question); include visual elements (diagrams, tables)."

Usage and Iteration Stage After obtaining a satisfactory result, the prompt can be generalized (formed into a template) and used to perform similar routine tasks. For the purpose of effective use, it is advisable to save successful prompts for further use. In the process of reusing the prompt, there may be a need to improve it based on the accumulated experience. Experimenting with different prompt formulations is a process of iteration. This stage is not mandatory and is performed as needed.

2.1.1 Prompt Creation Algorithm

The prompt creation process can be presented as the following algorithm:

1) Formulation of the basic query with additional details (if necessary):
   - Clear requirement/question.
   - Topic and goal/objective.
   - Type and format of the response.
   - Target audience.
   - Additional details.
2) Evaluation of the quality of the response to the basic query:
   - Relevance of the query.
   - Completeness of the response.
   - Correctness of the information.
   - Practical applicability.
3) Optimization (if necessary):
   - Adding clarifications.
   - Breaking down into subtasks.
   - Providing examples.
   - Indicating chain-of-thought prompting.
   - Assigning the role of AI.
4) Checking the result (after optimization):
   - Relevance of the request.
   - Completeness of the answer.
   - Correctness of the information.
   - Practical applicability.
5) Iteration (if necessary):
   - Clarifying the requirements.
   - Adding details.
   - Changing the format.
   - Correcting the structure.

For educators new to creating prompts for AI systems, it is recommended to start with a basic prompt and gradually add contextual details. Using standardized templates can facilitate this process. The following template is suggested:

1) Task: [clearly formulate the required action or result].
2) Topic: [specify the subject area or topic of the task].
3) Target audience: [describe the characteristics of the audience (e.g., grade, level of knowledge, age)].
4) Format: [determine the desired format of the output (e.g., list, table, essay)].
5) Additional requirements: [specify specific requirements such as volume, style, level of detail, visual elements].

This template helps structure the request and provides necessary information for correct AI interpretation. Avoid the following mistakes:

1) Overly general query. Vague wording leads to ambiguous results. For example, instead of "Create a math problem," use "Create three problems to find the area of a rectangle for 5th graders."
2) Conflicting requirements. Contradictory instructions can confuse the AI and lead to incorrect answers.
3) Overloading with details. Too many details at the initial stage can complicate the prompt creation process. Add details gradually.
4) Examples of optimized prompts. Overly general wording:
   - Incorrect: "Create a math problem."
   - Optimized: "Create 5 math problems: Topic: adding fractions, Grade: 6, Difficulty level: medium, Problem types: 2 on calculations, 3 text problems."
5) Conflicting requirements:
   - Incorrect: "Solve the problem quickly, but explain each step in detail."
   - Optimized: "Solve the problem of finding the area of a triangle with sides 3, 4, 5. Use Heron's formula. Show intermediate calculations. Write down the answer with an explanation."

Note that various prompt optimization techniques can also be used at the stage of designing the basic prompt, depending on the task at hand:

1) Step-by-step process description: chain-of-thought prompting:
2) Sample response: few-shot prompting.
3) Specific role or expertise: role prompting.
4) Complex structure: structuring + chain-of-thought.
5) Creative task: role prompting + few-shot.
6) Solving complex problems: chain-of-thought + few-shot.

This approach allows teachers to master prompt engineering skills and effectively use AI in educational activities. The proposed algorithm optimizes the process of creating prompts for AI, ensuring relevant and useful results for routine educational tasks.

2.2 Methodology

2.2.1 Research Design

The study follows a three-phase approach:

1) Preparatory stage. Literature review, identification of routine educational tasks, and development of structured prompts.
2) Experimental stage. Testing prompts across Claude, GPT, and Copilot chatbots, documenting response quality.
3) Analytical stage. Comparative evaluation of chatbot responses using predefined assessment criteria.

2.2.2 Research Methods

At the preparatory stage, existing approaches to prompt engineering and typical errors when creating prompts were studied using the following methods:

1) Analysis of scientific publications, articles, reports, and API documentation of chatbots using search engines: Web of Science, Scopus, Google Scholar, Researchgate, and other scientific databases.
2) Analysis of the teacher's pedagogical activity to identify typical routine tasks in the lesson, develop prompt templates for different types of activities and stages of the lesson, and model the teacher's activity in developing tasks from different subject areas (e.g., mathematics, Ukrainian language).

At the experimental stage, the effectiveness of the developed technology was assessed by comparative analysis of the responses of different chatbots to the same prompts. Experimental testing involved applying identical prompts to different chatbot systems and further analysis of the generated responses with subsequent optimization and iteration.

Meta-prompts were developed to solve certain routine teacher tasks, and data on the responses of different chatbots to the same prompts were collected.

At the analytical stage, quantitative processing of the obtained results was carried out, comparative analysis of the responses of different chatbots by each criterion was conducted, and mathematical statistics methods were applied to identify statistically significant differences between the results.

Preparatory Stage of the Study. At this stage, chatbot models were selected for further testing. The results of a survey of teachers at Ukrainian universities [8], publications by scientists from other countries [1], [5] showed the high popularity of Chat GPT among Ukrainian educators, correlating with global trends. Given the integration of the Copilot chatbot into the Microsoft 365 corporate suite, widely used in Ukrainian universities, the choice of Chat GPT and Copilot for testing the developed prompt creation technology is justified.

Based on the survey results and subjective preferences, the Claude chatbot was chosen as the third object of the study [8]. At the time of the study, limited scientific publications were found on the spread of the Claude chatbot among teachers in other countries, while most research focuses on Chat GPT.

Thus, the result of the preparatory stage was the selection of three chatbots for experimental testing: Chat GPT, Copilot, and Claude.

A set of test tasks in mathematics and the Ukrainian language was created, for which the corresponding basic prompts were developed. The teacher's activities in organizing students' educational and cognitive activities in the lesson were analyzed based on the generally accepted structure of a combined lesson, which includes the following stages:

1) the motivational stage of the lesson;
2) the stage of updating knowledge and methods of activity;
3) the stage of formation of new knowledge and methods of action;
4) the stage of consolidation and formation of skills and abilities;
5) the stage of lesson results and reflection of educational and cognitive activity.

Within each stage, various routine tasks are implemented, related to different types of student activities (e.g., oral survey, mathematical or spelling dictation, individual survey, homework check) and different forms of organization of educational activity (collective, pair, group, individual). Considering the characteristics of modern students as representatives of the digital generation, one task a teacher can solve with AI is creating interactive learning environments, such as lesson shells or individual stages of lessons in the format of journeys, quests, competitions.

The following criteria were defined to assess the effectiveness of the prompt:

1) Relevance (K1) – alignment with the initial prompt.

2) Completeness (K2) – depth of the response.
3) Accuracy (K3) – consistency with educational standards.
4) Practicality (K4) – direct applicability in lesson planning.
5) Structuredness (K5) – logical organization of the response.

A five-point evaluation scale is used for these criteria. For example, for criterion K1 (Request compliance):

- 5 – full compliance with all prompt requirements;
- 4 – compliance with the main requirements with minor deviations;
- 3 – partial compliance with the requirements;
- 2 – significant deviations from the requirements;
- 1 – minimal compliance;
- 0 – complete non-compliance.

At the preparatory stage, a form was developed for recording the results of the responses of the three studied chatbots – Chat GPT, Copilot, and Claude. This form allows systematizing and comparing the obtained data, ensuring the objectivity and scientific validity of the study.

Using the developed technology, 12 generalized prompts were created to solve routine tasks that may arise in the professional activities of teachers of any subject. At this stage of the study, experimental testing was conducted, which consisted of applying these prompts to different chat models. For this, the generalized prompts were detailed on the material of two academic subjects: mathematics and the Ukrainian language.

2.2.3 Analytical Stage of the Study

At this stage, a quantitative analysis of the results was conducted, including a comparative evaluation of responses generated by different chatbot models based on the prompt evaluation criteria. Additionally, statistical processing of the testing results for the prompt creation technology was performed. This enabled the interpretation of the experimental data and the formulation of conclusions regarding the effectiveness of the developed technology.

To assess the differences in responses produced by various chatbot models (ChatGPT, Copilot, Claude) based on prompts designed using the proposed methodology, the Kruskal-Wallis H test was applied at a significance level of $\alpha = 0.05$. The selection of this statistical test is justified by its capability to determine variations in response evaluations when switching between chatbot models.

The general hypothesis of the study was formulated as follows: using the developed technology, prompts were generated that allow obtaining relevant responses from different chatbot models according to predefined criteria.

The following statistical hypotheses were tested at a significance level of $\alpha = 0.05$:

1) H0 (null hypothesis). There is no statistically significant difference in the scores of responses (by a specific criterion) generated by different chatbot models based on prompts developed using the proposed technology.
2) H1 (alternative hypothesis). There is a statistically significant difference in the scores of responses (by a specific criterion) generated by different chatbot models based on prompts developed using the proposed technology.

Kruskal-Wallis H test calculations were performed separately for responses to prompts used in Ukrainian language lessons and separately for mathematics lessons.

According to the results presented in Table 1, the null hypothesis was rejected in all cases, confirming that there are statistically significant differences in the evaluation of responses (by specific criteria) generated by different chatbot models based on prompts developed using the proposed technology at a significance level of $p \leq 0.01$.

Pairwise comparisons between chatbots for each criterion were also performed. The results indicated statistically significant differences ($p \leq 0.01$) in the assessment of prompts between Claude and Copilot for criteria K1 (Relevance to request) and K3 (Information accuracy). Additionally, significant differences were observed between Claude and ChatGPT for criterion K4 (Practical applicability), as well as between Claude and ChatGPT ($p \leq 0.01$) and Claude and Copilot ($p \leq 0.05$) for criterion K5 (Response structure). In the evaluation by criterion K2 (Completeness of the response) insignificant differences between Claude and Copilot were found ($p \leq 0.02$).

Table 1: Kruskal-Wallis test results (K) Ukrainian language.

Criterion (K)	Test statistic, asymptotic significance (2-sided)
K1	9.564, p= 0.08
K2	6.230, p=0.044
K3	7.582, p=0.023
K4	10.404, p=0.006
K5	9.406, p=0.009

In the case of testing prompts for mathematics lessons (Table 2) according to criteria K1, K2, K3, and K4, the null hypothesis (H_0) was accepted, indicating no statistically significant differences in responses generated by different chatbot models based on the developed prompts. Therefore, pairwise comparisons were not conducted for these criteria.

Table 2: Kruskal-Wallis test results for mathematics.

Criterion (K)	Test statistic, asymptotic significance (2-sided)
K1	2.722, p=0.256
K2	2.191, p=0.334
K3	1.444, p=0.486
K4	0.105, p=0.949
K5	6.106, p=0.047

However, for criterion K5 (Response structure), the null hypothesis was rejected ($p \leq 0.05$), confirming significant differences in the evaluations of chatbot responses. Pairwise comparisons revealed significant differences between Claude and Copilot ($p \leq 0.05$). The results of the calculations are shown in Table 3.

The statistical analysis confirmed the effectiveness of the developed prompt creation technology for chatbot responses. For Ukrainian language prompts, significant differences were found among chatbot responses for all evaluation criteria, particularly between Claude and Copilot, as well as Claude and ChatGPT. In contrast, for mathematics prompts, significant differences were observed only for the criterion of response structure (K5), specifically between Claude and Copilot.

Table 3: Pairwise comparisons for K5 (mathematics prompts).

Comparison	Test statistic, SE, z, p-value
Copilot - ChatGPT	4.083, SE = 2.958, z = 1.380, p = 0.167
Copilot - Claude	7.292, SE = 2.958, z = -2.465, p = 0.014
ChatGPT - Claude	3.208, SE = 2.958, z = -1.085, p = 0.278

These findings suggest that the effectiveness of the prompt creation methodology may vary depending on the subject area. Future research could explore refinement strategies for prompt engineering to enhance response consistency across different AI chatbot models.

3 CONCLUSIONS

The study confirmed the effectiveness of the developed technology for creating educational prompts across various subjects, including mathematics and the Ukrainian language. Statistical testing using the Kruskal-Wallis H test demonstrated that different chatbot models (ChatGPT, Copilot, Claude) exhibit varying abilities in generating responses.

For Ukrainian language prompts, significant differences were found in responses based on multiple criteria. For Criterion K1 (Response Relevance), differences were observed between Claude and Copilot (p = 0.002). Criterion K2 (Response Completeness) showed differences between Claude and Copilot (p = 0.02). For Criterion K3 (Correctness of information) differences are also observed between Claude and Copilot (p = 0.018). In the evaluation by criterion K4 (Practical applicability) there are statistical differences between Claude and ChatGPT (p = 0,002) and Claude and Copilot (p = 0,021).

Criterion K5 (Response Structure) revealed distinctions between Claude and ChatGPT (p = 0.005) and Claude and Copilot (p = 0,015).

For mathematics prompts, differences were mostly insignificant, except for Criterion K5 (Response Structure), where Claude and Copilot exhibited a statistically significant difference (p = 0,014). These findings suggest that while the prompt creation technology is effective, chatbot selection should align with educational objectives.

The results indicate that teachers can utilize this technology to streamline routine tasks and select chatbot systems based on response relevance, completeness, accuracy, practical applicability, and structure. However, further research is needed to:

1) Increase the number of experts evaluating chatbot responses.
2) Extend testing to additional educational subjects.

Our findings align with research by Yurchak et al. [10], which highlighted chatbot-specific strengths, weaknesses, and prompt optimization strategies. The study also confirms the necessity of integrating professional expertise with AI-generated outputs, as noted by Li et al. [6].

The proposed technology applies known prompt optimization techniques, supporting Bozkurt’s [2] thesis on operational design as key to enhancing human-AI communication. This aligns with Cain’s [4] conclusions on the role of prompts in fostering personalized and equitable learning experiences. The study confirms that structured, reusable, and adaptive prompts can reduce teacher workload and improve lesson planning efficiency, supporting the flexibility required for diverse educational contexts.

REFERENCES

[1] D. Baidoo-Anu and L. Owusu Ansah, "Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning," SSRN, 2023. [Online]. Available: https://dx.doi.org/10.2139/ssrn.4337484.

[2] A. Bozkurt, "Tell me your prompts and I will make them true: The alchemy of prompt engineering and generative AI," Open Praxis, vol. 16, no. 2, pp. 111–118, 2024. [Online]. Available: https://doi.org/10.55982/openpraxis.16.2.661

[3] A. Bozkurt and R. C. Sharma, "Generative AI and prompt engineering: The art of whispering to let the genie out of the algorithmic world," Asian Journal of Distance Education, vol. 18, no. 2, pp. i–vii, 2023. [Online]. Available: https://doi.org/10.5281/zenodo.8174941.

[4] W. Cain, "Prompting Change: Exploring Prompt Engineering in Large Language Model AI and Its Potential to Transform Education," TechTrends, vol. 68, no. 1, pp. 47–57, 2024. [Online]. Available: https://doi.org/10.1007/s11528-023-00896-0.

[5] G. Cooper, "Examining science education in ChatGPT: An exploratory study of generative artificial intelligence," Journal of Science Education and Technology, vol. 32, no. 3, pp. 444–452, 2023. [Online]. Available: https://doi.org/10.1007/s10956-023-10039-y.

[6] Y. Li, J. Shi, and Z. Zhang, "An approach for rapid source code development based on ChatGPT and prompt engineering," IEEE Access, vol. 12, 2024. [Online]. Available: https://doi.org/10.1109/ACCESS.2024.3385682.

[7] S. Schulhoff, M. Ilie, N. Balepur, K. Kahadze, A. Liu, C. Si, ... and P. Resnik, "The prompt report: A systematic survey of prompting techniques," arXiv preprint, arXiv:2406.06608, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2406.06608.

[8] S. Skvortsova, T. Symonenko, T. Britskan, O. Onopriienko, and R. Romanyshyn, "Artificial intelligence in the professional activity of a university lecturer in Ukraine: Realities and prospects," Proceedings of the Second International Workshop on Artificial Intelligent Systems in Education, co-located with 23rd International Conference of the Italian Association for Artificial Intelligence (AIxIA 2024), vol. 3879, 2024. [Online]. Available: https://ceur-ws.org/Vol-3879/AIxEDU2024_paper_14.pdf.

[9] J. Wei, X. Wang, D. Schuurmans, M. Bosma, E. Chi, Q. Le, and D. Zhou, "Chain of thought prompting elicits reasoning in large language models," arXiv preprint, arXiv:2201.11903, 2022. [Online]. Available: https://arxiv.org/pdf/2201.11903.

[10] I. Y. Yurchak, O. O. Kychuk, V. M. Oksentyuk, and A. O. Khich, "Capabilities and limitations of large language models," Computer Systems and Networks, vol. 6, no. 2, pp. 286–300, 2024. [Online]. Available: https://doi.org/10.23939/csn2024.02.286.

Classification of the Topicality and Relevance of Evaluation Tools for VR Applications

Caroline Schon [1], Rong Huang[2], Hanna Hessenmüller[1], Simon Przybyl[1] and Johannes Tümler[1]
[1]*Department Electrical Engineering, Mechanical Engineering and Engineering Management, Anhalt University of Applied Sciences, Bernburger Str. 55, 06366 Köthen, Germany*
[2]*Department Architecture, Facility Management and Geoinformation, Anhalt University of Applied Sciences, Franz-Mehring Str. 21, 06846 Dessau, Germany*
{*caroline.schon, rong.huang, hanna.hessenmueller, johannes.tuemler*}*@hs-anhalt.de, simon.przybyl@student.hs-anhalt.de*

Keywords: Virtual Reality, Evaluation, Questionnaire Analysis, User Involvement

Abstract: Virtual reality (VR) has emerged as one of the fastest growing technologies in recent years, offering transformative opportunities in various disciplines such as education, medicine and industry. Classifying the advantages and disadvantages of such VR simulations must be done using appropriate evaluation tools. Despite these rapid technological developments, the evaluation of VR experiences lags behind technological advances. This article examines the discrepancy between the rapid technical evolution of VR and the slow progress in the development of suitable evaluation methods. Existing evaluation approaches are analyzed and their limitations in the context of new VR applications are identified. In addition, potential solutions and future research directions will be identified to enable a more appropriate assessment of the user experience, effectiveness and impact of VR technologies. The aim is to promote an integrated evaluation framework that can keep pace with technological innovations and thus support the sustainable use of VR in different application areas.

1 INTRODUCTION

With the promise of offering experiences that immerse users in simulated worlds, virtual reality (VR) has created applications that were previously unimaginable. The potential applications of VR are diverse and range from education and healthcare to architecture and entertainment. This opens up new horizons for innovation and interaction. The technical development of VR is taking place at a rapid pace. Advances in graphics, sensors and computing have led to significant improvements in the performance and accessibility of VR systems. Current VR headsets offer very good image quality and interactive options that were considered unattainable just a few years ago. These technological innovations have promoted the acceptance and integration of VR in various industries and defined new standards for immersive experiences.

Effective evaluation methods are essential to optimize the user experience, evaluate effectiveness in specific applications and identify potential risks or undesirable effects. Although the technical possibilities are developing at a rapid pace, the evaluation of VR experiences is still challenging. The majority of existing evaluation methods were originally designed for the evaluation of traditional media formats and therefore suggests that they are unable to comprehensively capture the specific characteristics and potential of VR. The need to develop adequate evaluation methods that take into account the complexity and multilayered nature of VR is becoming increasingly apparent.

The praxwerk project[1] of Anhalt University of Applied Sciences focuses on the digitalization of higher education, with the integration of VR into teaching and learning processes as a central research focus. As part of the project, the virtual learning application VR-BioTech-House® has been further developed to present complex biotechnological processes in a practical and interactive manner. This bridges theoretical knowledge with practical application, thereby enhancing the achievement of learning objectives. However, a key challenge is to systematically evaluate the impact of VR, particularly in terms of its effectiveness and its contribution to specific learning objectives. Evidence-based analysis is essential to fully unlock the potential of VR in higher education and guide future developments.

The aim of this paper is to provide both re-

[1]https://hs-anhalt.de/praxwerk

searchers and practitioners working on similar applications such as the BioTech-House, with an overview of the current state of evaluation in VR research and to provide impetus for future developments.

2 LITERATURE

In a meta analysis 14 different articles were reviewed that dealt with the evaluation of mixed reality (xR) and especially VR applications in a wide variety of areas with a focus on applications with a purpose of education.

Springer, Google Scolar and the Discovery Service of the Library of Anhalt University of Applied Sciences were used as search platforms for the analysis with variations of the search terms: “evaluation XR application”, “evaluation VR application” and “quality assessment VR application”. The articles found through this method were checked to see whether they dealt sufficiently with the explicit phrasing of different categories relevant to the evaluation. First and foremost, the research was based on articles that questioned the evaluation itself, but in order to obtain a larger number of studies, additional articles were included that evaluated specific applications as long as they explicitly mentioned the criteria used for evaluation.

Consequently, seven overarching categories were identified from the examined articles, which will be discussed in more detail below. These are:

- user experience;
- user assistance;
- game (or application) mechanics;
- quality factors;
- graphics;
- VRise (VR sickness or VR induced symptoms and effects);
- motivation.

Each of these categories contains smaller factors, although the number of these varies depending on the category and not each of them has been mentioned in every source. However, it should be noted that it is difficult to clearly delineate these categories, as the quality of the graphics, for instance, can have a direct influence on the user experience.

Factors that can be categorized as user experience were mentioned in a total of ten articles [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] whereby the extent to which this category was subdivided into smaller factors varies greatly depending on the source material. For example, user experience is generally listed twice as a factor [5, 6]. Overall, the following factors were found, that have an influence on the factor user experience: an adequate level of immersion [1], pleasant VR experience [1], high quality sounds [1]/ audio [2], suitable hardware (head mounted display (HMD) and computer) [1], information gift [2] (the way information is presented), language [2], ease to use [2, 3, 7], similarity to reality [3] and factors that depend on the user themselves: cognitive skills - e.g., visuospatial abilities etc., [4], participants attitude [4], levels of trust/acceptance of VR/MR tools [4, 9], motivation during usage [4, 7, 9], previous experience with VR (XR) [4, 5, 8] and levels of presence and engagement [4, 6, 10].

The greatest importance is attributed to ease of use, the user factor levels of presence and engagement, previous experience and motivation during usage, which are each mentioned in three different sources, as well as audio quality and the user factor levels of trust/acceptance of VR/MR tools with two mentions each. All other factors were only found in one source each.

The second identified category is user assistance. Factors belonging to this category were mentioned in three sources [1, 2, 6] and describe in which way the user is introduced to or guided through the program. These include the following factors: digestible tutorials [1], helpful tutorials [1], adequate duration of tutorials [1], (helpful in-game) instructions [1, 2], helpful in game prompts [1], goal and task design [2], feedback [2], information transmission [2] and didactics of in-game instructions [6].

As a result, (helpful in-game) instructions are assigned the greatest importance. This factor is mentioned directly in two of the three sources, although the factor of the last source relevant here, didactics of in-game instructions, also intersects this factor.

The next category, game mechanics, includes factors from six sources that determine movement and interaction in the application. These are the following: a suitable navigation system (e.g. teleportation) [1] / locomotion [2] / navigation operation [11], availability of physical movement [1], naturalistic picking/ placing of items [1], naturalistic use of items [1], naturalistic 2-handed interaction [1], interaction elements [2], user interface [2, 7], operation of interaction functionalities [2], design/functionalities of interaction functionalities [2], relevance assessment of interaction functionalities [2], menu control [2], control systems [2], room design [2], tracking [2], haptic interaction [3] and interaction regulation [12].

The factor most frequently acknowledged is the way in which the navigation system is implemented, having been mentioned in three different sources. The factors user interface and haptic interface were also referred to repeatedly with two instances each.

The quality factors of a VR application largely include factors that are difficult to measure in units and are often rather subjective. The basic factors referred to are dependability, validity, objectivity, reliability, effectivity, knowledge, utility and VR technology, all of which are discussed by Napetschnig [2]. Utility and knowledge are also dealt with by Tsivitanidou [6] and the latter by Antonopoulos[5]. Other factors are system performance [3, 8], system reliability [3] and technical aspects and tools features - e.g., effect of designed features, expected and experienced system functioning [2, 6].

The greatest importance is therefore assigned to the knowledge factor, i.e. the quality in which the application conveys knowledge to the user, which was discussed by Chardonnet [11]. This is followed in order of importance by utility, VR technology, system performance and the technical aspects and tools features, each of which are mentioned by Yoon [13].

The quality of the graphical representation [2] includes all aspects of an application that relate to the graphical representation. It includes the sub-aspects of software, degree of accuracy, structure, fineness/detail, concept, design, texture, value, claim and 3D character or character appearance. These aspects can be found in Napetschnig [2], whereby degree of accuracy can also be assigned to Kojić [7] and 3D character or character appearance, design and texture to Wienrich [8]. Finally, the aspects of the 2D model display and the 3D display, which are assigned to Ye [3], should also be mentioned.

The greatest importance was placed on the aspects of degree of accuracy [2, 7], design [2, 8], texture [2, 8] and 3D character/avatar appearance [2, 8]. In general, it can be seen that realism and accuracy of the representation play a major role among users. It is important to mention that the aspects of graphical representation dealt with were not always clearly definable, particularly in the case of Wienrich [8], as some of the evaluating participants made inaccurate descriptions of their experience with the simulation. An attempt was made to classify this information as well as possible in the given categories.

VRise (VR sickness or VR induced symptoms and effects) is considered here under the aspect of negative physiological reactions to VR applications. In the course of this, all of the following aspects of Kourtesis [1] are dealt with, as well as absence or insignificant presence of nausea, absence or insignificant presence of disorientation and cybersickness by Wang [14]. The latter is the only part of this list to be found in Chardonnet [11] and the absence or insignificant presence of dizziness in Yoon [13]. The remaining parameters are absence or insignificant presence of fatigue and absence or insignificant presence of instability.

The greatest importance can obviously be attributed to cybersickness, which includes many of the other sub-aspects. It is therefore mentioned directly in three sources, while the absence or insignificant presence of nausea [1, 14], absence or insignificant presence of disorientation [1, 14] and absence or insignificant presence of dizziness [1, 13] are only mentioned in two sources each.

Finally, motivation in use will be discussed. It describes the aspects of a VR application that motivate its use. Napetschnig [2], Borsci [4], Wienrich [8] and Kojić [7] address these aspects. Napetschnig [2] also refers to the subcategories logging, configuration, adaptability/customizability for different needs and interests, which is also mentioned by Kojić [7], practicability [8, 9], user-friendliness [7, 8] and usability [3, 8]. Learnability is only mentioned by Ye [3].

Practicability [2, 8, 9], user-friendliness [2, 7, 8] and usability [2, 3, 8] were mentioned the most of these factors, apart from the factor of motivation in use in general, which means that they are assigned the greatest importance.

It is obvious that the aspects for evaluating VR applications are not only numerous, but also cover a spectrum of sub-areas. Not only are issues as the visual and acoustic representation of the immersive world important, but also the effect of this world on the user, which in turn can be broken down into different aspects such as physiological reactions, the learning effect or sensory impressions. However, the question arises as to whether this selection of criteria is sufficient for the rapidly advancing state of VR technology or whether the questionnaires are already outdated compared to the demand for the evaluation of VR applications.

3 EVALUATION METHOD

The question posed will be answered using a qualitative research approach. This is particularly suitable for the present study, as it has a small sample size and an explorative character. The latter aims to gain deeper insights into user perceptions.

The data was collected as part of an experiment conducted during a science camp with pupils at the age of 13-15 years. Two groups of twelve technology and science enthusiasts took part in the study. These participants represent the target group that will increasingly come into contact with VR technologies

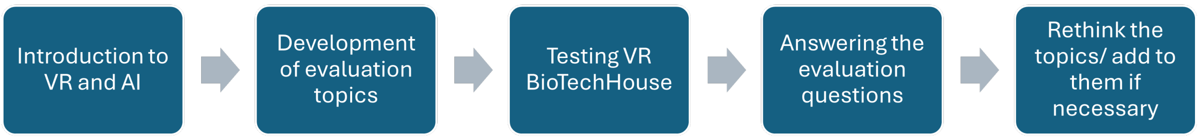


Figure 1: Procedure of the survey.

in the future. The application VR-BioTech-House was tested by the young people in an educational application with the goal of developing categories for evaluating the application. The implementation was part of a workshop on the topic of xR and AI. Firstly the students were introduced to xR and AI, then they were tasked with the development of evaluation criteria with the help of ChatGPT in groups of two. Only after that they had the oportunity to test the VR application BioTech-House which is presented below. Finally after the exploration phase, the criteria of all groups were discussed and if necessary amended or more added as can be seen in Figure 1.

The VR-BioTech-House application is an educational tool in the field of biotechnology that allows learners to explore everyday applications of biotechnology within a virtual simulation of an ordinary house. Participants can engage in hands-on activities to produce biotechnology products, such as yogurt [15], red wine, and yeast for bread-making [16]. In this experiment, the yogurt-making scenario was selected as the test case.

The central instruments of this study were the VR scenario, which served as the object of the experiment, and the online notepad, which the student groups used to record their criteria. The evaluation criteria thus found are the data of this study. The duration of the testing phase of the VR-application for each person was 25 minutes. During that, they could explore the application freely and did not get any outside instructions unless they asked for it. The goal was a successful completion of the scenario.

4 RESULTS

The results of the qualitative analysis revealed diverse user needs and expectations for VR applications, focusing on hardware, software, and other contributing factors. These insights emphasize the importance of aligning the design and evaluation of VR systems with these requirements to enhance user satisfaction and functionality.

The first thematic block focuses on hardware, as illustrated in Figure 2. An important point here is robustness, which refers to the hardware's resistance to physical stress. The sensor accuracy or tracking is also relevant concerning the precision with which the sensors detect and reproduce movements. Another aspect to consider is sustainability and fair trade, which aim to ensure environmentally friendly manufacturing and fair production conditions. The connections, in particular connectivity, are also relevant. The latter includes the type and number of connections as well as possible wireless connections. The field of vision provides information about the area that users can see through the hardware. The term comfort covers both the ergonomics and the wearing comfort of the devices. Another topic is the controllers. In this context, functionality is of central importance, whereby the operation and input options of the hardware are of crucial relevance. Finally, the sensory impressions are also relevant, which are influenced by audio and graphics, especially color, and determine the sensory perception of the user.

In the second thematic block, the results on the topic of software are sumarized and can be seen in Figure 3. A key aspect is user-friendliness, which refers to the ease of use and intuitive operability of the software. The software's ability to support multiple users simultaneously is addressed in the multiplayer category. The ability to customize the setting and content to meet the specific requirements and needs of the user is also of particular relevance. Adaptation to the respective learning level and consideration of different learning types are particularly important here. Another aspect to consider in this context is the performance of the software. It is of interest how smoothly it runs and how stable it is in use. The integration of visual or auditory effects and surprises can have a positive influence on the user experience. The ability of users to actively interact with the software is another important aspect that is subsumed under the term interactivity. The quality of the content offered by the software is recorded under the Content category. In this context, the fun of using the software and the topicality of the content, which is ensured by regular updates, are also emphasized. Furthermore, the focus on a specific target group is relevant, as adapting the software to the needs of this user group is essential.

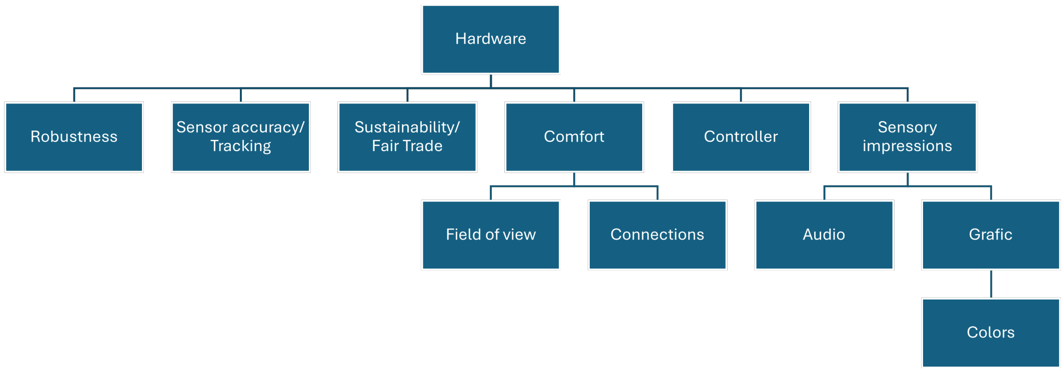


Figure 2: Results with a focus on hardware.

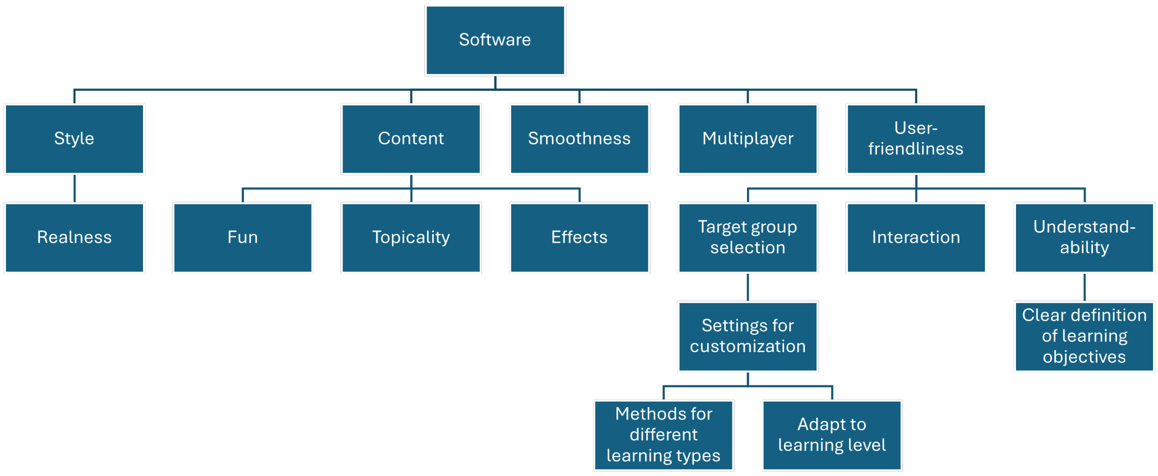


Figure 3: Results with a focus on software.

The aspect of comprehensibility of the application includes the adequate explanation of the functions and content of the software. In this context, the precise definition of the learning objectives is of crucial importance. Finally, the style of the software is also relevant, as is the extent to which the application is designed realistically.

In the last block, all topics that relate to neither hardware nor software are assigned. These are summarized in Figure 4. In the context of the use of VR, the aspect of health is also relevant, whereby the effects of the use of VR on health must be examined. It is also necessary to determine how efficient the use is, in particular the suitability of the experience in the context of learning and its potential to promote long-term engagement. Another aspect is the feeling or gaming experience evoked by the app, as well as the difficulty of use and integration into everyday teaching. The next aspect discussed was accessibility, with a particular focus on adjustable language options.

After testing, most participants did not add any new categories. However, a few added topics such as task and learning management as well as language and room boundaries. One group wrote: “We would definitely use other criteria now, e.g. content, audio”.

Compared with the findings of the literature review, it is notable that most criteria found during the review were mentioned by the different student groups in one way or another. The depths of the description may have varied, but the general implication was present. Notable for these similarities are for example user-friendliness as found in the evaluation with the criteria ease of use and intuitive operability of the software, with the corresponding criteria that appeared during the literature review in the categories user-experience (ease of use) and Motivation in use (user-friendliness). Notable differences between the literature review and the BioTech-House evaluation are that different areas were emphasized. During the evaluation, a clear category of hardware was formulated while during the literature review the criteria that are influenced by hardware are sub-aspects of bigger categories. Furthermore, the evaluation put more emphasis on a game-characteristic (Feel-

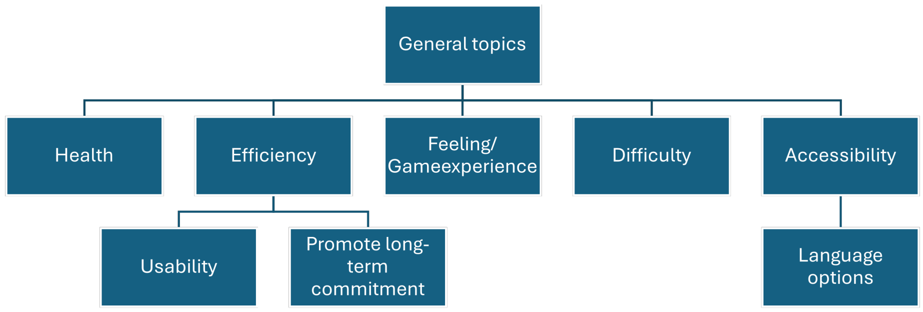


Figure 4: Results with a focus on general topics.

ing/Gameexperience, Fun), while the closest criteria of the literature review is motivation in use. The only two categories unique to the evaluation are the criteria multi-player and sustainability.

The findings provide key insights into the future development of BioTech-House. Hardware considerations, such as robustness, comfort, and precise tracking, are essential for creating a reliable and immersive learning experience. On the software side, user-friendly interfaces, adaptive learning features, and high-quality content are crucial to making virtual experiments both engaging and pedagogically effective. The participants also stressed the importance of interactivity, real-time feedback, and accessibility to ensure inclusivity and usability in diverse educational contexts.

In general, these findings offer practical guidance for the evaluation and improvement of BioTech-House. By addressing user needs in both hardware and software design, the application can optimize its effectiveness as a virtual learning tool. Furthermore, the results contribute to a broader understanding of how to develop VR systems that are both engaging and educationally impactful.

5 DISCUSSION

This analysis provides a comprehensive summary of user needs and expectations regarding the BioTech-House educational application, offering clear directions to optimize hardware, software, and overall user experience. In particular, the discussion of hardware emphasizes the importance of comfort, graphical and audio presentation, and robustness. These attributes are indispensable for tools used in educational settings, especially by students, where durability and reliability under varying conditions are critical.

Insights on software usability provide actionable guidance for developers working to enhance the application. Key factors include user-friendliness, support for multiplayer interactions, and the ability to customize educational content. Additionally, efforts to improve content engagement and maintain high-quality learning materials are essential to meet the needs of university students and other learners. Addressing aspects such as tailored content, adaptive difficulty levels, and seamless interactivity is crucial to ensure the software's educational effectiveness.

The results of this qualitative analysis align closely with the practical requirements of the BioTech-House project. These findings offer a roadmap for refining the user experience and strategically integrating features desired by users, thereby supporting the next stages of development. By addressing these considerations, BioTech-House can better fulfill its goal of delivering an engaging and effective virtual learning tool.

6 CONCLUSIONS

This study investigates the relevance of existing evaluation categories of VR experiences. It highlights the need for diverse evaluation parameters that encompass immersion, spatial perception, and psychological reactions. The literature review suggests that existing methods take these aspects into account, but the results indicate that this study may be limited by biases due to AI systems or participant influence. Furthermore, due to the specific background of the participating students as people with an interest in science and technology it is possible that they did not think of difficulties people with a lesser interest would have or put a greater emphasis on the abilities of hardware.

To improve the evaluation, the development of an integrative framework combining methods from different disciplines such as psychology, computer science and design is proposed. Interdisciplinary approaches seem particularly suitable for comprehensively mapping the complex dimensions of VR experiences. Future research directions could investigate how machine learning can be used to personalize VR experiences and adapt the evaluation in real time, they could explore if a greater focus on the impact of the used hardware in the evaluation could be helpful.

A sustainable and adaptable development of these evaluation strategies is seen as crucial to realize the full potential of VR technologies. Close collaboration among survey designers, developers, and end-users is essential to bridge the gap between technological advancements and user feedback.

ACKNOWLEDGMENTS

We thank the German "Stiftung Innovation in der Hochschullehre" for partially funding this work.

REFERENCES

[1] P. Kourtesis, S. Collina, L. A. A. Doumas, and S. E. MacPherson, "Validation of the virtual reality neuroscience questionnaire: Maximum duration of immersive virtual reality sessions without the presence of pertinent adverse symptomatology," Frontiers in Human Neuroscience, vol. 13, no. 417, pp. 1-15, Nov 2019, [Online]. Available: https://doi.org/10.3389/fnhum.2019.00417.

[2] A. Napetschnig, K. Brixius, and W. Deiters, "Development of a core set of quality criteria for virtual reality applications designed for older adults: Multistep qualitative study," Interactive Journal of Medical Research, vol. 12, pp. 1-15, Sep 2023, [Online]. Available: https://doi.org/10.2196/45433.

[3] J. Ye, S. Badiyani, V. Raja, and T. Schlegel, "Applications of Virtual Reality in Product Design Evaluation," in Human-Computer Interaction. HCI Applications and Services, J. A. Jacko, Ed., pp. 1190-1199, 2007, [Online]. Available: https://doi.org/10.1007/978-3-540-73111-5_130.

[4] S. Borsci, G. Lawson, and S. Broome, "Empirical evidence, evaluation criteria and challenges for the effectiveness of virtual and mixed reality tools for training operators of car service maintenance," Computers in Industry, vol. 67, pp. 17-26, Feb 2015, [Online]. Available: https://doi.org/10.1016/j.compind.2014.12.002.

[5] P. Antonopoulos, E. Fokides, and G. Koutromanos, "Understanding learning and learning experience in immersive virtual reality," Tech Know Learn, pp. 1-20, Aug 2024, [Online]. Available: https://doi.org/10.1007/s10758-024-09764-z.

[6] O. E. Tsivitanidou, Y. Georgiou, and A. Ioannou, "A learning experience in inquiry-based physics with immersive virtual reality: Student perceptions and an interaction effect between conceptual gains and attitudinal profiles," Journal of Science Education and Technology, vol. 30, no. 6, pp. 841-861, Dec 2021, [Online]. Available: https://doi.org/10.1007/s10956-021-09924-1.

[7] T. Kojic, D. Ali, R. Greinacher, S. Möller, and J.-N. Voigt-Antons, "User experience of reading in virtual reality - finding values for text distance, size and contrast," in 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), pp. 1-6, 2020, [Online]. Available: https://doi.org/10.1109/QoMEX48832.2020.9123091 .

[8] C. Wienrich, N. Döllinger, S. Kock, K. Schindler, and O. Traupe, "Assessing User Experience in Virtual Reality - A Comparison of Different Measurements," in Design, User Experience, and Usability: Theory and Practice, A. Marcus and W. Wang, Eds., pp. 573-589, 2018, [Online]. Available: https://doi.org/10.1007/978-3-319-91797-9_41.

[9] N. Pares and R. Pares, "Towards a model for a virtual reality experience: The virtual subjectiveness," Presence: Teleoperators and Virtual Environments, vol. 15, no. 5, pp. 524-538, Oct 2006, [Online]. Available: https://doi.org/10.1162/pres.15.5.524.

[10] P. Grabarczyk and M. Pokropski, "Perception of affordances and experience of presence in virtual reality," AVANT. The Journal of the Philosophical-Interdisciplinary Vanguard, vol. 7, no. 2, pp. 25-44, Dec 2016, [Online]. Available: https://doi.org/10.26913/70202016.0112.0002.

[11] M. A. Mirzaei, F. Merienne, and J.-R. Chardonnet, "Influence of navigation parameters on cybersickness in virtual reality," Virtual Reality, vol. 25, no. 3, pp. 565-574, Sep 2021, [Online]. Available: https://doi.org/10.1007/s10055-020-00474-2.

[12] A. Halbig and M. E. Latoschik, "A systematic review of physiological measurements, factors, methods, and applications in virtual reality," Frontiers in Virtual Reality, vol. 2, pp. 1-20, Jul 2021, [Online]. Available: https://doi.org/10.3389/frvir.2021.694567.

[13] H. J. Yoon, J. Kim, S. W. Park, and H. Heo, "Influence of virtual reality on visual parameters: immersive versus non-immersive mode," BMC Ophthalmology, vol. 20, no. 1, pp. 1-10, Dec 2020, [Online]. Available: https://doi.org/10.1186/s12886-020-01471-4.

[14] G. Wang and A. Suh, "User adaptation to cybersickness in virtual reality: A qualitative study," in Proceedings of the 27th European Conference on Information Systems (ECIS), pp. 1-12, Jun 2019, [Online]. Available: https://aisel.aisnet.org/ecis2019_rp/191.

[15] R. Huang, J. Rödigh, A. Berger, and J. Tümler, "Ein virtuelles biotech-haus für das lehren in der biotechnologie," in Proceedings of DELFI Workshops 2022, pp. 77-85, 2022, [Online]. Available: https://doi.org/10.18420/delfi2022-ws-17.

[16] J. Tümler, M. Alptekin, K. Krylova, R. Huang, and J. Rödigh, "Weiterentwicklung des vr-biotech-house: Wein- und hefeherstellung," in Proceedings of DELFI Workshops 2024, pp. 1-10, 2024, [Online]. Available: https://doi.org/10.18420/delfi2024-ws-04.

Challenges and Strategies of Remote Foreign Language Learning for Students with Autism Spectrum Disorder

Iryna Haidai, Liudmyla Suvorova, Nataliia Pankovyk, Yuliia Herasymchuk and Olha Zadoienko
Department Pedagogical Technologies and Language Training, Zhytomyr Polytechnic State University, Chudnivska Str. 103, 10005 Zhytomyr, Ukraine
org_gio@ztu.edu.ua, ktpl_slk@ztu.edu.ua, kptmp_pnm@ztu.edu.ua, kptmp_gyup@ztu.edu.ua, kptmp_zoi@ztu.edu.ua

Keywords: Foreign Language Competence, Distance Learning, Inclusive Environment, Gamification, Virtual Socialization, Communicative Skills.

Abstract: People with cognitive impairments constantly face a number of modern challenges, including a pandemic, martial law, and instability of psychoemotional functions, which already complicates the education of students with autism. Today, inclusive education in Ukraine is only beginning to develop effectively, but global crises significantly limit full-fledged education. The difficulty lies in adaptation - a kind of «adjustment» of educational materials for people with disabilities. A gadget, which was previously perceived as a means of entertainment, a tool for obtaining pleasure or «reward», must now become an educational center, a source of knowledge accumulation. Therefore, the goal was to study how the acquired digital skills and abilities can positively affect the education of students on various distance platforms. The results of the observation showed that the advantage of students with autism is their unique ability – to be able to work perfectly with a gadget. The study offers practical and proven effective ways of teaching students with ASD. The object of study was distance learning tools tested on students of the transitional educational level – grades 4-6, who studied in an inclusive form of education. The results showed that the organization of education of students with autism in a distance format has a number of both disadvantages and advantages. Among the disadvantages: unhealthy virtual socialization; inability to adapt educational materials; blurred social roles; sensory overload; technical problems with video and connection; distraction by clips; limited opportunities for forming strategies for psycho-emotional communication. On the other hand, the studies showed that individual topics were quite effectively mastered using online platforms. These are a case-block approach to presenting material using distance technologies, game practices, and programs that help in learning a foreign language. Such a well-developed experience will become a practical find in developing inclusion at the level of professional pre-higher and higher education in Ukraine in the future.

1 INTRODUCTION

In August 2024, as part of the organization of the "Parents and Early Intervention" program dedicated to the problem of autism, a rather active three-day discussion on the topic of children's disability took place in the context of the deployment of military operations in Ukraine. This event was held with the support of Daria Herasymchuk, the Commissioner of the President of Ukraine for Children's Rights and Child Rehabilitation, and the UN Office for the Protection of Human Rights [1]. The crisis situations in which Ukraine found itself – first the pandemic, and then the war – demanded a quick search for optimal ways to teach; therefore, the distance format became the environment that needed to be studied, deeply mastered methodically, and at the same time to look for the most effective mechanisms that would work on the result [2].

On April 26, 2022, the Cabinet of Ministers of Ukraine adopted and approved a number of changes to the organization of inclusive education in Ukraine in conditions of war (On amendments to the procedures approved by resolutions of the Cabinet of Ministers of Ukraine dated April 10, 2019 No. 530 and dated September 15, 2021 No. 957 CMU Resolution No. 483 of April 26, 2022) [3].

An important nuance was that schools and other educational institutions began to function as a shelter for families from regions of active hostilities, and desk-based learning was replaced by an online learning format. The above-mentioned document (Decree No. 483) also outlined a number of tips and practical recommendations for organizing the education of children with special educational needs under martial law. Some of them were developed two years before the start of a full-scale invasion, that is, during the pandemic.

2 PROBLEM STATEMENT

It is important to note the obstacles that inhibit acquiring knowledge in emergency situations. First of all, there is the problem of access to technology because effective learning requires a stable Internet connection and appropriate digital devices, which does not always seem possible in the conditions of systematic rocket attacks, lack of light, and network. Secondly, the distance format is a completely undeveloped form of learning organization for a student with disabilities because the home environment is less structured than a class or an auditorium.

Virtual education of children with autism involves, first of all, a radical revision of social roles: parents, at a certain point, have to play the role of a teacher, accompanying a student with ASD during an online lesson. As research shows, this reorganization of social functions leads to emotional exhaustion, a state of helplessness in parents who already find themselves alone with the problem of autism [4]. In addition, limited support for assistants and teachers can lead to a decrease in the effectiveness of the educational process, a decrease in motivation, and a deterioration in the quality of the teaching process. The organization of psychological support for children and their families is also a necessary step in creating favorable learning conditions in crisis conditions [5]. Therefore, it is essential to review the strategy of organizing the education of children with autism through the prism of teamwork, taking into account online cooperation.

Distance learning, which has become a new field in the knowledge acquisition system, has not been worked out sufficiently regarding inclusion. Due to the variability of gnosis (even autism has a relatively wide panorama of syndromes and their variations) and the different complexity of the degree of damage to cognitive abilities, each student had to find his own algorithm for organizing learning in crisis conditions.

A separate problem turned out to be the fact that not all parents have knowledge of a foreign language (both from the standpoint of language knowledge and teaching methods), which, again, complicated the learning process. Parents change roles with the teacher, thus becoming an important "channel of communication" for students with ASD, but it is extremely difficult for them to independently organize the child's education, where he could acquire knowledge without additional support [5].

The study of the problems of teaching students with autism is currently carried out at the level of scientific conferences in the context of scientific and methodological developments, forums, and collective studies. Even during the pandemic, a number of countries faced the need to find ways to preserve the learning process [6].

In choosing effective ways of working, it is necessary to consider the clinical features of the diagnosis first. The methodological component should be built taking into account the behavioral mechanisms of students with ASD, with an appeal to the "education system" in its close relationship with "health care, in order to ensure proper rights and services for children” [4]. The methodological base for students with disabilities is currently not properly developed in Ukraine, although the situation is improving yearly – and at least electronic versions of textbooks can be found on the Modernization of Education Content website [7].

While analyzing textbooks for learning the English language, it was found that the content of the proposed textbooks is somewhat complicated, especially for high school. The junior school has sufficient methodological support but shows a limited variety of educational materials. This can have a negative impact on students' motivation and language acquisition, as the use of monotonous textbooks does not contribute to the development of their cognitive skills. In turn, in high school, particularly for students of grades 9, 10, and 11, there is a significant shortage of methodological materials, and their content often does not consider students' cognitive capabilities. The same problem can be observed in higher education institutions, emphasizing the need to revise and improve educational materials and approaches to teaching English in senior and higher educational institutions. In view of this, there is an urgent need to develop more diverse and adapted textbooks that meet the modern requirements and needs of students.

The lack of knowledge and experience in the organization of distance-inclusive education leads to people with limited cognitive functions finding themselves "overboard" of educational activities, adversely affecting their psycho-emotional state [8].

Order of the Ministry of Education and Science of Ukraine, which is a separate component of the document "On approval of the Regulation on Distance Learning" dated 04/25/2013. No. 466, registered with the Ministry of Justice of Ukraine on April 30, 2013. under No. 703/23235, it is clearly defined that distance learning is characterized by taking into account the mechanism of individualization regarding the acquisition of knowledge, abilities, skills, and methods of cognitive activity with the organization of a specialized environment that functions on the basis of modern psychological-pedagogical and information-communication technologies [9].

Distance learning in the languages of the crisis sets the tone for the principle of inequality in education because students with autism found themselves without adequate methodical support [6]. Parents and teachers faced the challenge of a difficult situation: before a number of questions – how to teach, how to choose methodological tools for the class, which platforms can be effective, and whether the family has access to them. Virtual education does not involve selectivity in explanation or individualization: the teacher cannot physically approach a student with limitations and additionally explain to him [10]. The teacher explains in the same way for everyone, so a student with ASD needs to optimize the learning process so that the learning environment meets all the requirements of "justice in education" [11]. Psychological help in its usual format has changed to the so-called "virtual, remote teletherapy" [6]. Therefore, the direct and indirect interaction between the teacher and the pupil with ASD is lost [12]. The gradual mastery of digital tools in teaching still gave some positives. The mechanism of virtual organization of training contributed to the variability of the work, from independent mastering of the material to team interaction with the teacher in the format of an online meeting. It is about synchronous and asynchronous work options. Asynchronous mastering of the material makes the educational platform available anytime to discuss "learning system management" [2]. The distance format is characterized by relative accessibility to materials and asynchrony in mastering them. However, the use of digitalization in an inclusive educational environment is a significant challenge. The main reason is the lack of motivation in students with limited cognitive abilities, for whom a gadget from a game tool, entertainment, or sensory object of relief should turn into a learning environment. Such rethinking is quite difficult to implement when working with children with an autism spectrum disorder.

Barriers and limitations in a social sense, adaptation to new learning conditions, and spontaneous nature – are other aspects of the difficulties that arise during the acquisition of knowledge of children with autism [13], [14].

3 METHOD

The study of the mechanism of remote organization of learning for students with autism, particularly their acquisition of a foreign language, is qualitative. The collection of diagnostic and statistical data was carried out by means of an anonymous survey, implemented through the agreement with the city administration of the Department of Education, Zhytomyr (Ukraine). A questionnaire was previously created using a Google Form. Creating a Google Form for collecting relevant data facilitated quick feedback from parents and teachers. Here is the structure of such a questionnaire:

1) General information:
   - Full name of the respondent: (text field)
   - Social role (parent, teacher)
   - Student's grade: (text field)
2) Use of technology:
   - Which platforms do you use for learning? (checkboxes: Zoom, Google Meet, Microsoft Teams, etc.)
   - Do you experience difficulties with using technology? (yes/no)
3) Teaching methodology:
   - How do you assess the effectiveness of distance learning for the student? (scale: 1-extremely unsatisfactory to 5-excellent)
   - Which learning methods proved to be the most effective? (text field)
4) Student's needs:
   - What special needs does the student have during learning? (text field)
   - What difficulties arise during distance learning? (text field)
5) Support and resources:
   - Do you receive sufficient support from teachers and employees of the resource-inclusive center? (yes/no)

- What additional resources or support do you consider necessary? (text field)

6) Feedback:
   - Additional comments or suggestions: (text field)
   - Completion of the form
   - Thank you for your participation!

This online data collection tool allowed us to quickly engage a wide range of respondents: 180 parents (only one parent could complete the questionnaire) who have children with special educational needs, as well as 227 teachers. It is worth noting that the majority of parents were those whose children were diagnosed with ASD, communication behavior and socialization disorders, and mental development disorders with elements of autism (see Table 1).

Table 1: Individual psychological characteristics of students in quantitative terms.

Individual characteristics of the student	Quantitative assessment
Diagnosed with early childhood autism, established disability	18
Mental development disorders with elements of ASD	29
Behavioral, communicative, and social problems	86
Diagnosed with autism spectrum disorder without disability status	38
The child has social-communicative disorders not diagnosed by specialists but noted by parents and teachers	9

The age scale of these students covered grades 4-6. It should be clarified that the survey also included parents whose children study in special classes at general education schools (for example, such classes are organized at Lyceum No. 14 in Zhytomyr). Among the interviewed teachers and parents of students with SEN -- from lyceums No. 1, 4, 5, 7, 8, 10, 12, 15, 17, 19, 20, 21, 22, 23, 24, 26, 28, 30, 34, 35. The survey was conducted in 2020, and the implementation of effective methods and practices was checked during 2020-2023. The results are presented below in the main part of this publication.

The research took place in several stages (see Table 2).

A motivational rating scale was used to measure the state of success in learning a foreign language, taking into account the peculiarities of the nature of autism, which is a poorly researched socio-psychological problem. Mechanisms of both short-term and long-term development prospects were also taken into account.

When modeling the motivational rating scale, the following components were considered (see Table 3).

Table 2: Stages of research.

Stage	Description
Preparatory and evaluation	Development of a questionnaire for conducting an anonymous survey questionnaire
Diagnostic and prognostic	Study of survey results with determination of possible risks and achievements in learning a foreign language
Effective and practical	Methodical development and search for remote forms, methods, techniques, and practices in foreign language learning by students with ASD with subsequent adaptation/modification of educational materials
Problem-evaluative	Outline and development of evaluation criteria for students with autism who studied according to the selected methods during the research stage
Discussion and problem	Determination of problematic issues, difficulties, and formulation of educational, didactic, and scientific perspectives.

The analysis of the data provided for the determination of a range of urgent and priority problems guided the selection of effective, efficient methods, techniques, and platforms, with the help of which learning a foreign language would be as successful as possible. An overview of the most effective methodologies and a description of their effectiveness are demonstrated below.

4 RESULTS

4.1 Survey Results

Questions for teachers related to difficulties in working with children with SEN, support from inclusive resource centers. In addition, the issues of the principles of drawing up a program for the individual development of a child with SEN, finding means to increase the effectiveness of learning, and difficulties that arise during distance learning of students with cognitive limitations were raised.

Table 3: Percentage ratio of motivational assessment mechanisms in working with students with autism spectrum.

Mechanism	Content	Effectiveness in practice (%)
Individual approach	Consideration of individual characteristics and abilities of the child. Selection of materials according to the capabilities and interests of the child	100%
Clear goals	Establishing clear and achievable learning goals. Step-by-step breakdown. It is important that goals are specific and measurable	95%
Positive reinforcement	Use of positive reinforcing stimuli such as praise or reward for achievements. This encourages the child to further efforts. However, some children reacted aggressively to incentive stimuli (10%)	90%
Visual aids	Use of cards, pictures, or graphs that can help the child understand what is expected of them. However, some children did not understand the meaning of visual cues	85%
Regular monitoring	Conducting regular assessments and tracking progress. This helps to timely adjust the curriculum according to the child's needs (keeping an observation diary by the assistant parents)	100%
Interaction with parents and specialists	Coordinated work with parents and other specialists to ensure consistency of approaches at home and at school; however, 30% of parents did not always follow the necessary recommendations and did not interact in a team with teachers and specialists	70%

According to the interviewed teachers working in inclusive classes, the problem of the difficulties they faced was the most relevant. First of all, it is about instability in the behavioral reactions of students during the lesson (76.4%), the need to allocate additional time in the lesson to work with SEN (10.2%), as well as the problems of preparing and developing special tasks taking into account the individual development program (11.8%) – the results are presented in Figure 1.

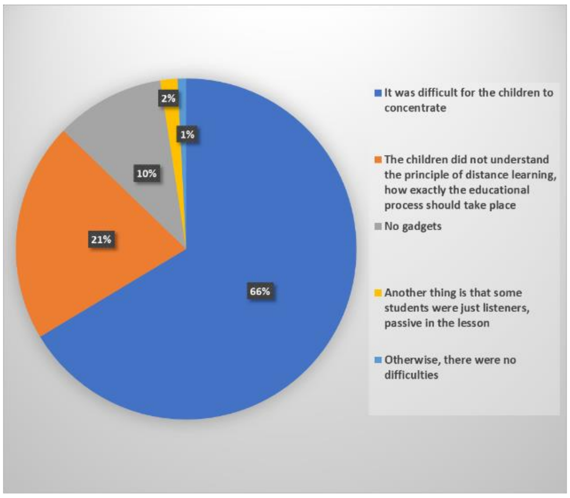


Figure 1: Difficulties in working with children with SEN during distance learning.

As we can see from the chart, the biggest difficulty during distance learning was that it was difficult for children to concentrate on online lessons - this was answered by 65.4% of respondents. This is due to the inability of children with autism to allocate attention rationally. It was also difficult for them to understand the very principle of distance learning because it was necessary to understand the reevaluation of social roles: the teacher is behind the screen of the gadget, and this is not a game but a full-fledged lesson, in addition, and dad or mom now play the role of an assistant (20.5%). It happened that students did not have individual gadgets or had to share them with their parents, brothers, or sisters (10.2%). There were isolated answers related to students' activity in the lesson – some students were passive listeners (1.6%). Only 0.8% of respondents had no difficulties.

A survey of parents on the same question about difficulties in distance learning showed the following results, presented in Figure 2.

From the parents' answers, it can be noted that the lack of pedagogical education inhibits the optimization of the process; for them, it is extremely important to organize an educational environment at home. In addition, most parents face the problem of the child not perceiving parents as teachers. This social reorientation of roles caused protests on the part of children and reluctance to cooperate. Distance learning limits the support of the teacher's assistant, the class teacher, and classmates as an environment for successful socialization.

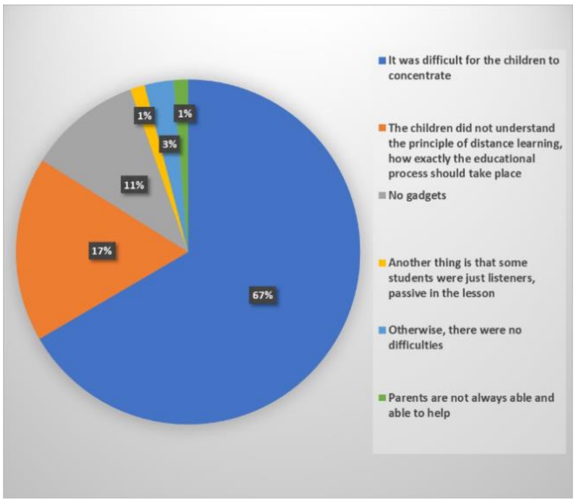


Figure 2: Difficulties in children with SEN during distance learning according to parents.

An educational problem also arose: if earlier, before the introduction of quarantine restrictions, parents did not allow or limit the use of gadgets, then with the introduction of the distance learning format, they would have to give up their beliefs for the sake of children's education.

The presented diagram emphasizes the fact that the most common problem during distance learning for students with ASD is the low level of concentration of attention.

Lack of understanding of the structure and organization of distance learning complicates the learning process for many students. It is worth noting that every child with autism is unique, so every educational plan must be unique. Studying at home with the support of a certain family member eliminates the rules and methods of work prescribed in the individual development program.

The results of surveys of teachers and parents indicate that teaching children with ASD requires a personalized approach, taking into account their strengths and peculiarities. Hence, there is a problem for teachers and parents regarding the correct organization of distance learning for children with ASD. Teachers support children with SEN at school, and parents take on this role at home. Failure to perceive mom or dad as a teacher reduces motivation and distorts the very desire to learn.

However, Dr. Stuart Shanker urges parents to approach challenges with compassion and understanding, not frustration: "Autistic children 'don't bother you,' they are 'difficult' [15]. Dr. also says that success in education is always possible. Temple Grandin: "How we see children with autism shapes how they see themselves" [16].

Teamwork can still be ensured even when teaching children with ASD remotely, provided there is constant dialogue and interaction between all participants in the educational process.

4.2 Adaptation Steps in Distance Learning of a Foreign Language by Students with ASD

Having received the results regarding the existing difficulties, we had the opportunity to develop and offer those formats of learning a foreign language, which, in our opinion, could be as effective as possible. Successful learning outcomes depend on specific teaching methods and support, positive dynamics, student engagement, and self-esteem. This is especially important in distance learning environments where students with autism can feel isolated [6].

Therefore, the methodology includes the assessment of both academic progress and individual psychological factors, students' motivation, a sense of self-efficacy, and the ability to master language material. Therefore, the methodology includes the assessment of both academic progress and consideration of individual psychological factors, student motivation, a sense of self-efficacy, and the ability to master language material.

First of all, it was important to study the methodological base that is currently available and available in school libraries. Not all middle school classes are equipped with such textbooks, although they can be found in electronic access (Textbooks for people with special needs. Institute for the Modernization of the Content of Education) [7].

Some textbooks, especially for 6th-grade students, seem too complicated. Children with autism have different degrees of brain damage and different cognitive abilities. High-functioning autism allows you to master a foreign language more deeply, but students with intellectual disabilities cannot cope with the material that the authors of the textbook offer [16]. Optimizing the educational process for students with ASD is impossible without a mechanism for adapting and modifying educational material [17]. It is not only about the development but also the very mechanism of effective submission. We have developed a system of steps that, in our opinion, will contribute to efficiency in learning a foreign language (Figure 3).

During the teachers' survey, it was found that children were not very active in online lessons. This problem was solved using the mechanisms described below.

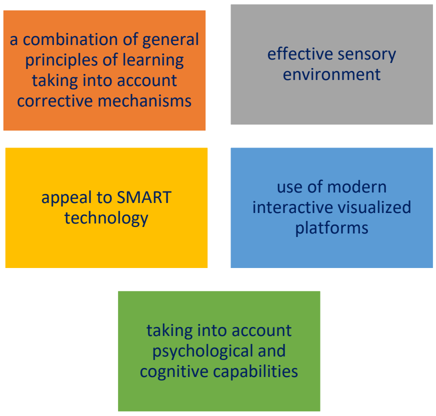


Figure 3: Mechanisms of effective teaching of students with ASD.

It is extremely important to be able to combine general pedagogical principles of learning with corrective pedagogy, applying this symbiosis in remote work. The technology of "small steps" (SMART) makes it possible to break down the task into stages, in portions, with the calculation of the final result. Individual and psychological aspects should be taken into account during the assessment of knowledge and during the development of an individual program for a student with ASD.

At the same time, maintaining an existing focus while searching for effective learning strategies should help avoid overreliance on diagnostic labels.

Adaptation of educational material for children with SEN should be carried out at all stages of work: at the level of selection of teaching methods, at the level of principles of using materials and resources, and at the level of evaluation. Thus, adaptation at the level of teaching methods occurs through the use of visualization, appeal to the capabilities of the Miro virtual board (Figure 4), and the completion of tasks, with a reduction in the amount of material.

According to the survey of teachers and parents, the problem was revealed that it was difficult for children to concentrate. In order to solve this problem, small steps were applied.

Tasks were divided into smaller ones, and thus, children could complete them more efficiently. For better understanding, the teacher can repeat the task and explain problematic points. Also, the teacher can give the student a task using a step-by-step algorithm (SMART principle).

Working with a mixed-ability class during a clothing vocabulary lesson, the teacher utilized an interactive whiteboard Miro to teach "There is/There are" structures. Using a "progressive disclosure" feature, the teacher motivated the class by displaying clothing images and encouraging students to construct correct sentences: This is a dress. /These are trousers.

Figure 4: Case study: Adapted exercise for clothing categorization using "There is/There are".

Adaptation for children with SEN at the assessment level is carried out by providing the student with the opportunity to complete the task in an arbitrary format, providing additional time to complete the task. It is also possible to allow the student to complete a smaller number of tasks on a test or exam, to allow the involvement of another person for help in completing written tasks, and to take notes instead of the child. It is worth noting that the assessment of a child with SEN should gradually be equated to the assessment of normal children – and by the end of school, the threshold for adaptation and modification will gradually decrease, which will contribute to the development of independent learning skills.

Another type of adaptation for students with SEN is adaptation at the level of materials and resources, including the use of materials of different levels of complexity, the use of printed and other materials, resources created by the teacher for students, alternative teaching materials, adaptation devices, and online platforms. For students with SEN, auxiliary alternative cards, diagrams, and line indicators can be used; the text can be marked with different colors, and audio and visual support can be provided.

Often there is a problem that children do not have their own gadgets and are forced to use them together with their parents and siblings. To resolve this issue in Ukraine, parents of children with special needs can apply to the "Lumos" fund – the Ukrainian Network for Children's Rights.

As part of the foundation's work, tablets were issued to children from the People's Liberation Army who needed psychological support due to military operations on the territory of Ukraine [19].

4.3 Digital Tools as a Component of Learning Individualization for Students with ASD

In the context of modern challenges for education caused by the pandemic and martial law in Ukraine, the use of digital tools to ensure individualized learning for students with autism spectrum disorders is of particular importance. Based on our research, we have identified the most effective digital platforms and tools that have been successfully adapted to meet the special educational needs of students with ASD during foreign language learning.

4.3.1 Miro Virtual Board as a Tool for Increasing Academic Resilience

One of the most effective digital tools we have implemented in the educational process is the Miro virtual board. Experience using this platform in teaching English to students with ASD has demonstrated several advantages in the context of learning individualization and creating a favorable socio-emotional environment (Table 4).

The application of Miro for students with ASD was adapted, taking into account the three levels of emotional design described by D. Norman, according to which a product should be not only functional and effective but also pleasant to use and evoke positive emotions [21]. Our research confirmed that this approach contributes to better student engagement, increased motivation, and overall satisfaction with the learning process.

The results of using Miro showed that 78% of students with ASD demonstrated positive emotional reactions while working with the platform. Particularly important is the observation that more reserved students with autism, who usually participate less in group discussions, made significant contributions to task completion on the virtual board without feeling social pressure from classmates.

4.3.2 Specialized Applications in the Formation of Language Competencies

Based on foreign experience and our own experimental research, we have implemented a series of specialized applications designed to consider the special needs of students with ASD. These applications were organically integrated into the foreign language learning process and adapted according to the individual characteristics of students (Table 5).

Table 4: Features of using the miro virtual board in teaching students with ASD.

Functional capabilities	Adaptation for students with ASD	Results of application
Creation of personalized workspaces	Ability to change board color, add personal elements, and customize the visual environment according to the student's sensory needs	Reduced anxiety, increased motivation, creation of a comfortable learning environment
Real-time collaboration	Structured interaction with clearly defined roles and tasks, the possibility of asynchronous work	Development of social skills, overcoming communication barriers
Use of emotional pictograms	Inclusion of visual elements for expressing emotional state	Development of emotional intelligence, improved reflection, establishing emotional connection
Creation of project spaces	Clearly structured tasks with visual instructions and timeframes	Development of planning skills, sequential task completion

Table 5: Specialized applications and their adaptation for students with ASD.

Application name	Functional features	Adaptation for students with ASD	Level of effectiveness
Autism iHelp - WH?	Development of language skills through questions with answers based on real-world visualization	Individual pace, topic selection according to special interests, gradual task complexity	High (85% of students demonstrated positive dynamics)
Autism iHelp Language Concepts	Development of basic language concepts and vocabulary expansion	Use of real-life photographs, categorization, and sorting, an adaptation of difficulty level	Medium (67% of students showed improvement in language concept acquisition)
Proloquo2Go	AAC system for alternative communication	Individual symbol selection, interface adaptation, vocabulary customization according to needs	Very high (91% of non-verbal students began using the application for communication)

Table 6: Model of digital tool integration for learning individualization.

Learning stage	Digital tools	Adaptation strategies	Results
Introduction of new material	Miro, Padlet, Visual Timelines	Structured presentation, use of color coding, visual cues	Improved perception and understanding of new material
Reinforcement and practice	Quizlet, Wordwall, Autism iHelp	Individual pace, repetition, visual support, gradual complexity	Increased level of material mastery, skill formation
Communicative practice	Miro (for projects), Google Meet with visual support	Structured dialogues, visual aids, clear roles and scenarios	Development of communication skills reduced social anxiety
Assessment	Miro, Quizizz, individualized tests	Adapted format, additional time, visual support, alternative response forms	Objective progress assessment, reduced test anxiety

The implementation of these applications in the educational process occurred gradually, with constant monitoring of their effectiveness and necessary adjustments. An important component of successful adaptation was a comprehensive approach that involved using applications not only during classes but also incorporating them into homework and therapeutic sessions.

4.3.3 Integration of Digital Tools into the Learning Process

For maximum effectiveness in individualizing learning for students with ASD, we developed a comprehensive approach to integrating digital tools into the educational process. This approach involves combining different platforms and tools according to individual student needs and educational goals (Table 6).

An important component of effective digital tool integration is continuous monitoring and evaluation of their impact on the learning process. For this purpose, we developed a system of indicators that allows tracking the progress of students with ASD in five dimensions of engagement: behavioral, cognitive, collaborative, emotional, and social, as proposed by Redmond et al. [22].

4.3.4 Results of a Digital Tool Implementation

The results of our research confirm the high effectiveness of implementing digital tools in the process of individualizing learning for students with ASD. Based on survey data and observations, the following positive trends were identified:

- Increased motivation: 82% of students with ASD demonstrated increased interest in foreign language learning when using adapted digital tools.
- Development of digital skills: Integration of Miro and other digital platforms contributed to the development of students' digital competencies, which is an important component of their future professional adaptation.
- Overcoming access barriers: Digital tools ensured accessibility of the educational process even in remote learning conditions, guaranteeing educational continuity in complex social conditions.
- Inclusive assessment practices: The use of digital platforms allowed the implementation of fair and unbiased forms of assessment that take into account the individual characteristics of students with ASD.
- Positive learning atmosphere: The emotional aspect of using digital tools, especially Miro with emotional pictograms, contributed to creating a positive learning environment.

Comparative analysis of the academic performance of students with ASD before and after implementing the complex of digital tools showed significant improvement in foreign language acquisition. In particular, the average indicator of vocabulary mastery increased by 34%, grammatical structures by 27%, and communication skills by 21%.

It is important to note that the effectiveness of digital tool implementation largely depends on a systematic approach that involves:

- Continuous teacher professional development in digital technology use
- Tool adaptation according to each student's individual needs
- Regular monitoring and effectiveness evaluation
- Close collaboration with parents and other support team specialists

The obtained results indicate that the integration of digital tools is an effective mechanism for individualizing learning for students with ASD, especially in the context of foreign language learning. Implementation of such tools not only increases the effectiveness of the educational process but also creates a favorable socio-emotional environment, which is critically important for students with autism.

5 DISCUSSIONS

It is worth noting that many difficulties arise from working with children with ASD. Among the contradictions that can be outlined:

1) the contradiction between the need to use special educational and methodological support for children with ASD and its lack because this issue has not yet received sufficient coverage;
2) contradiction in the approaches to the assessment of students with SEN because, on the one hand, such children need to be assessed from the position of creating a situation of success so that the child feels motivated to study. On the other hand, the entrance exams to higher education institutions are not adapted for children with SEN, and there is a gap between real assessment and knowledge;
3) in some places, there are difficulties with the correct determination of the degree of support because there are currently 5 of them.

Teaching languages to students with special educational needs is not without its challenges. However, over time, there has been much discussion and debate within the language teaching community about what might constitute a viable pedagogy for these learners.

Any consideration of interactive resources must take place in the context of such appropriate pedagogy, as resources alone will not deliver results. The main question should always be, "How will this resource improve learning?" There is no point in using resources just because they are available or because you can.

Turning to the LD Online resource, we highlight a seven-step approach to creating a foundation for language learning by students with learning disabilities, presented in Figure 5 [20].

Figure 5: The main stages of learning a foreign language by students with SEN.

6 CONCLUSIONS

A key element of successful foreign language learning by students with ASD in a distance format is a seven-step approach to creating a foundation for language learning by students with learning difficulties. The individual approach, integration of materials and situations from real life, development of functional academic skills and use of visual materials, and creation of adapted educational and methodological support are also important here.

To educate children with ASD, it is necessary to adapt traditional teaching methods to the distance format. This approach will meet the educational needs of children with ASD, as well as promote their social development and emotional well-being.

It is also important for such children to create a situation of success in learning, to provide challenging tasks, in particular, to divide tasks into small steps, to apply a structured approach to education, and to model correct answers. A structured learning environment is a key element of successful distance learning for students with ASD.

Research prospects:

1) Studying the experience of distance education will contribute to a more effective search for strategies, methods and means for teaching children with special educational needs.

2) Monitoring data obtained on the basis of the survey will allow us to outline possible risks, problems and difficulties in the future, as well as indicate effective forms of organizing online learning.
3) Since the war is ongoing in Ukraine, it is important to provide different approaches to organizing learning, including distance learning, because in the territories close to the front, there is no possibility of studying in person.
4) The study helped to identify the shortcomings of online learning, group them, describe them and look for alternative ways to solve problem situations in the future.

Attention to the education of children with disabilities is growing every year, which is due to those factors of a stressful, environmental, crisis-boundary nature. Therefore, the educational space must be prepared not only for different learning formats, but also for compensating for educational losses caused by destructive factors.

REFERENCES

[1] Ukrainian Parliament Commissioner for Human Rights, “World Autism Awareness Day.” Ombudsman.gov.ua, 2023. [Online]. Available: https://ombudsman.gov.ua/uk/news_details/vsesvitnij-den-poshirennya-informaciyi-shchodo-problem-autizmu.

[2] M. Supratiwi, M. Yusuf, and F. K. Anggraini, "Mapping the Challenges in Distance Learning for Students with Disabilities during COVID-19 Pandemic: Survey of Special Education Teachers,"International Journal of Pedagogy and Teacher Education, vol. 5, no. 1, p. 11, Apr. 2021. [Online]. Available: https://doi.org/10.20961/ijpte.v5i1.45970.

[3] Osvita.ua, "On Amendments to the Procedures Approved by the Resolutions of the Cabinet of Ministers of Ukraine of 10 April 2019 No. 530 and of 15 September 2021 No. 957," Apr. 26, 2022. [Online]. Available:https://osvita.ua/legislation/Ser_osv/86366.

[4] J. G. L. Andrade, A. C. F. do Carmo, A. C. Tamanaha, and J. Perissinoto, “Eficácia de Programas Educativos à Distância para pais de crianças com diagnóstico de Transtorno do Espectro Autista: uma revisão integrativa,” CoDAS, vol. 36, no. 5, 2024, doi: 10.1590/2317-1782/20242023291pt.

[5] T. Bratyshko, “Organization of distance learning for students with special educational needs in an inclusive education” Pedagogy of the formation of a creative person in higher and secondary schools, vol. 1, no. 70, pp. 51–55, 2020, doi: 10.32840/1992-5786.2020.70-1.9.

[6] C. K. Syriopoulou–Delli, E. Papaefstathiou, K. Sarri, A.-E. Filiou, and E. Gkiolnta. “The use of distance learning and E-learning in students with autism spectrum disorder: A review,” Nucleation and Atmospheric Aerosols, vol. 2909, no. 1, Jan. 2023. [Online]. Available: https://doi.org/10.1063/5.0181795.

[7] Electronic versions of textbooks, imzo.gov.ua. [Online]. Available: https://lib.imzo.gov.ua/yelektronn-vers-pdruchnikv.

[8] L. Safta-Zecheria, “Challenges posed by COVID-19 to the health of people with disabilities living in residential care facilities in Romania,” Disability & Society, vol. 35, no. 5, pp. 837–843, Apr. 2020, doi: 10.1080/09687599.2020.1754766.

[9] Ministry of Education and Science of Ukraine, "Organisation of distance learning at school. Guidelines, May 2020." [Online]. Available: https://surl.li/lbvoig.

[10] F. Fahrudin, A. Ansari, and A. S. Ichsan, “Pembelajaran Konvensional dan Kritis Kreatif dalam Perspektif Pendidikan Islam,” Hikmah, vol. 18, no. 1, pp. 64–80, Sep. 2021, doi: 10.53802/hikmah.v18i1.101.

[11] W. Zhang, Y. Wang, L. Yang, and C. Wang, “Suspending Classes Without Stopping Learning: China’s Education Emergency Management Policy in the COVID-19 Outbreak,” Journal of Risk and Financial Management, vol. 13, no. 3, p. 55, Mar. 2020, [Online]. Available: https://doi.org/10.3390/jrfm13030055.

[12] A. N. Bahasoan, W. Ayuandiani, M. Mukhram, and A. Rahmat, “Effectiveness of Online Learning In Pandemic Covid-19,” International Journal of Science, Technology & Management, vol. 1, no. 2, pp. 100–106, Jul. 2020, doi: 10.46729/ijstm.v1i2.30.

[13] O. S. Homdijah, E. Heryati, Ehan, and M. P. Annisi, “Distance Learning for Children with Autism Spectrum Disorders,” JPI (Jurnal Pendidikan Indonesia), vol. 12, no. 2, pp. 367–375, Jun. 2023, [Online]. Available: https://doi.org/10.23887/jpiundiksha.v12i2.45794.

[14] C. T. Moody and E. A. Laugeson, “Social Skills Training in Autism Spectrum Disorder Across the Lifespan,” Psychiatric Clinics of North America, vol. 43, no. 4, pp. 687–699, Dec. 2020. [Online]. Available: https://doi.org/10.1016/j.psc.2020.08.006.

[15] S. Shanker and T. Barker. Self-reg: How to help your child (and you) break the stress cycle and successfully engage with life. New York: Penguin Press, 2016.

[16] T. Grandin. The way I see it: a personal look at autism & Asperger’s. Arlington, Tx: Future Horizons Inc, 2015.

[17] D. Costa and M. Williams, “English”, May 2018.

[18] L. Hordyk, “Modifications for Students With Autism: How Can Teachers Support Students With Autism in General Education Classrooms,” 2022. [Online]. Available:http://hdl.handle.net/20.500.12680/tx31qq470.

[19] ‘Lumos Foundation,’ Ukrainian Child Rights Network, Apr. 05, 2024. [Online]. Available: https://childrights.org.ua/members/https-www-wearelumos-org.

[20] LD Online. “All about learning disabalities”. [Online]. Available: https://www.ldonline.org.

[21] D. Norman, “Emotional Design : Why We Love (or Hate) Everyday Things”. New York: Basic Books, 2004.

[22] P. Redmond, L.Abawi, A.Brown, R. Henderson, and A.Heffernan, “An online engagement framework for higher education. Online learning”. vol. 22. no 1. p. 183-204, 2018. [Online]. Available: https://doi.org/10.24059/olj.v22i1.1175.

The Significance of Personality-Oriented Technology in Higher Education in Ukraine

Olha Dienichieva, Iryna Biliak, Natalia Plakhotniuk, Tetiana Verhun and Kateryna Parnus
State University "Zhytomyr Polytechnic", Chudnivska Str. 103, 10005 Zhytomyr, Ukraine
{olhadienichieva, nat.pl2002, katemarchuk38}@gmail.com, {kim_vtm, ktpl_biv}@ztu.edu.ua

Keywords: Personality-Oriented Technologies, Self-Directed Learning, Responsive Learning Environment, Lifelong Education.

Abstract: This paper addresses the growing significance of personality-oriented technologies in higher education, particularly in times of crisis such as martial law in Ukraine. The research focuses on analyzing the application of digital tools, adaptive learning platforms, and AI-based educational resources to enhance individualized learning processes and support student engagement. The main aim of the study is to assess the impact of these technologies on students' motivation, academic outcomes, and adaptability during challenging periods. A mixed-methods approach was employed, combining quantitative and qualitative methods, including the analysis of survey data collected from 300 first-year students at Zhytomyr Polytechnic State University. The results show that while students widely utilize digital content and interactive platforms, the application of adaptive learning technologies and AI-driven tools remains limited. Students prefer independent learning modes and show moderate interest in mental health support services. The findings underline that the integration of personality-oriented technologies can significantly enhance learning efficiency, motivation, and self-directed learning skills. However, the implementation faces challenges related to limited engagement with adaptive tools and varying levels of digital literacy. The practical relevance of the study lies in offering pathways for educational institutions to better align digital solutions with diverse student needs. The theoretical contribution includes expanding the understanding of how personality-oriented technologies function in crisis-affected higher education environments and how they can be optimized for future resilience.

1 INTRODUCTION

The use of personality-oriented technology in education has gained considerable significance due to its ability to address the students' individual learning needs. Personality-oriented technology encompasses a range of pedagogical forms and methods, and digital tools, including adaptive software and artificial intelligence (AI) designed to develop and enhance individualized learning strategies and techniques. Pedagogical technologies are designed to modernise the content of educational material, considering varying pace of information processing and individual abilities and students' professional orientation, thereby fostering a more individualized approach to learning. Personality-oriented technologies enhance motivation and contribute to more effective educational outcomes by focusing on the development of students' professional competences and identifying areas for self-development and self-improvement [1; 2; 3]. They have become especially acute when access to education is significantly restricted. For instance, following the military invasion in Ukraine, educational institutions were compelled to transition to distance learning using various digital educational platforms (e.g. Zoom, Google Meet, etc.), which had a significant impact on the students' intrinsic motivation to learn [4].

Conventional pedagogical methods, which rely on face-to-face communication have frequently proven inadequate during times of various crises Personality-oriented technologies offer a solution by providing flexible, adaptable and accessible learning options to alleviate stress during educational process. Among the options that support lifelong learning, there are mobile learning applications and adaptive educational platforms individuals can use in challenging circumstances [5]. Modern technologies enhance learning outcomes, allowing students to progress at their own pace at the same time receiving professional support, consequently mitigating the impact on their educational experience [6]. Thus, personality-oriented technologies reflect societal trends to provide an opportunity to build a robust education system.

The role of modern technologies in education has shifted from delivering general information to providing personalised content, considering the processes of individualization and differentiation. The implementation of personality-oriented technology in education enables the identification of students' interests and abilities, for the purpose of fostering professional development of future specialists. Under these conditions, teachers are able to create a favourable educational environment by utilizing technological tools to stimulate external and internal factors that promote engagement in educational activities.

Philosophical and pedagogical frameworks emphasize the importance of recognizing the holistic nature learners: encompassing their mind, emotions, and personality. Traditionally, the tendency to adopt a one-size-fits-all approach to personality development may overlook the diverse ways in which individuals process information. The content of modern human development concepts in society does not restrict the use of individualized methods to unlock students' potential. Drawing on psychology and cognitive sciences to support professional development, the individualization and differentiation of the educational process aim to provide a wide range of task types for the professional development of the individual.

The particular importance of the studies of international scholars such as Dora Y. Kovrei, a representative of the Transylvanian Hungarian College in Berehove, who collaborates with Ukrainian researchers in the implementation of personality-centered strategies in Eastern Europe. Her work reflects the broader international interest in this issue, evidenced by France's national policy of learning individualization since 1989. Meta-analyses confirm the influence of personality traits on academic performance, while research from Asia demonstrates the effectiveness of using machine learning algorithms to personalize education.

Innovative technologies such as artificial intelligence, adaptive learning platforms, and interactive digital tools have enabled educators to put personality-oriented education into practice. These technologies can identify students' learning patterns, adapt content to their individual development, and offer personalized feedback, thereby transforming the teaching and learning process.

2 RESEARCH GOALS

The purpose of the study is to analyse the role of personality-oriented technologies in higher education, highlighting the importance of using digital tools to provide students with pedagogical support and maintain their motivation to education during challenging times. The research focuses on analyzing the characteristics of personality-oriented technology and identifying the challenges of its implementation in the educational process during challenging times. The objectives of the research are to explore the impact of personality-oriented technology implementation on the student engagement and assess the students' adaptability to the use of learning technologies in education. The practical significance of the research lies in identifying the ways to adapt educational practices to foster responsive learning environment.

3 RESEARCH METHODOLOGY

The following methods were employed in the research: analysis, synthesis, and generalisation to explore the issue of personality-oriented technology in education; analysis of the students' surveys, observation and systematisation to identify the relationships between learning goals, educational needs and learning outcomes during times of crisis.

To ensure the reliability and clarity of the survey, a pilot test was conducted with a sample group of students. Based on their feedback, ambiguous items were refined. The final version was reviewed by academic experts to ensure alignment with research objectives and internal consistency of the tool.

Methodology. The term "technology" is derived from Greek words that comprises two words: "techne" and "logos". The concept of technology in education is understood as a means to integrate various processes, techniques, methods to create favourable conditions for the individual development, address their individual educational needs and determine their learning objectives [7]. The following pedagogical technology should be highlighted as personality-oriented technology. This technology aims to develop the student's potential, enhance their skills and abilities, and provide professional experience.

It is essential to examine personality-oriented technology in the context of a crisis from three perspectives: social, educational, and psychological. In the social aspect, the technology is used to assess the educational needs of an individual in response to specific life situations, and to study the impact of social, political, and economic factors on personal development during a crisis. In the field of education, the use of personality-oriented technology ensures an individualised approach to learning. Considerable attention is given to the organization of the educational process in alignment with the analysis of labour market requirements, differentiation and individualization of learning. In the psychological field, significant emphasis is placed on the development of the student's individual potential, cognitive skills, psychological and pedagogical support in stressful situations and during rehabilitation after such experiences.

Personality-oriented technologies encompass a broad range of digital tools and methods designed to address the learners' educational needs. The implementation of this technology is grounded in three core principles: mobility, responsiveness, and learner-centered design [8; 9; 5; 2]:

Mobility refers to the ability of educational technology to provide multiple ways of learning, allowing students to access information in a way that suits their individual circumstances, schedule and learning environments. For instance, personality-oriented technologies facilitate asynchronous learning, where participants can work with resources at their own pace.

Responsiveness is another fundamental principle that emphasizes the need for educational tools to adapt content and teaching strategies according to real-time feedback from learners, while considering the environments in which they are learning. This dynamic approach ensures that learners receive appropriate support and guidance to foster their educational development.

The principle of Learner-centered design emphasizes the importance of tailoring the educational experience to each individual learner, considering their unique experiences, interests and learning styles. By prioritizing the needs of the learner, personality-oriented technologies foster an inclusive learning environment that enhances engagement and motivation.

The implementation of personality-oriented technologies in the educational process aims to enhance the effectiveness of students' learning. It is viewed as a means to develop students' critical thinking abilities, foster their learning habits and cultivate the ability to adjust everyday activities in stressful situations. The realisation of the educational process through educational platforms and the use of digital tools to provide personalised content enhances the accessibility of resources tailored to individual students' needs. This not only enhances the relevance of content, modernises educational programmes and updates curriculum, but also increases student internal motivation. For example, during martial law in Ukraine, educational platforms introduced the personalised content to address the educational challenges faced by displaced students, offering them learning resources to support their academic continuity.

The integration of Information and Communication Technologies (ICT) in learning process enhances language skills development while fostering critical thinking, creativity, self-expression, and personal growth. Additionally, leveraging tools such as YouTube, mind maps, and educational blogs enables students to unlock their academic potential, access up-to-date professional information, and engage in collaborative learning and communication [10].

It is important to emphasise that the use of innovative pedagogical technologies and digital tools, which allow students to gain professional and practical experience in challenging social conditions is aimed at the formation of personal values that are essential for their self-development. Personality-oriented technology focuses on creating favourable conditions for the realisation of creative skills. The primary objective of this technology is to facilitate the psychological adaptation to learning by enabling the mastery of professional training [11; 12].

Another important aspect of personality-oriented technology is the availability of communication. This means that learning platforms facilitate teacher-student interaction, instant assessment of completed tasks allowing for timely correction of written work and adherence to deadlines. This immediate feedback fosters intrinsic motivation to study and encourages students to take responsibility for their education [13; 14].

In addition, with the development of artificial intelligence technologies, the implementation of personality-oriented technology also implies providing personalised learning support, which reinforces the concept of adopting a personality-oriented approach to modernise the content of educational programmes to update the curriculum to enhance students' interest in education in time of crisis. Researchers recognize the potential of such technologies, emphasising that they contribute to improving student learning outcomes when traditional teaching methods and technologies may prove ineffective during times of crisis.

4 RESULTS

To understand the complex nature of personality-oriented technology applications in the educational process, a mixed-methods approach was employed: combining both quantitative and qualitative methods to collect and analyze the results of students' surveys to assess the effectiveness and potential of the implemented technologies. A structured survey was developed, consisting of three sections and 35 questions: the first section 'Frequency of Technology Use' (Table 1), focuses on the frequency of technology use in studying; the second section, 'Types of Personality-Oriented Technologies Applied' (Table 2), and the third section 'Impact on Learning Outcomes' (Table 3), examined the influence of digital tools on students' motivation and learning outcomes. A total of 300 survey respondents, including first-year university students from various disciplines at Zhytomyr Polytechnic State University, were informed about the purpose of the research. The first-year university students of various disciplines at Zhytomyr Polytechnic State University were informed about the purpose of the research. The survey was distributed via email and was conducted anonymously and confidentially.

The use of modern pedagogical technologies and digital resources modernizes the educational process, updates the curriculum content, and creates engaging learning environment where students can learn through interactive exercises and real-life communication situations. It is important to note that digital tools and online resources are widely integrated into the educational process (Table 1). It is emphasized that digital tools such as apps, educational videos and interactive technologies are extensively used, indicating the availability and relevance of these resources.

As a result, students demonstrate a moderate interest in adaptive and individualized learning features, with a clear preference for interactive content and some inclination toward gamification. However, their engagement with AI-based tools and social platforms remains limited (Table 2). The findings further indicate that while students recognize the advantages of digital and interactive content, the adoption of adaptive learning technologies and AI-driven educational tools, such as AI-powered language learning applications (e.g., Duolingo, WordUp, Learn English Grammar) and automated writing assistants (e.g., Grammarly), is not widespread. Independent learning is generally preferred over collaborative methods, such as peer-assisted learning, group discussions, and project-based learning. Moreover, mental health tools and adaptability in learning receive only moderate to low prioritization.

Table 1: Frequency of technology usage in students' learning process.

Survey issues	Results and outcomes
Usage of devices:	The vast majority (97.7%) of students use digital devices for learning on a daily basis.
Access to academic online resources:	Most participants frequently access online resources: 63.6% use them weekly and 22.7% use them daily.
Use of educational applications:	Educational applications are used moderately: 13.6% of students use them frequently, 34.1% use them from time to time, and 29.5% never use them.
Digital technologies for note-taking/memorization/lecturing:	The use of digital notes is balanced: 31.8% of students take notes often or sometimes, and 29.5% rarely.
Interaction with learning groups/ communities:	Involvement is varied: 31.8% of students participate daily, 15.9% engage weekly, and 34.1% rarely.
Making use of digital tools to increase efficiency:	Only a minority (11.4%) of respondents use productivity tools very often, 27.3% use them rarely, and 18.2% never.
Educational content:	Most students watch educational videos: 38.6% rarely, but 25% sometimes and 20.5% often.
Self-assessment technology:	The level of regular self-assessment is weak: 45.5% of participants rarely take tools such as tests, while 9.1% of students use them frequently.
Online learning/support/consultation:	Online learning/support/consultation is used infrequently: 45.5% of students rarely use it and only 9.1% of them often.

Table 2: Application of personality-oriented educational technologies.

Survey issues	Results and analyses
1. Use of Learning Platforms	Usage varies, with 22.2% of students using the tools frequently, 25% occasionally, and 38.9% rarely, indicating some interest, but limited widespread use.
2. Digital Content	Interactive content is the most preferred (44.4%), followed by visual content (16.7%) and text-based content (11.1%), highlighting a clear reference for interactivity.
3. AI-based EducationalTools	Only 13.9% of students use AI tools regularly, while half use them rarely, indicating limited application of AI in educational contexts.
4. Digital Individual Assessments	The use of personality assessments is low, with 8.3% of participants using them frequently and 38.9% using them rarely. This reflects a limited level of engagement with these tools.
5. Prioritizing adaptive educational resources	Most students are neutral (58.3%) regarding adaptive tools, while 16.7% strongly agree. The results indicate limited enthusiasm for the use of adaptive tools.
6. Individualization in the educational environment	Individuals show a slight preference for customization, with 25% strongly agreeing and 47.2% agreeing on its importance.
7. Use of Online Platforms	The research reveals a low level of participation, with 30.6% of respondents never using online platforms for community engagement, and only 8.3% using them daily, indicating that such platforms are not a top priority.
8. Use of Gamified learning resources	The usage of gamification is limited, with 11.1% of participants using it frequently, and 27.8% never using them, suggesting that gamification is not a widely preferred approach.
9. Key features of educational tools	Social sharing (25%) and usage (22.2%) are the most valued, while adaptive learning holds less importance (8.3%). Social communication emerges as a key factor.
10. Educational Recommendations	16.7% strongly agree, 36.1% agree that recommendations are useful, while 41.7% remain neutral. The responses are divided with substantial proportion of neutral opinions.
11.Personality-Oriented Learning Methods	Independent learning is preferred by 38.9% of respondents, followed by collaborative tasks at 16.7%. This indicates a stronger preference for autonomy in learning.
12. Immediate Feedback Resources	19.4% of students frequently use feedback, but 50% never utilize them. This indicates that although some participants value feedback tools, it is generally underutilized.
13. Focus on Mental Health Tools	The responses vary: 16.7% strongly agree, the majority remain neutral (52.8%). Although mental health tools are of notable concern, they are not considered a top priority.
14. Significance of responsiveness	8.3% of respondents consider adaptability to strengths and weaknesses to be very important, while 41.7% remain neutral. This suggests that flexibility is not a strong priority.
15. Use of Self-Reflection and Self-Assessment Tools	38,9 % of participants sometimes use learning technologies that encourage self-assessment tools, 25 % rarely use them, 19,4 % never use self-reflection tools and only 8,3 % of students frequently use these tools.

According to the survey, personality-oriented technologies enhance learning efficiency, however; a certain percentage of respondents still view them as merely useful (Table 3). A significant number of students maintain a neutral attitude towards these technologies, focusing on their overall effectiveness and primary outcomes in training.

It is important to highlight that a significant portion (46.9%) of students agreed that technology helps them understand complex topics (Table 3). At the same time half of the respondents suggested that the role of technology in a detailed examination of complex material may vary depending on the type of technology or individual learning style/pace. Alongside, 53.1% reported that technology had a positive impact on their motivation, half of the respondents reported a contradictory response. The discrepancy observed in the middle of the survey suggests that the use of personality-oriented technologies enhances students' motivation to acquire new knowledge, while other half may not view the use of technologies as significant factor influencing the motivation level. Most students agreed that digital technologies aid in time management, with 34.4% reporting a significant effect (Table 3).

Table 3: Effectiveness of digital technologies on student learning.

Survey Issues	Results %
Impact of Technology Use on Learning Efficiency	Significantly improved: 18,8% Slightly improved: 25% Neutral: 56,3%
Deepening comprehension of complex topics	Agreed: 46,9% Neutral: 53,1%
Information Storage	Often: 56,2 % Sometimes: 43,8%
Motivation to Education	Increased: 52% Neutral: 48% Decreased:0%
Ability to carry out evaluations	Improved: 41,8% Slightly improved: 37% No change: 21,2%
Managing time while studying	Significantly: 28,1% A little: 37,5% No impact: 34,4%
The impact of digital education on academic performance	Agree: 57,3% Neutral: 37,5% Disagree:5,2%
Reducing academic stress with technology	Frequently: 31,3% Sometimes: 34,4% Rarely: 25% Never: 9,3 %
Keep focused and organized	Agree: 25% Neutral; 46,9% Disagree: 28,1%
Improvement in learning outcomes	Improved: 51,5% Slightly improved; 42,4 % No impact: 6,1%

This suggests an ambiguous response to whether digital technologies consistently facilitate time management during training. More than half of the participants believe that personality-oriented technologies, innovative teaching methods and digital tools contribute to academic success, while a significant portion of the respondents remain neutral (Table 3). Many students report that stress reduction is achieved through a combination of various content technologies, methods, and techniques designed to mitigate its effects. It is noted that some of the students either agreed or remained neutral regarding the influence of technology on attention and the organization of the educational process. This result may suggest that the technologies provide some organizational tools for effective learning, but their impact is individualized.

5 CONCLUSIONS

Personality-oriented technology has proven to be a transformative approach in modern education, particularly during crises such as martial law in Ukraine. By integrating adaptive learning platforms (e.g., Canvas Network, Coursera, UA EduHub), AI, and interactive tools (e.g., Kahoot!, Quizlet, Google Classroom, Duolingo, BBC Learning English), this approach addresses individual learning needs and promotes flexibility, motivation, and skill development.

Survey data confirm high engagement with digital resources: according to Table 1, 97.7% of students use digital devices daily, and over 85% regularly access academic content. Table 2 highlights a strong preference for interactive materials (44.4%) and independent learning (38.9%), while use of AI tools (13.9%) and gamified resources (11.1%) remains limited. Table 3 supports the positive impact of these tools on efficiency, comprehension, and motivation, though student responses vary, reflecting diverse learning preferences.

Despite the benefits, challenges persist, including low engagement with adaptive tools, online collaboration, and mental health resources. Further refinement is needed to align these technologies with students' needs. Initiatives such as adaptive platforms (e.g., EdEra, Prometheus) have proven effective in supporting displaced students during wartime, illustrating the resilience of personality-oriented technologies. To maximize their impact, institutions must invest in teacher training, advanced tools, and curriculum integration, ensuring a more adaptive and inclusive learning environment. These adaptive strategies provide concrete pathways for shaping a responsive educational environment aligned with students' psychological, professional, and emotional development. Educational practices can be adapted by embedding flexible learning trajectories, providing real-time personalized feedback, and ensuring that instructional content evolves dynamically with learners' progress and contexts.

REFERENCES

[1] R. Jayadurga and S. Rathika, "Significance and Impact of Artificial Intelligence and Immersive Technologies in the field of Education," International Journal of Recent Technology and Engineering (IJRTE), vol. 12, pp. 66-71, 2023, [Online]. Available: https://doi.org/10.35940/ijrte.B7802.0712223.

[2] S.K. Mangal and U. Mangal, Essentials of Educational Technologies. New Delhi: PHI Learning, 2009.

[3] T. Bogachenko and R. Burke, "Educational 'exchange rates' in (re)settlement: The use of formal and informal learning resources by displaced people from Ukraine in Australia," International Journal of Educational Research, vol. 127, 102412, 2024, [Online]. Available: https://doi.org/10.1016/j.ijer.2024.102412.

[4] IEA, "Education and Science of Ukraine under Martial Law," Kyiv, 2023, [Online]. Available: https://mon.gov.ua/storage/app/media/zagalna%20serednya/serpnevakonferencia/2023/22.08.2023/informanalytic.zbirnosvita.v.umovah.voyennogo.stanuvykl.rozv.povoyen.perspekt.22.08.2023.

[5] M. Hrynova, Personality-oriented Technology of Education and Upbringing, 2015, [Online]. Available: http://acup.poltava.ua/wp-content/uploads/2015/03/GrinovaM.

[6] O. Dienichieva, M. Komogorova, S. Lukianchuk, L. Teletska, and I. Yankovska, "From Reflection to Self-Assessment: Methods of Developing Critical Thinking in Students," International Journal of Education and Information Technologies, vol. 15, pp. 227-236, 2021, [Online]. Available: http://doi.org/10.46300/9109.2021.15.23.

[7] R. A. Buchanan, "History of Technology," [Online]. Available: https://www.britannica.com/technology/history-of-technology.

[8] D. Bekh, "Personality-Oriented Model of Education as a Scientific Construct," Psychological and Pedagogical Problems of Rural School, vol. 3, pp. 143-156, 2000, [Online]. Available: https://library.udpu.edu.ua/library_files/psuh_pedagog_probl_silsk_shkolu/3/vupysk_21.pdf.

[9] K. Arar, A. Saiti, and M. Guajardo, "Redesigning and recomputing the future of education: The role of technology, the learning process, personality traits, and diversity in learning systems," Power and Education, vol. 15, no. 2, pp. 243-258, 2023, [Online]. Available: https://doi.org/10.1177/17577438221117346.

[10] N. Plakhotniuk, O. Chernysh, and O. Makarevych, "The Use of ICT for the Development of Foreign Language Communicative Competence in Hospitality Industry Students," Proceedings of International Conference on Applied Innovation in IT, vol. 12, no. 1, pp. 41-50, 2024.

[11] M. Borrego, "Innovation and ICTs in ESL: Exploring the Role of the Podcast as a Learning Enhancing Tool in Innovation Projects," Integrated Journal for Research in Arts and Humanities, vol. 3, no. 2, pp. 70-77, 2023, [Online]. Available: https://doi.org/10.55544/ijrah.3.2.12.

[12] M. Xodjayeva, R. Kabulov, R. Djuraeva, D. Alimjon, and A. Mukhammadiev, "Personality-Oriented Learning Technologies," International Journal of Recent Technology and Engineering (IJRTE), vol. 8, pp. 7706-7708, 2019, [Online]. Available: https://doi.org/10.35940/ijrte.D5369.118419.

[13] R. Sajja, Y. Sermet, M. Cikmaz, D. Cwiertny, and I. Demir, "Artificial intelligence-enabled intelligent assistant for personalized and adaptive learning in higher education," arXiv preprint arXiv:2309.10892, 2023, [Online]. Available: https://doi.org/10.48550/arXiv.2309.10892.

[14] C.-H. Tseng, H.-C. K. Lin, H. A. C.-W. Huang, and J.-R. Lin, "Personalized programming education: Using machine learning to boost learning performance based on students' personality traits," arXiv preprint arXiv:2501.00825, 2025, [Online]. Available: https://doi.org/10.48550/arXiv.2501.00825.

Peculiarities of Forming the Professional Culture of a Mathematics Teacher by Means of Digital Technologies in Higher Education Institutions of Ukraine

Mariana Kovtoniuk, Olena Soia, Olena Kosovets and Liubov Tiutiun
Vinnytsia Mykhailo Kotsiubynskyi State Pedagogical University, 32 Ostrozki Str., Vinnytsia, 21000, Ukraine
kovtonyukmm@vspu.edu.ua, soia.om@vspu.edu.ua, kosovets.op@vspu.edu.ua, tiutiun.la@vspu.edu.ua

Keywords: Mathematics Teacher, Professional Culture of the Teacher, Mathematical Culture, Scientific and Methodological Culture, Pedagogical Culture, Mathematical Thinking, Mathematical Language, Self-Development, Digital Technologies.

Abstract: The article substantiates the professional culture of a mathematics teacher as an integrated dynamic property of a personality that projects his/her general culture in the field of profession, and is a synthesis of mathematical, scientific, methodological and pedagogical cultures of a teacher, realized in a synergistic educational space using digital technologies. The purpose of the article is to carry out a scientific and theoretical analysis of the peculiarities of the formation of the professional culture of a mathematics teacher and a future mathematics teacher, to define the concept of professional culture of a mathematics teacher, to model its main components (mathematical, scientific, methodological and pedagogical cultures), to establish their structure, to select modern digital technologies for the implementation of professional culture, to conduct a pedagogical study in higher education institutions of Ukraine and to outline possible ways of its formation in a post-industrial society. Mathematical culture includes mathematical competencies, mathematical thinking, and mathematical language. Scientific and methodological culture involves scientific research, implementation of innovative teaching and learning methods, development of the author's educational and methodological product. The pedagogical culture of a mathematics teacher is his/her ability to self-analyze, professional development; self-education, self-development, self-management, pedagogical skills; motivation, communication, professional morality, reflection, adaptability to the professional environment. The Spearman's rank correlation coefficient r_s was used to statistically test the research hypothesis. The interpretation of the size of the effect of correlation dependence was performed in accordance with the methodological recommendations of J. Cohen.

1 INTRODUCTION

The multifaceted nature of the concept of “professional culture of a teacher” leads to different approaches to its interpretation and requires in-depth research to formulate a common understanding of the content and structure of the concept of “professional culture of a mathematics teacher”.

The analysis of scientific research allows us to identify a set of systemic elements that characterize professional culture as a phenomenon, namely: professional competencies, professional thinking, professional interest, professional experience, professional outlook, the degree of readiness of an individual for a particular type of activity, the ability to research, professional skill, adaptability to the professional environment, professional morality, etc. This concept includes the entire spiritual potential of an individual, intellectual, emotional and practical components of his or her consciousness.

Although foreign research emphasizes the importance of cultural factors in teacher training and performance [1, 2, 3, 4, 5, 6, 7, 8], there is limited practical experience of how culture moderates the relationship between worldview, value orientations and professional identity of teachers, especially in the Ukrainian context.

At the same time, the combination and generalisation of the key features of the culture of society, the culture of the individual and the culture of social integration in addition to and in conjunction with each other create a holistic picture of understanding culture as a scientific concept.

2 ANALYSIS OF CURRENT RESEARCH

L. Huberskyi argues that certain results of activity acquire the status of “culture” only to the extent that they contribute to the development of the essential forces of a person, the creative capabilities of the individual [9, p. 21]. Considering the activity through the prism of the diversity of its interpretations, we note that any significant result of the activity affects the formation and development of a person, and the accumulation of new achievements and values leads to an increase in the overall level of culture of society.

According to S. Honcharenko, “culture (lat. cultura – education, development) is a set of practical, material and spiritual achievements of society that reflect the achieved level of development of society and human. Culture is understood as the level of education and upbringing of a person, as well as mastery of a certain field of activity” [10, p. 182].

In our research, we will take as a basis the following definition of culture as a complex (totality) of material and spiritual achievements of society, symbolic products of personality development (dispositions, knowledge, abilities, creative forces, value orientations, etc.) and technologies for the implementation of any purposeful activity in a specific human environment [11].

I. Glazkova et al. [12, pp. 144-161] analyze the phenomenon of professional pedagogical culture as a socio-pedagogical phenomenon that integrates historical and cultural experience into coordinated pedagogical activity.

The article [13, pp. 42-53] aims to explore the relationship between educational practices, professional culture and sustainable development goals, in particular in the context of the Ukrainian national educational system and professional community.

The article [14, pp. 17-25] highlights the theoretical foundations of the professional culture of future specialists in accordance with the global challenges of the information society, concludes that the main components of the content of professional culture are motivation to master special professional knowledge, skills, professionally significant values, which in the process of forming professional culture should become personally significant.

The works of [15, 16, 17, 18, 19] are of great importance for our research.

The purpose of the article is to define the concept of professional culture of a mathematics teacher, model its main components (mathematical, scientific, methodological and pedagogical cultures), establish their structure, select modern digital technologies for the implementation of professional culture, conduct a pedagogical study on the peculiarities of the formation of professional culture of a mathematics teacher and a future mathematics teacher in higher education institutions of Ukraine and outline possible ways of its formation in a post-industrial society.

The structure of the study is based on the principle from the general to the specific, which allows us to consistently consider both the theoretical foundations of the problem and its practical aspects.

3 DIGITAL TOOLS IN THE FORMATION OF PROFESSIONAL CULTURE OF A MATHEMATICS TEACHER

The formation of professional culture of a mathematics teacher and a future mathematics teacher is possible with the use of system-activity, synergistic, competence, prognostic, personal development, educational, research and information approaches that allow improving the quality of education and personal awareness.

The basis of the methodological approach to the study of the personality of a high-level mathematics teacher is to substantiate and verify the peculiarities of the formation of his/her professional culture while working in a higher education institution.

Using the cultural approach [9, 10], it is possible to assess the professional culture of a mathematics teacher through a system of criteria that reflects his or her ability to cultural self-identification and integration into the professional community.

For us, the study of H. Mykhalin is especially valuable, as he revealed the content of the concept of “professional culture of a mathematics teacher” as a set of practical, material and spiritual achievements that determine the quality of the teacher's professional activity [20, p. 30]. The scientist considers mathematical, methodological, pedagogical, psychological, informational, linguistic and moral cultures to be the main components of the professional culture of a mathematics teacher and notes that each of these components of the professional culture of a specialist, in particular a mathematics teacher, with the exception of moral culture, consists of general and special parts, and its content significantly depends on the specifics of the specialist's activity.

Based on the analysis of scientific literature, we understand the professional culture of a mathematics teacher as an integrated dynamic property of a personality that projects his or her general culture in the field of the profession, is a synthesis of mathematical, scientific, methodological and pedagogical cultures, and is realized in a synergistic educational space using digital technologies. Since digital technologies play a fundamental role, we include them in all components of the professional culture of a mathematics teacher (Figure 1).

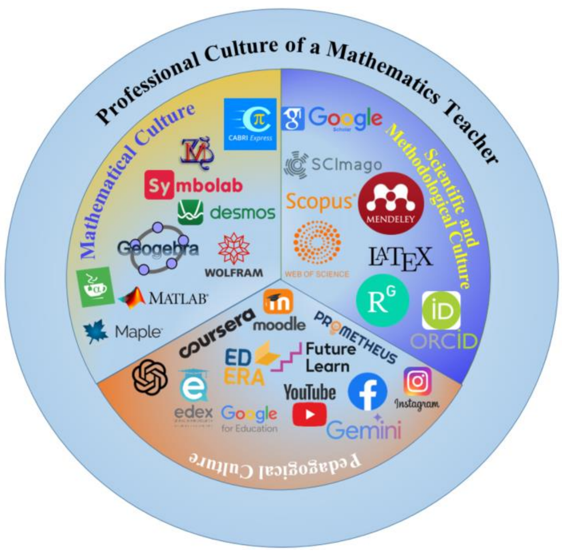


Figure 1: Professional culture of a mathematics teacher (author's development).

Let's consider the first dynamic property of the professional culture of a mathematics teacher - mathematical culture, which reflects the formation of a system of mathematical knowledge and skills, mathematical thinking and mathematical language, developed ability to self-education, which form the professional worldview of a mathematics teacher.

The problem of mathematical culture is addressed in the articles of both mathematicians who considered it from the mathematical aspect and scientists-pedagogues who studied it from the methodological aspect. In particular, Ye. Lodatko interprets the mathematical culture of society as "a complex social formation that is formed under the influence of mathematical traditions, an established system of mathematical education and mathematical achievements" [21, p. 78]. H. Zinchenko emphasizes that "the mathematical culture of the future mathematics teacher is determined not only by the high level of mastery of mathematical knowledge, the ability to use it in practice, mathematical language and speech, but also by the teacher's own system of values, his/her general worldview erudition and, most importantly, the ability to form this culture in students" [22, p. 92]. The analysis of the structural components of mathematical culture is carried out in [23, pp. 52-56].

The mathematical language allows mathematics educators to describe the world with unique precision and conciseness. Unlike natural languages, which are prone to ambiguity and contextual nuances, mathematics offers a rigorous formalism where each concept has a clear definition and statements can be proven or disproven through logical reasoning. This precision makes mathematics indispensable in science, technology, and many other areas of human activity.

Back in 1927, H. Williams wrote that mathematics is both a source of truth and a special language. The mathematical language is more defined and abstract than our usual means of thinking and communicating [24].

The main skills related to the mathematical language of a mathematics teacher: to know mathematical terminology, in particular in Ukrainian and English; to present educational material competently; to be able to highlight the main thing in mathematical sentences; to justify mathematically correct problem solving; to use computer mathematics systems.

Mathematical competence includes a culture of logical and algorithmic thinking, knowledge and ability to apply mathematical (numerical and geometric) methods to solve applied problems in various fields of activity, the ability to understand and use mathematical models and the ability to build such models to solve problems [25].

The ability of a mathematics teacher to work with digital mathematical tools such as Maple, MATLAB, GeoGebra, Maxima, MathCha, Symbolab, Wolfram Alpha, MathWorld, Desmos, etc. ensures the formation of his/her mathematical culture.

According to O. Maksymovych, “mathematical thinking is a complex process in which students master high-level thinking skills, namely: they learn to analyse, synthesise, generalise, classify, compare, etc. Mathematical thinking is one of the values of a person, and it is formed due to the mathematical activity of a person in the process of learning mathematics” [26, pp. 125-129].

Let's consider the scientific and methodological culture, which is a multifaceted component of the professional culture of a mathematics teacher and combines his/her scientific activity, implementation of modern teaching and learning methods, and development of his/her own educational and methodological product.

The formation of the scientific activity of mathematics teachers is ensured by the use of academic search engines (Google Scholar, Scopus, Web of Science, ORCID, Scimago Journal & Country Rank), professional document preparation systems (LaTeX), bibliographic managers (Mendeley, Zotero, EndNote), scientific communities (ResearchGate, Academia. edu) for writing and publishing articles, monographs, preparing and defending dissertations, projects and grants, obtaining titles of protection (patents, copyright certificates) for intellectual property rights.

Teaching and learning methods include the ability of a mathematics teacher to combine traditional and innovative approaches, methods, techniques, forms and means through modern digital technologies and digital competencies. As co-authors of the article [27, pp. 93-106], we propose a foresight wheel of digital technologies for use in the educational process, which represents a classification of teaching aids depending on the type of classes.

The educational and methodological product of a mathematics teacher consists of writing, editing, reviewing textbooks, teaching guidelines, workbooks, developing lectures, practical and laboratory classes, work programs and silabuses, tests and control works, which are placed in virtual mathematical educational environments. The design and development of such educational environments is carried out by means of content management systems (Moodle, WordPress, Drupal), testing systems (Kahoot!, Quizizz), online conferencing platforms (Zoom, Google Meet, Microsoft Teams), etc.

The article [28, pp. 51-57] investigates the potential of modeling a virtual mathematical educational environment on the example of a website, and identifies which tools and technologies can be used to create it.

The article [29] reveals the structure of digital competences through the following areas of activity: professional engagement, digital resources, teaching and learning, assessment, experiential learning, and promotion of students' digital competence.

In our professional activities, we pay special attention to the modern digital textbook in mathematical disciplines in the preparation of a bachelor of mathematics. After all, a a high-quality digital textbook ensures the formation of not only the mathematical culture of the future specialist, but also the formation of such basic competencies as the ability and willingness to self-learn, apply knowledge, skills and abilities to work with computer mathematics systems, self-education and future professional activities. The use of a digital textbook integrated into the learning technology designed and implemented by the teacher allows him or her to choose their own creative strategy and methodology for teaching students.

The pedagogical culture of a mathematics teacher, as an integral part of his/her professional culture, in a post-industrial society combines the ability to self-analyze his/her own professional activity and the desire to improve skills; the level of self-education, self-development, self-management; the presence of motivation, communication, professional morality and reflection of the teacher as components of the individual trajectory of his/her professional development.

The analysis of the essence of the concepts of “culture”, “independent work”, “professional culture” in the context of philosophical, cultural and pedagogical knowledge, various methodological approaches allowed to define the culture of independent work of a mathematics teacher as an integrative property of his personality aimed at accumulating general and professional competencies, motivational characteristics and volitional qualities during independent work in the conditions of professional activity.

Online courses on the Coursera, edX, Udemy, Prometheus, Future Learn platforms; task and project planning tools Google Calendar, Trello, Asana; social networks for professional networking LinkedIn, Facebook, Instagram; online forums, discussions, personal websites, blogs, etc. will be useful for teachers' independent work.

Each component of a mathematics teacher's professional culture is a prerequisite for success in teaching. However, not every mathematician who has succeeded in his or her research is a good teacher even in the field of his or her narrow specialty [20], just as a specialist in methodological or pedagogical research can become a highly qualified mathematics teacher.

The formation of the mathematical culture of a mathematics teacher begins in the master's program, where the corresponding additional specialty is provided. And the discipline "Methods of teaching mathematics in higher education" sets one of the tasks to familiarize master's students with the peculiarities of the methodology of teaching mathematics courses in higher education institutions of different levels and professional orientation; to reveal the goals, content, methods and means of studying individual mathematical disciplines.

Let us consider the example of the topic "Methods of teaching functions in mathematical analysis. The Limit of a Function", how these concepts are formed, starting from the function of one variable, two and many variables to the concept of mapping, operator and functional in a metric space. We use the methods of analogy, comparison, highlighting the main and differences in the definition of this concept in different metric spaces.

If we use the method of analogy and generalization, we see that in all cases of j-value functions of one variable, many variables, and mapping to metric space, keywords are preserved: "one for each".

If you use the comparison method, the objects of study change. In some cases, these are numbers, in others, points of the Euclidean space, and in metric spaces, points can be functions, numerical sequences, and other objects. We point out that objects of different nature can appear in the definition domain and the value domain:

Function of single variable (Figure 2).

$$\text{for each } x \in X \subset \mathbb{R} \xrightarrow{f} \text{ single } y \in Y \subset \mathbb{R}.$$

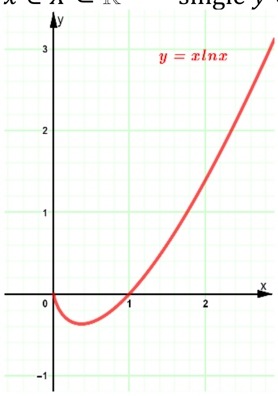


Figure 2: Graph of a single variable function (built in MathCha).

Function of many variables (Figure 3).

$$\text{every point of } x = (x_1, x_2, \dots, x_n) \in E \subset \mathbb{R}^n \xrightarrow{function} \text{ single number } y \in \mathbb{R}.$$

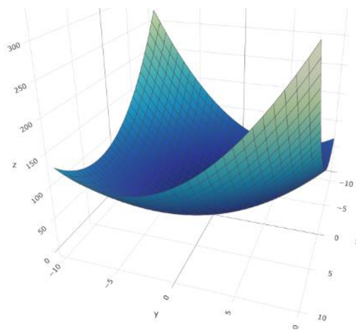


Figure 3: Graph of the function of two variables (built in MathCha).

Display in metric space. A mapping f, which corresponds to each point of the metric space X to one point of the metric space Y, is called a mapping from the metric space X to the metric space Y.

Similarly, to introduce the concept of the limit of a function, we can also use the method of analogy, generalization, and comparison, to show that the Heine definition, regardless of the nature of objects, is formally written in the same way:

$$\left(A = \lim_{x \to x_0} f(x)\right) \overset{df}{\Leftrightarrow} \left(\forall (x_n) \subset O^*(x_0) \cap X: \lim_{n\to\infty} x_n = x_0 \Rightarrow \lim_{n\to\infty} f(x_n) = A\right).$$

The definition of the boundary of a function at a Cauchy point depends significantly on how the distance is denoted in a particular metric space.

The limit of a function of a single variable:

$$\left(A = \lim_{x \to x_0} f(x)\right) \overset{df}{\Leftrightarrow} (\forall \varepsilon > 0 \exists \delta > 0, \forall x \in X: 0 < |x - x_0| < \delta \Rightarrow |f(x) - A| < \varepsilon).$$

In the definition of the boundary of a function of many variables, the points $x_0 = (x_{01}, x_{02}, \dots, x_{0n})$ of the Euclidean space $\mathbb{R}^n$ and numbers appear, so we pay attention to the distance notation:

$$\left(A = \lim_{\substack{x \to x_0 \\ x \in E}} f(x)\right) \overset{df}{\Leftrightarrow} (\forall \varepsilon > 0 \exists \delta > 0, \forall x \in E: 0 < d(x, x_0) < \delta \Rightarrow |f(x) - A| < \varepsilon).$$

Here it is important to realize that the point M_0 can be approached in different ways (in fact, there are many such ways), and this will significantly affect the further construction of the theory of functions of many variables, in particular, its differentiability.

To define a mapping boundary in the metric spaces X and Y, we take into account that objects can have different natures, so we denote their distances by $d_1(x, x_0)$ and $d_2(f(x), y_0)$, respectively. Thus, we conclude that in metric spaces, the construction of the theory of mapping boundaries depends on the nature of the objects and there may be interesting surprises.

A mathematics teacher must have the appropriate competencies to explain the fundamental concepts of “function” and “limit of function” to higher education students in a methodologically competent manner.

4 DISCUSSION

The authors of the article conducted a pedagogical study in higher education institutions of Ukraine on the peculiarities of forming the professional culture of a mathematics teacher and a future mathematics teacher. Respondents ranked the components of professional culture at the level of their structures using Google Form. The results of the primary data and intermediate calculations based on the survey of mathematics teachers and future mathematics teachers are presented in Table 1.

Table 1. Table of primary data and intermediate calculations based on the results of the survey of mathematics teachers (X_i) and future mathematics teachers (Y_i).

n	Component structure	X_i	Y_i	d_i	d_i^2
1	Mathematical competencies	2	2	0	0
2	Mathematical thinking	1	1	0	0
3	Mathematical language	3	3	0	0
4	Scientific research	9	7	2	4
5	Teaching and learning methods	4	5	-1	1
6	Educational and methodological product	8	9	-1	1
7	Self-analyze, professional development	7	4	3	9
8	Self-education, self-development, self-management, pedagogical skills	6	6	0	0
9	Motivation, communication, professional morality, reflection, adaptability to the professional environment	5	8	-3	9
	Sums			0	24

In order to establish a statistical relationship between the indicators of the significance of the components of professional culture at the level of their structures in mathematics teachers (48 people) and future mathematics teachers (53 people), Spearman's rank correlation coefficient r_s was used, which is determined by the formula: $r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$, where d_i is the rank difference; n is the number of features (Table 1) [30, pp. 59-63]. The study was conducted in higher education institutions of Ukraine with maximum respect for the privacy and confidentiality of the subjects, in compliance with the ethical norms and standards of the American Psychological Association (APA), the recommendations of the Ethical Code of the Scientist of Ukraine.

The primary data on the two parameters are presented in an order scale. The ranks are not repeated, so the empirical value of Spearman's r_s correlation coefficient will be determined by the general formula. We set the probability level of the first kind of error and formulate the null and alternative hypotheses for $\alpha = 0.01$:

- H_0: the indicators of the importance of professional culture components at the level of their structures for mathematics teachers and future mathematics teachers are not interrelated ($r_s = 0$);
- H_1: the indicators of the importance of professional culture components at the level of their structures for mathematics teachers and future mathematics teachers are interrelated ($r_s \neq 0$).

Calculate d_i – rank differences by the formula: $d_i = R_{X_i} - R_{Y_i}$; d_i^2 – squares of rank differences; sums of d_i^2 for the sample of subjects.

We calculate the empirical value of the Spearman's rank correlation coefficient using the formula r_s:

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{6 \cdot 24}{9 \cdot (9^2 - 1)} = 0.8.$$

This value indicates a direct correlation of the average force between the variables. Let's check its statistical significance.

According to the table of critical values of Spearman's correlation coefficient r_s [30, p. 130] for $df = n = 9$ and a given $\alpha = 0.01$, we find $r_{\text{crit}} = 0.798$. Since $|r_{\text{emp}}| > r_{\text{crit}}$ $(0.8 > 0.798)$, the hypothesis H_0 is rejected. The size of the standardized effect according to J. Cohen's classification is large ($r_s = 0.8$).

Thus, there is a statistically significant relationship between the indicators of the significance of the components of professional culture at the level of their structures among mathematics teachers. According to the statistics of Spearman's r_s criterion ($r_s = 0.8$; $p > 0.01$; $n = 9$), there are no grounds for accepting the null hypothesis.

There are common points of view on the ranking of components of professional culture of teachers and future teachers of mathematics. The significance of the indicators coincides with the ranks 1 (Mathematical thinking), 2 (Mathematical competencies), 3 (Mathematical language), 6 (Self-education, self-development, self-management, pedagogical skills). The significance of the indicators “Teaching and learning methods” and “Educational and methodological product” has a minimal difference (difference), represented by ranks 4-5 and 8-9, respectively.

The results of the study have shown that the indicators of the significance of the components of professional culture at the level of their structures in mathematics teachers and future mathematics teachers are interrelated. The statistically significant reliability of the study was confirmed by the Spearman's rank correlation coefficient r_s, performed in accordance with the methodological recommendations of J. Cohen.

5 CONCLUSIONS

The authors of the article have carried out a definitional analysis of the key concepts of the study. It is shown that the professional culture of a mathematics teacher combines three main components: mathematical, scientific, methodological and pedagogical cultures, which, in turn, are complex integrative formations and are closely interrelated. The formation of professional culture is possible only with the use of digital tools in all its components. We propose to use the wheel of digital technologies modeled in the article, which identifies the main directions of implementation of modern tools in the professional activity of the teacher.

By the professional culture of a mathematics teacher, we mean an integrated dynamic property of a personality that projects his or her general culture in the field of the profession, is a synthesis of mathematical, scientific, methodological and pedagogical cultures, and is realized in a synergistic educational space using digital technologies. Digital technologies have become an integral part of the modern world, and education is no exception. Their integration into all components of the professional culture of a mathematics teacher is fully justified and necessary. The conducted pedagogical research in higher education institutions of Ukraine has shown that the indicators of the significance of the components of professional culture at the level of their structures in mathematics teachers and future mathematics teachers are interrelated. To statistically test the hypothesis, the study used Spearman's rank correlation coefficient r_s. Interpretation of the size of the effect of correlation dependence was performed in accordance with the methodological recommendations of J. Cohen.

Prospects for further research are seen in a detailed analysis of the structural components of the professional culture of a mathematics teacher, the definition of local and global goals, objectives and methods of their implementation.

REFERENCES

[1] Y. Li and A. Z. Khairani, "Assessing factors influencing the formation of preservice teacher professional identity: Can commitment be related to self-image and task perception as a mediator?," Teaching and Teacher Education, vol. 160, p. 105021, 2025, [Online]. Available: https://doi.org/10.1016/j.tate.2025.105021.

[2] W. Yue, L. Yan, and Y. Yang, "How exactly do teachers' identities develop in the study travel? - A grounded theory study from China," Teaching and Teacher Education, vol. 144, p. 104605, 2024, [Online]. Available: https://doi.org/10.1016/j.tate.2024.104605.

[3] P. Mongkolhutthi, "Exploring 30 years of teacher work culture research (1993-2023): A bibliometric analysis," Review of Education, vol. 13, p. e70057, 2025, [Online]. Available: https://doi.org/10.1002/rev3.70057.

[4] C. Hongboontri and Y. Liao, "Teacher cultures: In search of their contents and forms," LEARN Journal: Language Education and Acquisition Research Network, vol. 14, no. 2, pp. 529-573, 2021, [Online]. Available: https://so04.tci-thaijo.org/index.php/LEARN/article/view/253280.

[5] C. Meredith, N. Moolenaar, C. Struyve, M. Vandecandelaere, S. Gielen, and E. Kyndt, "The importance of a collaborative culture for teachers' job satisfaction and affective commitment," European Journal of Psychology of Education, vol. 38, no. 1, pp. 43-62, 2023, [Online]. Available: https://doi.org/10.1007/s10212-022-00598-w.

[6] W. M. Smith, R. Funk, K. Quaisley, and E. McMillon, "Building a culture that empowers and motivates mathematics teacher leaders as change agents," School Science and Mathematics, pp. 1-13, 2025, [Online]. Available: https://doi.org/10.1111/ssm.18329.

[7] T. Tatik, H. T. Mai Nguyen, and T. Loughland, "Exploring factors influencing Indonesian novice teachers' achievement of teacher professional standards during induction: A cultural-historical activity theory perspective," Teaching and Teacher Education, vol. 159, p. 105013, 2025, [Online]. Available: https://doi.org/10.1016/j.tate.2025.105013.

[8] A. Acosta and M. L. Belmonte, "Attitudes of future pedagogues towards educational innovation in the university context," Research in Education and Learning Innovation Archives, vol. 34, pp. 91-104, 2025, [Online]. Available: https://doi.org/10.7203/realia.34.29172.

[9] L. V. Huberskyi, V. P. Andrushchenko, and M. I. Mykhalchenko, Culture. Ideology. Personality: methodological and ideological analysis. Kyiv: Znannia Ukrainy, 2002.

[10] S. U. Honcharenko, Ukrainian Pedagogical Dictionary: A Textbook. Kyiv: Lybid, 1997.

[11] I. A. Ziaziun, L. V. Kramushchenko, and I. F. Kryvonos, Pedagogical excellence: textbook. Kyiv: Vyshcha shkola, 2004.

[12] I. Glazkova, S. Khatuntseva, and L. Yaroshchuk, "Professional Pedagogical Culture: Historical Culturological Aspect," Revista Romaneasca Pentru Educatie Multidimensionala, vol. 12, no. 3, pp. 144-161, 2020, [Online]. Available: https://doi.org/10.18662/rrem/12.3/314.

[13] L. Kravchenko, R. Vynnychuk, O. Ilchenko, A. Lytvynenko, Ye. Kulyk, O. Palekha, and N. Sukhovii, "Professional culture of future specialists within the framework of ukraine's sustainable development concept," Brazilian Journal of Education, Technology and Society (BRAJETS), vol. 17, no. 1, pp. 42-53, 2024, [Online]. Available: http://dx.doi.org/10.14571/brajets.v17.nse1.42-53.

[14] L. Pankiv, T. Bodrova, N. Borshchenko, M. Chabaiovska, and N. Ihnatenko, "Theoretical bases of professional culture of future specialists in accordance with global challenges of the information society," Amazonia Investiga, vol. 12, no. 61, pp. 17-25, 2023, [Online]. Available: https://doi.org/10.34069/AI/2023.61.01.2.

[15] V. Halchenko, T. Skoryk, I. Bartienieva, O. Nozdrova, T. Shtainer, and T. Snyatkova, "The Technology of Forming the Professional Culture of Future Teachers: from Reflection to Creativity," Revista Romaneasca Pentru Educatie Multidimensionala, vol. 14, no. 4 Sup. 1, pp. 36-57, 2022, [Online]. Available: https://doi.org/10.18662/rrem/14.4Sup1/658.

[16] O. M. Voznyuk, "Definitive device research of the formation of professional culture of future specialists," Science and education, pp. 47-50, 2013.

[17] N. Mordovtseva, O. Slobodian, I. Hryhorieva, I. Uzhchenko, O. Tsalapova, and I. Makarenko, "Communication Culture as a Key Professional Skill," Revista Romaneasca Pentru Educatie Multidimensionala, vol. 15, no. 3, pp. 55-71, 2023, [Online]. Available: https://doi.org/10.18662/rrem/15.3/753.

[18] R. Shulyhina, M. Kochubei, V. Lashkul, Y. Malyk, and S. Titarenko, "Rhetorical and virtual communication: innovative characteristics of future professionals' communication culture," Amazonia Investiga, vol. 13, no. 75, pp. 32-44, 2024, [Online]. Available: https://doi.org/10.34069/AI/2024.75.03.3.

[19] I. T. Zotova and O. O. Kaidanovska, "Professional pedagogical culture of the teacher as a condition for the successful preparation of a future specialist," Science and education, vol. 3, pp. 71-75, 2013.

[20] H. O. Mykhalin, Professional training of a mathematics teacher in the process of teaching mathematical analysis: monograph. Kyiv: NPU imeni M. P. Drahomanova, 2003.

[21] Ye. O. Lodatko, Mathematical culture of primary school teacher as a basis of professional worldview. Sumy: VVP «Mriia» TOV, 2012.

[22] H. Zinchenko, "Mathematical Culture as an Innovative Component of the Future Mathematics Teacher's Professional Competence," Pedahohichni nauky, vol. 64, pp. 88-95, 2015.

[23] M. Kovtoniuk, "Formation of mathematical culture of bachelors of mathematics in the conditions of mixed form of education," Innovatsiina pedahohika XXI stolittia: novi kompetentnosti vykladacha zakladu vyshchoi osvity, Vinnytsia: VDPU, pp. 52-56, 2024, [Online]. Available: https://doi.org/10.31652/3041-1211-2024-52-56.

[24] H. B. Williams, "Mathematics and the Biological Sciences," Bulletin of the American Mathematical Society, vol. 33, no. 3, 1927.

[25] "The concept of a new Ukrainian school," 2016, [Online]. Available: https://mon.gov.ua/static-objects/mon/sites/1/zagalna%20serednya/nova-ukrainska-shkola-compressed.pdf.

[26] O. M. Maksymovych, "Modern aspects of mathematical thinking of a personality," Zbirnyk naukovykh prats: filosofiia, sotsiolohiia, psykholohiia, Ivano-Frankivsk, vol. 5, pp. 125-129, 2000.

[27] M. M. Kovtoniuk, O. P. Kosovets, O. M. Soia, and I. M. Leonova, "Architecture of digital technologies in the educational environment of the teacher as a transfer of innovations to the economic space of the state," Suchasni informatsiini tekhnolohii ta innovatsiini metodyky navchannia u pidhotovtsi fakhivtsiv: metodolohiia, teoriia, dosvid, problemy, Vinnytsia: TOV «Druk plius», vol. 68, pp. 93-106, 2023, [Online]. Available: https://doi.org/10.31652/2412-1142-2023-68-93-106.

[28] O. Kosovets, M. Kovtoniuk, O. Soia, and D. Koval, "Integration of digital technologies in modeling the educational environment of a bachelor in the conditions of martial law," Proceedings of International Conference on Applied Innovation in IT, vol. 12, no. 1, pp. 51-57, 2024, [Online]. Available: https://doi.org/10.25673/115641/.

[29] I. Barbashova, N. Bakhmat, I. Marynchenko, M. Ponomarova, and T. Holinska, "Neurotechnologies and artificial intelligence in forming the professional culture of pedagogical field specialists," Ad Alta: Journal of interdisciplinary research, vol. 13, no. 2, pp. 74-81, 2023, [Online]. Available: http://surl.li/ldodxb.

[30] V. F. Bosniuk, Mathematical methods in psychology: a course of lectures: A Multimedia training edition. Kharkiv: NUTsZU, 2020.

SECTION 2

DATA ANALYSIS AND PROCESSING

Proceedings of the 13th International Conference on Applied Innovations in IT (ICAIIT)

APRIL 2025

Quantum Field Tensor Model of Telecommunication Network Objects Interaction Based on Lie Groups

Victor Tikhonov[1], Yevhen Vasiliu[1], Eduard Siemens[2], Oksana Vasylenko[2], Olena Tykhonova[1], Kateryna Shulakova[1,2] and Olexandr Demchenko[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*
{victor.tykhonov, olena.tykhonova}@suitt.edu.ua, {ye.vasiliu, katejojo29, lex7dem}@gmail.com, oksana.vasylenko@hs-anhalt.de

Keywords: Quantum Physics, Telecommunication Networks, Tensor Modelling, Lie Groups, Anisotropic Information Flows, Quantum Entanglement, Network Performance, Flow Asymmetry, Network Simulation.

Abstract: This research explores the application of quantum physics methodologies to analyse digital flows within telecommunication networks. This study introduces a novel tensor model, grounded in the SU(2) Lie group, designed to simulate symmetric and asymmetric information interactions between network objects within a two-dimensional Euclidean complex space. The proposed model innovatively decomposes the tensor into three fundamental components: metric, torsion, and curvature tensors. The metric and torsion tensors are combined to form a complex vector system, effectively representing the intrinsic interaction dynamics between network objects. The curvature tensor, on the other hand, models the potential asymmetry introduced by an external observer, simulating third-party influences on network interactions. This approach allows for the representation of closed-time cyclic experiments, such as evaluating interactions between network nodes, as a continuous tensor field on a quantized topological circle. This framework not only provides a comprehensive perspective on information processes in data transmission networks but also draws parallels with elementary particle interactions in quantum physics. Furthermore, the research includes a statistical analysis using simulations in the NS3 environment, validating the model's effectiveness in identifying key characteristics of information flows. The analysis demonstrates the model's ability to detect and quantify the impact of external observers, the effects of traffic asymmetry, and changes in network dynamics through quantum entanglement entropy. The potential practical applications of this model, including network performance analysis, security enhancement, and routing optimization, are also discussed, highlighting its relevance to both theoretical and applied aspects of telecommunications and quantum physics.

1 INTRODUCTION

The world of physics and telecommunications might seem miles apart, but there's a fascinating bridge between them: the use of mathematical tools like tensors and Lie groups. These tools, originally developed to understand the symmetries and interactions of particles [1 - 8], are now finding new life in describing the complex flow of information in networks.

Think of it like this: just as particles interact and transform in predictable ways, so too does the data flowing through our networks. Researchers have been using tensor models to represent these interactions, drawing parallels between electrical circuits and data flows [9 - 12, 13 - 15]. They've even started to explore ideas like "cybernetic conductivity" and "information flow impedance", mirroring concepts from electrical engineering [15].

This approach has been particularly useful in understanding heavily loaded mobile networks and prioritizing data packets [9, 15]. More recently, researchers have been applying these techniques to wireless communications [16, 17] and even exploring the concept of "entanglement entropy" in network data [18, 19, 20].

However, there's a whole other level of physics that hasn't been fully tapped into yet: the world of quantum mechanics. This is where things get really interesting, with concepts like symmetry groups and the quirky behaviour of particles at the subatomic level [21, 22].

2 MOTIVATION AND OBJECTIVES OF THE WORK

The application of quantum mechanics to telecommunications is significant, as the mathematical frameworks used to describe quantum phenomena, such as Lie groups, provide a powerful approach for modelling the complex relationships between network components, including routers and switches.

Imagine being able to predict and control the flow of information with the same precision that physicists predict the behaviour of particles. This could lead to more efficient, robust, and secure networks.

Recent research indicates that quantum concepts like entanglement may enhance our understanding of network information flow [23]. This work builds upon these findings, aiming to develop a quantum field tensor model of anisotropic network object interactions, using Lie groups SU(2)/U(1), to explore how advanced quantum mechanics tools can improve telecommunication network modelling.

To achieve this, the following objectives are set.

1) Substantiation the model of anisotropic network relationships as a complex tensor function of discrete time for objects binary interaction traced by an external viewer.
2) Decomposition the complex tensor of anisotropic network relationships on the three components (metric tensor, torsion tensor and curvature tensor).
3) Presentation the manifold of anisotropic network complex tensor using the SU(2) Lie group.
4) Construction a quantum field model of network objects interaction on the U(1) Lie group.

3 THE COMPLEX TENSOR MODEL OF ANISOTROPIC NETWORK RELATIONSHIPS

Let NW an open anisotropic telecommunication network depicted in Figure 1; A, B – arbitrary objects (e.g., IP routers) with regular (conditionally "strong") information relationships; V – an external observer tracing the objects A, B through "weak" interaction with minimal influence on them; $P_{\Delta t}(x_1, x_2)$ – an abstract power of network objects binary interaction defined for a time quantum Δt (e.g., the number of IP-packets sent from x_1 to x_2 within the time Δt).

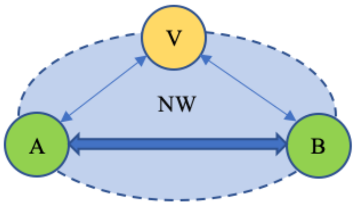


Figure 1: The graph on network objects (A, B) interaction by viewer tracing (V, A), (V, B).

Let us represent the graph (Fig.1) as a matrix function **P**(V, A, B) of ternary relationships between V, A and B, where each pair $\{P_{\Delta t}(x_1, x_2), P_{\Delta t}(x_2, x_1)\}$ reflects the correspondent binary relationships in an anisotropic network NW. Figure 2 shows an example of such a matrix P in integers.

P	V	A	B
V	0	1	6
A	3	0	10
B	2	4	0

Q	A		B	
A	3	1	10	
B	4		2	6

Figure 2: The matrix functions of objects (A, B) interaction by viewer tracing (V, A), (V, B).

The interaction data matrix P can be transformed, through a bijective mapping, into a specialized square matrix Q, characterized by a double diagonal structure, as illustrated in Fig. 2. This transformation preserves the information content, as both matrices, P and Q, possess 6 independent elements (specifically, the numerical values 1, 6, 3, 10, 2, and 4). This matrix Q, which we will refer to as the Matrix of External Observation (MEO), serves as a representation of the information interaction between two network objects, A and B, as perceived by an external observer V. This transformation simplifies subsequent mathematical operations and allows for a clearer visualization of the interaction dynamics.

Let us call the matrix **Q** in Fig. 2 the *matrix of external observation* (MEO) the information interaction of two arbitrary network objects A and B by the given viewer V.

Next, let us represent the matrix **Q** depicted in Figure 2 as the sum of symmetric matrix **R** and skew-symmetric matrix **S** as it is shown in Figure 3. It is clear, that partial matrices **R** and **S** has exactly 6 independent elements (here 2, 4, 7, +1, –2, ±3), and therefore, the mapping chain **P**→**Q**→(**R**+**S**) is strictly reversible (bijective).

Q	A	B
A	3 1	10
B	4	2 6

R	A	B
A	2	7
B	7	4

+

S	A	B
A	+1	+3
B	–3	–2

Figure 3: Decomposition the matrix of external observation **Q** on the symmetric and skew-symmetric parts.

Now, we finalize the above bijective mapping chain **P**→**Q**→(**R**+**S**) into equivalent presentation by the three parts **M**, **T**, **C** in Figure 4:

$$\mathbf{P} \rightarrow \mathbf{Q} \rightarrow (\mathbf{R} + \mathbf{S}) \rightarrow (\mathbf{M} + \mathbf{T} + \mathbf{C}) \qquad (1)$$

The algorithm of the ultimate transformation is the following. Each diagonal element of matrix **M** is the half-sum of correspondent rows and columns elements of matrix **R**, e.g., **M**(A,A) = 2+7=9; **M**(B,B) 7+4=11; each non-diagonal element of **M** equals to that in **R**. Matrix **T** takes the non-diagonal elements of S with zero-diagonal ones; instead, **C** obtains the diagonal elements of **S** with zero non-diagonal ones.

R	A	B
A	2	7
B	7	4

+

S	A	B
A	+1	+3
B	–3	–2

M	A	B
A	9	7
B	7	11

+

T	A	B
A	0	+3
B	–3	0

+

C	A	B
A	+1	0
B	0	–2

Figure 4: Presentation the matrix of external observation **Q** in tensor form.

It is easy to show, that due to its construction, the matrix **M** in Fig. 4 satisfies the Riemann metric tensor requirements for 2-dimensional Euclidian vector space. Matrix T is a torsion tensor of anisotropy in objects relationships [24]. Matrix **C** is curvature tensor of viewer-objects relation asymmetry.

Let us assemble two real matrices **M**, **T** (Fig. 4) into one Hermitian complex matrix **H** of metric tensor with torsion but without a curvature:

$$\mathbf{H} = \mathbf{M} + \mathrm{i} \cdot \mathbf{T}\,. \qquad (2)$$

Tensor H in (2) corresponds to a system of two vectors in a complex Euclidean space, in which vectors lengths are positive real numbers (diagonal elements of **H**). Instead, the Hermitian matrix C in Fig. 4 can have both positive and negative real diagonal elements, and therefore, can't be presented in a complex Euclidean or pseudo-Euclidean vector space.

In general, the external observation matrix Q in (2) can be represented as a combination of metric, torsion and curvature in a 2-dimensional non-Euclidean complex space:

$$\mathbf{Q} \rightarrow \{\mathbf{H}, \mathbf{C}\}. \qquad (3)$$

It is clear, that all three components **M**, **T**, **C** in Fig. 4 are always real or complex Hermitian (2×2) matrices. For unified representation of such matrices let's use the special unitary Lie group SU(2), that includes three unitary traceless Pauli matrices σ_1, σ_2, σ_3, (having determinant "–1")

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix};\ \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix};\ \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad (4)$$

along with Euclidian matrix σ_0 having determinant "+1" and trace "+2" (identity matrix I):

$$\sigma_0 = E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I \quad . \qquad (5)$$

Matrices $\sigma_0 \div \sigma_3$ form a basis on the set of all 2×2 Hermitian matrices, i.e. any one of them can be given as a linear combination L($\sigma_0, \sigma_1, \sigma_2, \sigma_3$). Matrices σ_1, σ_2, σ_3 are SU(2)-generators; each one defines a particular symmetry type and related subset on continuous transformations (phase rotations).

Suppose that a third-party viewer observation of network objects interaction is a cyclically closed process in time. The topological equivalent of a closed time is a circle O^1, and Cartesian product O^n is the only topological space to construct everywhere continuous vector/tensor field. The circle O^1 can be mapped onto the unitary Lie group U(1). Let

$$\mathrm{F} = \mathrm{L}(\sigma_0, \sigma_1, \sigma_2, \sigma_3) \times \mathrm{U}(1)\,. \qquad (6)$$

The product F in (6) is the set of all the non-singular (2×2) Hermitian matrices defined on the time-circle. It can be used as a quantum field tensor model of discrete network object's binary interaction. The set of 2-dimensional vector spaces L() in (6) can be studied using the Lie group SU(2).

4 STATISTICAL ANALYSIS OF NETWORK OBJECT INTERACTIONS

4.1 Methodology of Statistical Analysis

To validate the effectiveness of the proposed quantum field tensor model for network object interactions, a statistical analysis of key characteristics of information flows was conducted. The study focused on three main aspects: Network performance analysis: average packet delay, packet loss, and bandwidth; Identification of information flow asymmetry between nodes; Calculation of quantum entanglement entropy in information flows.

Data were collected through simulations in the NS3 environment, modeling the tensor dynamics of flows in the network. NS3 (Network Simulator 3) is an advanced discrete-event network simulation tool widely used in research and development for studying communication protocols and network performance. It enables the modeling of various network layers, including physical, link, and transport, making it an ideal tool for analyzing the interaction dynamics of network objects. NS3 supports packet-level tracing, statistical data collection, and realistic network behavior emulation, providing reliable insights into system performance [9]. Three scenarios were tested:

1) Symmetric interaction between nodes without external interference.
2) Asymmetric interaction, where one node receives significantly more traffic.
3) Impact of an external observer, modifying the tensor model metrics.

4.2 Research Results

4.2.1 Network Performance Analysis

Table 1 presents the measurement results for packet delay and loss across the three scenarios.

These values align with previous studies on network traffic modelling, which demonstrated similar packet loss patterns and delay fluctuations under asymmetric traffic loads [13].

4.2.2 Analysis of Flow Asymmetry

The study of information flow asymmetry revealed that, in the presence of an external observer, the asymmetry coefficient (calculated as the ratio of inbound to outbound traffic) increased by 1.8 times compared to the standard network. This is consistent with recent findings in tensor-based traffic modeling, which highlight significant distortions in data transmission due to asymmetric node interactions [16].

Table 1: Network performance metrics under different interaction scenarios.

Parameter	Symmetric network	Asymmetric network	Observer influence
Average delay (ms)	12.4	18.7	25.3
Delay variance (ms^2)	2.1	5.3	8.7
Packet loss (%)	0.5	2.3	4.8
Bandwidth (Mbps)	96.5	85.2	72.4

4.2.3 Quantum Entanglement Entropy

The entanglement entropy values obtained for different network interaction scenarios are presented in Table 2.

Table 2: Quantum entanglement entropy in different interaction scenarios.

Scenario	Quantum entropy
Symmetric interaction	0.72
Asymmetric interaction	1.34
Observer influence	2.01

These results align with recent studies on network entanglement entropy, where similar entropy growth patterns were observed in complex telecommunication networks with external interference [18].

The conducted analysis confirmed that the tensor model based on Lie groups effectively identifies key characteristics of information flows in telecommunication networks. The key conclusions are:

- The influence of an external observer significantly increases delay and packet loss [15]. Traffic asymmetry substantially alters the topology of information interaction [16].
- Quantum entanglement entropy can be used to detect changes in network dynamics [24].

These results can be applied to improve network performance and security, particularly in quantum communications and adaptive routing systems. The findings also suggest that tensor models could be further extended for optimizing real-world telecommunications infrastructure [21].

5 PRACTICAL APPLICATIONS (INTEGRATED)

The quantum field tensor model proposed in this paper has several potential practical applications within telecommunications networks.

Firstly, it can be utilized for network performance analysis. Tensor analysis allows for the identification of complex correlations within network traffic, indicating bottlenecks or inefficiencies. For example, the metric tensor can evaluate delays and packet loss, while the torsion tensor can analyze traffic asymmetry.

Secondly, it can enhance network security. The quantum field approach detects anomalies suggesting malicious activity. Sudden changes in the curvature tensor might signify a DDoS attack or intrusion.

Thirdly, it can optimize network routing and resource allocation. Tensor analysis discovers optimal routes and allocates resources efficiently. For example, it enables dynamic routing based on network load.

Fourthly, it can contribute to quantum communication protocol development. The quantum field approach describes complex quantum effects, enhancing security and efficiency.

To realize these applications, further research is needed, including algorithm development for tensor data analysis and real-world network simulations and experiments.

6 DISCUSSION

This work advances the application of quantum field theory to telecommunication networks by introducing a tensor model based on Lie groups SU(2)/U(1). This model aims to simulate the nuanced interactions between network objects under the influence of an external observer, drawing parallels with fundamental interactions in quantum physics.

The current landscape of networking technologies exhibits a growing interest in tensor modeling for analyzing information flows, reflecting a trend seen in quantum physics where symmetry and entanglement entropy are pivotal. The ability to represent internal network asymmetries using a torsion tensor within a Euclidean complex space, as demonstrated in Section 4, highlights the potential of this approach.

However, the challenge of integrating internal network dynamics with external observer influences, represented by the curvature tensor, remains a significant hurdle. This mirrors the complexities of unifying gravity with other fundamental forces in quantum physics, suggesting a deep commonality in the underlying mathematical structures.

Addressing this challenge may necessitate the development of a vector space capable of accommodating both positive and negative vector lengths, requiring a refinement of the classical scalar product concept. Additionally, a shift towards open system models, which account for external environmental impacts, might offer a more comprehensive framework for analyzing network behaviors.

7 CONCLUSIONS

This research's primary achievement is the development of a quantum field tensor model for anisotropic telecommunications networks, leveraging Lie groups SU(2)/U(1). This model pioneers a novel approach to simulating binary object interactions under weak third-party observation, effectively bridging the gap between information processes in networks and elementary particle interactions in quantum physics.

The statistical analysis conducted in this study has validated the effectiveness of the proposed tensor model in identifying key characteristics of information flows within telecommunications networks. Specifically, we have demonstrated that:

- The influence of an external observer significantly increases packet delay and loss, confirming the sensitivity of network dynamics to external factors.
- Traffic asymmetry notably alters the topology of information interactions, demonstrating the model's capability to reflect anisotropic relationships.
- Quantum entanglement entropy can be utilized to detect changes in network dynamics, opening prospects for applying quantum concepts in network monitoring and management.

These results have significant implications for enhancing network performance and security, particularly in the context of quantum communications and adaptive routing systems. The findings also suggest that tensor models can be further expanded to optimize real-world telecommunications infrastructure.

While the model represents a substantial theoretical advancement, it also points to areas requiring further exploration. The integration of internal and external network anisotropies, along with the development of suitable vector spaces and open system models, are crucial for enhancing the model's applicability.

Future research will focus on unifying the model of open information networks with physical quantum systems based on group theory. Our aim is to refine the model's capabilities and explore its practical implications in real-world scenarios, with a particular emphasis on developing algorithms for real world implementation.

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. Lewis and G. Wood, "Quantum Technology: Applications and Implications," CSIS report, Washington, USA, 2023, pp. 1-11. [Online]. Available: https://csis-website-prod.s3.amazonaws.com/s3fs-public/2023-05/230526_Lewis_QuantumTechnology.pdf?VersionId=iCOWm7k02Ms846I0Eb5DLeyD6dZN8K5F

[2] D. Valdano, "Why is Symmetry So Important in Physics?," 2016. [Online]. Available: https://medium.com/physics-as-a-foreign-language/why-is-symmetry-so-important-in-physics-1a7ac304770.

[3] J. Albert, "Group Theory in Particle Physics," 2007. [Online]. Available: https://webhome.phy.duke.edu/~jba10/grouptheoryreport.pdf.

[4] K. Landskron, "Matrix Representations of Symmetry Operations," in Inorganic Coordination Chemistry, 2024. [Online]. Available: https://chem.libretexts.org/Bookshelves/Inorganic_Chemistry/Inorganic_Coordination_Chemistry_(Landskron)/02%3A_Symmetry_and_Group_Theory/2.03%3A_Properties_and_Representations_of_Groups.

[5] M. Lloyd, "Matrices Lie: An introduction to matrix Lie groups and matrix Lie algebras," [Online]. Available: https://www.whitman.edu/documents/academics/majors/mathematics/2015/Final%20Project%20-%20Lloyd.pdf.

[6] A. Einstein, "Relativity. The Special and General Theory," Elegant Ebooks, 1920, pp. 1-165. [Online]. Available: https://www.ibiblio.org/ebooks/Einstein/Einstein_Relativity.pdf.

[7] R. J. Petti, "Derivation of Einstein–Cartan theory from general relativity," arXiv:1301.1588v11, 2014.

[8] G. Lisi, "An Exceptionally Simple Theory of Everything," 2007. [Online]. Available: https://www.researchgate.net/publication/2217412_An_Exceptionally_Simple_Theory_of_Everything.

[9] A. V. Lemeshko et al., "Research on tensor model of multipath routing in telecommunication network with support of service quality by greate number of indices," Telecommunications and Radio Engineering, vol. 73, no. 15, pp. 1339-1360, 2014. [Online]. Available: https://www.dl.begellhouse.com/journals/0632a9d54950b268,6044cfe37be7d325,39948ca75239d78d.html.

[10] C. Fernandes et al., "Applications of Tensor Models in Wireless Communications and Mobile Computing," 2020. [Online]. Available: https://www.researchgate.net/publication/339408331_Applications_of_Tensor_Models_in_Wireless_Communications_and_Mobile_Computing.

[11] M. Yevdokymenko, "Enhancement of the adaptive routing tensor model in the infocommunication network with providing quality of experience by the R-factor," Technology audit and production reserves, vol. 3, no. 2(53), 2020. [Online]. Available: https://journals.uran.ua/tarp/article/view/206031.

[12] C. A. Pacca et al., "Advanced Signal Modeling and Processing for Wireless Communications using Tensors," Preprint, 2025. [Online]. Available: https://www.researchgate.net/publication/388556719_Advanced_Signal_Modeling_and_Processing_for_Wireless_Communications_using_Tensors.

[13] A. E. Petrov, "Tensor Network Model of Logistics System," Modelling and Data Analysis, vol. 12, no. 3, pp. 5-24, 2022.

[14] G. Kron, "Tensors for circuits," Dover Publ., N.Y., 1959, pp. 1-250. [Online]. Available: https://archive.org/details/in.ernet.dli.2015.198349.

[15] I. I. Pasechnikov, "Metodologia analiza i sinteza predelno nagruzenyh informacionnyh setey," 2004, pp. 1-250.

[16] V. I. Tikhonov, "Postroenie tensornoy modeli asimmetrichnyh zifrovyh potokov v complexnom prostranstve," Problemy telecomunicaciy, no. 2 (4), pp. 42-53, 2011.

[17] Y. Lin et al., "Tensor-Based Algebraic Channel Estimation for Hybrid IRS-Assisted MIMO-OFDM," IEEE Transactions on Wireless Communications, 2021. [Online]. Available: https://www.researchgate.net/publication/348896986_Tensor-Based_Algebraic_Channel_Estimation_for_Hybrid_IRS-Assisted_MIMO-OFDM.

[18] K. Fitter et al., "A Max-Flow approach to Random Tensor Networks," Quantum Physics, 2024, arXiv:2407.02559 [quant-ph].

[19] R. Watanabe and H. Ueda, "Automatic Structural Search of Tensor Network States including Entanglement Renormalization," Quantum Physics, 2024, arXiv:2405.06534 [quant-ph].

[20] C. H. Wu and C. C. Yen, "The expressivity of classical and quantum neural networks on entanglement entropy," Eur. Phys. J., vol. 84, no. 192, 2024. [Online]. Available: https://link.springer.com/article/10.1140/epjc/s10052-024-12558-3.

[21] M. Razavy, "An Introduction to Inverse Problems in Physics," University of Alberta, Canada, 2020, pp. 1-388. [Online]. Available: https://www.worldscientific.com/worldscibooks/10.1142/11860#t=aboutBook.

[22] Deep Learning-Based C14 Pile-Up Identification in the JUNO Experiment. ACAT2024 Report of Contributions, pp. 1-244. [Online]. Available: https://indico.cern.ch/event/1330797/contributions/contributions.pdf.

[23] S. Kumar, "Entanglement Entropy, Negative Information and QED And QCD Correlations of Fields and Spaces," SSRN, 2024, pp. 1-10. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4815896.

[24] "Introducing Riemannian Geometry," pp. 91-139. [Online]. Available: https://www.damtp.cam.ac.uk/user/tong/gr/three.pdf.

A Novel Approach for Rapid Detection of Forest Degradation and Diseases Through Anomaly Analysis of Sentinel-2 Spectral Data

Sofiia Drozd[1,2], Nataliia Kussul[1,3] and Hanna Yailymova[1]
[1]*Department of Mathematical Modelling and Data Analysis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" 37, Prospect Beresteiskyi, 03056 Kyiv, Ukraine*
[2]*Department of Electrical Engineering, Mechanical Engineering and Engineering Management, Anhalt University of Applied Sciences, Bernburger Str. 57, 06366 Koethen, Germany*
[3]*Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA*
sofi.drozd.13@gmail.com, kussul@umd.edu, anna.yailymova@gmail.com

Keywords: Satellite Data, Anomaly Detection, Sentinel-2, Forest Degradation.

Abstract: Forest degradation is an ongoing global issue, with significant environmental impacts that necessitate efficient monitoring and management. This paper presents a simple yet effective method for detecting forest degradation using freely available Sentinel-2 satellite data and an anomaly detection approach. The aim of this study was to develop an accessible and reliable technique that could match the performance of more complex algorithms while using minimal computational resources. The research focused on spectral bands with 10-20 m resolution and vegetation indices (NDVI, NDMI, GCI, PSSRa) to analyze forest damage in the Harz region. The method involved identifying anomalies in the spectral data relative to randomly selected reference points from healthy forest areas, which were verified with high-resolution imagery from Google Earth Pro. The results demonstrated that specific Sentinel-2 bands, particularly B3 and B5, were the most informative for detecting damaged forests, while vegetation indices were less effective. By analyzing anomalies in these bands, we successfully tracked forest degradation from 2020 to 2024, revealing a significant increase in damage between 2020 and 2021, with a total of 68.1 thousand hectares of forest lost by 2024. The theoretical relevance of this study lies in the development of a cost-effective and straightforward method for forest monitoring, while the practical relevance is evident in its potential for large-scale forest management and conservation. This method provides an efficient tool for monitoring forest health with minimal data requirements and computational effort, offering a promising solution for forest managers and conservationists worldwide.

1 INTRODUCTION

Healthy forests maintain ecosystem balance, purify the air, and provide economically important resources for people. However, forests constantly face numerous threats, such as diseases, pests, illegal logging, and climate change, which jeopardize their existence. The forested area of the Harz Nature Park in Germany, for example, has decreased by over 47% from 2001 to 2023, largely due to global warming and poor forest management practices, according to Global Forest Watch estimates [1]. Therefore, it is crucial to regularly monitor forest health to identify issues promptly and take necessary actions for their preservation.

Traditional forest monitoring methods, including visual inspections, are limited and labor-intensive, often failing to detect problems in a timely manner, especially in remote or hard-to-reach forest areas. Forest rangers may miss signs of tree diseases or illegal logging, making it difficult to respond quickly to threats.

In contrast to traditional methods, Earth observation satellites are an effective tool for monitoring forests over large areas [2]. They allow for real-time collection of detailed information on vegetation health, soil moisture, surface temperature, and other parameters. Satellite data is successfully applied for monitoring the consequences of natural disasters [3, 4] and human-induced impacts [5], detecting damage in agricultural fields [6], etc. Satellite imagery can quickly identify changes in forest cover, assess tree health [4], detect illegal logging [7, 5], pest activities [8, 9], and identify other violations. Additionally, satellite imagery enables long-term analysis of forest dynamics [4, 10], helping to predict future issues and take timely measures.

Various methods are used to detect forest damage through satellites, including monitoring vegetation indices such [11, 12], time series analysis [13, 4, 10], and machine learning algorithms [14, 15], including convolutional neural networks or transformers [7, 16, 17]. While these methods are effective in detecting changes in forests, they often require large amounts of data and computational resources, and may need significant processing time.

In our study, we aim to develop a simple and rapid algorithm for accurately detecting forest damage and diseases with minimal data, time, and resource consumption, while maintaining high accuracy compared to deep learning methods. By using free Sentinel-2 satellite imagery, we will create an efficient tool for forest monitoring that allows for rapid responses to changes, even with limited resources and large areas.

2 DATA AND MATERIALS

2.1 Study Area

For our study, we selected the protected areas of the Harz Natural Park (Fig. 1), located in central Germany at the border of Thuringia and Saxony-Anhalt. Selected area covers 2,756 km², most of which is covered by forest. However, due to adverse weather conditions, including droughts, storms, and pest activity, the Harz forest has suffered significant damage in recent years [1].

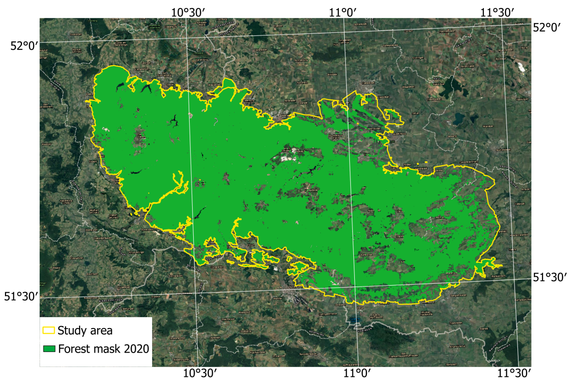


Figure 1: Study area.

To delineate the area of interest, we utilized vector data of Germany's protected areas provided by Protected Planet [18].

To distinguish forested areas from other land cover types, we employed the 2020 forest map with a spatial resolution of 10 meters, created based on Sentinel-2 imagery with support from the Joint Research Centre and the European Commission [19].

Consequently, 2020 was selected as the starting point for our study, focusing on identifying areas that have experienced deforestation or contain weakened and diseased forests since that time.

2.2 Satellite Data Used

To detect forest damage and diseases (unnatural vegetation decline), we used free Sentinel-2 satellite harmonized data, available with an update frequency of approximately five days from 2017 to the present.

The Sentinel-2 satellite has 13 spectral bands, including four bands at 10 meter resolution, six bands at 20 meter resolution, and three bands at 60 meter resolution. For our study, we will focus only on the bands with spatial resolutions of 10 and 20 m because bands with lower resolutions, such as those at 60 m, would not provide the level of detail necessary for our analysis of forest damage and disturbances.

For each year from 2020 to 2024, we selected a time series of images from the months of late spring and early summer (May–June), as trees are actively vegetating during this period. We masked clouds using Sentinel-2 Cloud Probability data with a cloudiness threshold of 20%. A composite was constructed from the cloud-free images, assigning each pixel the median reflectance value.

Additionally, as an extra source for validating the obtained results, we used open high-resolution imagery available for viewing in Google Earth Pro.

3 METHODOLOGY

To detect forest damage and diseases, we propose an anomaly detection method that identifies unusual or unexpected patterns in satellite imagery, deviating from typical forest conditions.

To assess the effectiveness of the proposed approach, we will analyze both individual Sentinel-2 spectral bands (Table 1) and their combinations — commonly used vegetation indices for forest monitoring [20] (Table 2). Specifically, the study will evaluate indices such as Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation Index (GNDVI), Normalized Difference Moisture Index (NDMI), Green Chlorophyll Index (GCI) and Plant Senescence Stress Reflectance Index (PSSRa), along with spectral bands with spatial resolutions of up to 20 m, which can provide insights into vegetation changes.

Table 1: Sentinel-2 spectral bands used.

Band	Spacial Resolution (m)	Wavelength (nm)	Description
B2	10	490	Blue
B3	10	560	Green
B4	10	665	Red
B5	20	705	Red Edge 1
B6	20	740	Red Edge 2
B7	20	783	Red Edge 3
B8	10	842	NIR
B8a	20	865	Narrow NIR
B11	20	1610	SWIR1
B12	20	2190	SWIR2

Table 2: Vegetation indices used.

Vegetation index	Formula	Application
NDVI	$\frac{B8-B4}{B8+B4}$	Vegetation health, density, and photosynthetic activity.
NDMI	$\frac{B8-B11}{B8+B11}$	Monitoring vegetation and soil moisture, drought assessment.
GCI	$\frac{B9}{B3}-1$	Chlorophyll content estimation.
PSSRa	$\frac{B7}{B4}$	Detection of vegetation stress and senescence.

First, we select several areas with damaged forests, identified using freely available high-resolution imagery from Google Earth Pro dated 26.06.2023.

We randomly select n points for healthy forest and n points for damaged forest across different areas. Then, using Sentinel-2 image for the corresponding date, we calculate the average values within the spectra of the studied bands and vegetation indices separately for pixels of healthy and damaged forest (1, 2). Based on the values obtained, we calculate the relative difference between the average values of healthy and damaged forest pixels in different spectra (3) and thus identify the most sensitive spectra as the bands/indices with the highest relative difference.

$$Mean_h = \frac{1}{n}\sum_{i=1}^{n} V_h(x_i, y_i) \tag{1}$$

$$Mean_d = \frac{1}{n}\sum_{i=1}^{n} V_d(x_i, y_i) \tag{2}$$

$$RelDiff = \frac{Mean_h - Mean_d}{Mean_h} \tag{3}$$

where $V_h(x_i, y_i)$ and $V_d(x_i, y_i)$ are the pixel values for healthy and damaged forest at positions (x_i, y_i).

In addition, we visually assess which spectral bands or indices are the most informative for detecting forest damage.

Once these indicators are identified, we apply the following approach:

1) **Defining reference areas**.
Randomly select n points of healthy forest to calculate the reference spectral characteristics, against which anomalies will be detected.
2) **Sampling and averaging spectral values**.
We extract pixel values from selected points and calculating the average value of these pixels for each band or index.
3) **Detecting anomalous areas**.
We use (4), where anomalies are defined as pixels with values differing from the average healthy forest by more than a threshold coefficient (k), which we set as 10%, in the studied spectrum (band or index).

$$Anomaly_i = \begin{cases} 1, & \text{if } |V_{p_i} - \overline{V_{h_i}}| > k \cdot \overline{V_{h_i}} \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where V_p is the value of the pixel in the studied spectrum i, $\overline{V_\mathrm{h}}$ is the average value of the healthy forest pixels, k is the threshold coefficient, set to 10%.
4) **Combining detected anomalies across different spectra**.
To reduce false-positive results, we combine anomalies from the bands/indices that are most sensitive to forest damage. Thus, areas with anomalous values in all spectra simultaneously are classified as damaged or diseased forest (5):

$$Forest_{damage} = \prod Anomaly_i \tag{5}$$

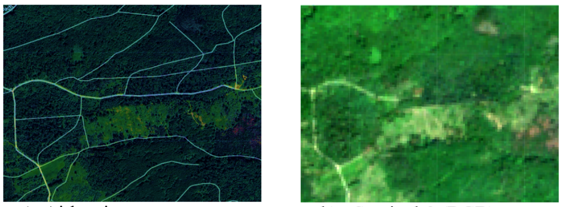
4 RESULTS

4.1 Determining the Most Informative Spectral Data of Sentinel-2 for Detecting Forest Damage

Figure 2 presents an image of a selected area of interest from June 2023 in ultra-high resolution from Google Earth Pro (Fig. 2a) and in high resolution from Sentinel-2 (Fig. 2b). This area was chosen to visually illustrate the appearance of damaged forest patches in different spectral ranges of Sentinel-2.

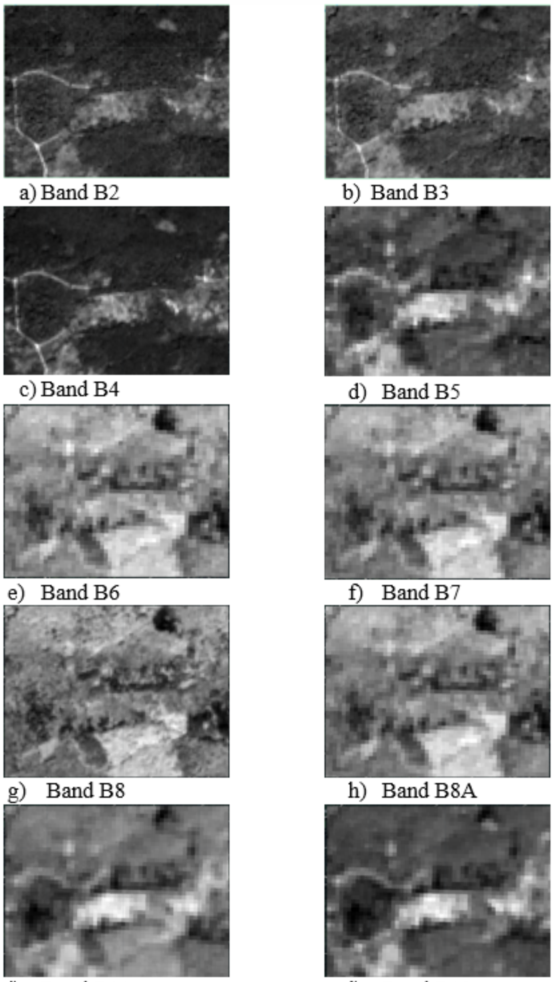
Figure 3 displays this area in various spectral bands of Sentinel-2, while Figure 5 shows its representation through vegetation index spectra.

From Figure 5, it is evident that damaged areas are most clearly distinguishable in spectral bands B2, B3, B4 (10 m) and B5, B11, B12 (20 m). Other bands do not provide such effective visual identification of damaged forests.



a) Airbus image b) Sentinel-2 (RGB)

Figure 2: Sample area with damaged forest, coordinates: 51°48’41.68” N, 10°51’06.94” E.



a) Band B2 b) Band B3 c) Band B4 d) Band B5 e) Band B6 f) Band B7 g) Band B8 h) Band B8A i) Band B11 j) Band B12

Figure 3: Sample area in the spectrum of the Sentinel 2 bands.

The bar chart in Figure 4 presents a numerical comparison of the mean values of the bands for healthy and damaged forest areas, while the line graph represents the relative difference between these values. Notably, for all bands, the pixel values of damaged forests exceed those of healthy forests. The largest discrepancies are observed in bands B3 (-40%) and B5 (-49.7%), making them the most informative spectral bands for detecting forest damage using Sentinel-2 data.

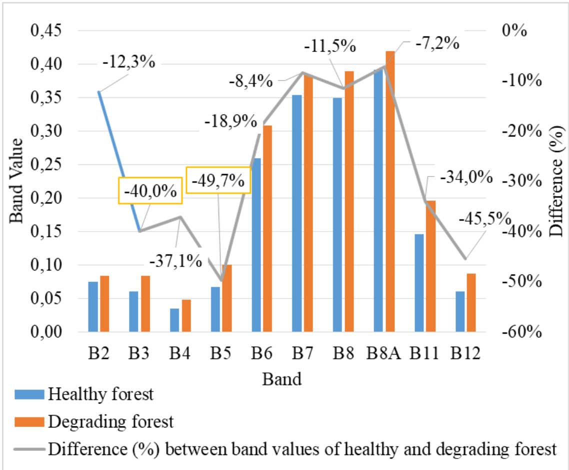
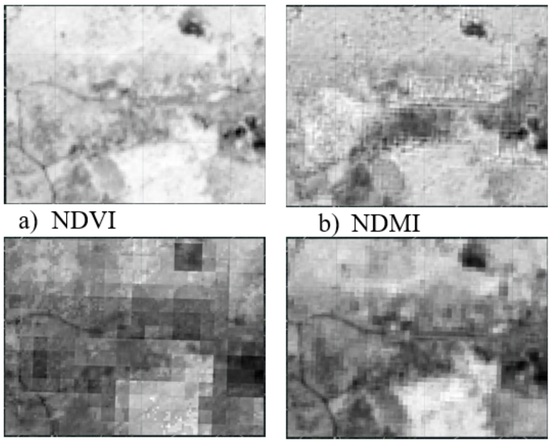


Figure 4: Histogram of average pixel values of Sentinel-2 bands for healthy and damaged forest.



a) NDVI b) NDMI c) GCI d) PSSRa

Figure 5: Sample area in the spectrum of the Sentinel 2 vegetation indices.

Regarding vegetation indices, visual analysis of the images (Fig. 5) indicates that they are not sufficiently effective for clearly distinguishing between pixels of healthy and damaged forests. However, numerical analysis reveals that the vegetation index values for damaged forests are significantly lower than those for healthy forests, particularly for GCI (-47%) and PSSRa (-21.4%) (Fig. 6). Nevertheless, due to the low visual separability of damaged forest areas from healthy ones using vegetation indices, we decided to focus solely on spectral bands B3 and B5.

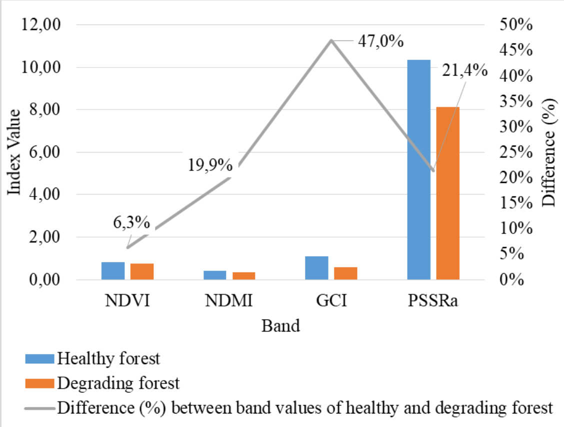
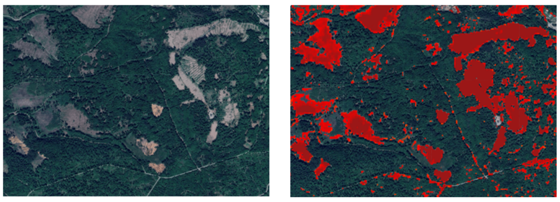


Figure 6: Histogram of average pixel values of Sentinel-2 vegetation indices for healthy and damaged forest.

Thus, for further analysis and detection of damaged forests, we will utilize the intersection of anomalies identified in bands B3 and B5.

4.2 Identifying Damaged Forest Areas Through Anomaly Detection

Figure 7 shows an example of detecting damaged forest areas using the proposed approach based on anomaly analysis in spectral channels B3 and B5. As shown in the figure, the combination of anomalies detected in these channels allows for clear differentiation between healthy and damaged forest areas, almost without missing any damaged areas and minimizing the number of false positives.

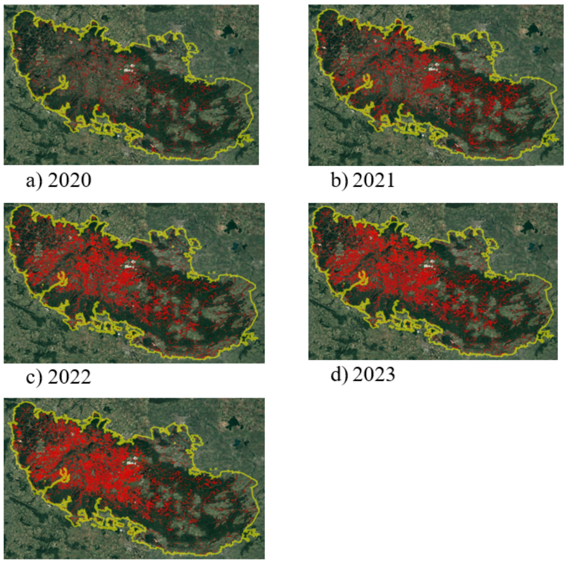


a) Airbus image (04.06.2023) b) Detected damaged forest

Figure 7: Example of detecting damaged forests. Coordinates: 51°34'59.36" N, 10°59'47.50" E.

By applying this method, we evaluated the dynamics of forest damage from 2020 to 2024 (Fig. 8) and calculated the areas of degraded land.

As seen in Figure 9, the area of damaged forest sharply increased between 2020 and 2021 (from 25.13 to 49.41 thousand hectares) and continues to grow each year, although the rate of increase is slowing. The highest concentration of degraded forest is observed in the central and western parts of the Harz, in the regions of Thuringia and Lower Saxony, while to the east, in Saxony-Anhalt, the extent of the damage is smaller. As of 2024, the total area of degraded forest in the Harz amounts to 68.1 thousand hectares.



a) 2020 b) 2021 c) 2022 d) 2023 e) 2024

Figure 8: Detected damaged forests on the territory of the Harz, 2020-2024.

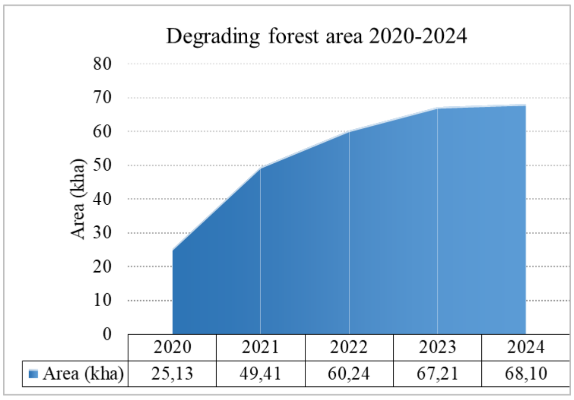


Figure 9: Dynamics of changes in the area of degraded forests in the Harz, 2020-2024.

5 CONCLUSIONS

In this study, we proposed a simple method for detecting degrading forests based on free Sentinel-2 satellite data and an anomaly detection approach, analyzing the forest areas in the Harz region. We examined individual Sentinel-2 channels with spatial resolution of 10-20 m and four vegetation indices—NDVI, NDMI, GCI, and PSSRa—on test plots with damaged forest, using high-resolution imagery from Google Earth Pro

(2023) as auxiliary data for verification. It was found that specific Sentinel-2 spectral channels were more effective for detecting degrading forests than vegetation indices. The most informative Sentinel-2 channels were identified as B3 (green) and B5 (red edge 1), with B4 and B12 also proving to be useful.

By defining anomalous pixel values in the B3 and B5 channels relative to randomly selected reference areas (healthy forests), we successfully detected damaged forest areas with high accuracy and minimal time, data, and computational resource costs. Using the developed method, we were able to track the dynamics of forest loss in the Harz region and calculate the areas of damage. It was determined that the largest losses occurred between 2020 and 2021 (approximately 24.27 thousand hectares of forest were lost), after which the trend gradually slowed down, though it remained negative. By 2024, the total loss of forest in the Harz region amounted to 68.1 thousand hectares.

Thus, this study demonstrates that anomaly detection in Sentinel-2 imagery can serve as an effective, cost-efficient method for monitoring forest degradation, enabling the assessment of forest loss over time and providing valuable information for forest management and conservation efforts. The method can be adapted to other regions, facilitating broader applications in forest health monitoring.

ACKNOWLEDGMENTS

This research was supported by the European Union and Land Sachsen-Anhalt through the ReSeDiUm project: Algorithms and Methods for Land Area Measurement to Estimate Land Areas, as well as the Horizon Europe projects funded by the European Union: SWIFTT (Satellites for Wilderness Inspection and Forest Threat Tracking, 2022–2025, grant agreement No. 101082732) and FUTUREFOR (Copernicus Applications for Next-Generation Forest Monitoring, 2024–2027, grant agreement No. 101180278).

REFERENCES

[1] G. F. Watch, "Harz, Sachsen-Anhalt, Germany Deforestation Rates Statistics — GFW," [Online]. Available: https://www.globalforestwatch.org/dashboards/country/DEU/13/7/?category=forest-change.

[2] J. Wang, T. W. Sammis, V. P. Gutschick, M. Gebremichael, S. O. Dennis, and R. E. Harrison, "Review of satellite remote sensing use in forest health studies," Open Geogr. J, vol. 3, no. 1, pp. 28-42, 2010.

[3] F. Knutzen, P. Averbeck, C. Barrasso, L. M. Bouwer, B. Gardiner, J. M. Grünzweig, S. Hänel, K. Haustein, M. R. Johannessen, S. Kollet et al., "Impacts on and damage to European forests from the 2018–2022 heat and drought events," Natural Hazards and Earth System Sciences, vol. 25, no. 1, pp. 77-117, 2025.

[4] M. Lange, S. Preidl, A. Reichmuth, M. Heurich, and D. Doktor, "A continuous tree species-specific reflectance anomaly index reveals declining forest condition between 2016 and 2022 in Germany," Remote Sensing of Environment, vol. 312, p. 114323, 2024.

[5] A. Mujetahid, M. Nursaputra, and A. S. Soma, "Monitoring illegal logging using google earth engine in Sulawesi Selatan Tropical Forest, Indonesia," Forests, vol. 14, no. 3, p. 652, 2023.

[6] N. Kussul, S. Drozd, H. Yailymova, A. Shelestov, G. Lemoine, and K. Deininger, "Assessing damage to agricultural fields from military actions in Ukraine: An integrated approach using statistical indicators and machine learning," International Journal of Applied Earth Observation and Geoinformation, vol. 125, p. 103562, 2023.

[7] L. Shumilo, N. Kussul, and M. Lavreniuk, "U-Net model for logging detection based on the Sentinel-1 and Sentinel-2 data," in 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 4680-4683, 2021.

[8] J. Zhang, S. Cong, G. Zhang, Y. Ma, Y. Zhang, and J. Huang, "Detecting pest-infested forest damage through multispectral satellite imagery and improved UNet++," Sensors, vol. 22, no. 19, p. 7440, 2022.

[9] R. Hall, G. Castilla, J. White, B. Cooke, and R. Skakun, "Remote sensing of forest pest damage: A review and lessons learned from a Canadian perspective," The Canadian Entomologist, vol. 148, no. S1, pp. S296-S356, 2016.

[10] M. Decuyper, R. O. Chávez, M. Lohbeck, J. A. Lastra, N. Tsendbazar, J. Hackländer, M. Herold, and T.-G. Vågen, "Continuous monitoring of forest change dynamics with satellite time series," Remote Sensing of Environment, vol. 269, p. 112829, 2022.

[11] F. Haghighian, S. Yousefi, and S. Keesstra, "Identifying tree health using sentinel-2 images: a case study on Tortrix viridana L. infected oak trees in Western Iran," Geocarto International, vol. 37, no. 1, pp. 304-314, 2022.

[12] M. Piragnolo, F. Pirotti, C. Zanrosso, E. Lingua, and S. Grigolato, "Responding to large-scale forest damage in an alpine environment with remote sensing, machine learning, and web-GIS," Remote Sensing, vol. 13, no. 8, p. 1541, 2021.

[13] F. Giannetti, M. Pecchi, D. Travaglini, S. Francini, G. D'Amico, E. Vangi, C. Cocozza, and G. Chirici, "Estimating VAIA windstorm damaged forest area in Italy using time series Sentinel-2 imagery and continuous change detection algorithms," Forests, vol. 12, no. 6, p. 680, 2021.

[14] T. Molnár and G. Király, "Forest disturbance monitoring using cloud-based Sentinel-2 satellite imagery and machine learning," Journal of imaging, vol. 10, no. 1, p. 14, 2024.

[15] N. S. Podoprigorova, G. A. Savchenko, K. R. Rabcevich, A. I. Kanev, A. V. Tarasov, and A. N. Shikohov, "Forest damage segmentation using machine learning methods on satellite images," in

International Conference on Neuroinformatics, pp. 380-388, 2023.

[16] D. John and C. Zhang, "An attention-based U-Net for detecting deforestation within satellite sensor imagery," International Journal of Applied Earth Observation and Geoinformation, vol. 107, p. 102685, 2022.

[17] J. Anwander, M. Brandmeier, S. Paczkowski, T. Neubert, and M. Paczkowska, "Evaluating Different Deep Learning Approaches for Tree Health Classification Using High-Resolution Multispectral UAV Data in the Black Forest, Harz Region, and Göttinger Forest," Remote Sensing, vol. 16, no. 3, p. 561, 2024.

[18] UNEP-WCMC and IUCN, "Protected planet: The world database on protected areas (wdpa)," 2025, [Online]. Available: https://www.protectedplanet.net.

[19] EU Forest Observatory, [Online]. Available: https://forest-observatory.ec.europa.eu/forest/rmap.

[20] Radmin, "Sentinel-2 data and vegetation indices," [Online]. Available: http://www.eo4geo.eu/training/sentinel-2-dataand-vegetation-indices/.

A New Hybrid Metaheuristic Model for Image Edge Detection

Adil Khalil, Adel Abed Al Wahaab and Ahmed Abbas
[1]*Department of Computer Science, College of Education for Pure Science, University of Diyala, Al-Muradiyah Str. 30, 32001 Diyala, Iraq*
dr.adil.khalil@uodiyala.edu.iq, purecomp.adeel.abdulwahab@uodiyala.edu.iq, dr.ahmed.k.abbas@uodiyala.edu.iq

Keywords: Image Processing, Hybrid Model, Edge Detection, Bat Algorithm, Canny.

Abstract: Image edge detection is a vital process in various applications, such as medical image analysis, computer vision, and security systems. Several models were proposed to determine image edges. However, each method has some limitations in finding the best edges, such as choosing the ideal parameters or the presence of noise. New techniques, such as nature-inspired optimization, have emerged as a promising approach in several domains. These techniques may have the potential to provide advanced capabilities to improve the image edge detection process. A metaheuristic model such as the Bat algorithm may offer appropriate parameters for the edge detection algorithm. Therefore, the primary focus of this study is to build a new hybrid model that integrates the Bat algorithm and Canny filter to enhance the output of image edge detection process. To achieve the goal of this study, the hybrid model has been applied to JPG images. Notable improvements in edge detection were observed during the application of the proposed system on the tested images compared to the traditional Canny algorithm. The improvement rate in the performance of the proposed system reached 30%. It was concluded from this study that modifying the parameters of the Canny filter using the proposed dynamic model leads to optimizing the image edge detection processes. Therefore, integrating other algorithms, such as deep learning techniques, is recommended to study the parameters and performance of edge detection operators.

1 INTRODUCTION

In the applications of computer vision and image processing, edge detection is considered an essential step in image interpretation[1]–[3]. Although traditional edge detection methods offer acceptable performance, they face some limitations. The main difficulty is identifying the correct and closed edges. For this reason, several algorithms were proposed for identifying image optimal edges [4][5].

Methods that process images to detect edges are divided into two groups: 1) The first group does not use prior knowledge about the scene or image. This type is limited to the image to be examined and is based on the local processing of neighbouring pixels. For example, the Canny operator is one of the most popular edge detection methods [4], [6]. This algorithm consists of several stages, including image smoothing, gradient calculation, non-maximal peak suppression, and hysteresis concatenation. 2) The second group uses prior knowledge about the edges of the image. The most common techniques in this field are the algorithms inspired by nature, that is called optimization algorithms. This group of techniques is used to find the optimal values of the basic parameters to increase the system's efficiency.

Traditional edge detection models provide efficient results for limited purposes. However, these methods have difficulty in dealing with noise or enhancing edges in some complex cases. For example, the effectiveness of the Canny algorithm depends on the selection of appropriate parameters, such as the upper and lower thresholds [5], [7], [8]. The Canny filter is very sensitive to the choice of these parameters, especially when the selection of these parameters is human intervention, which may lead to the loss of some essential edges of the image.[9], [10].

Optimization algorithms can analyze the problem to find optimal values for the system parameters [11], [12]. Therefore, optimization algorithms are widely used to overcome such problems. The Bat algorithm is an ideal search model inspired by the echolocation behavior of bats[13]. This algorithm is based on the principle of searching for the best solutions in a way inspired by nature. This algorithm is characterized by its ability to explore the solution space efficiently and search for optimal solutions in a way similar to nature.

This paper aims to present a new hybrid model based on the Bat algorithm and the Canny filter to enhance the image edges detection processes. The proposed model was applied to 100 different images, and the results were compared with the edges detected using the regular Canny filter. The ground truth file for each tested images and metrics such as Structural Similarity Index Measure (SSIM) and Dice Similarity Coefficient (Dic) were used to measure system efficiency.

2 MATERILAS AND METHOD

2.1 Image Dataset

The dataset contains 100 color JPG images. Each image has a corresponding image containing the real edge data (ground truth data). The images varied between natural scenes, objects, and complex backgrounds, making them suitable for testing edge detection algorithms.

2.2 The Proposed Model

The following steps were followed to implement the proposed model:

1) Upload the image and convert it to grayscale.
2) Apply a Canny filter to extract raw edges from the image. Basic Canny parameters (minimum and maximum) were determined to obtain the edges.
3) Refine edges using the Bat algorithm. The Bat algorithm determines the optimal Canny parameters (low and high threshold) that provide the best performance in edge detection.
4) Evaluate the models using SSIM and Dice metrics to measure the edge-detected performance of the proposed hybrid model compared to the Canny algorithm using ground truth data as a reference.

Figure 1 illustrates the workflow of the proposed hybrid model.

2.3 The Mechanism of Bat Algorithm

Micro bats use echolocation to find food, navigate barriers, and roost. The Bat algorithm uses this information to create its mechanism [14]. Micro bats move around an octave using brief, frequency-modulated sounds f, with pulse emission rates accelerating when homing on prey. The Bat algorithm's foundation is based on three principles.

1) All bats use echolocation in some magical way to perceive background barriers, distance, and the distinction between food and prey.
2) Bats fly randomly with velocity υ_i, frequency f, and loudness A_o to find prey. They can change the frequency and wavelength of their pulses as they come closer to their target, allowing them to change their speed appropriately.
3) We suppose that the loudness extends from a significant positive A_o to a minimal constant value A_{mi}, regardless of many conceivable fluctuations.

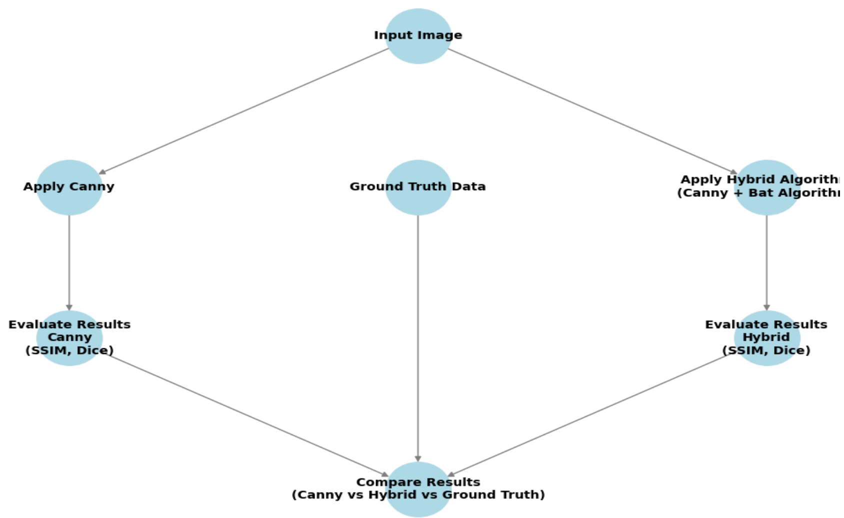


Figure 1: The proposed system diagram.

At iteration t, each bat is linked to a location x_i^t and a velocity v_i^t in a d-dimensional search or solution area. Right now, x^* is the best solution out of all the bats. As seen below, the locations x_i^t and velocities v_i^t can thus be found by translating the three rules mentioned above:

$$f_i = f_{min} + (f_{max} - f_{min})\beta, \quad (1)$$

$$v_i^t = v_i^{t-1} + (x_i^{t-1} - x^*)f_i. \quad (2)$$

$$x_i^t = x_i^{t-1} + v_i^t. \quad (3)$$

where a random vector $\beta \in [0,1]$ is selected from a uniform distribution. As was previously noted, relying on the field size of the issue relevant, we will employ either wavelengths or frequencies for implementation; $f_{min} = 0$ and $f_{min} = 1$. First, a uniformly chosen frequency from [f_{min}, f_{max}] is assigned to each bat at randomly. This makes the bat method a frequency-tuning algorithm that offers a well-balanced mix of exploitation and exploration. According to [15], [16], Loudness and pulse emission rates allow automatic management and auto-zooming into prospective solutions. A random walk with direct exploitation is used in the local search that modifies the current optimal solution in line with the equation:

$$x_{new} = x_{old} + \partial_{A^t}, \quad (4)$$

where A^t is the average loudness of all the best at this time step, and $\partial \in [-1,1]$ is a random value.

2.3.1 Differences in Loudness and Pulse Rates

To regulate exploration and exploitation stages, we must adjust loudness and pulse emission rate during iterations. If $A_{min} = 0$, loudness can be set between A_{min} and A_{max}, as loudness drops after finding prey and pulse emission increases. Assuming these things, we have:

$$A_i^{t+1} = \alpha A_i^t, \quad (5)$$

and

$$r_i^t = r_i^o\ [1 - \exp(-\gamma t)], \quad (6)$$

where α and ϒ are constants. The loudness at iteration t, initial pulse emission rate r_i^o, and cooling factor α are used to calculate the pulse emission rate at iteration t, with $0 < \alpha < 1$ and $\Upsilon > 0$. [15]–[17].

2.4 Study Evaluation Metrics

In this study, two metrics were used to analyze the quality of the proposed model. The first one is the Structural Similarity Index Measure (SSIM). This measure is used to analyze the quality of images or videos by comparing the structure of the original image with a reference image (ground truth) [18]. The SSIM is computed using the following equation:

$$SSIM(a,b) = \frac{(2\mu_a\mu_b + c_1)(2\sigma_{ab} + c_2)}{(\mu_a^2 + \mu_b^2 + c_1)(\sigma_a^2 + \sigma_b^2 + c_2)}, \quad (7)$$

where $\boldsymbol{\mu_a}$ and $\boldsymbol{\mu_b}$ are the average values of images a and b respectively. The contrast between images a and b is illustrated in $\boldsymbol{\mu_a}$ and $\boldsymbol{\mu_b}$ respectively. C values are constant to prevent division by zero.

The second metric is the Dice Similarity Coefficient (Dice). This metric is used in the field of image processing to evaluate the similarity between two image data, especially in the edge detection process. The mathematical representation of this measure is shown below:

$$Dice = \frac{2|A \cap B|}{|A| + |Y|}, \quad (8)$$

Where A and B are the set of points in the original image and ground truth image respectively. The number of similar points between the two images is represented by $A \cap B$

3 RESULTS AND DISCUSSION

Figure 2a shows the original image after converting it to grayscale format. Figure 2b shows the traditional Canny results, and Figure 2c presents the results of Canny after optimizing its parameters using the Bat algorithm. In addition, this Figure presents the ground truth image which is related to the original image as shown in Figure 2d. The comparison between Canny model and the hybrid model results was performed by studying the values of SSIM and Dice metrics based on the ground truth image as a reference for measurement. Therefore, from this Figure, one can notice that the proposed model achieved higher SSIM and Dice values compared to the traditional canny

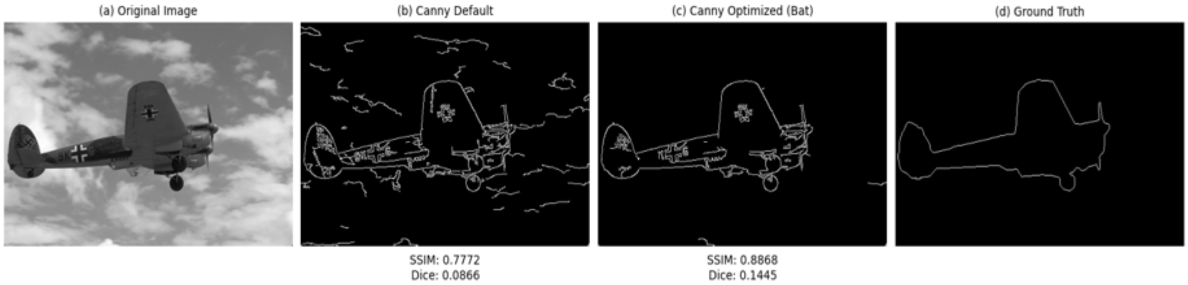


Figure 2: Enhancing image edges detection using the Bat hybrid model with Canny filter. (a) Original image (b) Canny result (c) Hybrid model result (d) Ground Truth image.

model. The hybrid model provides more accurate edges in important areas, which may improve performance in applications such as computer vision. In addition, it is clear from Fig. 2 that the hybrid model provides more accurate details than the traditional model, especially within the main regions of the image.

Table 1 provides the SSIM and Dice metrics values of Canny algorithm and the proposed hybrid model for 10 images. From this table, we can observe that the performance results of our hybrid model outperformed the canny algorithm regarding SSIM and Dice values.

Table 1: Performance analysis of the hybrid model compared to the Canny filter using SSIM and Dice metrics for 10 images.

Image no.	canny		Hybrid (canny + Bat alg.)	
	SSIM	Dice	SSIM	Dice
1	0.3447	0.0392	0.5036	0.0477
2	0.7777	0.0866	0.8868	0.1445
3	0.4374	0.0514	0.0672	0.0803
4	0.6546	0.0442	0.7562	0.0502
5	0.6448	0.0757	0.7879	0.1524
6	0.6321	0.1157	0.7967	0.1875
7	0.3442	0.0901	0.5934	0.1422
8	0.5115	0.0700	0.8538	0.2113
9	0.7582	0.1411	0.8755	0.2531
10	0.4784	0.0146	0.6851	0.0918

In all cases, the hybrid algorithm achieved higher SSIM values, indicating that the extracted edges are more similar to the ground truth image. Furthermore, the values increased in most images, indicating a reduction in the differences between the extracted edges and the real edges. In contrast, Some images showed only slight improvement (e.g. Figure 2c and 2d), which may mean that the algorithm may need to adjust some parameters to handle different types of images.

It can be said that the hybrid algorithm is better in the case of applications that require high accuracy in identifying edges. Therefore, in medical image analysis, the hybrid method can be better because it provides finer edges, which makes it easier to detect tumors or different tissues. In contrast, in image and video surveillance, the method may help improve the detection of objects or subtle movements.

Based on the experiments performed on the test images, it was found that the algorithm provided high performance and achieved high index values when applied to images with uniform backgrounds and low contrast, where the edges are clearer and less complex. Therefore, the proposed algorithm can be applied to images with blurry backgrounds such as images taken with wide aperture cameras. The algorithm can also be applied to images with objects that stand out clearly from the background such as wildlife photographs or products displayed on a uniform background, document and drawing scans – where the background is white or of a uniform color with prominent lines, and medical imaging images (such as X-rays or MRIs) especially when there is significant contrast between tissues.

Figure 3(a) shows a comparison view of SSIM values between Canny filter and the hybrid model (Canny + Bat Algorithm) across 100 images. Form this Figure, one can observe that the hybrid model achieved higher SSIM values than the traditional Canny for all tested images. This indicates that the extracted edges using the hybrid model are closely matched to the ground truth compared to the Canny filter. Therefore, these results demonstrate the efficiency of the proposed system in reducing noise and improving the detection of image edges when compared with the traditional Canny filter.

In contrast, Figure 3(b) shows accuracy edge detection result of Canny and the hybrid model using the Dice metric. From this figure, it is clear that the proposed hybrid system achieved higher Dice values than the Canny filter. This findings indicate that the hybrid system achieved more accurate edges that are close to the original image edges.

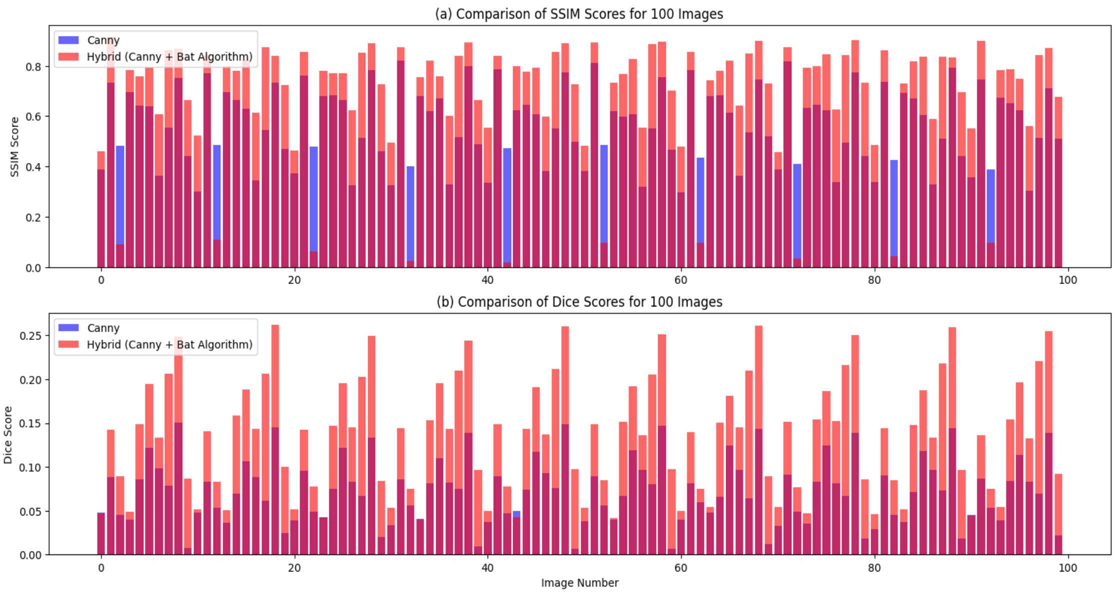


Figure 3: Evaluation metrics results. (a) SSIM values and (b) Dice values.

4 CONCLUSIONS

A hybrid model to optimize Canny parameters using the Bat algorithm to improve the image edge detection process was presented in this paper. The statistical metrics used in this study showed that the hybrid model outperformed the Canny algorithm in detecting image edges. The study confirmed that choosing the Canny filter parameters automatically contributes significantly to improving the filter efficiency compared to the fixed parameters that are pre-determined for the traditional filter operation. According to this finding, the hybrid model is useful in applications that require high accuracy such as medical computer vision or security systems. Furthermore, the proposed model of this study can be used in other applications that require more accurate edge detection, such as autonomous robotics or advanced medical data analysis. Finally, algorithms such as deep learning can be integrated to study edge detection performance.

REFERENCES

[1] M. Ali, L. Goyal, C. M. Sharma, and S. Kumar, "Edge-computing-enabled abnormal activity recognition for visual surveillance," Electronics (Switzerland), vol. 13, no. 2, 2024, [Online]. Available: https://doi.org/10.3390/electronics13020251.

[2] R. C. Gonzalez and P. Wintz, Digital Image Processing. Reading, MA, USA: Addison-Wesley, 1987.

[3] M. M. Kabir, J. R. Jim, and Z. Istenes, "Terrain detection and segmentation for autonomous vehicle navigation: A state-of-the-art systematic review," Information Fusion, vol. 113, p. 102644, Jan. 2025, [Online]. Available: https://doi.org/10.1016/j.inffus.2024.102644.

[4] Z. Xu, X. Ji, M. Wang, and X. Sun, "Edge detection algorithm of medical image based on Canny operator," in J. Phys.: Conf. Ser., vol. 1955, p. 012080, 2021, [Online]. Available: https://doi.org/10.1088/1742-6596/1955/1/012080.

[5] H. M. Zangana, A. K. Mohammed, and F. M. Mustafa, "Advancements in edge detection techniques for image enhancement: A comprehensive review," Int. J. Artif. Intell. Robot. (IJAIR), vol. 6, no. 1, pp. 29-39, 2024, [Online]. Available: https://doi.org/10.25139/ijair.v6i1.8217.

[6] N. D. Lynn, A. I. Sourav, and A. J. Santoso, "Implementation of real-time edge detection using Canny and Sobel algorithms," IOP Conf. Ser.: Mater. Sci. Eng., vol. 1096, no. 1, p. 012079, 2021, [Online]. Available: https://doi.org/10.1088/1757-899X/1096/1/012079.

[7] N. Tawfeeq and J. Harbi, "Using the Canny method with deep learning to detect and predict river water level," J. Al-Qadisiyah Comput. Sci. Math., vol. 16, no. 2, pp. 135-150, 2024, [Online]. Available: https://doi.org/10.29304/jqcsm.2024.16.21565.

[8] W. A. Sasmita, H. Mubarok, and N. Widiyasono, "Agricultural path detection systems using Canny-edge detection and Hough transform," IAES Int. J. Robot. Autom., vol. 13, no. 3, pp. 247-254, 2024, [Online]. Available: https://doi.org/10.11591/ijra.v13i3.pp247-254.

[9] T. Saini, P. S. Shiakolas, and C. McMurrough, "Evaluation of image segmentation methods for in situ quality assessment in additive manufacturing,"

Metrology, vol. 4, no. 4, pp. 598-618, 2024, [Online]. Available: https://doi.org/10.3390/metrology4040037.

[10] X. D. Sun and X. F. Sun, "An edge detection algorithm based upon the adaptive multi-directional anisotropic Gaussian filter and its applications," J. Supercomput., vol. 80, no. 11, pp. 15183-15214, Jul. 2024, [Online]. Available: https://doi.org/10.1007/s11227-024-06044-6.

[11] K. H. Choi and J. E. Ha, "An adaptive threshold for the Canny edge with actor-critic algorithm," IEEE Access, vol. 11, pp. 67058-67069, 2023, [Online]. Available: https://doi.org/10.1109/ACCESS.2023.3291593.

[12] H. Harish and A. S. Murthy, "Edge discerning using improved PSO and Canny algorithm," in Commun. Comput. Inf. Sci., pp. 192-202, 2023, [Online]. Available: https://doi.org/10.1007/978-3-031-43140-1_17.

[13] M. Shehab et al., "A comprehensive review of bat inspired algorithm: Variants, applications, and hybridization," Arch. Comput. Methods Eng., vol. 30, no. 2, pp. 765-797, Mar. 2023, [Online]. Available: https://doi.org/10.1007/s11831-022-09817-5.

[14] X. S. Yang, "A new metaheuristic bat-inspired algorithm," in Stud. Comput. Intell., pp. 65-74, Springer, 2010, [Online]. Available: https://doi.org/10.1007/978-3-642-12538-6_6.

[15] X. S. Yang, "Bat algorithm: Literature review and applications," Int. J. Bio-Inspired Comput., vol. 5, no. 3, pp. 141-149, 2013, [Online]. Available: https://doi.org/10.1504/IJBIC.2013.055093.

[16] P. Kora and S. R. Kalva, "Improved bat algorithm for the detection of myocardial infarction," SpringerPlus, vol. 4, no. 1, pp. 1-18, Dec. 2015, [Online]. Available: https://doi.org/10.1186/s40064-015-1379-7.

[17] X. S. Yang, "Bat algorithm and cuckoo search: A tutorial," in Artificial Intelligence, Evolutionary Computing and Metaheuristics, pp. 421-439, Springer, 2013, [Online]. Available: https://doi.org/10.1007/978-3-642-29694-9.

[18] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in Proc. Asilomar Conf. Signals, Syst. Comput., pp. 1398-1402, 2003, [Online]. Available: https://doi.org/10.1109/ACSSC.2003.1292216.

A Comprehensive Method for Anomaly Detection in Complex Dynamic IoT Systems

Andrii Liashenko and Larysa Globa
Institute of Telecommunication Systems, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Beresteiska Avenue 37, 03056 Kyiv, Ukraine
andrey.lyashenko44@gmail.com, lgloba@its.kpi.ua

Keywords: Anomaly Detection, Temporal Graphs, Temporal Graph Neural Networks, Autoencoder, Graph Neural Networks, Reconstruction Error, Dynamic Systems, Transportation Networks.

Abstract: Modern dynamic systems, such as transportation networks and IoT infrastructures, generate massive volumes of interrelated temporal data represented as temporal graphs. Conventional methods – like clustering, statistical thresholds, and classical time series analysis – often fail to account for the spatial-temporal dependencies inherent in these systems, leading to high false positive rates or missed complex anomalies. In this paper, we propose a novel anomaly detection approach that combines Temporal Graph Neural Networks (TGNN) with Autoencoders. The method utilizes TGNN to extract robust node representations by capturing both local connectivity and temporal evolution, while an autoencoder is trained to reconstruct normal node behavior. Anomalies are subsequently identified through significant reconstruction errors, which serve as indicators of deviations from typical patterns. Experimental evaluations on the real-world PeMSD7 dataset demonstrate that the proposed TGNN + Autoencoder method improves detection accuracy by 17.33% compared to traditional methods, reduces false positives by 4.71%, and achieves a 6.02% higher F1-score relative to using TGNN or autoencoder individually. These results underline the practical relevance of our approach for real-time monitoring of transportation networks, while also contributing theoretically to the integration of spatial and temporal features in anomaly detection.

1 INTRODUCTION

Dynamic network systems such as transportation networks, the Internet of Things (IoT), financial markets, and cybersecurity generate vast amounts of interconnected temporal data. For example, according to Statista (Fig. 1), the number of IoT devices in the world is projected to reach 39.6 billion by 2033 [1].

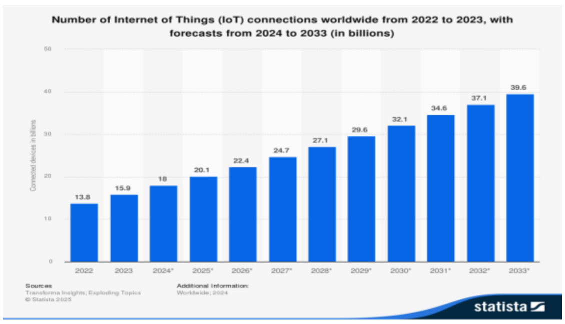


Figure 1: Number of IoT devices by 2033 [1].

Data generated by such devices form dynamic graph structures with nodes and edges that change over time. Anomalies in these networks may indicate transport congestion, IoT failures, financial fraud, or cyberattacks, making early detection critical to prevent accidents, inefficiencies, or financial losses. However, traditional methods like clustering, statistical thresholding, and machine learning have significant limitations – they often ignore the inherent graph structure needed to capture complex network dynamics [2], lack adaptability as approaches such as One-Class SVM or Isolation Forest require pre-training and fail to accommodate new patterns [3], and rely on fixed thresholds that lead to high false positive rates in dynamic environments [4]. Thus, new adaptive approaches that consider both temporal and structural dynamics of graphs are essential for more accurate anomaly detection.

The remainder of the paper is organized as follows. Section 2 describes the methodology, detailing Temporal Graph Neural Networks and Autoencoders for anomaly detection Section 3 reviews traditional methods for anomaly detection. Section 4 defines the problem addressed in this study. Section 5 details the proposed method of anomaly detection, which combines Temporal Graph Neural Networks with Autoencoders. Finally, Section 6 presents the experimental results.

2 METHODOLOGY

The use of Graph Neural Networks (GNN) has significantly improved the analysis of graph data, allowing the creation of deep representations of nodes and edges. However, standard GNNs do not take into account temporal dependencies. Therefore, Temporal Graph Neural Networks (TGNNs) were proposed for dynamic systems (Fig. 2), which take into account the evolution of graphs over time [5]. Autoencoders (AEs) have proven themselves well in anomaly problems, especially in unsupervised learning (Fig. 3). AEs learn to reconstruct data, and nodes that have a high reconstruction error are considered anomalous [6].

Node: In TGNN, a node is a dynamic entity (e.g., a traffic sensor) in a temporal graph, representing time-varying data. *Characteristic (Feature):* These are the measurable, time-dependent attributes of a node, such as sensor readings or traffic speed. *Latent Representation:* A compact, low-dimensional encoding of a node's features, capturing its essential spatial and temporal patterns.

In this paper, we propose a method that combines TGNN with an AE for anomaly detection in temporal graphs. The approach creates latent representations of nodes using a TGNN trained to predict their states over time, then trains an autoencoder to reconstruct normal node states, with anomalies identified by high reconstruction error.

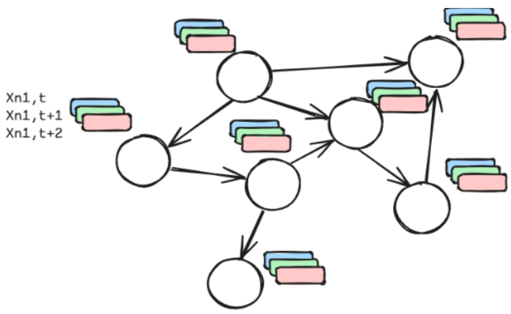


Figure 2: Timestamp graph.

TGNNs extend classical GNNs to process dynamic graphs by incorporating a temporal component. Unlike traditional GNNs, which model static node relationships, TGNNs update node states at each time step based on both neighboring nodes and historical information, often using recurrent mechanisms such as Gated Recurrent Units (GRU) or Long Short-Term Memory (LSTM) [7, 8].

The classical formula for updating the representation of a node in GNN is as follows:

$$h_i^{(t)} = \sigma\left(W \sum_{j\in\mathcal{N}(i)} \alpha_{ij}^{(t)} h_j^{(t-1)}\right)$$

where:

- $h_i^{(t)}$ – d-dimensional state vector of node i at time t, where each element represents a specific feature (e.g., speed, traffic, etc.).,
- *N(i)* – the set of neighbors of node i,
- $\alpha_{ij}^{(t)}$ – the attention weight between nodes i and j
- W – a trainable weight matrix that performs a linear transformation (projection) of the input node features,
- σ – the activation function (for example, ReLU or Sigmoid).

At each time step t, the state vector $h_i^{(t)}$ of node i is computed by aggregating the state vectors of its neighbors *N(i)*, weighted by the attention coefficients $\alpha_{ij}^{(t)}$ – determine the influence of each neighboring node j on node i. They are typically computed using a small neural network to evaluate the "importance" of each neighbor, and then normalized (e.g., via softmax) so that the contributions sum to one. This mechanism enables the model to focus on the most relevant neighbors when aggregating information, effectively capturing complex spatial-temporal relationships.

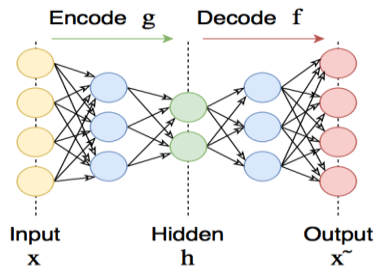


Figure 3: Autoencoder architecture.

The encoder transforms the input data x_t into a compact representation [9] h_t

$$h_t = g(x_t)$$

where:

- x_t – the input data at time t,
- h_t – the hidden feature vector or compressed representation of the data in the hidden space,
- g – the encoder function, which typically includes several layers of neural networks (for example, linear transformations or activation functions).

The decoder reconstructs the input data $\hat{x}_t$ of the hidden representation h_t, trying to recreate the original input data:

$$\hat{x}_t = f(h_t)$$

where:

- $\hat{x}_t$ – the reconstructed output data for time t,
- f – the decoder function, which typically includes several layers of neural networks to reconstruct the output data from the compact representation.

This transformation allows the autoencoder to reduce the dimensionality of the input data, preserving only the most important information about its structure and temporal dynamics.

To train the autoencoder, we use the loss function (Mean Square Error (MSE)) $L(x_t, \hat{x}_t) = \|x_t - \hat{x}_t\|$.

When working on new data, if the model is unable to accurately reconstruct the input data, this may indicate anomalous changes in the graph. Therefore, an anomaly is defined as a large reconstruction error:

$$\Delta_t = \|x_t - \hat{x}_t\|$$

The proposedmethod was tested on PeMSD7 - a real dataset of transport networks [10].

3 RELATED WORK

Traditional time series techniques–such as ARIMA, Holt-Winters Exponential Smoothing, and Hidden Markov Models–model node changes over time and predict future graph states. Yet, they require transforming graph data into independent time series, resulting in lost structural information, high data requirements, and sensitivity to model parameters [12].

Clustering and machine learning methods (e.g., K-Means, DBSCAN, Isolation Forest, One-Class SVM) also have been used to detect anomalies by grouping nodes or identifying outliers. Their main drawbacks are the need to manually set parameters and the lack of temporal and relational context, limiting their effectiveness in complex graphs [13].

In contrast, modern approaches leverage deep neural networks, particularly GNNs and TGNNs, which capture both spatial and temporal dependencies [14]. This work proposes a combination of TGNN with AEs: the TGNN models local and temporal connections, while the AE learns to reconstruct normal node patterns–nodes with high reconstruction error are marked as anomalous [15, 16].

This integrated approach significantly improves anomaly detection accuracy by reducing false positives and enhancing adaptability to new data.

Table 1: Comparison of anomaly detection methods.

Method	Description	Advantages	Disadvantages
Statistical Methods (Z-score, Grubbs' Test)	An anomaly is defined as a deviation from the mean. Nodes or edges that significantly differ from the norm are labeled as anomalous.	Simplicity, computational efficiency	Does not consider graph structure, sensitive to noise
Time Series Methods (ARIMA)	Utilize time-series models to predict node behavior. Anomalies are flagged when the actual value substantially deviates from the predicted value.	Perform well with periodic trends	Do not account for structural changes, sensitive to parameters
Clustering Methods (K-Means, DBSCAN)	Nodes are clustered based on similarity, and those that lie far from the main groups are considered anomalous.	Effective for group anomalies	Require manual selection of cluster count, do not incorporate temporal data
Machine Learning Methods (Isolation Forest, One-Class SVM)	Models are trained on normal data and classify any deviation as anomalous (one-class learning).	Can operate without extensive manual parameter tuning.	Unsuitable for dynamic graphs without adaptation
GNN (Graph Neural Networks)	Graph Neural Networks analyze structural relationships among nodes but do not account for temporal dynamics.	Capture inter-node connections	Do not model the time-evolving behavior of nodes

According to the Table 1, we can conclude that traditional methods are insufficient for analyzing dynamic graphs because they do not take into account the graph structure and the temporal dynamics of changes. This limits their ability to effectively detect anomalies in complex networks where both spatial and temporal relationships are important.

4 PROBLEM DEFINITION

The paper addresses the challenge of detecting anomalies in temporal graphs representing dynamic systems (e.g., transportation networks) where traditional methods fail to capture the intricate spatial-temporal dependencies, leading to high false positives. The goal is to develop a TGNN-based approach combined with Autoencoders to accurately model node behavior over time and identify deviations via reconstruction errors.

5 PROPOSED COMPLEX METHOD OF ANOMALY DETECTION

Before detailing our method, we define the research problem: given a temporal graph $G = (V, E, X, T)$ – where V are nodes (e.g., traffic sensors), E are time-varying spatial connections, X are time-dependent features (e.g., speed measurements every 5 minutes), and T comprises historical data reflecting normal node behavior–the task is to identify nodes that deviate significantly from these expected patterns, by leveraging both local spatial relationships and long-term temporal dependencies.

This task involves analyzing the dynamic evolution of each node's state, which is influenced not only by its own historical data but also by the states of its immediate neighbors, as encoded by the adjacency matrix and attention coefficients. The challenge lies in integrating these local interactions with the long-term trends present in the node features over extended periods. To address this, the proposed method (Fig. 4) employs a TGNN to jointly capture spatial and temporal information, along with an AE to learn compact latent representations. Anomalies are then detected by identifying significant deviations in the reconstruction error, which signal abnormal behavior.

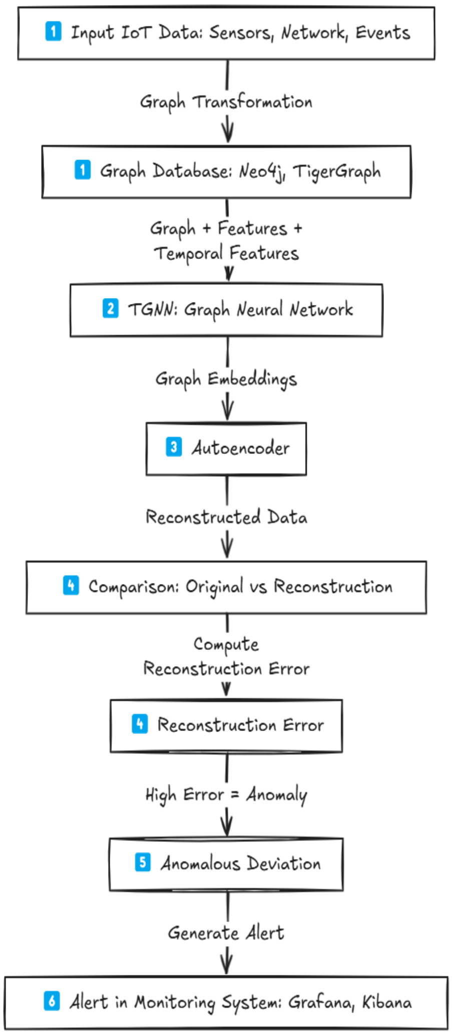


Figure 4: Proposed complex method.

This proposed IoT anomaly detection system operates through six key stages:

1) Incoming IoT data (sensors, network events) are converted into graph structures and transferred to the Graph Database.
2) TGNN takes as input a graph, node characteristics and timestamp history. It creates graph embeddings to represent relationships in the data.
3) Autoencoder receives these embeddings, tries to reconstruct the original data.
4) Comparison of the original and reconstructed data → reconstruction error is calculated.
5) If the error is high → anomaly is detected.
6) Anomaly deviation triggers an alert in the Monitoring System (Grafana, Kibana).

We designed our TGNN architecture using a three-layer A3TGCN model combined with a single-layer GRU to efficiently capture both spatial and temporal dependencies in dynamic IoT networks.

The model consists of three graph convolutional layers with 64 hidden units each, using ReLU activation and an attention-based aggregation mechanism to refine node representations by weighting the influence of neighboring nodes. The three-layer A3TGCN model was selected to ensure that nodes incorporate information from both direct neighbors and second-order connections, effectively modeling localized interactions in dynamic graphs. Increasing the depth beyond three layers resulted in diminishing improvements while increasing computational costs, making deeper architectures inefficient for real-time IoT applications. A shallower model, on the other hand, did not provide sufficient contextual information for anomaly detection. The use of attention-based aggregation further refines node representations, ensuring that structurally important nodes have a greater impact on anomaly detection.

To model temporal dependencies, we integrate a single-layer GRU instead of more complex recurrent architectures like LSTM. GRUs provide a faster and more memory-efficient alternative to LSTMs while maintaining comparable performance in capturing long-term dependencies. Unlike simple RNNs, GRUs effectively retain relevant historical information while dynamically controlling memory updates, making them well-suited for large-scale, continuously evolving IoT networks. The combination of TGNN for spatial learning and GRU for sequential modeling allows the system to distinguish between normal fluctuations and true anomalies over time.

Finally, a fully connected output layer maps the learned node embeddings to a single scalar value per node, indicating its predicted state. The model is trained using the Mean Squared Error (MSE) loss function, which evaluates the reconstruction error between predicted and actual states.

The encoder network consists of three fully connected layers with 128, 64, and 32 neurons, each followed by a LeakyReLU activation function to introduce non-linearity while preserving small gradient updates for low-activation values. The final layer of the encoder maps the data into a latent space of 16 dimensions, providing a compressed representation of node embeddings while maintaining key structural and temporal information.

The decoder network mirrors the encoder, consisting of three fully connected layers (32, 64, 128 neurons), using LeakyReLU activation in the hidden layers and a linear activation in the final layer to reconstruct the original input. This symmetrical structure ensures effective reconstruction while preserving node-specific features.

To improve generalization and prevent overfitting, we apply dropout (0.2 probability) and batch normalization after each hidden layer. The AE is trained using MSE loss, which quantifies the difference between reconstructed and actual node embeddings, helping the model learn normal patterns in graph data.

Correctly setting the threshold τ is crucial: too low increases false positives, while too high leads to false negatives. In this work, we adopt a quantile method that automatically adapts τ to the reconstruction error distribution by setting it at the 95th percentile.

$$\tau = Q_{95},$$

where Q_{95} is the value above which 5% of nodes with the largest reconstruction error are located. This approach allows for dynamic detection of anomalies, reducing the risk of false positive detections, since the model adapts to changes in the graph structure and the dynamics of its nodes.

The 95th percentile is chosen empirically because it reduces false positives by avoiding the misclassification of normal nodes, provides noise immunity by adapting to the specific error distribution rather than relying on a manually set threshold, and effectively captures rare events by identifying the top 5% of the most deviant nodes. To validate this choice, we performed a threshold sensitivity analysis, summarized in the table above. Selecting the 90th percentile increases recall to 0.88 but significantly raises false positives due to lower precision. Conversely, the 98th percentile improves precision to 0.91 and reduces false positives, but recall drops sharply to 0.74, causing many anomalies to be missed. The 95th percentile offers the best balance between precision and recall with an F1-score of 0.88, ensuring both effective anomaly detection and a manageable false positive rate.

Unlike fixed-threshold methods, the quantile approach adapts automatically to the graph's state: in stable conditions, the error distribution is narrow and τ is low, while significant changes adjust τ to new data. Our method captures the entire process in one formula: the state vector of node i at time t is computed by aggregating neighbor states (weighted by attention coefficients), transforming them via a trainable weight matrix and activation function, and then passing the result through an encoder-decoder to reconstruct the original state. This entire process is encapsulated in the following expression:

$$\hat{h}_t(t) = f_{DEC}\left(f_{ENC}\left(\sigma\left(W\sum_{j\in N(i)}\alpha_{ij}(t)h_j(t-1)+b\right)\right)\right)$$

In this formulation, $h_i(t)$ denotes the state vector of node i at time t, $\alpha_{ij}(t)$ represents the attention coefficient reflecting the influence of neighbor j on node i, and $W$is a trainable weight matrix that projects the aggregated features into a new space before applying the non-linear activation σ. The functions f_{ENC} та f_{DEC} correspond to the encoder and decoder of the autoencoder, respectively, which learn a compact latent representation and reconstruct the original signal.

$$Anomaly(i,t) = (\left|\left|h_i(t) - \hat{h}_i(t)\right|\right|^2 > \tau).$$

The anomaly indicator is then determined by comparing the reconstruction error $\left|\left|h_i(t) - \hat{h}_i(t)\right|\right|^2$ a threshold τ; if the error exceeds τ, the node is flagged as anomalous. This compact representation captures the entire method's essence, seamlessly integrating spatial and temporal dynamics for effective anomaly detection.

Variables and notation:

- $h_i(t)$ – is the node representation from the TGNN at time t;
- f_{ENC} and f_{DEC} – are the encoder and decoder functions of the autoencoder;
- σ – is an activation function (e.g., ReLU);
- W and b trainable parameters;
- $\alpha_{ij}(t)$ – represents the attention weight between node i and its neighbor j;
- τ – is the threshold for anomaly detection.

Thus, the proposed approach allows not only to train high-quality representations of nodes in temporal graphs, but also to effectively identify anomalous nodes using the reconstruction error. The use of the quantile threshold selection method ensures the adaptability of the model, which allows avoiding problems associated with excessive sensitivity to noise in the data. The proposed method combines the advantages of TGNN in training representations with the advantages of the autoencoder in detecting deviations, which makes it an effective tool for analyzing anomalies in dynamic graphs.

6 EXPERIMENT

This section evaluates the proposed TGNN+AE method for anomaly detection on the PeMSD7 dataset, which contains temporal graphs of a transport network (nodes = road sensors, edges = spatial connections). The goal is to detect traffic anomalies (e.g., accidents, congestion) by analyzing reconstruction errors and comparing results with traditional methods (Isolation Forest, One-Class SVM, K-Means). Traffic speeds were normalized using Z-score, and sequences of 24 previous values were used to predict the current state. The TGNN extracts spatiotemporal features, while the autoencoder compresses these into a latent space, with the 95th percentile of the reconstruction error used as the anomaly threshold. The experimental results show that the TGNN+AE method improves detection accuracy by 17.33%, reduces false positives by 4.71%, and increases the F1-score by about 6% compared to using each method separately.

Table 2: Comparison of results.

Method	Precision	Recall	F1-score
Isolation Forest	0.72	0.68	0.70
One-Class SVM	0.76	0.64	0.69
K-Means	0.78	0.72	0.75
Autoencoder (AE)	0.81	0.76	0.79
TGNN	0.85	0.81	0.83
TGNN + AE	0.89	0.85	0.88

From Table 2, it can be seen that TGNN + AE improves the anomaly detection accuracy by 10.67% compared to the best traditional method (K-Means):

$$Percentage\ Increase = \frac{F1_{TGNN+AE} - max\ (F1_{Traditional})}{max\ (F1_{Traditional})} \times 100 = \frac{0.88 - 0.75}{0.75} = 17.33\%.$$

TGNN + AE provides 6.02% higher F1-measure compared to using Autoencoder or TGNN alone:

$$Percentage\ Increase = \frac{F1_{TGNN+AE} - max\ (F1_{AE}, F1_{TGNN})}{max\ (F1_{AE}, F1_{TGNN})} \times 100 = \frac{0.88 - 0.83}{0.83} = 6.02\%.$$

TGNN + AE reduces the number of false positives by 4.71% compared to TGNN due to improved Precision:

$$\textit{False Positive Reduction} = \frac{Precision_{TGNN+AE} - Precision_{TGNN}}{Precision_{TGNN}} \times 100 = \frac{0.89 - 0.85}{0.89} = 4.71\%.$$

To assess the efficiency, the distribution of the reconstruction error and the definition of the 95th percentile as the anomaly threshold were used. The graph (Fig. 5) shows how the threshold and the error distribution were calculated:

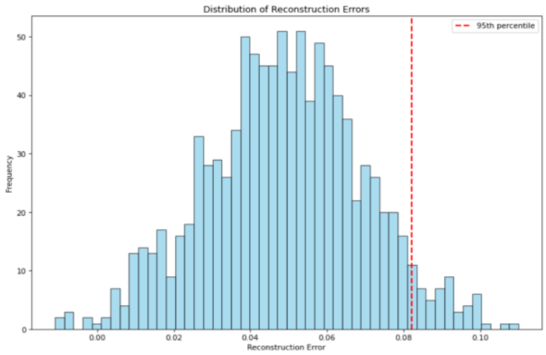


Figure 5: Distribution of reconstruction error.

This graph shows the distribution of the reconstruction error and the anomaly threshold (red dashed line), which is determined by the 95th percentile. Nodes with a reconstruction error greater than this threshold are classified as anomalous. Comparative analysis proved that the proposed complex method outperforms traditional approaches, providing higher accuracy and lower number of false positive detections.

7 CONCLUSIONS

This paper proposes a comprehensive method for anomaly detection in temporal graphs based on a combination of TGNN and a graph AE.

The TGNN captures spatiotemporal relationships, while the AE learns hidden node representations and measures deviations through reconstruction error. Experimental results show that the TGNN+AE approach improves detection accuracy by 17.3%, reduces false positives by 4.71%, and increases the F1-measure by 6.02% compared to using each method alone. A key element is setting the anomaly threshold at the 95th percentile of the reconstruction error, which adaptively identifies anomalous nodes without manual tuning, thereby enhancing stability. Overall, this integrated method effectively predicts node dynamics and identifies anomalies, paving the way for further research into adaptive anomaly detection in complex dynamic systems.

REFERENCES

[1] Statista, "Number of IoT-connected devices worldwide 2024–2033," [Online]. Available: https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/, [Accessed: Apr. 2024].

[2] Y. Wu, H. N. Dai, and H. Tang, "Graph neural networks for anomaly detection in industrial IoT," IEEE Internet of Things Journal, vol. 8, no. 3, pp. 212-223, Mar. 2021.

[3] J. Yang and Z. Yue, "Learning hierarchical spatial-temporal graph representations for anomaly detection," IEEE Transactions on Knowledge and Data Engineering, early access, 2022.

[4] Y. Feng and J. Chen, "Full graph autoencoder for one-class group anomaly detection," IEEE Internet of Things Journal, vol. 9, no. 5, pp. 4021-4030, May 2022.

[5] Y. Lai, Y. Zhu, and L. Li, "STGLR: A spacecraft anomaly detection method using spatio-temporal graph learning," Sensors, vol. 25, no. 4, pp. 1123-1137, 2025.

[6] T. N. Kipf and M. Welling, "Variational Graph Autoencoders," 2016.

[7] J. Wan, L. Cao, J. Bai, and J. Li, "Learning multiple types of features on a hybrid neural network for blockchain transaction behavior detection," SSRN, 2024.

[8] M. W. Asres, C. W. Omlin, L. Wang, P. Parygin, and D. Yu, "Data quality monitoring through transfer learning on anomaly detection for the Hadron calorimeters," 2024.

[9] J. Liu, X. Han, and X. Shang, "Spatial-Temporal Memories Enhanced Graph Autoencoder for Anomaly Detection in Dynamic Graphs," 2024.

[10] PEMS-D7 Dataset, [Online]. Available: https://paperswithcode.com/dataset/pemsd7, [Accessed: Apr. 2024].

[11] J. Qi, C. Zeng, Z. Luan, S. Huang, S. Yang, and Y. Lu, "Beyond window-based detection: A graph-centric framework for discrete log anomaly detection," preprint, 2025.

[12] M. Wen, Z. H. Chen, Y. Xiong, and Y. C. Zhang, "LGAT: A novel model for multivariate time series anomaly detection with improved anomaly transformer and learning graph structures," Neurocomputing, vol. 554, pp. 126-138, 2025.

[13] R. Hosseini, F. Simini, V. Vishwanath et al., "A Deep Probabilistic Framework for Continuous Time Dynamic Graph Generation," 2024.

[14] F. Xiao, S. Chen, Y. Ma et al., "SENTINEL: Insider threat detection based on multi-timescale user behavior interaction graph learning," IEEE Transactions on Information Forensics and Security, vol. 19, pp. 432-445, 2024.

[15] J. Li, X. Deng, and B. Yao, "Enhanced anomaly detection of industrial control systems via graph-driven spatio-temporal adversarial deep support vector data description," Expert Systems with Applications, vol. 224, p. 119899, 2025.

[16] J. Cao, X. Di, J. Li, K. Yu, and L. Zhao, "IoVST: An anomaly detection method for IoV based on spatiotemporal feature fusion," Future Generation Computer Systems, vol. 145, pp. 409-421, 2025.

Multidimensional Model of Sebastian Unger’s Idiostyle in Poetic Creativity: Corpus Analysis and NLP Methods

Tetiana Hromko and Liudmyla Panchuk
Ukrainian-German Educational and Scientific Institute, Odesa Polytechnic National University, Shevchenko Avenue 1, 65044 Odesa, Ukraine
hromkotv@gmail.com, ludapanchuk6@gmail.com

Keywords: Idiostyle, Corpus Analysis, NLP, Topic Modeling, Automated Text Analysis, Emotional Modeling, Stylometry, Sebastian Unger.

Abstract: The article presents a multidimensional model of Sebastian Unger’s idiostyle based on corpus analysis and natural language processing (NLP) methods. The study is based on a structured approach to the analysis of authorial style, comprising text certification, thematic modeling, and stylometric evaluation. A subcorpus of Unger’s texts (SPU) was created and subjected to automated processing using such methods as word vectorization (Word2Vec, TF-IDF), topic modeling (LDA, BERT), syntactic and morphological analysis, and emotional modeling (Sentiment Analysis). The results of the analysis show the presence of clear stylistic markers in Unger’s work, including metaphorical structures, fragmentary composition, dominance of expressive vocabulary, and specific syntactic models. It is found that the author’s poetry tends to the categories of “nature”, “myth”, “philosophy”, which is confirmed by thematic clustering and analysis of key concepts. The proposed methodology of corpus research allows automating the identification of the author’s style, providing a quantitative assessment of his linguistic features and opening up new perspectives for digital stylometry and authorial attribution.

1 INTRODUCTION

Modern corpus analysis of literary texts in combination with natural language processing (NLP) methods opens up new possibilities for linguistic interpretation of the author's style that were previously beyond the scope of traditional methods of literary studies. The development of automated text analysis technologies, such as deep neural networks and statistical style models, allows us to identify hidden patterns in the structure of speech and stylistic dominants of writers [4]. The introduction of such methods as thematic modeling (BERT, GPT) [1], word vectorization (Word2Vec, TF-IDF), and stylometric analysis through attribution of authorship significantly expands the possibilities of studying the individual style of a writer [7].

The importance of this approach is due not only to the growing amount of textual data, but also to the need for new methods of processing them based on mathematical and computational models [4]. A corpus-based study of literary texts, including morphological, syntactic, semantic, and stylistic analysis, allows for accurate quantitative characteristics of the author's speech, contributing to an objective assessment of its structural features. A special role is played by the author's identification of texts by linguistic markers, which is made possible by combining stylometric methods with machine learning [5]. Authorship detection through automated analysis of textual characteristics demonstrates high efficiency in recognizing individual style and differentiating between authors even in large-scale corpora.

In this context, the corpus analysis of Sebastian Unger's works [3] is a promising direction that combines linguistic methods with algorithmic modeling and allows to form a multidimensional model of his idiostyle. This study involves the systematization of the corpus of S. Unger's texts, its linguistic certification and automated processing in order to determine frequency and stylistic characteristics. Particular attention will be paid to semantic analysis using clustering models of thematic domains (Word2Vec, LDA), as well as to the analysis of emotional coloring of texts through Sentiment Analysis. The use of these methods in the study of author's style will contribute to the development of stylometric research within digital humanities and computational linguistics, providing a new level of automation of text analysis processes [6].

2 RELEVANCE AND ANALYSIS OF THE TOPIC AREA

The study of a writer's idiostyle using corpus-based methods and natural language processing (NLP) technologies opens up new perspectives for the objective analysis of linguistic features, semantic dominants, and stylistic markers of a literary text. Within the framework of modern digital linguistics, the use of automated approaches allows not only to describe the structural regularities of an author's speech, but also to verify its language model, which is important for stylometric analysis, authorial attribution, and recognition of text patterns in large data corpora.

The corpus analysis of Sebastian Unger's poetic texts involves the creation of a structured subcorpus (Subcorpus of Poetry by Sebastian Unger - SPU), which was compiled on the basis of the author's collections [8, 9, 10, 11] undergoes automated processing using tokenization, lemmatization, syntactic and semantic parsing methods. This allows us to identify patterns in the distribution of lexical units, syntactic constructions, and stylistic devices. In particular, the thematic scope of his poetry covers the concepts of memory, spatial reference, metaphysical categories, and urban chronotopes, which can be isolated and formalized through distributional analysis.

The use of NLP methods allows for a comprehensive syntactic and semantic analysis of the text, including the construction of vectorized models, thematic clustering, and network analysis of the relationships between key concepts. The corpus analysis involves systematizing the author's texts, their certification, thematic classification, and stylistic evaluation using machine learning methods. The formation of the Subcorpus of Poetry by Sebastian Unger (SPU) and its automated processing allow for a detailed analysis of frequency characteristics, identification of stylistic markers, and establishment of semantic relations between key concepts. In particular, the thematic scope of his work covers a wide range of images related to memory, space, and metaphysical reflections, which create a complex system of symbolic meanings.

The identification of stylistic patterns, the frequency of metaphorical constructions and features of the author's syntax contributes to the construction of a multidimensional model of idiostyle, which, in turn, can be used in the tasks of automatic text classification and stylometric identification. Automated corpus certification, creation of thematic maps and graph models of lexeme interrelationships allows us to describe the author's idiostyle in quantitative terms, which is an important step in the development of digital humanities and computer-aided literature analysis.

3 METHODOLOGY

The research methodology is aimed at building a multidimensional model of idiostyle through corpus analysis and NLP technologies. A subcorpus of S. Unger's texts from the Lyrikline and Open Mike platforms was formed, and the collected texts were subjected to lemmatization and cleaning. Frequency analysis allowed us to identify key lexemes, and the TF-IDF, Word2Vec, BERT, and GPT models were used for semantic clustering and analysis of thematic relations. Network analysis revealed structural relationships between images, and graph modeling allowed us to reconstruct the cognitive organization of the texts. The stylistic analysis assessed the level of syntactic complexity, the frequency of complex subordinate constructions, and the use of rhetorical figures. The idiostyle was formalized through machine learning, and style vectorization allowed us to identify its unique linguistic markers. This approach helps to verify the author's style and expand the methodological basis for stylometric and computational linguistic research.

The corpus analysis of Sebastian Unger's idiostyle is carried out by means of a multi-level structuring of texts, which involves their certification, thematic classification, stylistic analysis and semantic modeling using NLP methods. Borametz's poem “Das pflanzliche Lamm” was chosen as a research unit, which allows us to formalize the author's linguistic patterns and determine his stylistic dominants. The reproduction of the stanza structure makes it possible to trace the patterns of rhythm and syntactic organization of the text, which are key to the formation of the author's style. To ensure the automated analysis, the poem was certified, which includes genre and typological characteristics, determination of structural features, frequency characteristics, stylistic markers and semantic parameters. The morphological composition of the text was analyzed, which showed the dominance of nouns and adjectives with a high degree of expressiveness, as well as verbs in the present tense, which contributes to the dynamism of poetic speech.

The automated text analysis was performed using NLP methods [2], in particular through POS-tagging, which allowed us to determine the part-of-speech distribution and key lexical dominants. Syntactic modeling revealed the predominance of simple and parallel constructions that ensure the rhythmic organization of the text. Thematic modeling with the help of transformer models confirmed that the poem belongs to the categories NATURE, MYTH, and PHILOSOPHY, and the analysis of the emotional coloring of the text showed its neutral-positive connotation. Additionally, a graph analysis of stylistic markers was performed, which revealed a network of relationships between the key concepts of the work. The introduction of multilevel analytics made it possible to identify linguistic patterns characteristic of the author's style and to formalize them within the framework of a corpus study.

The obtained results not only allow us to characterize the individual linguistic features of Unger's poetry, but also to test a methodological approach to automated corpus analysis of author's texts. The formation of a subcorpus of texts, their certification, and the use of natural language processing methods ensure the creation of a representative model of idiostyle, which is an important step in the study of authorial strategies and the development of digital stylometry.

The experimental part of the study involves the formation of a corpus of Sebastian Unger's texts and the use of automated methods for their processing to identify characteristic linguistic and stylistic features. To ensure the accuracy of the analysis, the Subcorpus of Poetry by Sebastian Unger (SPU) was created, containing selected poetic texts by the author, structured by genre, chronological and thematic parameters. Each text underwent preliminary linguistic processing, including lemmatization, cleaning of stop words and punctuation marks, data normalization, and preparation for further machine analysis. The formation of a structured corpus makes it possible to apply algorithmic analysis methods to verify the author's idiostyle.

The main stage of the corpus processing was POS-tagging, which allowed us to determine the partial-language distribution of the text, quantitative indicators of the frequency of lexical items and their distribution in the text structure. Grammatical categories were identified using the spaCy library, which provided detailed characteristics of each word in the context of its syntactic role. The analysis of the syntactic organization of the texts made it possible to trace the ratio of simple and complex subordinate constructions, to identify the author's preferences for using certain syntactic schemes and the level of their complexity. The semantic modeling of the texts was carried out by means of thematic clustering using BERT, which made it possible to identify key concepts, their interrelationships and semantic dominants characteristic of Unger's poetic language.

An additional parameter of the analysis was the identification of stylistic markers, including expressive vocabulary, frequency stylistic constructions, inversions, repetitions, and metaphorical structures. The introduction of graph analysis made it possible to model the connections between key lexemes, forming visual representations of the dominant images in the poetic corpus. Sentiment Analysis was used to assess the emotional coloring of the texts, which helped determine the overall tone of the text, its emotional accents, and connotations. The analysis of rhythmic and phonetic parameters was carried out by identifying patterns in the length of lines, the distribution of stop phrases and pauses, which are key to the versioning structure of the works.

The obtained results make it possible to identify the unique linguistic features of Unger's texts, formalize them within the framework of a multidimensional idiostyle model, and compare them with other authorial styles. The corpus approach in combination with NLP methods provides an automated determination of the author's stylistic dominants and thematic priorities, which contributes to the further development of the methodology of digital stylometry and authorial attribution of texts [5, 7].

4 APPROACH

The corpus analysis of Sebastian Unger's idiostyle is carried out by means of a multi-level structuring of texts, which involves their certification, thematic classification, stylistic analysis and semantic modeling using NLP methods. Borametz's poem “Das pflanzliche Lamm” was chosen as a research unit, which allows us to formalize the author's linguistic patterns and determine his stylistic dominants. The reproduction of the stanza structure makes it possible to trace the patterns of rhythm and syntactic organization of the text, which are key to the formation of the author's style. To ensure the automated analysis, the poem was certified, which includes genre and typological characteristics, determination of structural features, frequency characteristics, stylistic markers and semantic parameters. The morphological composition of the text was analyzed, which showed the dominance of nouns and adjectives with a high degree of expressiveness, as well as verbs in the present tense, which contributes to the dynamism of poetic speech.

The automated text analysis was performed using NLP methods, including POS tagging, which allowed us to determine the part-of-speech distribution and key lexical dominants. Syntactic modeling revealed the prevalence of simple and parallel constructions that ensure the rhythmic organization of the text. Thematic modeling with the help of transformer models confirmed that the poem belongs to the categories NATURE, MYTH, and PHILOSOPHY, and the analysis of the emotional coloring of the text showed its neutral-positive connotation. Additionally, a graph analysis of stylistic markers was performed, which revealed a network of relationships between the key concepts of the work. The introduction of multilevel analytics made it possible to identify linguistic patterns characteristic of the author's style and to formalize them within the framework of a corpus study.

The obtained results not only allow us to characterize the individual linguistic features of Unger's poetry, but also to test a methodological approach to automated corpus analysis of author's texts. The formation of a subcorpus of texts, their certification, and the use of natural language processing methods ensure the creation of a representative model of idiostyle, which is an important step in the study of authorial strategies and the development of digital stylometry.

5 EXPERIMENT

The experimental part of the study involves the formation of a corpus of Sebastian Unger's texts and the use of automated methods for their processing to identify characteristic linguistic and stylistic features. To ensure the accuracy of the analysis, the Subcorpus of Poetry by Sebastian Unger (SPU) was created, containing selected poetic texts by the author, structured by genre, chronological and thematic parameters. Each text underwent preliminary linguistic processing, including lemmatization, cleaning of stop words and punctuation marks, data normalization, and preparation for further machine analysis. The formation of a structured corpus makes it possible to apply algorithmic analysis methods to verify the author's idiostyle.

The main stage of the corpus processing was POS-tagging, which allowed us to determine the partial-language distribution of the text, quantitative indicators of the frequency of lexical items and their distribution in the text structure. Grammatical categories were identified using the spaCy library, which provided detailed characteristics of each word in the context of its syntactic role. The analysis of the syntactic organization of the texts made it possible to trace the ratio of simple and complex subordinate constructions, to identify the author's preferences for using certain syntactic schemes and the level of their complexity. The semantic modeling of the texts was carried out by means of thematic clustering using BERT, which made it possible to identify key concepts, their interrelationships and semantic dominants characteristic of Unger's poetic language.

An additional parameter of the analysis was the identification of stylistic markers, including expressive vocabulary, frequency stylistic constructions, inversions, repetitions, and metaphorical structures. The introduction of graph analysis made it possible to model the connections between key lexemes, forming visual representations of the dominant images in the poetic corpus. Sentiment Analysis was used to assess the emotional coloring of the texts, which helped determine the overall tone of the text, its emotional accents, and connotations. The analysis of rhythmic and phonetic parameters was carried out by identifying patterns in the length of lines, the distribution of stop phrases and pauses, which are key to the versioning structure of the works.

The obtained results make it possible to identify the unique linguistic features of Unger's texts, formalize them within the framework of a multidimensional idiostyle model, and compare them with other authorial styles. The corpus approach in combination with NLP methods provides an automated determination of the author's stylistic dominants and thematic priorities, which contributes to the further development of the methodology of digital stylometry and authorial attribution of texts.

6 RESULTS

The analysis of a subcorpus of Sebastian Unger's poetic works demonstrates the possibilities of automated stylometric text processing, in particular through vectorization, frequency analysis, and modeling of semantic relations, which allows us to highlight the characteristic linguistic features of his idiostyle.

For an example of subcorpus analysis, we present the text of Sebastian Unger's poem “Borametz - Das pflanzliche Lamm” in German in a graphic record [8]:

Borametz – Das pflanzliche Lamm
Die Wurzeln tief in der Erde,
Der Stamm fest und stark,
Blätter wie grüne Hände,
Die Sonne im Blick, den Himmel im Mark.
Ein Lamm, geboren aus Pflanzenfleisch,
Mit Wolle weich und rein,
Es wächst empor im Morgenkreis,
Ein Wunder der Natur allein.
Kein Blöken hört man, keinen Laut,
Doch Leben pulsiert im Blatt,
Ein Wesen, das dem Himmel traut,
In stiller, grüner Pracht.
So steht es da, das Pflanzenlamm,
Ein Rätsel dieser Welt,
Verbindet Wurzel, Blatt und Stamm,
Wie's uns im Buche erzählt.

The graphic recording demonstrates the ascending original structural organization of the text and provides material for traditional literary and linguistic analysis of its stylistic, rhythmic, syntactic, and cognitive features. The reproduction of the stanza structure allows us to trace the patterns of rhyme, intonation pauses, and features of versioning, which are key to the analysis of the poetic discourse of Sebastian Unger.

Table 1: Textological certification of the poetic text “Borametz - Das pflanzliche Lamm ” by S. Unger.

Parameter	Description	Example (Borametz - Das pflanzliche Lamm)
Author	Full name of the author.	Sebastian Unger.
Title of the work	Title of the poetic/prose text.	*Borametz - Das pflanzliche Lamm*.
Year of publication	Date of first publication (if available).	No exact publication date, published on the Lyrikline platform.
Source	Hyperlink to the online publication or ISBN of the book.	Lyrikline.
Genre	Poetry / prose / interview.	Poetry.
Type of text	Written / oral (interview, speech).	Written.
Length (word count)	Number of words or characters in the text.	82 words.
Language	Original language.	German (DE).
Place of publication	Journal, website, book, blog.	Website Lyrikline.
Publisher	Name of the publishing house (if a printed source).	Not applicable (online publication).
Structural features	Verse/prose form, presence of stanzas, rhymes, sections.	Verse form, 4 stanzas of 4 lines each, no strict rhyme or close to assonance, meter varies, with some approximation to classical rhythm in certain lines.
Keywords	Main themes of the text (ecology, human, nature, philosophy, etc.).	Nature, mythology, plant, animal, symbolism, unity, metaphysics.
Stylistic features	Use of metaphors, symbols, expressive vocabulary.	Extensive metaphors, allegory, symbols, inversion of syntactic constructions, expressive vocabulary, personification of nature.
Phonetic features	If an oral corpus is analyzed (repetitions, pauses, emphases).	No direct phonetic expressiveness (shift of focus from phonetics to semantics), internal rhythm through image repetitions and parallelisms.
Morphological analysis	Dominant parts of speech in the text.	Predominantly nouns (*Borametz, Lamm, Wurzeln, Stamm, Blätter, Sonne*), adjectives (*tief, fest, grün, weich*), verbs in the present tense (*wächst, pulsiert, verbindet*).
Syntactic analysis	Average sentence length, complex or simple constructions.	Short and medium-length sentences, minimal use of subordinate clauses, dominant simple predicate with extended attributes.
Emotional tone	Analysis using Sentiment Analysis (neutral, positive, negative).	Neutral-elevated, leaning towards a metaphysical and mythopoetic discourse (Sentiment Analysis suggests a mixed neutral-positive evaluation).
Accessibility	Open access / paid content.	Open access on the Lyrikline platform.
Accessibility	Open access / paid content.	Open access on the Lyrikline platform.

This approach makes it possible to assess the interaction of the formal parameters of the text with the semantics and symbolism of images, as well as to study the frequency of recurring motifs within the corpus analysis. The graphical structure helps to identify the author's stylistic markers and form a multidimensional model of the poet's idiostyle, which is one of the goals of this study.

The passportization of the poem “Borametz - Das pflanzliche Lamm” by Sebastian Unger is carried out by means of a formalized description of the text by structural, stylistic and semantic parameters, which allows standardizing it for corpus analysis using NLP methods. It includes identification data, genre-typological characteristics, distribution of parts of speech, syntactic features, stylistic markers, as well as thematic modeling, mentioned in Table 1, which allows us to identify key concepts and patterns of the author's idiostyle.

The analysis of Sebastian Unger's texts involves the use of both manual and automated methods of semantic classification. The key thematic domains are identified by manually categorizing text fragments, which allows us to separate the main conceptual groups. Among them are the themes of nature (NATURE), philosophical reflections (PHIL), human experience (HUMAN), mythological and fairy-tale images (MYTH), imaginary worlds (DREAM), and temporal changes (TIME). For example, the line “Die Wurzeln tief in der Erde” is labeled as {SEM: NATURE}, which indicates the dominant natural imagery. Similarly, the phrase “Ein Rätsel dieser Welt” is labeled {SEM: PHIL}, as it reflects a philosophical understanding of existence.

To automate the thematic analysis, we use an approach based on transformer models, such as BERT, GPT, and word vectorization methods such as Word2Vec and TF-IDF. The analysis algorithm includes several stages: first, the Unger corpus is loaded in PDF, TXT, or HTML formats, followed by tokenization and lemmatization. Text fragments are vectorized by converting words into multidimensional representations, which are then clustered using K-Means or Latent Dirichlet Allocation (LDA) algorithms.

The automatic detection of semantic domains is based on a transformational architecture that ensures that sentences are categorized into their most relevant topic groups. For example, using the BERT transformer model to analyze the phrase “Die Wurzeln tief in der Erde” leads to the following result, shown on Figure 1.

```
from transformers import pipeline

classifier = pipeline("zero-shot-classification", model="facebook/bart-large-mnli")

text = "Die Wurzeln tief in der Erde."
labels = ["Nature", "Philosophy", "Dream", "Myth", "Human", "Time"]

result = classifier(text, labels)
print(result)
```

Figure 1: Zero-shot (a Python code fragment).

The output of the model shows 85% correspondence to the nature theme, which confirms the correctness of the thematic markup.

Visualization of thematic clusters allows for spatial grouping of text fragments according to their semantic characteristics. This study uses Word2Vec in combination with t-SNE projection for multidimensional analysis of thematic relations, represented on Figure 2.

```
import matplotlib.pyplot as plt
from sklearn.manifold import TSNE
import numpy as np

# Умовні вектори слів (BERT embeddings)
vectors = np.random.rand(10, 768)  # Замінити на реальні BERT-вектори
labels = ["Nature", "Philosophy", "Dream", "Myth", "Human", "Time", "Abstract", "Emotion"

tsne = TSNE(n_components=2, random_state=42)
reduced = tsne.fit_transform(vectors)

plt.figure(figsize=(10, 6))
plt.scatter(reduced[:, 0], reduced[:, 1])

for i, label in enumerate(labels):
    plt.annotate(label, (reduced[i, 0], reduced[i, 1]))

plt.title("Thematic Clustering of Texts (Word2Vec/BERT)")
plt.show()
```

Figure 2: Visualization of thematic clusters of texts using t-SNE and vector representations (Word2Vec/BERT) (in Python).

The result of this analysis is the formation of a cluster structure that reflects the relationships between text fragments according to their semantic characteristics.

The automated analysis of Unger's corpus is carried out using NLP models that perform part-of-speech tagging (POS-tagging), syntactic and stylistic features, and thematic classification, shown on Fig. 3.

This approach allows not only to classify Unger's texts by thematic domains, but also to carry out an automated analysis of the stylistic and syntactic features of his work.

It is the use of transformational models and clustering algorithms in the corpus analysis of Sebastian Unger's texts that makes it possible to automate the process of identifying key themes, stylistic features, and cognitive dominants in his poetry and prose works. This opens up new perspectives in stylometric research and digital literary studies.

```
import matplotlib.pyplot as plt
from sklearn.manifold import TSNE
import numpy as np

# Умовні вектори слів (BERT embeddings)
vectors = np.random.rand(10, 768)  # Замінити на реальні BERT-вектори
labels = ["Nature", "Philosophy", "Dream", "Myth", "Human", "Time", "Abstract", "Emotion"

tsne = TSNE(n_components=2, random_state=42)
reduced = tsne.fit_transform(vectors)

plt.figure(figsize=(10, 6))
plt.scatter(reduced[:, 0], reduced[:, 1])

for i, label in enumerate(labels):
    plt.annotate(label, (reduced[i, 0], reduced[i, 1]))

plt.title("Thematic Clustering of Texts (Word2Vec/BERT)")
plt.show()
```

Figure 3: Thematic grouping of texts based on BERT representations and t-SNE visualization (in Python).

The model of multidimensional analysis of Sebastian Unger's idiostyle is based on a corpus-based approach that combines structured text certification, automated data processing, and analysis of stylistic and semantic features based on NLP methods. The basis of the study is an extended parameterization system that formalizes the text through unified markers that allow us to determine the morphological, syntactic, semantic, and stylistic properties of the corpus of Unger's works. In accordance with the parameterization table, presented below as Table 2, the analysis is carried out through POS-tagging, which includes partial language distribution and identification of key lexical units, syntactic classification, which determines the structural features of the text, semantic modeling, which allows to identify dominant themes and concepts, and the stylistic level, which reveals the figurative system of the works. idiostyle works.

Automated corpus profiling of Sebastian Unger allows for identifying the patterns of his idiolect through the analysis of linguistic element frequency, the distribution of syntactic structures, and stylistic markers. One of the key aspects of this model is its structured nature, which ensures a systematic description of the text through standardized categories. This approach facilitates an accurate analysis of the writer’s style, enabling not only qualitative description but also the examination of his texts within the framework of statistical models of stylometric analysis. The use of a machine-readable format allows the corpus to be integrated into various NLP applications such as NLTK, spaCy, and BERT, expanding the possibilities of natural language processing. Specifically, the automatic analysis system enables text clustering by thematic domains, the evaluation of syntactic patterns, and the identification of Unger’s stylistic features based on machine learning algorithms.

The flexibility of the model allows for the adaptation of the parameter system to specific research tasks, enabling the incorporation of new features, including stylistic and cognitive characteristics. Semantic analysis involves identifying the thematic categories of a text, covering concepts of nature, philosophy, social aspects, abstract notions, and spatial characteristics. The determination of structural and morphological parameters is carried out through the identification of syntactic constructions and the frequency distribution of parts of speech, which helps to distinguish the author's stylistic markers. The stylistic characteristics of texts are defined through the analysis of metaphorical and expressive devices, which contribute to the creation of a unique poetic writing style.

Table 2: Classification parameters for automated analysis of poetic texts of the corpus.

Category	Notation	Description
Morphological parameters (POS)	NOUN, VERB, ADJ, ADV, PRON, MOD, PREP	Parts of speech: noun, verb, adjective, adverb, pronoun, modal verb, preposition.
Syntactic features (SYN)	SIMPLE, COMPLEX, PARALLEL, PASSIVE, ELLIPSIS	Type of construction: simple sentence, complex sentence, parallel structure, passive voice, elliptical structures.
Semantic level of analysis (SEM)	NATURE, PHILOSOPHY, HUMAN, ABSTRACT, SPACE	Thematic categories of the text: nature description, philosophical motifs, human and society, abstract concepts, spatial characteristics.
Stylistic parameters (STY)	FIGMET, FIGSIM, RHET, EXPR	Key expressive means: metaphors, comparisons, rhetorical questions, expressive vocabulary.
Additional structural parameters	SENT, LEN (SHORT, MID, LONG), WC	Sentence length (in words), length categorization (short, medium, long), total word count in the text.

The development of a multidimensional model of idiolect involves integrating all the aforementioned parameters into a comprehensive analysis that includes corpus profiling, stylistic mapping, semantic modeling, and automated analysis of linguistic features using NLP algorithms. This approach not only delineates the individual linguistic traits of Sebastian Unger but also constructs a representative model of his creative style. The clustering of texts based on syntactic and stylistic features facilitates the identification of recurring constructions and distinctive linguistic techniques that shape his poetic idiolect. Such an approach enables an objective determination of the patterns underlying the formation of the author’s linguistic worldview, the identification of key elements of his creative methodology, and a comparative analysis of his poetry within the broader context of stylistic experimentation in contemporary German literature. The main functional advantages of the formalized approach to the stylistic analysis of texts in the corpus study of Sebastian Unger’s poetry are summarized in Table 3.

Table 3: Functional advantages of the formalized approach to the stylistic analysis of texts in the corpus study of Sebastian Unger’s poetry justify.

Characteristic	Description
Structuring	Systematic text description through standardized markers, enabling quick retrieval of generalized characteristics such as part-of-speech distribution (POS), types of syntactic constructions (SYN), thematic dominants (SEM), and stylistic features (STY). Ensures precise and representative text analysis.
Machine readability	Utilization of the format in NLP applications such as NLTK, spaCy, and BERT. Facilitates automated data extraction, word frequency analysis, syntactic structure identification, sentiment analysis, thematic group clustering, and stylometric studies using machine learning algorithms.
Flexibility	Ability to expand the parameter system according to research needs: adding stylistic features, phonetic characteristics, or cognitive markers. Adaptable to various methodological approaches (cognitive linguistics, computational stylistics, comparative literary studies, semantic modeling).

7 CONCLUSIONS

The study of Sebastian Unger's idiolect based on corpus analysis and NLP methods has enabled a multidimensional description of his linguistic and stylistic specificity. The use of modern digital technologies has facilitated the automated analysis of the author's lexicon, syntax, semantic dominants, and rhetorical devices, making it possible to formalize a model of the poet's idiolect. The main findings of the research include identifying the key lexico-semantic features of Unger’s texts through frequency analysis and thematic grouping of lexemes, detecting dominant syntactic structures that characterize his style – particularly a tendency toward complex syntactic constructions and inversions – as well as formalizing intertextual connections in his works, which indicate postmodernist traits in his poetry. Furthermore, the development of graph-based models of relationships between key concepts in the author's texts has allowed for the tracing of the cognitive structure of his creativity, while the verification of his idiolect using machine learning algorithms has confirmed its uniqueness within the contemporary German-language poetic landscape.

The obtained results demonstrate the potential of corpus analysis and NLP methods in the study of writers' idiolects, opening new prospects for automated stylometric analysis. The use of computational technologies in linguistic and literary research contributes to a more precise and objective understanding of an author's writing style, enables the comparison of stylistic characteristics across different writers, and simplifies the identification of unique linguistic patterns. Expanding this approach allows not only for the analysis of individual authors' works but also for the evaluation of broader trends in the development of literary styles in the digital age.

A promising direction for further research is the comparative analysis of Unger’s work alongside other representatives of contemporary German-language poetry, which would provide a deeper understanding of his artistic style within the broader context of literary tradition. Additionally, refining the methodology of corpus analysis and integrating more advanced neural networks for text analysis could further enhance the precision of identifying stylistic characteristics and cognitive features of an author’s writing. The incorporation of multimodal approaches, including the analysis of audiovisual aspects of poetic texts, also represents a valuable step in the study of idiolects. Thus, the integration of corpus linguistics, computational literary studies, and artificial intelligence has the potential to significantly expand the horizons of modern research on literary language.

REFERENCES

[1] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of NAACL-HLT, pp. 4171-4186, 2019.

[2] D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed. Pearson, 2021.

[3] KAS Literatur, "Sebastian Unger," [Online]. Available: https://www.kaschlit.de/autorinnen/sebastian-unger.

[4] I. Khomytska, V. Teslyuk, N. Kryvinska, and I. Bazylevych, "Software-Based Approach Towards Automated Authorship Acknowledgement - Chi-Square Test on One Consonant Group," Electronics, vol. 4, no. 7, p. 1138, Jul. 2020, [Online]. Available: https://doi.org/10.3390/electronics9071138.

[5] M. Koppel, J. Schler, and Sh. Argamon, "Authorship attribution in the wild," Language Resources and Evaluation, vol. 45, no. 1, pp. 46-52, 2011, [Online]. Available: https://doi.org/10.1007/s10579-009-9111-2.

[6] D. Madigan, A. Genkin, D. D. Lewis, Sh. Argamon, D. Fradkin, and Ye. Li, "Author Identification on the Large Scale," in AIP Conference Proceedings, vol. 803, pp. 509-5013, 2005, [Online]. Available: https://doi.org/10.1063/1.2149832.

[7] E. Stamatatos, "Authorship attribution using text distortion," in Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, vol. 1, pp. 1138-1149, 2017.

[8] S. Unger, "Borametz – Das pflanzliche Lamm," Lyrikline, [Online]. Available: https://www.lyrikline.org/de/gedichte/borametz-das-pflanzliche-lamm-10652.

[9] S. Unger, Das Pferd als sein eigener Reiter: Essays zum Ende der Natur. Berlin: Matthes & Seitz Berlin, 2024.

[10] S. Unger, "Die Tiere wissen noch nicht Bescheid," Open Mike, Apr. 18, 2018, [Online]. Available: https://www.openmikederblog.de/2018/04/18/new-readings-sebastian-unger-die-tiere-wissen-noch-nicht-bescheid/.

[11] S. Unger, Über die Dächer abwärts. Berlin: Matthes & Seitz Berlin, 2024.

Advanced Software Package for Estimatiom Boundary Trajectories of Electron Beams and Other Practical Applications

Igor Melnyk[1], Mykhailo Skrypka[1], Alina Pochynok[1] and Olga Demyanchenko[2]
[1]*National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Beresteiska Avenue 37, 03056 Kyiv, Ukraine*
[2]*Institute of Applied Mathematics and Fundamental Sciences of the National University "Lviv Polytechnic", Mytropolyta Andreia Str. 5, 79016 Lviv, Ukraine*
imelnik@phbme.kpi.ua, scientetik@gmail.com, alina_pochynok@yahoo.com, olgademyanchenko@gmail.com

Keywords: Simulation, Interpolation, Extrapolation, Ravine Dependence, Stiff Dependence, Root-Polynomial Function, Deviation.

Abstract: The article describes advanced computer software for numerical simulation, interpolation, extrapolation, and approximation of electron beam boundary trajectories. From a mathematical point of view, the distinguishing feature of the proposed software is the use of different-order root-polynomial functions for estimating the boundary trajectories of electron beams. The accuracy of estimation is very high. The level of interpolation and extrapolation errors is a range of a few percent, and in some cases, it is a fraction of a percent. For obtaining a minimal level of error estimation and guaranteeing the convergence of the choosing function, the method of including the positive deviation into root-polynomial functions is proposed and verified. Corresponding examples are also given. Particularities of realizing computer software in the Python programming language are also considered. Using manual input of digital data for interpolation and approximation is also possible. Elaborated computer software is supplied by a well-developed graphic user interface, and for simplicity of working with it, different simulation tasks are located in the separate bookmarks. Transferring numerical data between different tasks using the corresponding interface components is also possible. The program interface is created using advanced means of Python programming and software standards, and for this reason, it is easy for understanding by users and working with it. Therefore, elaborated computer software can also be used for solving other significant practical tasks connected with interpolation, approximation, and extrapolation of stiff ravine dependences.

1 INTRODUCTION

Solving the problems of interpolation, extrapolation and approximation of stiff ravine dependences is very important task today. Analyses of such dependences is often necessary for solving different optimization task in problems of physics, economy, psychology, social science and other. Therefore, analyzing of ravine dependences is the special problem in the data science and computer algorithms, including egression analyze [1 – 3].

A special task in this aspect is analyzing the boundary trajectories of Electron Beams (EB). Generally, this task is corresponded to basic laws of electro-physics and electron optics [4 – 10], and have been considered in papers [11 – 17]. Generally, this approach firstly has been applied to research and further development of high voltage glow discharge electron guns, which are widely applied in industry today in electronics [18 – 20] and mushing-building industry [21 – 33]. Generally, principle of operation such type of electron guns is given in papers [23 – 33].

The approach of estimation short-focus EB trajectories, proposed in papers [11 – 17], is based on using for interpolation, approximation and extrapolation of the boundary trajectory of electron beam, former by of high voltage glow discharge electron guns, by root-polynomial functions (RPF). General conception of interpolation by such functions is described in paper [11, 12], and it has been pointed out in this works, that using this approach by the suitable position of reference point in the symmetric ravine date sets obtaining high accuracy of interpolation, range of fraction of percent relatively to obtained numerical data, is really possible. Examples of data approximation using this approach are given in papers [13, 14]. Generally, it based on two different approaches: approximation by the reference points [14] and by the tangents [17]. In the papers [15, 16] the method of extrapolation of EB trajectories and advanced method of interpolation using slightly different location of reference points, generally taking from the extrapolation task [16], have been described.

Theoretical background of using proposed interpolation and approximation methods, based on the mathematical particularities of RPF and they derivatives, is given in paper [16]. In the work [18] on the base of proposed theoretical approach another possible application of RPF was described. The main applications of this approach, have been considered in [18], are probability theory tasks [34, 35] and fuzzy-logic tasks [36 – 42].

But as noted in [11, 12], the magnitude of the interpolation error strongly depends on the position of the base points in the numerical data sets, and if the value of the interpolated set near the minimum is close to zero, then the RPFs usually diverge. Basic structure of elaborated software package has been described in work [15, 16]. Some practical recommendations in terms of avoiding the problem of RPF divergence, which was pointed out in works [15, 16], have been given in [17], however this problem has not been completely solved till today.

Therefore, the purpose of this paper is to describe the included changes in the developed software package that allow obtaining guaranteed convergence of the algorithms for calculating RPF coefficients when solving practical problems related to the interpolation and extrapolation of stiff ravine digital data sets. Some corresponding examples are also given.

2 STATEMENT OF PROBLEM

As it has been pointed out in the papers [11, 12], in the general form RPF for the known set of basic points on longitudinal coordinate z and transversal coordinate r $\{P_1(z_1, r_1), P_2(z_2, r_2), \ldots, P_{n+1}(z_{n+1}, r_{n+1})\}$ is written as follows:

$$r_b(z) = \sqrt[n]{C_n z^n + C_{n-1} z^{n-1} + \ldots + C_1 z + C_0}, \tag{1}$$

where where n is the degree of the polynomial and the order of the root-polynomial function, and $C_0 - C_n$ are the polynomial coefficients.

Generally, the problem of defining polynomial coefficients is very simple and, with known coordinate of basic points **P** it led to solving the set of linear equation [11, 12]:

$$\begin{cases} C_n z_1^n + C_{n-1} z_1^{n-1} + \ldots + C_1 z_1 + C_0 = r_1^n; \\ C_n z_2^n + C_{n-1} z_2^{n-1} + \ldots + C_1 z_2 + C_0 = r_2^n; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots; \\ C_n z_n^n + C_{n-1} z_n^{n-1} + \ldots + C_1 z_n + C_0 = r_n^n; \\ C_n z_{n+1}^n + C_{n-1} z_{n+1}^{n-1} + \ldots + C_1 z_{n+1} + C_0 = r_{n+1}^n. \end{cases} \tag{2}$$

Clear, that the set of equation (2) is linear and can be simply solved analytically using well-known Gauss – Seidel methods of exclusion the variables [43, 44]. The analytical relations, which have been obtained for calculation the coefficients $C_0 - C_n$ for the RPF from second to sixth order are given in the works [14 – 17]. Generally, now namely these relations are used for calculation the polynomial coefficients in elaborated computer software Electron Beam Trajectories Interpolation, Approximation and Extrapolation (EBTIAE). Basic structure of these computer software is given in works [15 – 17].

But as have been pointed out in the works [15,16], for small values of r in the region of minimum proposed algorithms of calculation are usually divergence with the small values of r near the region of minimum. Corresponded result, has been obtained manually for the set of basic points (r; z): (0; 0.01), (0.01; 0.007), (0.02; 0.003), (0.03; 0.001), (0.04; 0.003), (0.05; 0.007), (0.06; 0.01), is shown in Fig. 1. In this dataset all values on z coordinate are given in meters, and on r coordinate – in millimeters. Clear that value r_4 = 0.001 mm = 10^{-6} m in the region on minimum is very small. The result, presented in Figure 1, has been obtained for sixth-order RPF.

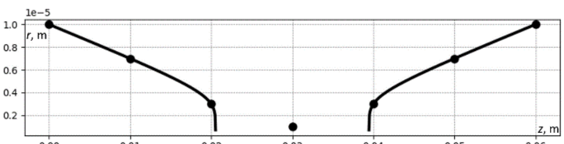


Figure 1: Illustration of divergence the algorithm of calculation the sixth-order RPF coefficients for ravine dataset with extra-small value near the minimum region.

In general, this problem is explained simply by the fact that in relation (1) the polynomial is under the root of degree n, and accordingly, for negative values of the coordinate r, the given problem generally has no solution. Therefore, it is obvious that the condition $r > 0$ is necessary for the convergence of the algorithms for calculating the RPF coefficients, but, unfortunately, it is not sufficient. As the conducted studies have shown, for values of r close to zero, the usually proposed computational algorithms still diverge. This problem has been generally formulated in the works [16, 17], but correct solution for its solving wasn't find rill today. It also should be pointed out, that generally for the stiff functions finding the minimum, which value is close to zero, is always the sophisticated problem for different numerical method and such task usually has to be solved very carefully with analyzing the precision of provided iterative calculations [45 – 49].

Therefore, advanced method for obtaining guaranteed convergence of calculation RPF coefficients by analytical relations, which have been obtained early, will be proposed and analyzed in the next section of this work.

3 USE DEVIATION AND ADVANCED FORM OF ROOT-POLYNOMIAL FUNCTION

3.1 Basic Conception and Corresponding Example

Really, the solution is very simple, when the relation (1) is rewritten in advanced form with positive deviation δ as follows:

$$r_b(z) = \sqrt[n]{C_n z^n + C_{n-1} z^{n-1} + \ldots + C_1 z + C_0} - \delta. \qquad (3)$$

In this case the set of basic points for calculation polynomial coefficients is rewritten with deviation as follows:

$\{P_1(z_1, r_1 + \delta), P_2(z_2, r_2 + \delta), \ldots, P_{n+1}(z_{n+1}, r_{n+1} + \delta)\}$.

In Figure 2 presented the result of interpolation for the same set of points, as in pervious section, but using deviation $\delta = 0.01$. It is clear, that in this case divergence of RPF is not observed and solution, which has been obtained, is generally correct.

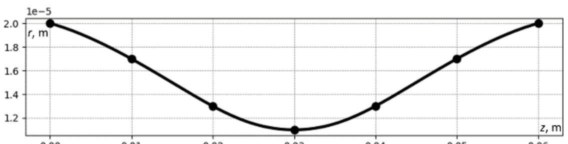


Figure 2: Illustration of convergence the algorithm of calculation the sixth-order RPF coefficients for ravine dataset with extra-small value near the minimum region in the case of using deviation $\delta = 0.01$.

Corresponding sixth-order RPF is written as follows:

$$r(z) = \sqrt[6]{\begin{array}{c} -7.2608025 \cdot 10^{-20} z^6 + 1.30694445 \cdot 10^{-20} z^5 - \\ -8.59362525 \cdot 10^{-22} z^4 + 2.4706836 \cdot 10^{-23} z^3 - \\ -2.10426 \cdot 10^{-25} z^2 - 3.616736445 \cdot 10^{-27} z + \\ +6.410 \cdot 10^{-29} \end{array}} - 0.01.$$

3.2 Use Deviation for Interpolation and Extrapolation the Boundary Trajectory of Electron Beams

Generally, the stiff ravine function with the values, which in the region of minimum are close to zero, are corresponding to the trajectory of EB in the case of high acceleration voltage and small beam current [4 – 10]. Certainly, using deviation in such case for guaranteed convergence of calculation RPF coefficients is necessary.

In the papers [15, 16], it has also been pointed out that a small error of interpolation using RPF is possible only in the case of a treatment-symmetric ravine dataset. In the case of asymmetry, the value of the error is, generally, greatly raised.

On the physical point of view, the boundary trajectory of short-focus EB, propagated in ionized gas, is described by following set if algebra-differential equations [4 – 10, 14]:

$$f = \frac{n_e}{n_{i0} - n_e}; C = \frac{I_b\left(1 - f - \beta^2\right)}{4\pi\varepsilon_0 \sqrt{\frac{2e}{m_e}} U_{ac}^{1.5}}; \frac{d^2 r_b}{dz^2} = \frac{C}{r_b}; \theta = \frac{dr_b}{dz} + \theta_s;$$

$$n_e = \frac{I_b}{\pi r_b^2};\ v_e = \sqrt{\frac{2eU_{ac}}{m_e}};$$

$$n_{i0} = r_b^2 B_i p n_e \sqrt{\frac{\pi M \varepsilon_0 n_e}{m_e U_{ac}}} \exp\left(-\frac{U_{ac}}{\varepsilon_0 n_e r_b^2}\right); \qquad (4)$$

$$\gamma = \sqrt{1 - \beta^2};\ \tan\left(\frac{\theta_{\min}}{2}\right) = \frac{10^{-4} Z_a^{4/3}}{2\gamma\beta^2};\ \tan\left(\frac{\theta_{\max}}{2}\right) = \frac{Z_a^{3/2}}{2\gamma\beta^2};$$

$$\overline{\theta^2} = \frac{8\pi\left(r_b Z_a\right)^2 dz}{n_e} \ln\left(\frac{\theta_{\min}}{\theta_{\max}}\right),\ \beta = \frac{v_e}{c},$$

where U_{ac} is the accelerating voltage, I_b is the EB current, p is the residual gas pressure in the region of EB propagation, z is the longitudinal coordinate, dz is the electron path in the longitudinal direction at the current iteration, r_b is the radius of the boundary trajectory of the electron beam, n_e is the beam electrons' concentration, n_{i0} is the concentration of residual gas ions on the beam symmetry axis, ε_0 is permittivity, v_e is the average velocity of the beam electrons, m_e is the electron mass, c is the light velocity, γ is the relativistic factor, f is the residual gas ionization level, B_i is the gas ionization level, $\theta_{\min}$ and $\theta_{\max}$ are the minimum and maximum scattering angles of the beam electrons, corresponding to Rutherford model, $\bar{\theta}$ is the average scattering angle of the beam electrons, Z_a is the nuclear charge of the residual gas atoms.

Corresponded analytical relation for error estimation is follows [11, 12]:

$$\varepsilon(z) = \frac{\left| r_{num}(z) - r_E(z) \right|}{r_{num}(z)} \cdot 100\%, \qquad (4)$$

where r_E is the result of estimation using relation (1) and set of equations (2), r_{num} is the result of solving the set of algebra-differential equation (4) using four order Runge – Kutt method [43 – 47].

Let’s considering now the tasks of interpolation and extrapolation boundary trajectory of EB with such simulation parameters: U_{ac} = 25 kV, I_b = 0.6 A, p = 7 Pa, start EB radius r_{b0} = 3 mm, start longitudinal coordinate z_0 = 0.1 m, start angle of EB convergence α = 12^0, last longitudinal coordinate z_0 = 0.15 m, number of calculated points N = 120000. Corresponding simulation result is given at Figure 3.

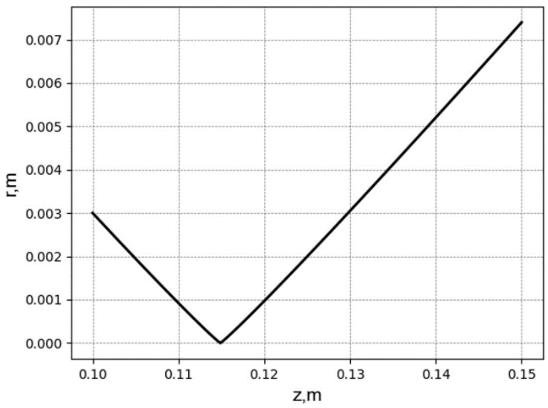


Figure 3: Illustration of solving simulation task for U_{ac} = 25 kV, I_b = 0.6 A, and r_{b0} = 3 mm.

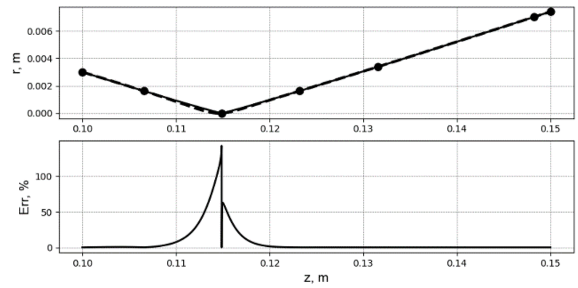
Clear, that EB trajectory in this case is very stiff and, generally, similar to the function of absolute value $r(z) = 0.2179 \cdot |z - 0.1148|$. But, in contrary on absolute value function, the derivative in the point of minimum $z_{\min}$ = 0.1148 m is $\frac{dr}{dz} = 0$.

The result of solving interpolation task for such parameters of EB is given in Figure 4a, the deviation is δ = 2.5 mm. Since ravine function is left-hand asymmetric, the significant error is existed in the region of minimum, range of 100 %. In contrary, for combined interpolation-extrapolation task [15, 16], minimal error is significantly smaller, range of 3 – 4 %, but it also corresponding to the region of EB focus position. The value of deviation for this task is δ = 1.4 mm. Graphic dependences, have been obtained for this combined interpolation-extrapolation task, are given in Figure 4b.

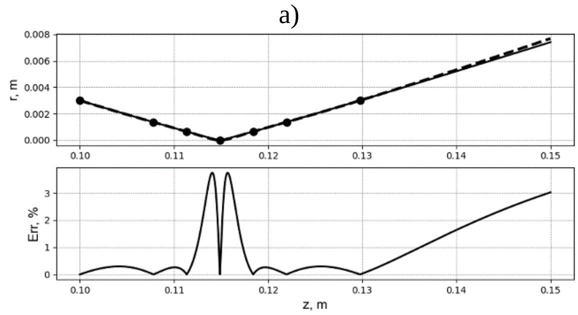
In the case of solving combined interpolation-extrapolation task obtained sixth order RPF is written as follows:

$$r(z) = \sqrt[6]{\begin{array}{c} 2.332142 \cdot 10^{-4} z^6 - 1.607568 \cdot 10^{-4} z^5 + \\ +4.6247388 \cdot 10^{-5} z^4 - 7.107475 \cdot 10^{-6} z^3 + \\ +6.1542275 \cdot 10^{-7} z^2 - 2.84664851 \cdot 10^{-8} z + \\ +5.495177647 \cdot 10^{-10} \end{array}} - 1.4.$$

Clear, that, in any case maximal relative interpolation error correspond to the region of minimum because the numerical values of function r_{num} in this region are very small. Therefore, such result is in agreement with relation (4).



a)



b)

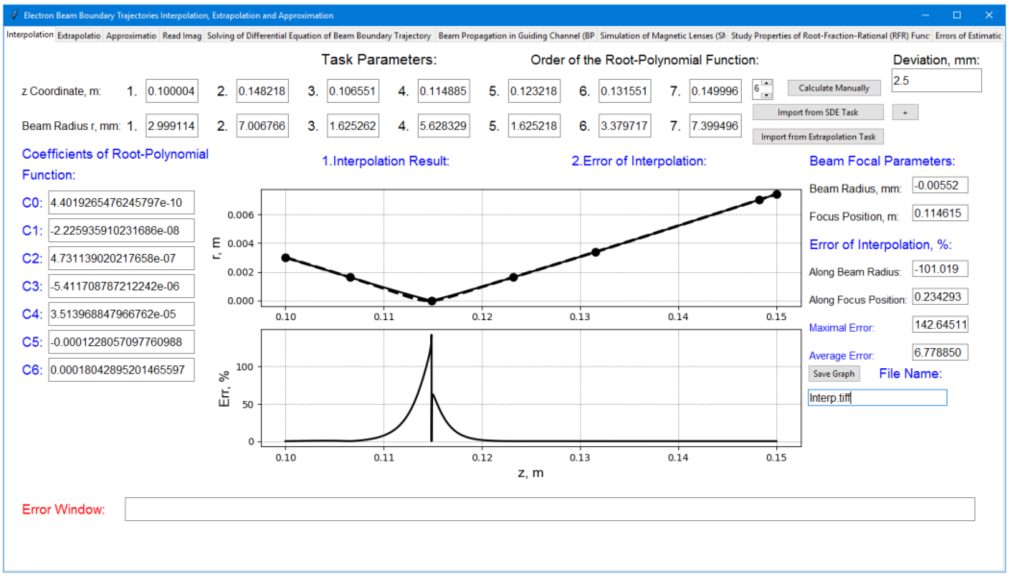
Figure 4: Illustration of solving interpolation task (a) and combined interpolation-extrapolation task (b) for EB boundary trajectory with simulation task parameters: U_{ac} = 25 kV, I_b = 0.6 A, and r_{b0} = 3 mm.

Another approach to refining the parameters of relativistic EB in powerful accelerators in the strong microwave electromagnetic fields has been proposed in papers [50, 51]. This approach is based on using microcomputers and careful statistical analyze of experimental results.

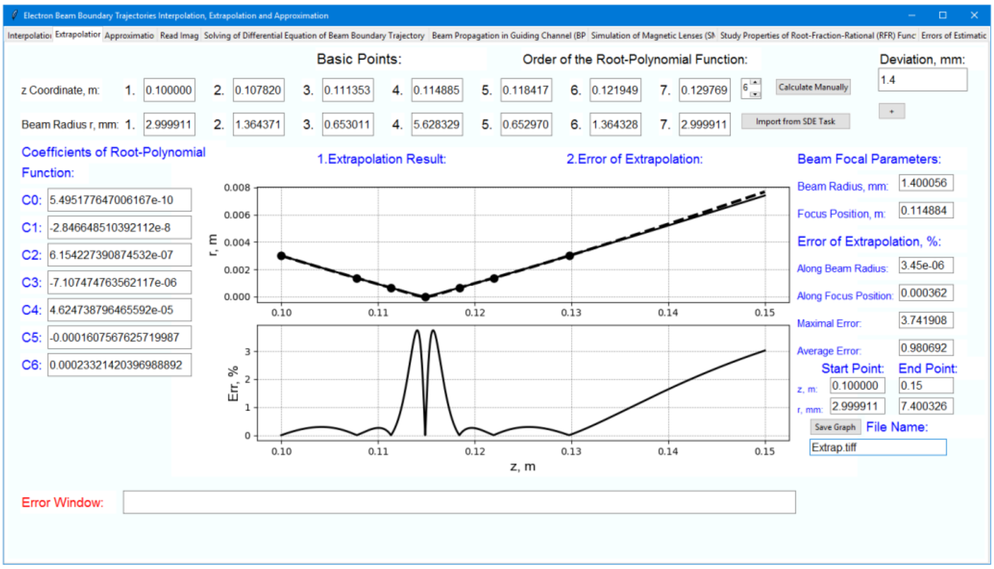
4 PARTICULARITIES OF ELABORATED COMPUTER SOFTWARE

Generally, the particularities of elaborated software package EBTIAE, created using the advanced means of Python programing language [52 – 55], just have been considered early in works [13 – 17]. The package EBTIAE is created by the means of functional programing in one module, and for providing sophisticated numerical calculations and creating scientific graphics advanced Python libraries, such as `numpy` and `matplotlib` has been used [52 – 55]. From the library `numpy` has been effectively used the advanced means of large digital arrays treatment, including methods of matrix programming. And for realizing data transferring between different tasks the means of describing global variables have been used.

But the main particularity of the elaborated EBTIAE package is the location of all independent tasks on the separate tabs of the canvas, formed in the graphic interface window. And in these tabs the specific textboxes are located for input numerical data and the buttons for providing calculations and data transferring [14 – 16]. Generally, such approach is fully corresponded to conception of functional programming [52 – 55]. And since new function of using deviation has been realized in the EBTIAE package, corresponding tabs “Interpolation” and “Extrapolation” have been slightly modified. The appearances of these two tabs on the computer screen with all the elements of the interface window are shown in Figure 5.



a)



b)

Figure 5: The tabs “Interpolation” (a) and “Extrapolation” (b) of elaborated computer software EBTIAE. Screen copies.

The main differences of these tabs from the figures, given in works [14 – 16], are the appearance of the textbox “Deviation” and the button “+”, which are located in both tabs in the upper right corner. In the textbox “Deviation” the corresponding value of deviation has to be typing, and in case of pressing on the button “+” the typing value of deviation is added to the basic points on *r* coordinate, which are located on the top of both tabs in the second row. But the button “+” button included only for providing manual calculations. Calculation the boundary trajectory of EB is provided by the same way, as early, namely, after events of pressing to the buttons “Import from SDE Task”, which are also located in both tabs in the upper right corner (see Figure 5a and 5b).

The deviation value is read from the text window and passed to the function for calculating the EB trajectory. If the value of this parameter is incorrect, a corresponding error message is issued. Thus, only minor changes have been introduced into the interface of the developed computer software, which simplifies the work of ordinary users with it.

5 CONCLUSIONS

The conducted studies have shown that, in general, with the correct choice of the deviation value, the minimum value of the interpolation error and extrapolation of the EB boundary trajectory can be achieved using the RPF of different orders, from the second to the sixth. In reality, the reduction in the interpolation error with using deviation can be by several orders of magnitude. But, most importantly, by correctly selecting the deviation, it was possible to completely solve the problem of the RPF divergence at radial coordinate values close to zero.

The corresponding interpolation and extrapolation RPF were obtained for all EB parameters, including for problems with a high degree of rigidity, which correspond to high values of the accelerating voltage and low beam current.

The use of functional programming tools in the Python language allowed us to make appropriate changes to the program interface with minor changes in the algorithms and computational procedures.

Interesting, that it is also possible to interpolate and extrapolate numerical data in manual mode, which significantly expands the range of tasks to be solved by elaborated computer software. In particular, using the developed software, it is possible to solve problems of probability theory and fuzzy logic, where functional dependencies with a high degree of rigidity are often used for accurate data processing.

The obtained theoretical and practical results are of great practical interest to a wide range of specialists in the fields of interpolation and extrapolation theory, computational algorithms, and processing of large data arrays.

ACKNOWLEDGMENTS

All scientific research described in this work has been carried out in the Laboratory of Electron Beam Technological Devices of the National Technical University of Ukraine "Igor Sikorsky Kiev Polytechnic Institute" as part of scientific and technical research works commissioned by the Ministry of Education and Science of Ukraine, state registration number is 0124U001525.

REFERENCES

[1] N. Draper and H. Smith, Applied Regression Analysis, 3rd ed. Wiley Series, 1998, 706 p. [Online]. Available: https://doi.org/10.1002/9781118625590.

[2] D. W. Hosmer Jr., S. Lemeshow, and R. X. Sturdivant, Applied Logistic Regression. Wiley, 2013, 528 p. [Online]. Available: https://doi.org/10.1002/9781118548387.

[3] D. P. McGibney, Applied Linear Regression for Business Analytics with R: A Practical Guide to Data Science with Case Studies. International Series in Operations Research & Management Science, vol. 337, 2023, 440 p.

[4] S. Schiller, U. Heisig, and S. Panzer, Electron Beam Technology. New York: John Wiley & Sons, 1982, 509 p.

[5] M. Szilagyi, Electron and Ion Optics. Springer Science & Business Media, 2012, 539 p.

[6] S. J. R. Humphries, Charged Particle Beams. Courier Corporation, 2013, 834 p.

[7] E. Kasper and P. Hawkes, Principles of Electron Optics: Applied Geometrical Optics. Elsevier Science, 1989, 1188 p.

[8] P. Grivet, P. W. Hawkes, and A. Septier, Electron Optics. Elsevier, 2013, 412 p.

[9] J. Orloff, Handbook of Charged Particle Optics. CRC Press, 1997, 528 p.

[10] Electron Beams, Lenses, and Optics, A. B. El-Kareh, Ed. Academic Press, 2012, 428 p.

[11] I. Melnyk, S. Tuhai, and A. Pochynok, "Interpolation of the Boundary Trajectories of Electron Beams by the Roots from Polynomic Functions of Corresponded Order," in 2020 IEEE 40th International Conference on Electronics and Nanotechnology (ELNANO), 2020, pp. 28-33. [Online]. Available: https://ieeexplore.ieee.org/document/9085228.

[12] I. V. Melnyk and A. V. Pochynok, "Study of a Class of Algebraic Functions for Interpolation of Boundary

Trajectories of Short-Focus Electron Beams," System Researches and Information Technologies, no. 3, pp. 23-39, 2020.

[13] I. Melnyk and A. Pochynok, "Basic Algorithm for Approximation of the Boundary Trajectory of Short-Focus Electron Beam Using The Root-Polynomial Functions of the Fourth and Fifth Order," System Research and Information Technologies, no. 3, pp. 127-148, 2023.

[14] I. Melnyk and A. Pochynok, "Theoretical justification of application possibility of different order root-polynomial functions for interpolation and approximation of boundary trajectory of electron beam," Radioelectronics and Communications Systems, vol. 66, no. 1, pp. 53-73, 2023. [Online]. Available: https://doi.org/10.3103/S073527272302002.

[15] I. Melnyk, A. Pochynok, and M. Skrypka, "Comparison of methods for interpolation and extrapolation of boundary trajectories of short-focus electron beams using root-polynomial functions," System Research and Information Technologies, no. 3, pp. 77-92, 2024. [Online]. Available: https://doi.org/10.20535/SRIT.2308-8893.2024.3.05.

[16] I. Melnyk, A. Pochynok, and M. Skrypka, "An advanced method of interpolation of short-focus electron beams boundary trajectories using higher-order root-polynomial functions and its comparative study," System Research and Information Technologies, no. 4, pp. 133-153, 2024. [Online]. Available: https://doi.org/10.20535/SRIT.2308-8893.2024.4.11.

[17] I. Melnyk, S. Tuhai, M. Skrypka, and A. Pochynok, "Membership Root-Polynomial Functions," in Digital Ecosystems: Interconnecting Advanced Networks with AI Applications, M. Klymash et al., Eds. Lecture Notes in Electrical Engineering, vol. 1198, Springer, 2024, pp. 364-390. [Online]. Available: https://doi.org/10.1007/978-3-031-61221-3.

[18] A. A. Druzhinin, N. S. Liakh-Kaguy, I. P. Ostrovskii, Y. M. Khoverko, and K. Rogacki, "Superconductivity and Kondo effect of PdxBi2Se3 whiskers at low temperatures," Journal of Nano- and Electronic Physics, vol. 9, no. 5, p. 05013, 2017. [Online]. Available: https://jnep.sumdu.edu.ua/en/full_article/2307.

[19] A. Druzhinin, I. Ostrovskii, Y. Khoverko, K. Rogack, and N. Liakh-Kaguy, "Superconductivity and weak localization of PdxBi2Se3 whiskers at low temperatures," Applied Nanoscience (Switzerland), vol. 8, no. 4, pp. 877-883, 2018. [Online]. Available: https://doi.org/10.1007/s13204-018-0671-6.

[20] A. A. Druzhinin, I. P. Ostrovskii, N. S. Liakh, and S. M. Matvienko, "Thermo-EMF in Si-Ge solid solution whiskers," Journal of Physical Studies, vol. 9, no. 1, pp. 71-74, 2005. [Online]. Available: https://physics.lnu.edu.ua/jps/.

[21] T. O. Prikhna et al., "Electron-Beam and Plasma Oxidation-Resistant and Thermal-Barrier Coatings Deposited on Turbine Blades Using Cast and Powder Ni(Co)CrALY(Si) Alloys I. Fundamentals of the Production Technology, Structure, and Phase Composition of Cast NiCrAlY Alloys," Powder Metallurgy and Metal Ceramics, vol. 61, no. 1-2, pp. 70-76, 2022.

[22] T. O. Prikhna et al., "Electron-Beam and Plasma Oxidation-Resistant and Thermal-Barrier Coatings Deposited on Turbine Blades Using Cast and Powder Ni(Co)CrAlY(Si) Alloys Produced by Electron-Beam Melting II. Structure and Chemical and Phase Composition of Cast CoCrAlY Alloys," Powder Metallurgy and Metal Ceramics, vol. 61, no. 3-4, pp. 230-237, 2022.

[23] H. Xu, X. Sang, B. Yang, Y. Peng, and J. Fan, "Development of Novel Cold-Cathode Gas-Discharge Electron Beam Gun: An Instrumentation Study," Chinese Journal of Vacuum Science and Technology, vol. 41, no. 3, pp. 284-290, 2021.

[24] J. Chang et al., "Characterization of Magnetic Focusing of Electron Gun Based on Structural Parameters of Magnetic Lens," Chinese Journal of Vacuum Science and Technology, vol. 44, no. 5, pp. 463-469, 2024. [Online]. Available: https://doi.org/10.13922/j.cnki.cjvst.202401009.

[25] L. Gu et al., "Simulation Analysis of Factors Affecting Penning Discharge in High Current Pulsed Electron Beam," Chinese Journal of Vacuum Science and Technology, vol. 44, no. 2, pp. 184-190, 2024. [Online]. Available: https://doi.org/10.13922/j.cnki.cjvst.202306003.

[26] Y. Qiu et al., "Electrostatic Focusing Characteristic of Electron Gun with Multi-Parameter Coupling," Chinese Journal of Vacuum Science and Technology, vol. 41, no. 11, pp. 1094-1100, 2021. [Online]. Available: https://doi.org/10.13922/j.cnki.cjvst.202101027.

[27] C. Deng et al., "Property Characterization of Electron Flood Gun with Self-Developed Test Platform," Chinese Journal of Vacuum Science and Technology, vol. 40, no. 9, pp. 847-852, 2020. [Online]. Available: https://doi.org/10.13922/j.cnki.cjovst.2020.09.09.

[28] J. Wang, "Applications of Resistance Spot Welding in Assembling Electron Gun for Travelling Wave Tube: A Review Study," Chinese Journal of Vacuum Science and Technology, vol. 40, no. 4, pp. 381-388, 2020. [Online]. Available: https://doi.org/10.13922/j.cnki.cjovst.2020.04.17.

[29] Y. Xiang et al., "Design Optimization of Planar Sheet Beam Magnetron Injection Gun: A Simulation Study," Chinese Journal of Vacuum Science and Technology, vol. 40, no. 3, pp. 226-232, 2020. [Online]. Available: https://doi.org/10.13922/j.cnki.cjovst.2020.03.08.

[30] Y. Wang et al., "Design Optimization of Low Energy Electron Gun: A Simulation Study with Hybrid Mesh Model," Chinese Journal of Vacuum Science and Technology, vol. 40, no. 1, pp. 1-6, 2020. [Online]. Available: https://doi.org/10.13922/j.cnki.cjovst.2020.01.01.

[31] Y. Gu et al., "Design of New Electron Gun Capable of Generating Uniform Electron-Beam with Large Width," Chinese Journal of Vacuum Science and Technology, vol. 39, no. 12, pp. 1109-1113, 2019. [Online]. Available: https://doi.org/10.13922/j.cnki.cjovst.2019.12.11.

[32] W. Huo et al., "Design and Properties of Cu-Based LaB6 Photocathode for Pseudospark Discharge in Electron Beam Source: An Experimental Study," Chinese Journal of Vacuum Science and Technology, vol. 39, no. 8, pp. 631-636, 2019. [Online]. Available: https://doi.org/10.13922/j.cnki.cjovst.2019.08.03.

[33] X. Fu et al., "Mechanism for Black Particle Formation on Gold Coatings Deposited by Electron Beam Evaporation: A Theoretical and Experimental Study,"

Chinese Journal of Vacuum Science and Technology, vol. 39, no. 5, pp. 396-400, 2019. [Online]. Available: https://doi.org/10.13922/j.cnki.cjovst.2019.05.07.

[34] E. Wentzel and L. Ovcharov, Applied Problems of Probability Theory. Mir, 2002, 432 p. [Online]. Available: https://mirtitles.org/2022/06/03/applied-problems-in-probability-theory-wentzel-ovcharov/.

[35] J. A. Gubner, Probability and Random Processes for Electrical and Computer Engineers. Cambridge, UK: Cambridge University Press, 2006. [Online]. Available: http://www.cambridge.org/gb/academic/subjects/engineering/communications-and-signal-processing/probability-and-random-processes-electrical-and-computer-engineers.

[36] L. A. Zadeh and R. A. Aliev, Fuzzy Logic Theory and Applications: Part I and Part II. WSPC, 2018, 610 p.

[37] R. Belohlavek, J. W. Dauben, and G. J. Klir, Fuzzy Logic and Mathematics: A Historical Perspective, 1st ed. Oxford University Press, 2017, 544 p.

[38] E. Trillas and L. Eciolaza, Fuzzy Logic: An Introductory Course for Engineering Students. Springer, 2015, 204 p.

[39] J. M. Mendel et al., Introduction to Type-2 Fuzzy Logic Control: Theory and Applications. John Wiley & Sons, 2014, 256 p.

[40] G. J. Klir and B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, 1st ed. Pearson College Div., 1995, 592 p.

[41] H. T. Nguyen and E. A. Walker, A First Course in Fuzzy Logic, 3rd ed. Chapman and Hall CRC, 2005, 440 p.

[42] E. H. Mamdani, "Application of fuzzy algorithms for control of simple dynamic plant," Proceedings of the Institution of Electrical Engineers, vol. 121, no. 12, pp. 1585-1588, 1974.

[43] I. N. Bronshtein et al., Handbook of Mathematics, 5th ed. Springer, 2007, 1164 p.

[44] Handbook on Mathematical Functions with Formulas, Graphs and Mathematical Tables, M. Abramovich and I. Stegun, Eds. National Bureau of Standards, Applied Mathematic Series, vol. 55, Washington, 1964, 1046 p.

[45] S. C. Chapra and R. P. Canale, Numerical Methods for Engineers, 7th ed. McGraw Hill, 2014, 992 p.

[46] M. K. Jain, S. R. K. Iengar, and R. K. Jain, Numerical Methods for Scientific and Engineering Computation. New Age International Pvt. Ltd., 2010, 733 p.

[47] J. H. Mathews and K. D. Fink, Numerical Methods Using MATLAB, 3rd ed. Prentice Hall, 1998, 680 p.

[48] G. M. Phillips, Interpolation and Approximation by Polynomials. Springer, 2023, 312 p.

[49] J. F. Epperson, An Introduction to Numerical Methods and Analysis, Revised Edition. Wiley-Interscience, 2007, 590 p.

[50] V. M. Lazurik, V. T. Lazurik, G. Popov, and Z. Zimek, "Two-parametric model of electron beam in computational dosimetry for radiation processing," Radiation Physics and Chemistry, vol. 124, pp. 230-234, 2016.

[51] V. Lazurik and V. Moskvin, "Monte Carlo calculation of charge-deposition depth profile in slabs irradiated by electrons," Nuclear Instruments and Methods in Physics Research, Section B: Beam Interactions with Materials and Atoms, vol. 108, no. 3, pp. 276-281, 1996.

[52] P. Robbins, Python Programming for Beginners. Independently published, 2022, 114 p.

[53] W. McKinney, Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Jupyter, 3rd ed. O'Reilly Media, 2023, 579 p.

[54] F. Chollet, Deep Learning with Python, 2nd ed. Manning, 2022, 509 p.

[55] M. Lutz, Learning Python, 5th ed. O'Reilly, 2013, 1643 p.

Statistical Analysis of the Three-Dimensional Data of Software Metrics RFC, CBO, and WMC that are not Normally Distributed

Sergiy Prykhodko[1,2], Lidiia Makarova[1] and Andrii Pukhalevych[1]
[1]*Department of Software for Automated Systems, Admiral Makarov National University of Shipbuilding, Heroes of Ukraine Aveneu 9, 54007 Mykolaiv, Ukraine*
[2]*Department of Software Engineering, Odesa Polytechnic National University, Shevchenko Avenue 1, 65044 Odesa, Ukraine*
sergiy.prykhodko@nuos.edu.ua, lidiia.makarova@nuos.edu.ua, andrii.pukhalevych@nuos.edu.ua

Keywords: Software Quality, Software Development, Statistical Analysis, Prediction Interval, Confidence Interval, Nonlinear Regression, OOD Metric, RFC, CBO, WMC, Multivariate Box-Cox Transformation, Outlier Detection.

Abstract: Empirical data of RFC (response for a class), CBO (coupling between object classes), and WMC (weighted methods per class) software metrics, that can be used for estimation of software quality, deviate from normality. These metrics unveil multivariate skewness and kurtosis that do not conform to a multivariate Gaussian distribution. At the same time, well-known statistical methods that assume data normality may not be appropriate for the analysis of non-Gaussian data. To detect the outliers in the three-dimensional data of RFC, CBO, and WMC metrics and to estimate the confidence and prediction intervals of nonlinear regressions for these metrics, we need to use three-variate normalizing transformations. For statistical analysis of RFC, CBO, and WMC metrics, their normalization using the three-variate Box-Cox transformation was applied. Mardia's test for the transformed data after applying the multivariate Box-Cox transformation points that the transformed dataset is Gaussian. A technique for detecting outliers in multivariate non-Gaussian data based on the squared Mahalanobis distance for normalized data was applied to ensure the removal of outliers. Three nonlinear regression models for each of the RFC, CBO, and WMC metrics were constructed. The confidence and prediction intervals of nonlinear regressions for each of the RFC, CBO, and WMC metrics were built. Well-known statistical characteristics PRED(0.25) and MMRE for both the primary and the test datasets show that the model quality is satisfactory. The confidence and prediction intervals of nonlinear regressions for these metrics can be used for estimation of the quality of the object-oriented design of the software.

1 INTRODUCTION

Statistical analysis of multivariate data plays an important role in many areas, including empirical software engineering [1]. Empirical software engineering studies apply various methods of multivariate statistical analysis. Assuring the validity of such methods and corresponding results is challenging and critical [2]. As it is known [3], many methods of multivariate statistical analysis are based on the assumption that the data is normally distributed. Also, we know [4] if the data are not normally distributed, it is misleading to draw conclusions based on the normal distribution.

The above also applies to the well-known software metrics RFC (response for a class), CBO (coupling between object classes), and WMC (weighted methods per class). Although these metrics, along with three others (DIT - depth of inheritance tree, LCOM - lack of cohesion in methods, and NOC – a class's number of children), were proposed by Chidamber and Kemerer back in 1991 [5] for measuring the three non-implementation steps in Booch's definition of the object-oriented design (OOD), they are still used today to solve and other problems [6-15], including software quality [16-21]. In [21] the author proposed to apply the confidence and prediction intervals of nonlinear regressions for the RFC, CBO, and WMC metrics for evaluating the quality of software systems from the point of view of their OOD. In [21] the three-variate Box-Cox normalizing transformation was used to clean data from outliers and build the nonlinear regression models, the confidence and prediction intervals for the nonlinear regressions for the metrics RFC, CBO, and WMC since, firstly, according to the Mardia test, the three-dimensional data of these metrics are not normally distributed and, secondly, the residuals distribution of the corresponding linear regression models is not Gaussian. However, the above use of the three-variate Box-Cox transformation was based on the data sample from 51 open-source apps in Java.

In this paper, we have extended the results to a larger amount of data of metrics RFC, CBO, and WMC. As in [21], to detect outliers we apply the technique based on the squared Mahalanobis distance for the multi-dimensional normalized data and to build the nonlinear regression models, the confidence and prediction intervals for the nonlinear regressions for the metrics RFC, CBO, and WMC we use both univariate and multivariate normalizing transformations.

2 MATERIALS AND METHODS

Existing researches show that the data of software metrics can deviate from normality. Empirical values from the data set published in [21], consisting of RFC, CBO, and WMC metrics, were not normally distributed. These metrics unveil multivariate skewness and kurtosis that do not conform to a multivariate Gaussian distribution. At the same time, well-known statistical methods that assume data normality may not be appropriate for the analysis. Therefore, it is required to perform data normalization using normalizing transformations, as in [21]. Corresponding to [22], the bijective multivariate normalizing transformation will be used to convert a not Gauss-distributed random vector $\boldsymbol{P} = \{X_1, X_2, \dots, X_m\}^T$ into a Gauss-distributed random vector $\boldsymbol{T} = \{Z_1, Z_2, \dots, Z_m\}^T$:

$$\boldsymbol{T} = \Psi(\boldsymbol{P}), \tag{1}$$

where m is the number of metrics.

A transformation inverse to (1) is the following:

$$\boldsymbol{P} = \Psi^{-1}(\boldsymbol{T}). \tag{2}$$

In this research, we use multivariate Box-Cox transformation (BCT) (three-variate BCT in this case to transform values of each of RFC, CBO, and WMC metrics, taking into the correlation between the metrics):

$$Z_j = \begin{cases} \left(X_j^{\lambda_j} - 1\right)/\lambda_j, & \text{if } \lambda_j \neq 0; \\ ln\left(X_j\right), & \text{if } \lambda_j = 0. \end{cases} \tag{3}$$

There λ_j is a parameter of BCT and Z_j is the Gauss-distributed variable; $j = 1,2, \dots, m$. The estimates of these parameters are calculated by the method of maximum likelihood as in [3].

For the Box-Cox transformation the log-likelihood function is the following:

$$L(\boldsymbol{\lambda}) = -\frac{N}{2} ln\left(det(\boldsymbol{S}_N)\right) + (\boldsymbol{\lambda} - 1) \textstyle\sum_{i=0}^{N} ln\,(X_i). \tag{4}$$

There $\boldsymbol{\lambda}$ is the vector of lambda values, $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \dots, \lambda_m\}$; N is the number of data rows; $\boldsymbol{S}_N$ is the sample covariance matrix:

$$\boldsymbol{S}_N = \frac{1}{N} \textstyle\sum_{i=1}^{N} (\boldsymbol{T}_i - \overline{\boldsymbol{T}})\,(\boldsymbol{T}_i - \overline{\boldsymbol{T}})^T. \tag{5}$$

There $\overline{\boldsymbol{T}}$ is the sample mean vector, $\overline{\boldsymbol{T}} = \{\overline{Z}_1, \overline{Z}_2, \dots, \overline{Z}_m\}^T$; $\overline{T}_j = \frac{1}{N} \sum_{i=1}^{N} Z_{j_i}$; $j = 1,2, \dots, m$.

To ensure the removal of outliers, we apply a technique based on the squared Mahalanobis distance for normalized data, as described in [22]. For each multivariate data point the squared Mahalanobis distance d_i^2 can be calculated as

$$d_i^2 = (\boldsymbol{T}_i - \overline{\boldsymbol{T}})^T \cdot \boldsymbol{S}_N^{-1} \cdot (\boldsymbol{T}_i - \overline{\boldsymbol{T}}). \tag{6}$$

There $\boldsymbol{S}_N$ is the sample covariance matrix as defined in (5).

Regarding to [21], data points having d_i^2 that are greater than $3(N^2 - 1)F_{3,N-3,0.005}/N(N-3)$ are treated as outliers. $F_{3,N-3,0.005}$ is the F-distribution quantile with 3 and $N - 3$ degrees of freedom and a significance level of 0.005.

All identified outliers must be removed if present.

After removing outliers it is possible to estimate the quality of data points. This requires constructing intervals of the prediction and of the confidence for non-linear regression models for each metric. As in [21], we will use the transformation inverse to (1) and regression analysis for the prediction interval construction:

$$\Psi_Y^{-1}\left(\hat{Z}_Y \pm t_{\frac{\alpha}{2},\nu} S_{Z_Y} \left\{1 + \frac{1}{N} + (\boldsymbol{z}_X^+)^T \boldsymbol{S}_Z^{-1} (\boldsymbol{z}_X^+)\right\}^{\frac{1}{2}}\right), \tag{7}$$

where Ψ_Y is the normalizing transformation component for the dependent variable Y; $\hat{Z}_Y$ is a result of prediction with the equation of linear regression, $\hat{Z}_Y = \hat{b}_0 + \hat{b}_1 Z_1 + \hat{b}_2 Z_2$; $t_{\frac{\alpha}{2},\nu}$ is a quantile of the Student's t-distribution having ν degrees of freedom; $\nu = N - 3$; $\boldsymbol{z}_X^+$ is a vector consisting from $Z_{1_i} - \bar{Z}_1$, $Z_{2_i} - \bar{Z}_2$ for row i; $\bar{Z}_j = \frac{1}{N}\sum_{i=1}^{N} Z_{j_i}$, $j = 1, 2, \ldots, k$; $S_{Z_Y} = \frac{1}{\nu}\sum_{i=1}^{N}\left(Z_{Y_i} - \hat{Z}_{Y_i}\right)^2$.

In the (7) $\boldsymbol{S}_Z$ is the covariance matrix of the predictor variables, defined as (4):

$$\boldsymbol{S}_Z = \begin{pmatrix} S_{Z_1 Z_1} & S_{Z_1 Z_2} \\ S_{Z_1 Z_2} & S_{Z_2 Z_2} \end{pmatrix}, \quad (8)$$

where $S_{Z_q Z_r} = \sum_i^N \left[Z_{q_1} - \bar{Z}_q\right]\left[Z_{r_i} - \bar{Z}_r\right]$, $q, r = 1, 2$; $\bar{Z}_j = \frac{1}{N}\sum_{i=1}^{N} Z_{j_i}$, $j = 1, 2$.

For constructing intervals of the prediction for each of the involved metrics, we sequentially should treat one of the normalized metrics as the dependent variable of (7), and the remaining k metrics as independent variables of (7).

Regarding [21], for data points that are located inside the range between lower and higher values of the interval of the confidence (for each RFC, CBO, and WMC metrics) software quality is medium. For data points that are located inside the range from a lower value of the interval of the prediction to a higher value of the interval of the confidence for each metric, software quality is high. For other data points software quality is low.

3 ANALYSIS OF METRICS OF SOFTWARE DEVELOPED IN JAVA

An illustration of how we can use the listed methods of statistical analysis on the multivariate data when the data are not normally distributed will be provided. This illustration will be done by the estimation of the OOD quality of the software. Like in [21], we will use software-level RFC, CBO, and WMC metrics to construct three corresponding regression models, and intervals of the prediction and the confidence. In addition to the data described in [21] (53 rows), we will use 35 more data rows with empirical values of RFC, CBO, and WMC metrics for the software developed in Java. Additional rows are provided in Table 1.

These values were collected for the software developed in Java and stored in the public GitHub repositories. The collection of these values was performed by the CK framework on the class level, and then the values were converted to the software level by averaging the number of classes.

Table 1: OOD Metrics for the software developed in Java.

Repository	RFC	CBO	WMC
3D-TETRIS	9.906	4.547	21.516
Chemtris	5.234	3.455	4.977
Cubes	7.479	5.212	8.506
DestinationSol	12.923	8.344	12.34
finisterra	8.812	5.954	7.342
GDX-RPG	9.267	4.449	11.096
GdxGame	11.408	5.942	9.777
kickoff	8.068	5.227	8
Klooni1010	9.471	6.059	8.824
Koru	6.587	3.985	5.726
lightblocks	9.838	5.063	10.515
mario-game	9	5.022	9.109
martianrun	8.475	4.459	9.918
marvelous-bob	7.085	6.641	5.41
mini2Dx	10.156	5.039	14.939
Norii	15.193	6.62	17.422
Novix	10.06	3.595	5.888
OasisGame	9.373	6.152	8.088
odb-naturally-selected-2d	6.137	6.874	6.263
OverblownGame	7.94	6.508	10.377
Particle-Park	10.029	5.086	5.714

The first 3 rows from Table 1 will be merged with 53 rows from the dataset described in [21] (56 rows in total). 32 more rows from Table 1 will be used as a test dataset.

Checking the merged dataset with Mardia's test shows that this dataset is not Gaussian (multivariate skew $\beta_1 = 11.66$ and multivariate kurtosis $\beta_2 = 27.17$). Therefore, as expected, it is not possible to apply the method based on the squared Mahalanobis distance for unnormalized data to remove outliers and use linear regression analysis (the residual distribution of the corresponding linear regression models is not Gaussian).

Following [21], it is needed to use multivariate normalizing transformations to remove outliers from the data and construct non-linear regression models, and intervals of the prediction and the confidence. In this illustration, we will use multivariate Box-Cox transformation (3). Here we sequentially replace Z_j, X_j, λ_j with Z_{RFC}, X_{RFC}, λ_{RFC}, then with Z_{CBO}, X_{CBO}, λ_{CBO}, and finally with Z_{WMC}, X_{WMC}, λ_{WMC}.

The estimation of $\boldsymbol{\lambda}$ for (3) was done using the corresponding log-likelihood function (4). Estimates were calculated using the Apache Math library implementation of the BOBYQA Optimizer (Bound Optimization BY Quadratic Approximation).

Piece of a computer program in Java to get estimations of $\boldsymbol{\lambda}$ with BOBYQA:

```
public static double[]
estimateLambda(double[][] data) {
  int dimension = data[0].length;
  double[] initialGuess = new
double[dimension];
  Arrays.fill(initialGuess, 1.0);

  MultivariateFunction objective =
lambda -> computeLogLikelihood(data,
lambda);

  BOBYQAOptimizer optimizer = new
BOBYQAOptimizer(2 * dimension + 1);

  PointValuePair result =
optimizer.optimize(
    MaxEval.unlimited(),
    MaxIter.unlimited(),
    new ObjectiveFunction(objective),
    GoalType.MAXIMIZE,
    new InitialGuess(initialGuess),
    SimpleBounds.unbounded(dimension)
  );

  return result.getPoint();
}
```

Multivariate Box-Cox transformation parameter estimates for 56 rows are the following: $\hat{\lambda}_{\mathrm{RFC}} = 0.2562$, $\hat{\lambda}_{\mathrm{CBO}} = 0.7704$, $\hat{\lambda}_{\mathrm{WMC}} = -0.2198$.

Multivariate distribution after applying the multivariate Box-Cox transformation (3) with estimated parameters has a multivariate skew $\beta_1 = 1.53$ and multivariate kurtosis $\beta_2 = 15.45$. Specified value points that the transformed dataset is Gaussian. Therefore, it is possible to apply regression analysis on the transformed dataset, having previously checked for outliers.

Outliers check was done using (6) by calculating the squared Mahalanobis distance for each of the 3-dimensional rows. Sample means for this data are $\bar{Z}_{RFC} = 6.903$, $\bar{Z}_{CBO} = 2.086$, $\bar{Z}_{WMC} = 4.038$. The corresponding sample covariance matrix is the following:

$$\boldsymbol{S}_N^{-1} = \begin{pmatrix} 0.451 & 0.384 & -0.997 \\ 0.384 & 46.262 & -12.904 \\ -0.997 & -12.904 & 6.747 \end{pmatrix}.$$

The largest value of the squared Mahalanobis distance calculated with (6) is $d_{max}^2 = 12.08$. This value is smaller than a quantile of the F distribution with degrees of freedom 3 and 53 and a significance level of 0.005. It means that the transformed dataset does not contain significant deviations that can be treated as outliers.

By sequentially treating one of the normalized metrics as the dependent variable and the remaining 2 metrics as independent variables, we constructed three nonlinear regression models for each of the RFC, CBO, and WMC metrics based on 3-variate BCT:

$$Y = \left[\hat{\lambda}_Y\left(\hat{Z}_Y + \varepsilon\right)\right]^{1/\hat{\lambda}_Y}. \quad (9)$$

There $\hat{Z}_Y$ is a result of prediction with the equation of linear regression, $\hat{Z}_Y = \hat{b}_0 + \hat{b}_1 Z_1 + \hat{b}_2 Z_2$ that depends on predictors Z_1 and Z_2 for the Gaussian data transformed with 3-variate BCT; is ε is a Gaussian random variable, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$.

The estimates of the parameters for the nonlinear regression models (9) for each of the RFC, CBO, and WMC metrics are provided in Table 2. The distribution of ε of each linear regression for normalized metrics was Gaussian regarding the Chi-square test with a significance level of 0.05. Residual distributions and the density of the corresponding Gaussian distributions are shown in Figure 1.

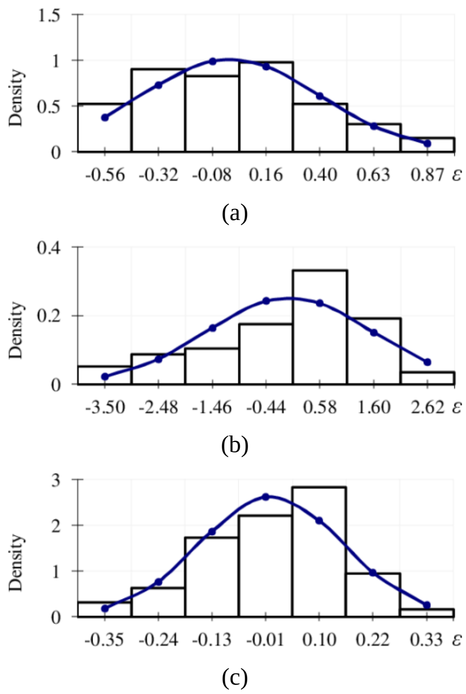


Figure 1: Residuals distributions of the linear regression models for the normalized values of RFC (a), CBO (b), and WMC (c) metrics.

Table 2: The estimates of the parameters for the nonlinear regression models.

No	Y	$\hat{b}_0$	$\hat{b}_1$	$\hat{b}_2$	σ_ε^2	MMRE	PRED
1	RFC	-0.972	0.148	1.913	0.389	0.16	0.839
2	CBO	-0.752	2.367	-0.913	1.555	0.241	0.679
3	WMC	1.017	0.279	-0.008	0.148	0.224	0.643

We used the popular characteristics, MMRE and PRED(0.25), to evaluate the predictive accuracy of nonlinear regression models (9) for each of the RFC, CBO, and WMC metrics. It is acceptable to have an MMRE of no more than 0.25 and a PRED(0.25) of no less than 0.75. The MMRE and PRED(0.25) values for the abovementioned models are shown in Table 2. These characteristics show that the model's quality is satisfactory.

To compute the intervals of the prediction and of the confidence for the nonlinear regression for the RFC metric the following values will be used in (7) and (8): $S_{Z_Y} = 0.157$, $\bar{Z}_1 = 6.903$, $\bar{Z}_2 = 2.086$, $\boldsymbol{S}_Z^{-1} = \begin{pmatrix} 0.005 & -0.027 \\ -0.027 & 0.385 \end{pmatrix}$.

To compute intervals of the prediction and the confidence for the nonlinear regression for the CBO metric the following values will be used in (7) and (8): $S_{Z_Y} = 2.510$, $\bar{Z}_1 = 4.038$, $\bar{Z}_2 = 2.086$, $\boldsymbol{S}_Z^{-1} = \begin{pmatrix} 0.078 & -0.214 \\ -0.214 & 0.820 \end{pmatrix}$.

To compute intervals of the prediction and the confidence for the nonlinear regression for the WMC metric the following values will be used in (7) and (8): $S_{Z_Y} = 0.0228$, $\bar{Z}_1$ =4.038, $\bar{Z}_2 = 6.903$, $\boldsymbol{S}_Z^{-1} = \begin{pmatrix} 0.056 & -0.016 \\ -0.016 & 0.007 \end{pmatrix}$.

To compute intervals of the prediction and of the confidence for the nonlinear regression for each of the RFC, CBO, and WMC metrics the following values will be used in (7) and (8): $N = 56$, $\nu = 53$, $t_{0.025,53}$=2.006.

For the test dataset consisting of 23 rows, the values of the MMRE and PRED(0.25) for the nonlinear regression models (9) for each of the RFC, CBO, and WMC metrics are provided in Table 3.

Table 3: Nonlinear regression models quality characteristics for the test dataset.

No	Y	MMRE	PRED
1	RFC	0.114	0.875
2	CBO	0.347	0.375
3	WMC	0.190	0.719

These characteristics for the test dataset also show that the model's quality is satisfactory.

4 DISCUSSION

For statistical analysis of RFC, CBO, and WMC metrics, we propose to apply their normalization using the three-variate Box-Cox transformation. This choice is due to the following. Firstly, the three-dimensional data of these metrics are not normally distributed and, secondly, the residuals distribution of the corresponding linear regression models is not Gaussian.

To detect the three-dimensional outliers in the data, we applied the appropriate technique [22] based on the multivariate normalizing transformations. The use of the three-variate Box-Cox transformation allows us to additionally take into account the correlation between RFC, CBO, and WMC metrics.

To build the confidence and prediction intervals for the nonlinear regressions for RFC, CBO, and WMC metrics for evaluating the quality of open-source apps in Java, we used a 0.05 significance level, as the appointed one usually, although this value may be discussed.

The statistical analysis of RFC, CBO, and WMC metrics based on the three-variate Box-Cox transformation demonstrates its capabilities. In the future, it is necessary to build corresponding mathematical models based on other data sets.

5 CONCLUSIONS

For statistical analysis of RFC, CBO, and WMC metrics, have proposed to apply their normalization using the three-variate Box-Cox transformation. To detect the outliers in the three-dimensional data of RFC, CBO, and WMC metrics and to estimate the confidence and prediction intervals of nonlinear regressions for these metrics, we need to use three-variate transformations. The constructed confidence and prediction intervals of nonlinear regressions for the above metrics may be applied to estimate the quality of open-source software in Java.

Moving forward, we plan to develop examples of statistical analysis of RFC, CBO, and WMC metrics that do not have limitations due to the programming language and the sample size.

ACKNOWLEDGMENTS

We are grateful to P. Fadeev, a bachelor's student at the Admiral Makarov National University of

Shipbuilding, for his assistance in collecting software metrics data, which allowed us to expand the dataset used in the data analysis.

REFERENCES

[1] D. Mendez, P. Avgeriou, M. Kalinowski, and N. bin Ali, Handbook on teaching empirical software engineering. Cham: Springer, 2024, doi: 10.1007/978-3-031-71769-7.

[2] J. Härtel and R. Lämmel, “Operationalizing validity of empirical software engineering studies,” Empir. Softw. Eng., vol. 28, Nov. 2023, Art. no. 153, doi: 10.1007/s10664-023-10370-3.

[3] R. A. Johnson and D. W. Wichern, Applied Multivariate Statistical Analysis. New Jersey: Pearson Prentice Hall, 2007.

[4] J. Osborne, Best Practices in Data Cleaning: A Complete Guide to Everything You Need to Do Before and After Collecting Your Data. SAGE Publications, Inc., 2013, doi: 10.4135/9781452269948.

[5] S. R. Chidamber and C. F. Kemerer, “Towards a metrics suite for object oriented design,” ACM SIGPLAN Notices, vol. 26, no. 11, pp. 197–211, 1991, doi: 10.1145/118014.117970.

[6] C. Haritha Madhav and K. S. Vipin Kumar, “A method for predicting software reliability using object oriented design metrics,” in Proc. Int. Conf. on Intelligent Computing and Control Systems (ICCS), 2019, pp. 679–682, doi: 10.1109/ICCS45141.2019.9065541.

[7] E. A. Alomar, M. W. Mkaouer, A. Ouni, and M. Kessentini, “On the Impact of Refactoring on the Relationship between Quality Attributes and Design Metrics,” in Proc. Int. Symposium on Empirical Software Engineering and Measurement (ESEM), Porto de Galinhas, Brazil, 2019, pp. 1-11, doi: 10.1109/ESEM.2019.8870177.

[8] M. Rizwan, A. Nadeem, and M. A. Sindhu, “Empirical Evaluation of Coupling Metrics in Software Fault Prediction,” in Proc. 17th Int. Bhurban Conf. on Applied Sciences and Technology (IBCAST), 2020, pp. 434–440, doi: 10.1109/IBCAST47879.2020.9044489.

[9] A. Vescan, C. Serban, and G. C. Crisan, “Software defects rules discovery,” in Proc. IEEE 14th Int. Conf. on Software Testing, Verification and Validation Workshops (ICSTW), 2021, pp. 101–109, doi: 10.1109/ICSTW52544.2021.00028.

[10] M. Begum, M. H. Shuvo, I. Ashraf, A. A. Mamun, J. Uddin, and M. A. Samad, “Software Defects Identification: Results Using Machine Learning and Explainable Artificial Intelligence Techniques,” IEEE Access, vol. 11, pp. 132750–132765, 2023, doi: 10.1109/ACCESS.2023.3329051.

[11] M. Ziobrowski, M. Ochodek, J. Nawrocki, and B. Walter, “Towards Reliable Rule Mining about Code Smells: The McPython Approach,” in Proc. 18th Conf. on Computer Science and Intelligence Systems (FedCSIS), 2023, pp. 65–66, doi: 10.15439/2023F2071.

[12] P. Ofem, B. Isong, and F. Lugayizi, “Metrics for Evaluating and Improving Transparency in Software Engineering: An Empirical Study and Improvement Model,” SN Computer Science, vol. 5, Art. no. 8, 2024, doi: 10.1007/s42979-024-03471-3.

[13] A. S. Prykhodko and E. V. Malakhov, “Determining object-oriented design complexity due to the identification of classes of open-source web applications created using PHP frameworks,” Radio Electronics, Computer Science, Control, 2024, no. 2 (69), pp 160–166, doi: 10.15588/1607-3274-2024-2-16.

[14] M.-A. Levasseur and M. Badri, “Prioritizing unit tests using object-oriented metrics, centrality measures, and machine learning algorithms,” Innovations in Systems and Software Engineering, 2024, doi: 10.1007/s11334-024-00550-9.

[15] D. A. Rebro, S. Chren, and B. Rossi, “Source Code Metrics for Software Defects Prediction,” in Proc. ACM Symposium on Applied Computing, 2023, pp. 1469–1472, doi: 10.1145/3555776.3577809.

[16] A.-J. Molnar, A. Neamţu, and S. Motogna, “Evaluation of Software Product Quality Metrics,” in Proc. Communications in Computer and Information Science, vol. 1172, 2020, pp. 163-187, doi: 10.1007/978-3-030-40223-5_8.

[17] M. M. A. Dabdawb and B. Mahmood, “A Network of Object-Oriented Software Metrics Parameters,” in Proc. 10th IEEE Int. Conf. on Communication, Networks and Satellite (Comnetsat), 2021, pp. 172–178, doi: 10.1109/COMNETSAT53002.2021.9530822.

[18] I. C. Nwandu, J. N. Odii, E. C. Nwokorie, and S. A. Okolie, “Evaluation of Software Quality in Test-driven Development: A Perspective of Measurement and Metrics,” Int. Journal of Information Technology and Computer Science, vol. 14, no. 6, pp. 13–22, 2022, doi: 10.5815/ijitcs.2022.06.02.

[19] I. M. A. Wikantyasa, A. P. Kurniawan, and S. Rochimah, “C K Metric and Architecture Smells Relations: Towards Software Quality Assurance,” in Proc. 14th Int. Conf. on Information and Communication Technology and System (ICTS), 2023, doi: 10.1109/ICTS58770.2023.10330874.

[20] S. Jin, Z. Li, B. Chen, B. Zhu, and Y. Xia, “Software Code Quality Measurement: Implications from Metric Distributions,” in Proc. IEEE Int. Conf. on Software Quality, Reliability and Security (QRS), 2023, pp. 488–496, doi: 10.1109/QRS60937.2023.00054.

[21] S. Prykhodko, “Evaluating Quality of Software Systems by the Confidence and Prediction Intervals of Regressions for RFC, CBO and WMC Metrics,” WSEAS Transactions on Systems, vol. 23, pp. 322–330, 2024, doi: 10.37394/23202.2024.23.36.

[22] S. Prykhodko, N. Prykhodko, L. Makarova, and A. Pukhalevych, “Application of the squared Mahalanobis distance for detecting outliers in multivariate non-Gaussian data,” in Proc. 14th Int. Conf. on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), 2018, pp. 962–965, doi: 10.1109/TCSET.2018.8336353.

Exploration of the Efficiency of SLM-Enabled Platforms for Everyday Tasks

Volodymyr Rusinov and Nikita Basenko
Faculty of Informatics and Computer Engineering, Igor Sikorsky Polytechnic Institute, Beresteiska Avenue 37, 03056 Kyiv, Ukraine
v.rusinov.io11f@kpi.ua, basenko.nikita@lll.kpi.ua

Keywords: AI, SLM, IoT, NLP, Model Evaluation Metrics.

Abstract: This study explores the potential of Small Language Models (SLMs) as an efficient and secure alternative to larger models like GPT-4 for various natural language processing (NLP) tasks. With growing concerns around data privacy and the resource-intensiveness of large models, SLMs present a promising solution for research and applications requiring fast, cost-effective, and locally deployable models. The research evaluates several SLMs across tasks such as translation, summarization, Named Entity Recognition (NER), text generation, classification, and retrieval-augmented generation (RAG), comparing their performance against larger counterparts. Models were assessed using a range of metrics specific to the intended task. Results show that smaller models perform well on complex tasks, often rivalling or even outperforming larger models like Phi-3.5. The study concludes that SLMs offer an optimal trade-off between performance and computational efficiency, particularly in environments where data security and resource constraints are critical. The findings highlight the growing viability of smaller models for a wide range of real-world applications.

1 INTRODUCTION

In the period of large language models (LLM) dominance in the market, some companies are thinking about the security of their data when developing their own solutions using open source LLM application programming interfaces(API). Data leaks, identity theft and other types of malicious activity may present a challenge when using these solutions, as company data may contain sensitive information such as personal documents, financial reports and so on. Given the specifics of the business area, companies are enticed to train their own custom LLMs or provide data to outside vendors for them to train the models, facing the same obstacles described earlier. [1] In addition, such systems require computational power that require outsourcing LLMs into Cloud, which cannot function without the Internet, as they send requests to APIs, saving the computing resources of the device.

It is also worth noting that sometimes the use of large models (such as GPT-4o, Claude Sonnet-3.6 or even Llama-70b) for simple tasks such as text classification or answers to household questions. This leads not only to an increase in the cost of such a system, but also to the execution time, which negatively affects the user experience directly.

2 OBJECTIVE AND TOOLS

This study aims to evaluate the efficacy of Small Language Models (SLMs) as viable alternatives to larger models for specific natural language processing tasks in resource-constrained environments. The primary challenge lies in applying such models to a range of tasks commonly employed by researchers to enhance productivity or automate routine processes. For instance, in fields like healthcare, where data security is critical, reliance on large corporations for data management may be inappropriate due to confidentiality concerns. It is therefore necessary to identify the core tasks that small language models (SLMs) must be capable of performing to effectively address the demands of contemporary research.

To comprehensively evaluate SLM performance, we have identified five core NLP tasks that represent common requirements in research and practical applications:

1) Translation - The conversion of text between major world languages, representing a fundamental yet well-defined linguistic task that serves as a baseline for model linguistic capabilities.

2) Summarization - The extraction and condensation of key information from longer texts, requiring semantic understanding and content prioritization abilities.
3) Named Entity Recognition (NER) - The identification and classification of named entities within text, representing a structured information extraction task that evaluates the model's ability to recognize semantic patterns.
4) Text Classification - The categorization of text into predefined or custom classes, requiring the model to understand context and apply flexible classification schemas.
5) Retrieval-Augmented Generation (RAG) - The integration of external context into generated responses, evaluating the model's capacity to process supplied information and incorporate it into outputs.

These tasks were selected to provide a comprehensive assessment of both fundamental and advanced linguistic capabilities required in contemporary research and practical applications.

The selection of models for evaluation was governed by the following criteria. The first one is the limitation of the number of parameters and the size of the model. In this research small models up to 2B parameters are observed, because such models are small enough to be able to run locally on any research hardware without affecting system performance, in addition chosen models should be up to 2.5GB. It is important to consider not only the average value but also specific metrics, such as IFEval (instruction-following evaluation, used for evaluation model's capabilities to follow instructions), MUSR (multistep soft reasoning, used for understanding large contexts and reasoning) and not necessarily but still it would be good BBH (big-bench Hard, used for general understanding of the world).

The evaluation incorporates multiple model families, including fine-tuned variations of Llama, Qwen [11] and its variations and also a bunch of different non NLP related models (different 3d, 2d visualizations which are not considered in this article). [13] And the SmolLM [1] family, whose capabilities prompted this investigation. Additionally, the Phi-3.5 model from Microsoft was incorporated to serve as a comparative benchmark against the other smaller models. Each model will be systematically evaluated across all six tasks using task-specific metrics that accurately reflect performance quality. The evaluation will be conducted within a controlled environment using identical prompts, configuration parameters, and system instructions to ensure comparative validity.

The experimental setup will utilize the distributed AI platform described in Section 3, which enables efficient deployment and testing of multiple models simultaneously while maintaining consistent evaluation conditions.

3 PLATFORM

For this study, a distributed AI platform was implemented using a network of Raspberry Pi devices, interconnected in a tree topology augmented with additional inter-level connections following a De Bruijn sequence. This configuration, combined with an MLOps (Machine Learning Operations) process, provides an efficient and cost-effective setup for deploying and testing SLMs for the purposes of this research. The goal is to utilize the advantages of the embedded systems and locally deployable nature of Raspberry Pi hardware to evaluate the performance of SLMs across a range of natural language processing (NLP) tasks, while ensuring efficient model training and deployment workflows. The platform is organized in a tree topology with additional DeBrujin connections, where one central node (managed by Cloud in this case) managed the coordination of tasks, while multiple Raspberry Pi devices handled SLM tasks. [6] This hierarchical structure allowed the system to scale easily, with additional nodes onboarding different models, demonstrated in this work (Fig. 1). This approach showcases the decentralized nature of Edge computing, with local deployment, which is beneficial for the systems that utilize SLMs.

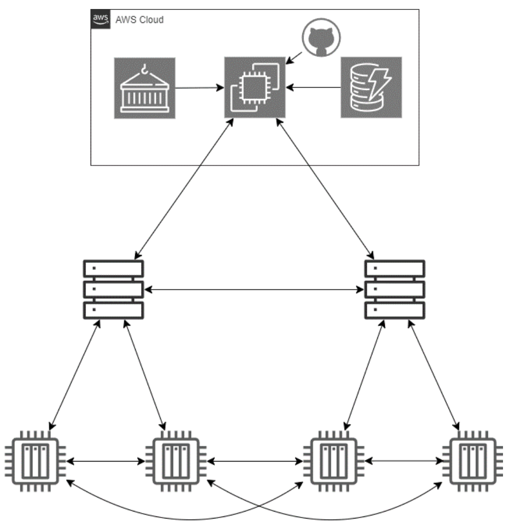


Figure 1: Example of an AI platform using the described approach.

In order to use this system optimally, MLOps is used in streamlining the machine learning lifecycle within this platform. The CI/CD pipeline was responsible for automating the workflow of data preparation, model training, validation, deployment, and monitoring. Specifically, these tools are used to automate the deployment of SLMs across the Raspberry Pi devices, ensuring that necessary updates are propagated across the network. In addition, this platform provides a framework for continuous integration and continuous delivery, allowing the new models to be iteratively improved based on real-time performance feedback and continuous training of SLMs. Each Raspberry Pi node was equipped with a lightweight containerized environment, which allowed for the isolated execution of different SLMs. This containerization ensured that each model could run independently on the devices without interfering with other processes, optimizing the overall performance and stability of the system. The nodes communicated through a message-passing interface, enabling them to share intermediate results, such as model predictions, across the distributed network in real time.

The distributed setup, coupled with the MLOps process, allowed for efficient parallel processing during model evaluation. Moreover, this process enabled continuous monitoring of the models' performance, with real-time metrics being collected and analyzed. This helped ensure that the models were functioning as expected and allowed for quick troubleshooting in the event of model drift or errors. In order to leverage the benefits of using such approach and to test new model parameters or sequences, LM Studio is used for the experiments. LM studio is integrated through the API with the user interface, furthermore, it allows for simple integration into continuous training process, which is beneficial for the platform's ability to rapidly reconfigure itself in various scenarios. The ability to quickly iterate on models and deploy them across the Raspberry Pi network made it possible to optimize their performance on specific tasks, all while maintaining a low resource footprint. For example, while one node handles a translation task, another could simultaneously work on summarization, and a third could be responsible for Named Entity Recognition (NER). This parallel processing capability significantly reduced the overall execution time for testing multiple models on various tasks, making the setup highly efficient and scalable for handling large volumes of data.

4 EXPERIMENTS

In order to test such models it is necessary to collect some data from these tasks. Additionally, all models were tested using the platform described in the previous paragraph, which is specifically designed for rapid reconfiguration and deployment of various models.

The same parameters (numbers – refer to screenshot 15.01) such as Seed value, temperature, context size were set for all models (see Table 1 for model specifications). The models were not provided with any prior interaction history, as the evaluation was conducted without the use of contextual cues or supplementary prompts. Additionally, all models were provided with the same system prompt, which is default for all models.

Table 1: Model overview.

Model name	Number of parameters	Model size	Quantization
google/gemma-2-2b-it	2b	1.52GB	q4
Qwen/Qwen2.5-1.5B-Instruct	1.54b	1GB	q4
meta-llama/Llama-3.2-1B-Instruct	1.24b	1.23GB	q4
Qwen/Qwen2.5-0.5B-Instruct	0.494b	0.65GB	q8
HuggingFaceTB/SmolLM2-1.7B-Instruct	1.71b	1GB	q4
HuggingFaceTB/SmolLM2-360M-Instruct	0.362b	0.7GB	q16
microsoft/Phi-3.5-mini-instruct	3.82b	2.23GB	q4

4.1 Model Comparison

In this section, we compare different models using a set of metrics, all specific to each task.

Translation. In this case, the reference translation of the text was compared with the translations generated by the models. For evaluation, COMET (Crosslingual Optimized Metric for Evaluation of Translation), a state-of-the-art tool for the automatic assessment of machine translation quality [2], was utilized. This metric is particularly appropriate, as it is grounded in a pre-trained XLM-RoBERTa language model, which facilitates the evaluation of both linguistic accuracy across multiple languages and the contextual coherence of the translated text. This metric was applied to the outputs generated by our models, yielding the following results (Fig. 2).

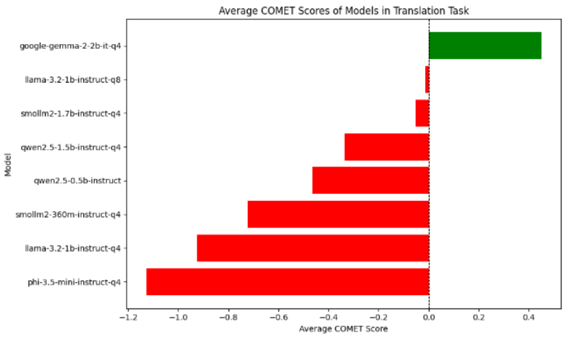


Figure 2: Comparison between different SLMs using COMET.

As shown on the graph (Fig. 2.), the model developed by Google outperforms the others, which can be attributed to the extensive multilingual training data used for the Gemma model family. This training approach enabled the model to achieve higher performance in understanding a wide range of languages, particularly those that are most commonly used. The worst values were obtained from the Phi-3.5 model because of its tendency to write verbose explanations, when it was not inquired to do so. In addition, this model tends to constantly write what words it translated and why. This affects both the evaluation and the user experience. The difference between variying degrees of quantization of Llama-3.2-1b models is evident, wtih int4 performing significantly worse than int8.

The largest model Phi-3.5 has a peculiarity to describe additional information, when it is not explicitly asked for.

Summarization In this traditional task, we will evaluate the models using the BERTScore metric, which, similar to COMET, leverages pre-trained language models such as BERT (Bidirectional Encoder Representations from Transformers) to assess the semantic similarity between the predicted text and the reference text. Because of the bidirectional representation of words and the transformation of texts into a vector space, BERTScore helps to better determine the degree of semantic similarity between texts. [3] Below are the results of the models in our tests:

When analyzing the test results (Fig. 3) it is noted that all the tested models are highly efficient for the explored tasks. It is worth mentioning that the SmolLM2 family models - both models (larger and medium sized), were able to perform significantly better than the largest model among all presented - Phi-3.5. It can also be concluded that for this task, smaller models such as SmolLM2-360m are more effective, as they make optimal use of the device's computing power—an advantage that is less prominent in the Phi-3.5 model.

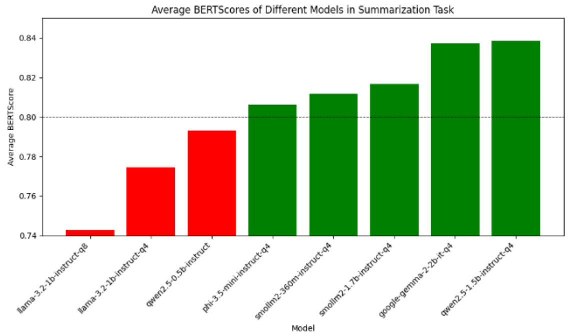


Figure 3: Comparison between different SLMs using BERT.

NER. In this instance, the models were evaluated based on their prediction accuracy using the following (1), which incorporates both the incompleteness of the model's response and a penalty for failure to adhere to the provided instructions:

$$Rbase=(Correct + 0.5 * Partial) / TotalTrueEntities. \quad (1)$$

In this equation, Correct - fully correct entities, Partial - partially correct entities. - TotalTrueEntities - total number of true entities.
Based on this (2), the models are evaluated:

$$Rfinal = if(ExtraSymbols = "\checkmark", Rbase / 2, Rbase), \quad (2)$$

where ExtraSymbols indicates the presence or absence of additional irrelevant content in the model's response, and Rbase represents the base accuracy value calculated from (1). The final rating (Rfinal) is penalized by a factor of 2 when the model's output contains extraneous symbols or information not requested in the query. This penalty mechanism was implemented to quantitatively account for the models' adherence to the instruction constraints, as superfluous output can negatively impact both computational efficiency and user experience in practical applications. This approach allows for a more comprehensive evaluation of model performance beyond mere entity recognition accuracy, incorporating the quality and precision of the generated output as essential evaluation criteria. A notable result of the study is the unexpectedly high performance of the Qwen-2.5-0.5b and SmolLM2-360m models, which demonstrated exceptional performance on a task that is traditionally considered to be more complex for models with a limited number of parameters. This result is of particular interest because it contradicts the commonly held assumption of a direct correlation between model size and the quality of its performance on complex linguistic tasks.

In the context of this study, the use of Large Language Models (LLMs) demonstrates significant advantages over traditional natural language processing tools such as Spacy. The key difference lies in the quality and structured nature of the output data. While LLMs generate logically organized and easily human interpretable output, Spacy-based solutions, despite their transformational architecture, often provide less coherent output consisting of a set of semantically similar tokens. This observation highlights the superiority of LLM in tasks requiring not only processing accuracy but also human-readable results.

Classification. This classification task stands out from the classic approach by employing custom user classes. User classes are named, defined and modified by the user. Accuracy metric is used as an expected model output, formatted either solely as a class or a structured output defining a class. Therefore, accuracy metric is the best fit for this task and it accurately displays the classification task.

$$\textit{Accuracy} = (\textit{Number of Correct Predictions}) / (\textit{Total Number of Predictions}) \quad (3)$$

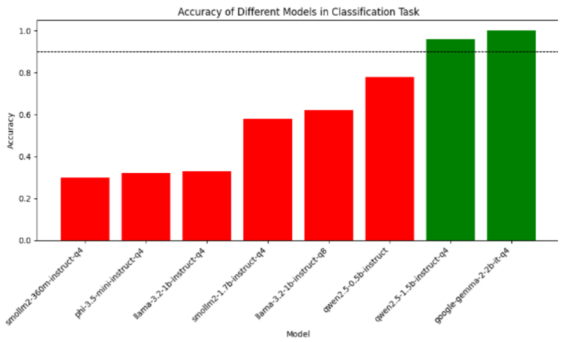


Figure 4: Comparison between different SLMs using Accuracy metric for classification task.

The task itself is complicated by the fact that the model, as previously mentioned in the earlier example, must guess a custom category based on the custom description. Figure 4 illustrates the performance comparison of different SLMs in this classification task using the Accuracy metric. These categories can be different and the text can be inexplicable. Gemma predicted all the test cases, which is not the case with the latest models. Phi-3.5 and Llama - like to hallucinate, giving a non-existent category or return the prompt itself that it was provided with. It is possible that by slightly modifying the prompt or system instructions, the performance of Qwen2.5-0.5b could be improved, potentially allowing it to achieve results comparable to its older counterpart or even outperform Gemma.

RAG. This task is evaluated using the RAGAS (Retrieval Augmented Generation Assessment) estimation method, which is based on model evaluation by means of a larger model. In a similar way, LLM models are trained by using a model-evaluator to evaluate the correctness of the model. The RAGAS approach in this experience evaluates any task but in this case it is applied to RAG tasks to observe how well the model responds. [4] The model estimator is openai-4o. The accuracy metric employed is implicitly consistent, assigning a value of 0 for incorrect responses and 1 for correct responses. The results of the evaluation can be observed below:

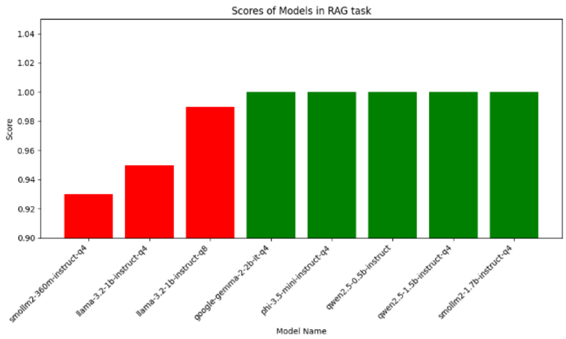


Figure 5: Comparison between different SLMs using RAGAS approach.

The analysis of the results (Fig. 5) demonstrates that all the models under study are capable of achieving a high level of performance. Special attention should be paid to the efficient performance of small-size models in particular Qwen2.5-0.5b and Llama-3.2-1b. Regarding the SmolLM-0.36b model, it is important to highlight its important characteristic: rather than providing short answers, it tends to reproduce a portion, or, in some cases, the entire input query within its own responses.

5 CONCLUSIONS

Based on a comprehensive analysis of the obtained results, conclusion can be reached that the Gemma-2-2b-instruct model has performed better than the other researched models, taking into account that it has the largest parametric characteristics among all the studied models. However, when considering the efficiency of computational resource utilization, special attention should be paid to the Qwen-2.5 series model (in particular, the version with 0.5 billion parameters), which demonstrates an optimal ratio between performance and resource intensity, despite the fact that it does not reach the maximum performance in terms of absolute metrics in this study.

A notable aspect of the study is the performance of the Phi-3.5 model, which, despite its larger parametric characteristics and theoretical potential to perform more efficiently, did not demonstrate the expected superiority over smaller models. Contrary to initial assumptions that this model was expected to deliver qualitatively superior results due to its extended architecture, its actual performance was significantly below the leading position in the comparison table. This observation underscores the important conclusion that increasing the size of a model does not always correlate directly with improved performance in specific applications.

REFERENCES

[1] "SmolLM - blazingly fast and remarkably powerful," Feb. 2024, [Online]. Available: https://huggingface.co/blog/smollm.

[2] R. Rei, C. Stewart, A. C. Farinha, and A. Lavie, "COMET: A Neural Framework for MT Evaluation," arXiv.org, Sep. 18, 2020, [Online]. Available: https://arxiv.org/abs/2009.09025.

[3] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "BERTScore: Evaluating Text Generation with BERT," arXiv.org, Apr. 21, 2019, [Online]. Available: https://arxiv.org/abs/1904.09675.

[4] S. Es, J. James, L. Espinosa-Anke, and S. Schockaert, "RAGAS: Automated Evaluation of Retrieval Augmented Generation," arXiv (Cornell University), Jan. 2023, [Online]. Available: https://doi.org/10.48550/arxiv.2309.15217.

[5] A. Grattafiori et al., "The Llama 3 herd of models," arXiv.org, Jul. 31, 2024, [Online]. Available: https://arxiv.org/abs/2407.21783.

[6] V. Rusinov, O. Honcharenko, A. Volokyta, H. Loutskii, O. Pustovit, and A. Kyrianov, "Methods of topological organization synthesis based on tree and dragonfly combinations," in Lecture notes on data engineering and communications technologies, 2023, pp. 472-485, [Online]. Available: https://doi.org/10.1007/978-3-031-36118-0_43.

[7] G. Team et al., "Gemma 2: Improving open language models at a practical size," arXiv.org, Jul. 31, 2024, [Online]. Available: https://arxiv.org/abs/2408.00118.

[8] A. Yang et al., "QWen2.5 Technical Report," arXiv.org, Dec. 19, 2024, [Online]. Available: https://arxiv.org/abs/2412.15115.

[9] "smollm/text/README.md at main · huggingface/smollm," GitHub, [Online]. Available: https://github.com/huggingface/smollm/blob/main/text/README.md.

[10] M. Abdin et al., "PHI-3 Technical Report: A highly capable language model locally on your phone," arXiv.org, Apr. 22, 2024, [Online]. Available: https://arxiv.org/abs/2404.14219.

[11] J. Bai et al., "Qwen Technical Report," arXiv.org, Sep. 28, 2023, [Online]. Available: https://arxiv.org/abs/2309.16609.

[12] Z. Wang et al., "Re-TASK: Revisiting LLM Tasks from Capability, Skill, and Knowledge Perspectives," arXiv.org, Aug. 13, 2024, [Online]. Available: https://arxiv.org/abs/2408.06904.

[13] J. Zhou et al., "Instruction-Following evaluation for large language models," arXiv.org, Nov. 14, 2023, [Online]. Available: https://arxiv.org/abs/2311.07911.

Engaging Students in the Digital Age: A Communicative-Pragmatic and AI-Driven Analysis of Online Prospectuses

Stepan Davydovych, Maryna Davydovych, Kateryna Parnus, Iryna Haidai and Liudmyla Suvorova
Department of Pedagogical Technologies and Language Training, Zhytomyr Polytechnic State University, Chudnivska Str. 103, 10005 Zhytomyr, Ukraine
kptmp_dss@ztu.edu.ua, kim_dms@ztu.edu.ua, katemarchuk38@gmail.com, org_gio@ztu.edu.ua, ktpl_slk@ztu.edu.ua

Keywords: Digital University Prospectuses, Linguistic Strategies, Higher Education Marketing, Natural Language Processing (NLP), Data-Driven Decision-Making, Chatbot-Assisted Guidance, User Experience (UX) In Education, Automated Content Generation, Online Student Recruitment.

Abstract: The shift from printed to digital university prospectuses has transformed higher education marketing, integrating artificial intelligence and interactive technologies to improve communication. As institutions adopt AI-driven personalization, chatbots, and automated content structuring, the linguistic strategies used in prospectuses evolve to maintain clarity, engagement, and persuasion. However, little research has explored how digitalization influences the communicative-pragmatic aspects of these materials. This study examines how universities implement various linguistic strategies to structure digital prospectuses, ensuring they remain informative, persuasive, and accessible. By analyzing contemporary digital formats, the research identifies key trends in content presentation, engagement techniques, and AI-enhanced communication. The findings suggest that while automation improves accessibility and responsiveness, human-centered linguistic frameworks remain essential for maintaining credibility and trust. This study contributes to the understanding of digital prospectuses as adaptive, multimodal texts that merge traditional rhetorical techniques with AI-driven engagement tools. The results highlight the need for universities to balance technological innovation with effective communication strategies, ensuring that digital prospectuses remain user-friendly, persuasive, and aligned with prospective students' expectations in an increasingly digital landscape.

1 INTRODUCTION

Higher education institutions have traditionally used prospectuses to provide essential information about academic programs, campus life, and admissions. These materials were originally distributed in print, offering prospective students a tangible resource to explore their options. However, with the rise of digital technologies and the growing reliance on online platforms, universities have largely replaced printed prospectuses with digital versions available on their official websites [1]. This transition has improved accessibility, allowed for real-time updates, and introduced interactive features that enhance student engagement [3].

In this digital environment, the communicative-pragmatic aspect of university prospectuses has become a key focus of research. Language choices, such as direct address, persuasive techniques, and interactive elements, are crucial in engaging prospective students and influencing their decisions [5, 7]. That's why active usage of AI-powered tools such as chatbots, automated recommendations, and interactive multimedia and understanding linguistic and communicative strategies is essential for making university marketing materials accessible but also engaging and persuasive.

2 RELEVANCE AND ANALYSIS OF THE TOPIC AREA

Existing studies highlight key differences between printed and digital prospectuses. Printed materials offer a static, text-heavy experience that some students prefer for its tangibility and structured presentation [10]. In contrast, digital prospectuses incorporate real-time updates, multimedia elements, and personalized content, enhancing engagement [1, 9].

AI-powered tools such as chatbots and recommendation systems have introduced new levels of interactivity, allowing students to receive instant responses and tailored suggestions [11]. Universities also use Natural Language Processing (NLP) techniques to refine the wording of their prospectuses, ensuring clarity and relevance [6]. While these advancements improve accessibility, they also raise new questions about how digital environments influence language use and whether persuasive strategies differ from those in traditional printed formats.

The communicative-pragmatic approach offers a valuable framework for examining digital prospectuses. Drawing on Speech Act Theory and Discourse Analysis, this study systematically identifies and categorizes specific linguistic features – such as directive speech acts (e.g., imperatives), lexical items signaling institutional authority, and rhetorical devices like repetition and contrast – that are commonly used in digital university prospectuses. Unlike traditional printed materials, digital prospectuses allow for dynamic, interactive communication, making it essential to understand how linguistic strategies evolve in these contexts.

This study explores how AI and digitalization influence the linguistic aspects of university prospectuses. Unlike previous research, which has mainly focused on technical functionality and student engagement metrics, this analysis considers how language use, discourse structures, and persuasive techniques evolve in digital contexts. By bridging linguistic and technological perspectives, this study provides insights into how universities can optimize their communication strategies to effectively engage students in the digital age.

3 METHODOLOGY

This study employs a mixed-methods approach, integrating linguistic analysis and survey-based research to examine communicative-pragmatic strategies in digital university prospectuses. The rationale for this approach lies in the need to investigate both the structural and functional aspects of language in prospectuses and the engagement patterns of prospective students interacting with these materials.

3.1 Survey Participants

The study collected responses from over 150 participants, primarily targeting prospective university applicants. The demographic composition was dominated by undergraduate students, who comprised 74% of the sample, followed by high school students (23%) considering university enrollment (Fig. 1). A smaller proportion consisted of postgraduate students and individuals in other categories. As shown in Figure 1, the majority of participants were undergraduate and high school students actively considering university enrollment.

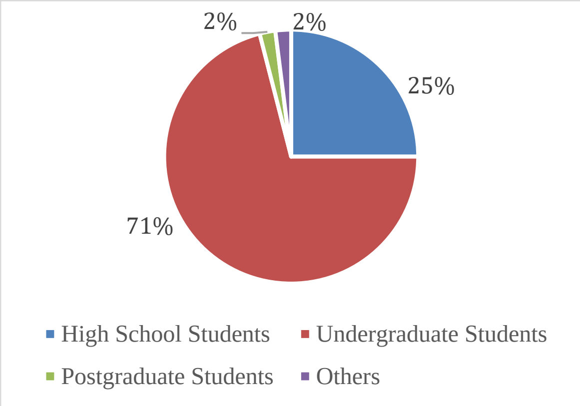


Figure 1: Students' current academic status.

3.2 Corpus Selection

This study analyzed a corpus of 20+ digital university prospectuses from British higher education institutions, collected from their official websites. The following criteria were applied in corpus selection:

1) Institutional Diversity: A mix of highly-ranked and mid-tier universities to reflect variations in communicative strategies.
2) Prospectus Format: Inclusion of both static PDFs and interactive digital prospectuses incorporating multimedia elements (videos, chatbots, AI-driven recommendations).
3) Field Accessibility: Prospectuses were selected from publicly available sources, ensuring transparency and replicability in the analysis.

By limiting the dataset to British universities, this study provides a focused examination of communicative-pragmatic strategies within a specific educational and cultural context. This approach allows for a detailed linguistic comparison while aligning with the broader aim of evaluating AI-driven engagement techniques in UK higher education marketing.

3.3 Research Design

This study employs a mixed-methods approach, integrating linguistic analysis and survey-based research. This approach enables a comprehensive examination of both the linguistic features embedded in university marketing materials and the actual engagement patterns of prospective students interacting with these digital resources.

A mixed-methods approach is particularly suitable for this study because:

1) The linguistic analysis focuses on identifying and coding instances of direct address, modal verbs, personal pronouns, and evaluative adjectives to determine how these features are used to construct persuasive and engaging messaging across different universities.
2) Survey responses validate these linguistic features by assessing their impact on student engagement and perception, offering empirical support for textual findings.
3) The combination of computational text analysis and user perception studies bridges the gap between theoretical linguistic frameworks and real-world digital engagement metrics.

By integrating these methods, this study advances the understanding of how language, technology, and user experience intersect in modern university marketing, contributing to the broader discussion on digital transformation in higher education.

4 RESULTS

While this study's linguistic analysis examined how language is used in digital prospectuses, the survey responses provide complementary insight into which language features and interactive elements students find most engaging. While official university websites remain the most trusted and widely used resource, other channels, such as social media, AI-driven tools, and personal recommendations, also play a crucial role in shaping perceptions and decision-making.

The findings confirm that digital platforms dominate the university research process, with official university websites serving as the primary information source. While social media plays an increasing role in shaping perceptions, traditional formats such as printed brochures have become largely secondary, reinforcing the shift toward online-first recruitment strategies (Fig. 2). These findings emphasize that universities must adopt a multi-channel approach, integrating persuasive digital content, AI-enhanced tools, and social engagement strategies to ensure that prospectuses remain a relevant and effective recruitment tool.

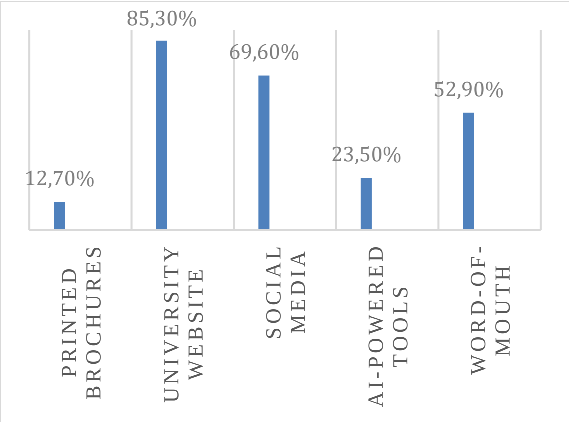


Figure 2: How students typically research universities.

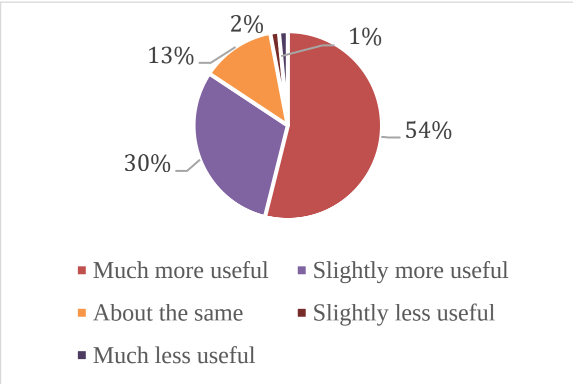


Figure 3: Comparative usefulness of digital and printed university prospectuses.

The findings confirm a clear shift toward digital-first university communication, as illustrated in Figure 3. The preference for digital formats is driven by greater adaptability, accessibility, and interactivity, reinforcing previous results that printed brochures are the least utilized source of university information. While a small portion of students still value physical copies, their role in recruitment is diminishing as universities prioritize dynamic and up-to-date digital resources.

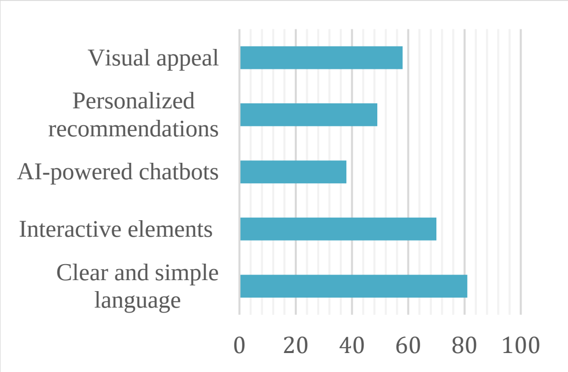


Figure 4: Most engaging features of digital prospectuses.

Students identified clarity of language as the most important factor in engagement (Fig. 4), highlighting the need for well-structured, accessible content in university marketing. Interactive elements, high-quality visuals, and AI-driven personalization were also valued, reflecting an increasing demand for engagement beyond passive reading. These results reinforce the importance of a user-centered approach, where universities combine clear, persuasive language with digital interactivity to enhance engagement and influence decision-making effectively.

The shift to digital university prospectuses has transformed how institutions communicate with prospective students. The integration of AI-driven tools, interactive elements, and data-driven personalization further enhances the following strategies:

1) The ***informative strategy*** ensures that university prospectuses provide concise, structured, and factual content, a crucial aspect in digital communication where students seek quick, easily navigable information. This strategy is realized through several tactics that are increasingly shaped by digital tools:
   1.1) ***Nominative Tactic***. Reinforces institutional identity through strategic university naming, aiding in brand recognition in search algorithms and digital marketing. The University of Manchester prominently features its name and achievements on its undergraduate prospectus homepage, reinforcing its brand identity: *"At Manchester, we've always been about pushing forward. Since 1824, great minds have come here to challenge assumptions, create the exceptional, and change the world* [15]."
   1.2) ***Detailing Tactic***. Structured course descriptions help in SEO optimization and AI-driven content summaries for personalized recommendations. The University of Cambridge offers detailed descriptions of its undergraduate courses, providing prospective students with comprehensive information: *"We offer over 30 undergraduate courses at Cambridge, covering more than 65 subject areas* [16]."
   1.3) ***Regulative Tactic***. Directs students through hyperlinks, AI chatbots, and interactive navigation tools, making information retrieval seamless. The University of York employs interactive navigation tools in its digital prospectus, directing students to relevant information efficiently: *"Our prospectus magazine gives you a deeper insight into what life on campus is really like, from the students already here* [22]."

2) The **argumentative strategy** strengthens a university's appeal by using data-driven arguments, testimonials, and rankings, elements that are increasingly automated and personalized in digital spaces. 39% of students prioritize data-driven language, indicating that AI-driven personalization and dynamic statistic updates enhance credibility:
   2.1) ***Recommendation Tactic.*** Testimonials and alumni success stories are frequently used in AI-driven chatbots and social proof elements on university websites. The University of Sussex showcases its global recognition and alumni achievements to enhance its appeal: *"We're highly ranked in the UK and world league tables, with our research, teaching and outreach recognised internationally* [20]."
   2.2) ***Quality Assurance Tactic.*** Employment statistics and rankings are now delivered dynamically through personalized dashboards and automated recommendation systems. Harper Adams University emphasizes its strong graduate employment rate: *"With cutting-edge facilities, the highest graduate employment rate in the UK (99.2%) and purpose-led education, Harper Adams is the perfect university for any applicant determined to make a lasting impact* [12]."

3) The ***positioning strategy*** establishes a unique institutional identity, crucial in a highly competitive digital space where universities must stand out. 69% of students engage more with interactive digital prospectuses, emphasizing the importance of AI-driven visual storytelling:
   3.1) ***Value Orientation Tactic.*** AI-generated content can dynamically adjust mission statements and value propositions based on user interest.
   3.2) ***Differentiation Tactic***. Machine learning models analyze student preferences to highlight unique features dynamically. For instance, if a prospective student frequently explores environmental science topics, the university's website might prominently feature its cutting-edge sustainability programs and research initiatives.
   3.3) ***Pictorial Visualization Tactic***. VR campus tours, infographics, and interactive media enhance engagement and reinforce branding. Many institutions offer virtual tours that provide immersive experiences. University of Cambridge offers a comprehensive virtual tour featuring 360-degree views of its historic architecture, libraries, and student accommodations, allowing prospective students to explore the campus remotely [17]. University of Warwick provides a 360° virtual tour that includes interactive panoramas of academic buildings, social spaces, and accommodation options, complemented by student ambassador videos [21].

4) The **persuasive framing strategy** subtly shapes student perceptions, and in digital formats, this is enhanced through AI-driven engagement tools, adaptive content, and behavioral targeting. While 33% of students prefer direct address, 39% favor neutral descriptions, suggesting that AI-generated engagement needs a balance between personalization and credibility:
   4.1) ***Intrigue Creation Tactic:*** Personalized email campaigns use automated intrigue-based messaging to boost engagement. The University of Dundee provides guidance on the use of AI in teaching and assessment, reflecting a commitment to integrating advanced technologies into their educational framework [18].
   4.2) ***Authority Appeal Tactic:*** AI curates expert-driven content to reinforce institutional credibility. The University of Liverpool provides comprehensive guidance on the appropriate use and referencing of AI in research, demonstrating leadership in the ethical application of AI technologies [19].
   4.3) ***Intimization Tactic:*** Chatbots and personalized video messages create a sense of direct connection. For instance, the University of Oxford employs chatbots to provide instant support to prospective students, enhancing their engagement with the institution. Additionally, personalized video messages from faculty or current students can create a welcoming atmosphere, making prospects feel valued and connected.

5) The **regulative strategy** focuses on directing student engagement, a role increasingly played by AI-driven tools. AI improves accessibility and engagement, but its effectiveness depends on the accuracy and relevance of recommendations:
   5.1) ***AI Chatbots for Instant Assistance:*** 73.8% of students find chatbots useful but require improvements in context-aware responses. Lancaster University introduced "Ask L.U.," a voice and chatbot interface that assists students with information about timetables, grades, and campus activities. This digital companion is accessible via mobile applications and Amazon Echo devices, enhancing the student experience through real-time assistance [14].
   5.2) ***Personalized AI-Driven Recommendations:*** 60.2% value program and scholarship recommendations, reinforcing the need for adaptive AI-driven engagement. Collaborating with over 600 higher education institutions, including the University of Cambridge and Imperial College London, Unibuddy connects prospective students with current students and alumni. This peer-to-peer engagement platform utilizes data-driven insights to match users based on interests and queries, facilitating personalized interactions [23].
   5.3) ***Automated Summaries and Virtual Tours:*** AI-generated content summaries and interactive campus tours enhance information processing. Keele University

developed an interactive campus map using ThingLink, incorporating audio introductions and 360-degree videos. This virtual tour enables prospective students to familiarize themselves with the campus layout and facilities, enhancing their decision-making process [13].

AI and digital tools have significantly enhanced the effectiveness of linguistic strategies in university prospectuses. Clear structure and organization are improved through AI-generated summaries, while credibility is reinforced with real-time data and testimonials. Personalized content and interactive storytelling help universities stand out, making digital prospectuses more engaging. AI-driven adaptation also allows for more responsive and targeted communication. However, AI should support rather than replace human-driven linguistic strategies, ensuring that digital prospectuses remain clear, trustworthy, and engaging for prospective students.

5 CONCLUSIONS

The shift from printed to digital university prospectuses has transformed how institutions communicate with prospective students. This study highlights how AI-driven personalization, interactive design, and linguistic strategies work together to optimize engagement and influence decision-making. Digital prospectuses are not just informational tools but dynamic, adaptive platforms that integrate clear structure, persuasive content, and personalized recommendations to enhance accessibility and credibility.

The findings confirm that students value clarity, interactivity, and data-driven content, with AI-powered tools playing an increasing role in guiding their choices. However, while AI enhances navigation, personalization, and engagement, it should complement rather than replace human-driven linguistic strategies. Universities must balance automation with authentic, persuasive communication, ensuring that their prospectuses remain informative, engaging, and responsive to student needs in the digital age.

REFERENCES

[1] J. Burdett, "Students as consumers: A commercial imperative for higher education?," Stud. High. Educ., vol. 42, no. 11, pp. 1958-1970, 2017, [Online]. Available: https://doi.org/10.1080/03075079.2015.1127908.

[2] N. Fairclough, Critical Discourse Analysis: The Critical Study of Language, 2nd ed. London, U.K.: Routledge, 2010.

[3] J. Hemsley-Brown and S. Goonawardana, "Brand harmonization in the international higher education market," J. Bus. Res., vol. 60, no. 9, pp. 942-948, 2007, [Online]. Available: https://doi.org/10.1016/j.jbusres.2007.01.019.

[4] J. Hemsley-Brown and I. Oplatka, Higher Education Consumer Choice. London, U.K.: Palgrave Macmillan, 2020.

[5] K. Hyland, Metadiscourse: Exploring Interaction in Writing. London, U.K.: Continuum, 2005.

[6] K. K. Kapoor, K. Tamilmani, N. P. Rana, P. Patil, Y. K. Dwivedi, and S. Nerur, "Advancements in artificial intelligence and implications for marketing: A review and research agenda," Int. J. Inf. Manag., vol. 57, p. 102225, 2021, [Online]. Available: https://doi.org/10.1016/j.ijinfomgt.2020.102225.

[7] G. Kress and T. van Leeuwen, Reading Images: The Grammar of Visual Design, 3rd ed. London, U.K.: Routledge, 2021.

[8] J. A. Lindholm, K. Szelenyi, S. Hurtado, and W. S. Korn, The American College Student: The Demographics of Student Engagement. Baltimore, MD: Johns Hopkins Univ. Press, 2020.

[9] A. Peruta and A. B. Shields, "Marketing higher education: A student recruitment model," J. Market. High. Educ., vol. 28, no. 2, pp. 226-245, 2018, [Online]. Available: https://doi.org/10.1080/08841241.2018.1457914.

[10] J. J. Selingo, The New U: Reshaping Higher Education in the Age of Digital Disruption. Washington, DC: The Chronicle of Higher Education, 2018.

[11] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education," Int. J. Educ. Technol. High. Educ., vol. 16, no. 1, pp. 1-27, 2019, [Online]. Available: https://doi.org/10.1186/s41239-019-0171-0.

[12] Harper Adams University, "Your Digital Prospectus," 2024, [Online]. Available: https://www.harper-adams.ac.uk/study/1011/your-digital-prospectus.

[13] Keele University, "Interactive Campus Map," 2024, [Online]. Available: https://www.thinglink.com/blog/keele-university-uses-thinglink.

[14] Lancaster University, "AI Chatbot 'Ask L.U.'," 2024, [Online]. Available: https://www.lancaster.ac.uk/news/lancaster-university-launch-pioneering-chatbot-companion-for-students.

[15] The University of Manchester, "Undergraduate Prospectus," 2024, [Online]. Available: https://www.manchester.ac.uk/study/undergraduate/prospectus.

[16] University of Cambridge, "Undergraduate Study," 2024, [Online]. Available: https://www.undergraduate.study.cam.ac.uk.

[17] University of Cambridge, "Virtual Tour," 2024, [Online]. Available: https://www.virtualtour.study.cam.ac.uk.

[18] University of Dundee, "AI in Teaching and Assessment," 2024, [Online]. Available: https://www.dundee.ac.uk/guides/ai-artificial-intelligence-teaching-and-assessment.

[19] University of Liverpool, "AI in Research," 2024, [Online]. Available: https://ukrio.org/ukrio-resources/ai-in-research.

[20] University of Sussex, "Rankings and Reputation," 2024, [Online]. Available: https://www.sussex.ac.uk/about/facts/rankings.

[21] University of Warwick, "Undergraduate Virtual Tour," 2024, [Online]. Available: https://warwick.ac.uk/study/undergraduate/virtualtour.

[22] University of York, "Undergraduate Prospectus," 2024, [Online]. Available: https://www.york.ac.uk/study/undergraduate/prospectus.

[23] Unibuddy, "Peer-to-Peer Student Engagement," 2024, [Online]. Available: https://unibuddy.com/blog/postgraduate-peer-to-peer/.

Automated Assessment of Student Queries in Redis

Yuliia Prokop
Department of Computer Science, Czech Technical University in Prague, Jugoslávských partyzánů 1580/3, 160 00 Prague, Czech Republic
prokoyul@fel.cvut.cz

Keywords: NoSQL, Redis, Key-value Store, Automated Assessment, RedisJSON, RediSearch, Education, Databases.

Abstract: In recent years, the popularity of NoSQL systems has grown significantly due to their flexibility and high performance when working with large volumes of data. Redis is one of the most popular key-value stores actively used in industry and education. However, automated approaches for NoSQL assignment evaluation, especially those involving advanced Redis modules, remain underdeveloped. This paper presents a web-based system for automated assessment of students' Redis queries, supporting basic structures (e.g., list, sorted set, and hash manipulation) and advanced features (RedisJSON and RediSearch). The system provides instant feedback on syntax errors, enabling students to correct mistakes and resubmit solutions in real-time. A pilot study with 42 master's students showed that about 78% successfully mastered the basics of Redis on the first try, while only 39% passed advanced assignments. With repeated attempts and targeted feedback, overall success on advanced tasks increased to 76%, highlighting the importance of continuous, automated guidance. The paper also discusses typical errors logged by the system (inconsistent or incorrect key naming, syntax errors when setting key expiration, incorrect index creation or JSON field references, etc.). The results demonstrate that integrating an automated Redis query assessment into educational programs can significantly enhance student engagement and learning efficiency. The flexible and modular design of the proposed system allows easy extension to other NoSQL databases and provides valuable data for instructors to refine teaching materials.

1 INTRODUCTION

Digital transformation drives the continuous development of data processing technologies and generates new technical challenges. Over the last decade, the increased demand for efficient handling of big, unstructured, and semi-structured data has led to the active development of NoSQL database management systems. For instance, the rapid growth of real-time analytics and IoT applications underscores the need for more flexible and scalable data solutions.

According to DB-Engines Ranking [1], as of February 2025, relational DBMSs still occupy about 72% of the market, while the remaining 28% are represented mainly by NoSQL solutions. At the same time, the demand for specialists familiar with non-relational DBMSs is growing, which calls for modernizing the curricula in technical universities. In response, many educational programs now integrate specialized modules on NoSQL technologies to better equip students for emerging industry demands. The key direction of modernization is the development of automated knowledge assessment systems that allow students to get real-time feedback when performing database tasks.

Despite developing tools for validating SQL and some document-oriented databases, key-value stores remain underrepresented in automated evaluation. Redis, one of the most popular in-memory solutions, is rarely considered a full-fledged teaching and validation object, especially for its advanced features (RedisJSON, RediSearch). This paper fills this gap by proposing a methodological and software toolkit for the automated assessment of Redis skills.

2 LITERATURE REVIEW

Many modern studies indicate a steady trend toward integrating NoSQL technologies into academic database programs [2 – 9]. The practice-oriented (hands-on) approach is the most widely used. It helps students to deepen their understanding of modern data storage and processing methods.

The modern ecosystem of non-relational DBMSs includes several major models (key-value stores, document stores, wide-column stores, and graph databases) [2, 10]. Redis is ranked 6th among all databases in the DB-Engines ranking and is widely used in the industry. It is often demonstrated to students as an in-memory NoSQL system showing the basic concepts: caching, message brokering, and real-time data processing. Some studies [2 – 4] confirm that integrating Redis into academic courses helps students better understand NoSQL's theoretical aspects and acquire skills relevant to modern industrial tasks.

Some authors study the problems of evaluating student queries in NoSQL. Thus, the paper [11] describes the peer correction technique: students assess their colleagues' queries, identifying syntactic and semantic errors in SQL and NoSQL. The authors show that the average accuracy of such evaluations reaches 83%, and the correction results correlate moderately with the overall performance. The authors conclude that peer correction contributes to a better understanding of one's own errors and the development of query writing skills.

The paper [12] presents a TriQL system designed to teach three types of databases at once: relational (MySQL), graph (Neo4j), and document-oriented (MongoDB). Student work is assessed through lab assignments in which students compare the results of their queries executed on native engines. Such an approach helps to analyze the differences in depth and realize the advantages and limitations of each model. The authors of the paper [13] explore the possibilities of automatically evaluating student queries in SQL, MongoDB, and Neo4j using a combination of the PrairieLearn platform and proprietary analysis algorithms.

The articles [14, 15] describe a similar mechanism. Students' solutions in MongoDB and graph databases are executed through the PrairieLearn online platform, which automatically matches the results to a reference solution. Works [16, 17] describe the integration of the NoSQL Data Adapter module into the Moodle system, allowing students to perform queries remotely on NoSQL databases and receive instant automatic feedback.

Analysis of current research shows that many authors focus on pedagogical aspects of learning and do not study query assessment automation. In addition, these works rarely address key-value systems.

Existing platforms for automatic evaluation of student queries predominantly lack built-in NoSQL support. Some specialized solutions exist, e.g., MongoDB University (https://learn.mongodb.com/), Neo4j Sandbox (https://neo4j.com/sandbox/), DataStax Academy for Cassandra (https://www.datastax.com/dev/academy), and Redis University (university.redis.com). They provide interactive tools to teach mostly only basic operations, and the possibilities to integrate them into the classroom are limited.

Thus, integrating NoSQL into university curricula and searching for efficient automatic evaluation of students' assignments are topical problems that attract the attention of modern researchers. However, key-value stores – particularly Redis – do not receive sufficient attention in most works. The methodological aspects of teaching and automatic assessment of Redis's advanced features (hash and JSON data indexing, full-text search, aggregation, etc.) have also been insufficiently studied and require further research.

This paper aims to develop and justify a specialized methodological and software toolkit for the automated assessment of student queries in Redis with instant feedback and the ability to correct errors in the same session.

3 METHODS AND MATERIALS

3.1 General Solution Architecture

As part of the research, we developed a web application to automate assessing students' knowledge of using Redis. Its main components are:

- Web interface (client part) for entering queries and viewing results;
- Server part, which processes incoming queries and interacts with Redis;
- Redis NoSQL server, where commands are executed (taking into account RedisJSON and RediSearch modules);
- A database (MongoDB) for logging and storing the validation results.

In designing the application, we emphasized a modular architecture that facilitates straightforward updates and expansions. Instructors can modify or add functionalities by ensuring each component is loosely coupled without impacting other system parts.

The architectural diagram of the system is presented in Fig. 1.

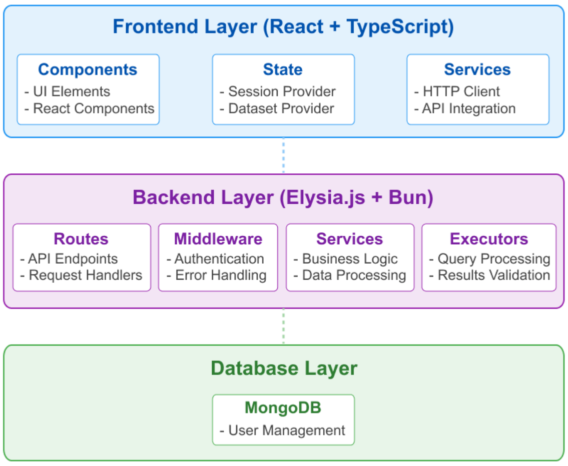


Figure 1: Architectural diagram of the system (source: author's development).

The interaction of components takes place according to the scheme “client-server-Redis-MongoDB.” The web interface accesses the server part, which forms and forwards requests to Redis. After that, the results are returned for display to the student, and at the same time, logs are written to MongoDB.

3.2 Materials and Data Preparation

A dataset covering all key Redis types – strings, lists, sets, sorted sets, hashes, bit fields, and geodata – has been developed to ensure comprehensive testing. It includes about 200 queries for inserting various information about stores, products, and customers, such as:

```
HSET product:1 name "Espresso" price
2.50 calories 3 caffeine 64 size "1 oz"
ZADD store:1:menu 1 "Espresso" 2
"Latte" 3 "Cappuccino" 4 "Americano" 5
"Mocha" 6 "Flat White" 7 "Cold Brew" 8
"Frappe" 9 "Green Tea" 10 "Chai Latte"
LPUSH store:1:recent_orders "Vanilla
Latte" "Espresso" "Caramel Macchiato"
"Americano" "Mocha" "Cappuccino" "Flat
White" "Cold Brew" "Frappe" "Green Tea"
SADD loyalty:Gold customer:1 customer:4
GEOADD store_locations -122.2585
37.8614 "CityBrew University"
```

Each task is stored as a JSON document and includes:

- task description (context and required actions);
- a reference solution (one or more correct queries);
- test query (optional – provided only when the operation does not return a result).

Queries are grouped in collections according to the topics to which they relate.

Example of a task JSON document (from the data selection collection):

```
{
"Question": "Get the store with the
lowest sales for the current month.",
"Solution": ["ZRANGE monthly_sales 0 0
WITHSCORES"],
"Test": ""
}
```

Example of a task from the RediSearch collection:

```
{
"Question": "Create an index and find
users with Gold loyalty status. Return:
name, email, loyalty_status. Sort by
'name' ascending. Use LIMIT 0 1000.",
"Solution": [
"FT.CREATE
idx:customer_loyalty_only_gold ON HASH
PREFIX 1 customer: SCHEMA
loyalty_status TEXT SORTABLE name TEXT
SORTABLE email TEXT SORTABLE",
"FT.SEARCH
idx:customer_loyalty_only_gold
'@loyalty_status:Gold' SORTBY name ASC
LIMIT 0 1000 RETURN 3 name email
loyalty_status"],
"Test": ""
}
```

The teacher creates tasks by combining different data structures and scenarios. The level of task complexity in the system is not directly fixed. If necessary, the teacher can flexibly control the “depth” and complexity of queries, forming a variety of task collections.

For more precise control over the progress of learning, the tasks are divided into two main groups: basic operations (creation, modification, selection, deletion, simple aggregates) and advanced functions (RedisJSON, RediSearch): index creation, full-text search, aggregation, and work with complex JSON structures.

3.3 Methods of Assessment and System Operation

The algorithm for giving and executing assignments consists of these steps:

1) Selecting an assignment group: The student indicates whether they want to do basic or advanced assignments in the web interface.
2) Task list generation: Based on the teacher's settings, the system randomly generates a set of assignments with the required number of tasks.
3) Task display: On the web application page, the student sees the text of the current task and an area to enter a Redis query (or queries).
4) Entering and executing: After entering the query text, the student clicks the "Execute" button; they can skip the question by clicking the "Skip" button.

The system performs the following steps automatically:

1) Parsing and splitting queries: The text the user sends as a single answer to the current question is split into separate Redis commands.
2) Resetting the test environment: Before each test query, the initial dataset in Redis is recreated anew, and all previously created indexes are deleted. This guarantees the same initial database state.
3) Student query execution: The system sends commands to the Redis server. If Redis returns a syntax error, the user is notified immediately, and the testing process is paused. The student can correct and send the query again for execution (return to step 1).
4) Comparison with the reference solution: If the query is executed without errors, the system resets the data to its original state and executes the reference query. The results (full text) are then compared with the student query.
5) Saving logs: The testing process (task description, reference solution and result, student's solution and result, and comparison result in the form of true or false) is recorded in MongoDB.
6) Final result output: The system provides the student with a message listing the number of correct queries, a list of incorrect queries, and a notification about the test's success.

The result of each task is considered correct if the final state of the data in Redis or the returned result matches the state of the data or the result of the reference solution.

The system uses a full-text comparison of the output results of two solutions (user and reference). Each task description is designed to avoid evaluation ambiguities so that the final database state or query result is clearly defined once all requirements are met. For example, the assignment explicitly specifies the sorting order of the output records, the list of fields to include in the result, the numeric data types for the fields in the created index, and the record limit. This ensures high evaluation accuracy.

However, despite its simplicity and reliability, this approach has certain limitations. To address these, it is planned to introduce structure-oriented comparison and more flexible matching rules (e.g., ignoring character case and the order of records in the result or supporting incomplete matches). These improvements will extend the system's functionality and allow a more accurate evaluation of students' queries, not limiting them to a rigid full-text check.

The block diagram of the algorithm for automatic query evaluation is shown in Fig. 2.

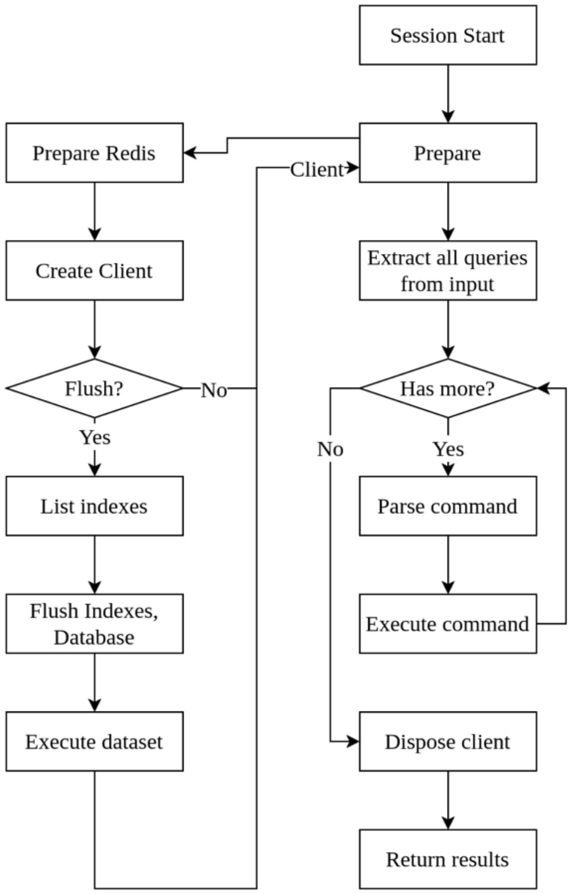


Figure 2: Block diagram of the algorithm for automatic query assessment (source: author's development).

The teacher sets the minimum number of correct solutions required to pass the test. Based on these settings, a notification about the test's success is generated. If the student falls short, they can retake the test.

3.4 User Interface

The web application implements authentication via login and password matching the credentials of the NoSQL server. From a cybersecurity perspective, this approach restricts access solely to enrolled university students, as external registration is disabled. By pre-registering accounts on the NoSQL server – managed by the instructor before the semester begins – the system minimizes the risk of unauthorized access and potential security breaches, ensuring that only verified and trusted users can interact with the testing environment.

After logging in, the following functionality is available to the students:

- selection of test type (“Redis basic” or “Redis advanced” buttons);
- list of current tasks (issued sequentially);
- field for entering a query;
- “Execute” button;
- “Skip” button (in case the input field is empty);
- “Open dataset” button;
- “Quit” button to abort the test;
- displaying Redis syntax errors (if any).

Figure 3 shows a screenshot of the window while testing Redis advanced queries. The “Execute” button is inactive until the answer is entered. The gray text in the answer field reminds users that queries should be separated by a blank line. The system also allows the assessment of queries in MongoDB, and this functionality is currently being tested.

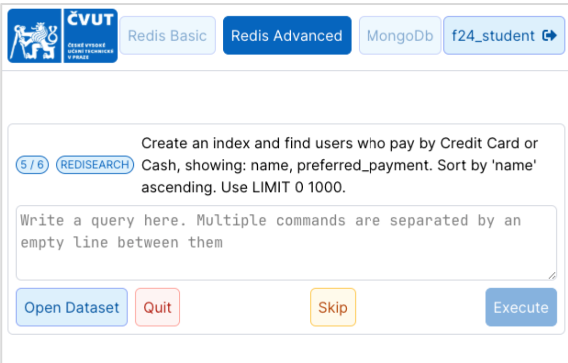


Figure 3: Screenshot of the user window when taking the test (source: author's development).

After the test is completed, the student receives final feedback. Skipped tasks are marked as incorrect and added to the list of incorrect queries.

The web interface of the teacher (Fig. 4) provides the following features:

- select a specific test kind and load a collection of tasks;
- configure the number of tasks from each collection that will be included in the test;
- delete a collection of tasks.

The administrator loads the data set (a set of commands for inserting data of different types). Logs can be exported in JSON format for review and analysis by the teacher.

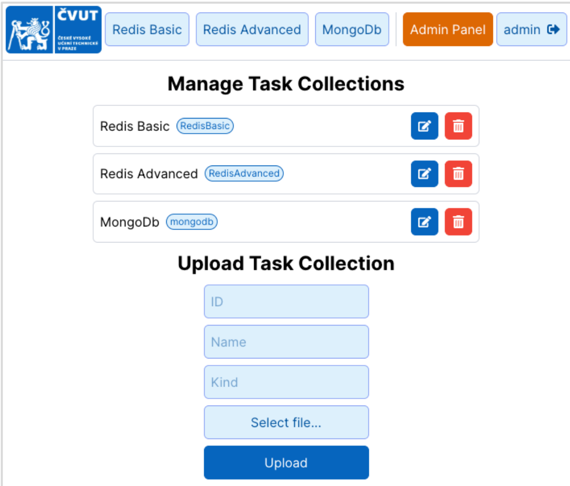


Figure 4: Screenshot of the instructor’s panel (source: author's development).

4 RESULTS AND DISCUSSION

For the pilot implementation of the system, 42 first-year master’s students of the specialty “Open Informatics” were selected. All of them were willing to take the proposed tests. Participation in testing was not a mandatory element of the course: students took the tests to better prepare for the exam and to get additional feedback. This approach, on the one hand, ensures high motivation among test takers. Still, on the other hand, it may lead to a particular sampling bias (more interested or prepared students participate predominantly). In the future, it is planned to make the test mandatory and include it in the system of admission to the exam.

Students performed two types of tasks: first, they mastered basic operations (creating, selecting, updating, and deleting data in Redis) and then moved on to advanced assignments related to the RedisJSON and RediSearch modules. Each participant took the test individually, receiving basic and advanced tasks randomly.

The results showed that most students (about 78%) successfully coped with basic tasks on the first attempt. The main errors were related to syntactic inaccuracies in queries and attempts to scan across the hash fields. Thanks to the possibility of instant checking and resubmitting in case of a mistake, the average percentage of completed basic tasks reached 93% after two or three attempts. This indicates the system provides the dynamic feedback needed to identify and correct errors in Redis basic commands quickly.

Figure 5 shows a test log fragment with a student’s successful answer.


```
▾ 11: Object
    kind : "Update"
    question : "Add 'Cinnamon Dolce Latte' to the end of the recent_orders list for st…"
  ▾ solution : Array (1)
      0: "RPUSH store:1:recent_orders "Cinnamon Dolce Latte""
    test : "LRANGE store:1:recent_orders -1 -1"
  ▾ userSolution : Array (1)
      0: "RPUSH store:1:recent_orders "Cinnamon Dolce Latte""
    response : "1"
    testResponse : "[
                     "Cinnamon Dolce Latte"
                    ]"
    userResponse : "1"
    userTestResponse : "[
                         "Cinnamon Dolce Latte"
                        ]"
    correct : true
```


Figure 5: Test log fragment (source: author's development).

A common issue arose when students were asked to create additional data structures (e.g., sets) to serve as reverse indexes linked to existing hashes. Although students demonstrated competence in using standard hash commands (HSET, HGET, etc.), they struggled with conceptualizing how to populate complementary sets for more advanced searching.

Students sometimes overlooked the specific naming conventions provided in the dataset. For instance, if the dataset used a particular prefix or a structured naming scheme (e.g., `store:1:equipment`), some students would create new keys without following the same format. This discrepancy could lead to mismatches with the reference solutions and to “incorrect result” flags in the automated checker.

Another issue was the misuse of spaces in key names. Students occasionally introduced unintended spaces or special characters, causing Redis to misinterpret the command and corrupt the intended data structure.

Another problem was the use of incorrect syntax when specifying the key “time to live”. Many students forgot to provide all the necessary parameters or specified the time in the wrong units.

More problematic was the case with advanced Redis features, including RedisJSON and RediSearch capabilities. Only 39% of students could complete the test at the end of the first pass.

Many students struggled to configure the index correctly when using RediSearch for full-text search or indexing JSON documents. Some omitted essential fields, causing incomplete search results, while others declared fields with the wrong data type. Even if the index was set up correctly, students often encountered errors when making search queries, such as improperly accessing fields in a JSON document.

However, with repeated attempts and analysis of the errors displayed by the system, the success rate increased to 76%. This confirms that continuous automatic feedback gives students a deeper understanding of advanced Redis features.

In addition to quantitative measures for automatic assessment, the system provided valuable information for further adjustments to the course materials.

Thus, the results of implementation demonstrate the high efficiency of the system and the relevance of its use in the educational process. The proposed solution can be easily integrated into NoSQL training courses, complementing both classical lecture-practical formats and distance learning programs.

In the future, we plan to expand the assignment bank to include more “real-world” scenarios and to refine the hint system so that it not only indicates when an error has occurred but also gives guidance on how to correct it.

The advantage of the developed systém is that it is easily extensible to other NoSQL databases. In particular, it currently allows testing queries in MongoDB and can be extended to Cassandra and Neo4j. Such improvements will further enhance the quality of training and student satisfaction with the results of their studies.

5 CONCLUSIONS

A literature review shows that modern NoSQL technologies are widely recognized as necessary for academic curricula. However, automatic assessment, especially for Redis queries, remains poorly studied. Therefore, a system must be developed to evaluate Redis queries automatically.

The automated assessment system for student queries in Redis discussed in this paper demonstrates its effectiveness in teaching fundamental NoSQL concepts. Its key feature is instant feedback, which is provided in real-time. This allows students to identify and correct any syntax errors made immediately. Another essential capability of the created system is the support of advanced Redis modules.

Unlike existing online courses and sandboxes for Redis, the system presented here allows the instructor to customize assignment types (including advanced and more complex tasks), define a passing threshold, and track all user activity in real time.

An essential advantage of the system is its modular architecture, which allows scaling the solution for any scenario and expanding the task bank. In addition, the system is easily extensible to other NoSQL databases (MongoDB, Cassandra, Neo4j).

The pilot implementation showed that students have the most difficulties due to an incomplete understanding of the principles of working with key-value databases. For example, they try to perform cross-hash scanning across all fields in basic queries. In addition, they have difficulties mastering advanced Redis features related to RedisJSON and RediSearch modules. The high error rate in tasks related to index creation and full-text search highlights the need to emphasize these topics in both the course's theoretical instruction and practical classes.

Storing tasks and reference solutions in JSON format facilitates the automatic checking of results and subsequent analysis of student's practical skills. The possibility of flexibly adjusting the number and types of tasks and randomization mechanisms opens prospects for more accurately matching test sessions to individual students' skill levels. This, together, increases the efficiency of the educational process.

The study results confirm that automated query assessment systems are promising tools for university courses focused on modern data management technologies. They can be effectively applied in traditional lecture-practice learning formats and online learning platforms.

A further research direction could involve the development of structure-oriented comparison methods and flexible matching rules – such as case-insensitive checks, record-order independence and support for partial matches – to overcome the limitations of the current full-text comparison, broaden the system's applicability and yield more accurate assessments of student queries.

REFERENCES

[1] DB-Engines, "DBMS popularity broken down by database model," DB-Engines, [Online]. Available: https://db-engines.com/en/ranking_categories. [Accessed: 19-Feb-2025].

[2] N. Tripathi, "Teaching NoSQL databases in higher education: A review of models, tools and methods," SSRN, 2024, [Online]. Available: http://dx.doi.org/10.2139/ssrn.4971813.

[3] M. Menzin, S. Mohan, D. R. Musicant, and R. Soori Murthi, "NoSQL in undergrad courses is no problem," in Proceedings of the 51st ACM Technical Symposium on Computer Science Education, 2020, pp. 962-963, [Online]. Available: https://doi.org/10.1145/3328778.3366909.

[4] S. Mohan, "Teaching NoSQL databases to undergraduate students: a novel approach," in Proceedings of the 49th ACM Technical Symposium on Computer Science Education, 2018, pp. 314-319, [Online]. Available: https://doi.org/10.1145/3159450.3159554.

[5] M. Greiner, "Teaching NoSQL data models: a tutorial," in Proceedings of the 2021 AMCIS, 2021, Article 19.

[6] S. Kim, "Seamless integration of NoSQL class into the database curriculum," in Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education, 2020, pp. 314-320, [Online]. Available: https://doi.org/10.1145/3341525.3387399.

[7] M. Lin, J. Jun, Y. Zhu, and C. Zhan, "Research on the teaching reform of database curriculum major in computer in big data era," in Proc. 2017 12th Int. Conf. Computer Science and Education (ICCSE), IEEE, 2017, pp. 570-573.

[8] B. Fowler, J. Godin, and M. Geddy, "Teaching case: introduction to NoSQL in a traditional database course," Journal of Information Systems Education, vol. 27, no. 2, p. 99, 2016.

[9] A. Bajaj and W. Bick, "The rise of NoSQL systems: research and pedagogy," Journal of Database Management, vol. 31, no. 3, pp. 67-82, 2020, [Online]. Available: https://doi.org/10.4018/JDM.2020070104.

[10] C. Costa and M. Santos, "Big data: State-of-the-art concepts, techniques, technologies, modeling approaches and research challenges," IAENG International Journal of Computer Science, vol. 44, pp. 285-301, 2017.

[11] W. Wu, "Assessing peer correction of SQL and NoSQL queries," in Proceedings of the 54th ACM Technical Symposium on Computer Science Education V.1, 2023, pp. 535-541.

[12] A. Alawini, P. Rao, L. Zhou, L. Kang, and P.-C. Ho, "Teaching data models with TriQL," in Proceedings of the 1st International Workshop on Data Systems Education, 2022, pp. 16-21.

[13] Z. Li, S. Yang, K. Cunningham, and A. Alawini, "Assessing Student Learning Across Various Database Query Languages," 2023 IEEE Frontiers in Education Conference (FIE), College Station, TX, USA, 2023, pp. 1-9, [Online]. Available: https://doi.org/10.1109/FIE58773.2023.10343409.

[14] R. Alkhabaz, S. Poulsen, M. Chen, and A. Alawini, "Insights from student solutions to MongoDB homework problems," in Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education V.1, 2021, pp. 276-282, [Online]. Available: https://doi.org/10.1145/3430665.3456308.

[15] M. Chen, S. Poulsen, R. Alkhabaz, and A. Alawini, "A quantitative analysis of student solutions to graph database problems," in Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education V.1, 2021, pp. 283-289, [Online]. Available: https://doi.org/10.1145/3430665.3456314.

[16] A. Werner, "Applying NoSQL data adapter with the learning paths mechanism for better knowledge transfer in the age of distance learning," Procedia Computer Science, vol. 207, pp. 3330-3339, 2022, [Online]. Available: https://doi.org/10.1016/j.procs.2022.09.424.

[17] A. Werner and M. Bach, "NoSQL e-learning laboratory-interactive querying of MongoDB and CouchDB and their conversion to a relational database," in Communications in Computer and Information Science, vol. 865, 2018, pp. 581-592, [Online]. Available: https://doi.org/10.1007/978-3-319-67792-7_56.

SECTION 3

CONTROL AND AUTOMATION

Proceedings of the 13th International Conference on Applied Innovations in IT (ICAIIT)

APRIL 2025

Calibration of the Open-Vocabulary Model YOLO-World by Using Temperature Scaling

Max Andreas Ingrisch, Subashkumar Rajanayagam, Ingo Chmielewski and Stefan Twieg
Department of Electrical, Mechanical and Industrial Engineering, HS Anhalt, Bernburger Str. 55, 06366 Köthen, Germany
{*max.ingrisch, subashkumar.rajanayagam, ingo.chmielewski, stefan.twieg*}*@hs-anhalt.de*

Keywords: Calibration, YOLO-World, Temperature Scaling, Expected Calibration Error, Open-Vocabulary Detection.

Abstract: In many areas of the real world, such as robotics and autonomous driving, deep learning models are an indispensable tool for detecting objects in the environment. In recent years, supervised models such as YOLO or Faster R-CNN have been increasingly used for this purpose. One disadvantage of these models is that they can only detect objects within a closed vocabulary. To overcome this limitation, research is currently being conducted into models that can also detect objects outside the known classes of the training data set. A model is therefore trained with base classes and can recognize novel, unseen classes – this is referred to as open-vocabulary detection (OVD). Novel models such as YOLO-World offer a solution to this problem, but they tend to over- or underestimate when calculating confidence values and are therefore often poorly calibrated. However, reliable determination of confidence values is a crucial factor for the use of these models in the real world to ensure safety and trustworthiness. To address this problem, this paper investigates the influence of the calibration method temperature scaling on the OVD model YOLO-World. The optimal T-value is determined by 2 calibration data sets (Pascal VOC and Open Images V7) and then evaluated on the LVIS minival dataset. The results show that the use of temperature scaling improved the Expected Calibration Error (ECE) from 6.78% to 2.31%, but the model still tends to overestimate the confidence values in some bins.

1 INTRODUCTION

Object detection plays an essential role in many areas and systems. These include robotics and autonomous driving, where it enables interaction with the environment (e.g. traffic sign recognition [1]) and helps to avoid unwanted collisions. In recent years, deep learning (DL) techniques have been used to develop models such as *YOLO* [2] and *Faster R-CNN* [3]. However, these models are limited to the detection of objects that were learned during training (closed-set object detection). To overcome this limitation, more research has been done on models that can also detect objects outside the training data set. This is called Open-Vocabulary Detection (OVD). [4] CLIP (Contrastive Language-Image Pre-training) [5], which involves the joint training of a text and image encoder, is an important step towards the realization of an OVD model. This approach attempts to extract text features and image features and map them in a common embedding space. Based on CLIP, OVD models such as *YOLO-World* [6] or *Det-CLIP* [7] were developed. There are a variety of other OVD models such as *Grounding DINO* [8], which uses a BERT text encoder instead of CLIP.

Despite the achieved generalization capability of these models, there is currently a problem that the calculated confidence values are not always reliable, as shown for Grounding DINO in [9]. This means that the models tend to be over- or underconfident – in some cases, true positives (TP) have too low confidence values, while false positives (FP) have too high. The models are therefore not well calibrated. Without a reliable determination of the confidence values, the use of such models in the real world is associated with some risks - for example, if a red traffic light is identified as green with 95% certainty in 90% of cases.

To overcome the problem of miss-calibration, there are numerous methods, which are presented in [10] and [11] for Neural Networks. In this paper, the calibration of YOLO-World is first examined by calculating the Expected Calibration Error (ECE) and displaying the corresponding Reliability Diagram. From this, it can be deduced how large the calibration error is and whether the model tends to over- or underestimate the confidence values. Subsequently, the simple calibration method *Temperature Scaling* is applied in accordance with [12] and the influence on the confidence values and the calibration error is determined. Two calibration data sets, Pascal VOC [13] and Open Images V7 [14, 15], are used to determine the optimal T-value. Furthermore, the influence of the calibration on the accuracy of the model is examined by calculating the mAP (mean Average Precision) for each determined T-value. YOLO-World was selected for this study because, according to [6], it outperforms other models, such as Grounding DINO and Det-CLIP, in terms of mAP (mean Average Precision) on the LVIS minival [16] dataset.

2 BACKGROUND

2.1 YOLO-World

YOLO-World [6] is an OVD model that can detect novel classes, i.e., objects that were not included in the training data set, in an image. For this purpose, YOLO-World uses a frozen text encoder (CLIP), which generates the text embeddings $W = TextEncoder(F) \in \mathbb{R}^{C \times D}$ from a given text F, where C is the number of classes and D is the dimensionality of the embeddings. Here, $w_j \in W$ represents the j-th text embedding. The YOLO backbone (YOLOv8) then extracts the image features of an image I. From the given input of image I and text F, K object embeddings $\{e_k\}_{k=1}^{K}$ with $e_k \in \mathbb{R}^D$ are then generated. The logit (non-probabilistic output of the network [11]) can thus be formulated as

$$l_{k,j} = \alpha \cdot \text{Batch-Norm}(e_k) \cdot \text{L2-Norm}(w_j)^T + \beta \ . \quad (1)$$

In equation 1, $l_{k,j}$ represents the logit, i.e., the object-text similarity, between the k-th object embedding and the j-th text embedding. The text and object embeddings are previously L2-normalized or Batch-normalized and the product is scaled by two constants α and β, which are learned during the training process. After that, the logits are activated by a sigmoid function [17], which is defined as

$$\sigma(l_{k,j}) = \frac{1}{1 + e^{-l_{k,j}}} \ . \quad (2)$$

In a sigmoid function, the probability $\sigma(l_{k,j})$ of each logit is calculated independently, in contrast to the softmax function.

2.2 Temperature Scaling

The *temperature scaling* calibration method is a method that is applied after the training process of an DL model. For this, the logits are scaled with a temperature value $T > 0$, i.e. a constant value. [10, 11] The optimal temperature value T must be learned using a separate calibration data set. According to [10, 11], this method is very efficient and performs better than other calibration methods in vision tasks. The scaled logits can be calculated using the equation

$$l_{k,j}^{cal} = \frac{l_{k,j}}{T} \ , \quad (3)$$

where $l_{k,j}^{cal}$ are the logits after applying temperature scaling.

2.3 Mean Average Precision (mAP)

The mAP (mean Average Precision) is a metric for determining the performance of a model. To calculate this metric, the AP (average precision) value is averaged across all classes K [18]:

$$mAP = \frac{1}{K} \sum_{i=1}^{K} AP_i \ . \quad (4)$$

To determine the AP, two further values, *recall* and *precision*, must first be determined, which can be derived from the number of true positives (TP), false positives (FP) and false negatives (FN). Further information can be found in [19]. For each confidence threshold, a recall and a precision value is calculated and displayed in a so-called *recall-precision curve*. The approximated area under this curve then yields the AP.

2.4 Reliability Diagram and Expected Calibration Error (ECE)

Reliability diagrams can be used to visualize the calibration of a model graphically. [10, 11] Let i be a sample consisting of the maximum confidence value $\hat{p}_i$, the associated predicted label $\hat{y}_i$, and the true class label y_i. First, all samples i are divided into M bins $(B_1, ..., B_M)$ based on their confidence values $\hat{p}_i$, where the interval size is $1/M$. B_m contains the set of samples that fall within the confidence interval $I_m = (\frac{m-1}{M}, \frac{m}{M}]$, where $m \in 1, ..., M$. According to this classification, two metrics can be derived that are essential for the visualization of the reliability diagram as well as for the calculation of the expected calibration error. The accuracy $acc(B_m)$, which is defined as

$$acc(B_m) = \frac{1}{|B_m|} \sum_{i \in B_m} \mathbf{1}(\hat{y}_i = y_i) \ , \quad (5)$$

determines the ratio of the number of samples that have correctly predicted the label to the total number of samples ($|B_m|$) of the bin B_m. In contrast to accuracy, the confidence of B_m,

$$conf(B_m) = \frac{1}{|B_m|} \sum_{i \in B_m} \hat{p}_i \, , \tag{6}$$

calculates the mean of all predicted confidence values $\hat{p}_i$ in the bin B_m.

In a reliability diagram, the accuracy and confidence are then displayed for each bin B_m, where a deviation from the diagonal $f(x) = x$ symbolizes a miss-calibration. [10, 12] In Figure 2, a blue bar indicates the $acc(B_m)$, while a red bar always indicates a gap between $acc(B_m)$ and $conf(B_m)$, i.e., a calibration error. For a perfectly calibrated model is $acc(B_m) = conf(B_m)$ for all $m \in 1,...,M$. Since the reliability diagram does not take into account the number of samples per bin ($|B_m|$), a confidence histogram is also illustrated. This provides an overview of the percentage of samples, based on the total number, in each bin B_m (see Figure 2).

The Expected Calibration Error (ECE) reflects a value that quantitatively captures the calibration of the model. To do this, the weighted mean of the difference between $acc(B_m)$ and $conf(B_m)$ is calculated,

$$ECE = \sum_{m=1}^{M} \frac{|B_m|}{n} \left| \mathrm{acc}(B_m) - conf(B_m) \right| . \tag{7}$$

Here, n is the total number of samples. For a perfectly calibrated model, $ECE = 0$.

3 METHODOLOGY

The aim of this paper is to examine the calibration of YOLO-World and to analyze the influence of the calibration method *temperature scaling* while maintaining the accuracy and performance of this model (mAP).

The YOLO-World model used in this study is the *YOLO-Worldv2-L* model (size: 1280), which was pretrained using the *Objects365* [20] and *GoldG* [16] datasets (see [17]). A maximum of 300 detections are considered per image. For the evaluation, the LVIS minival [16] dataset (a subset of LVIS [21]) with 5000 images and 1203 object classes is used for all investigations. Furthermore, the localization of the objects is done using bounding boxes.

For the implementation, the ECE and the reliability diagram of the non-calibrated YOLO-World are calculated first. The number of bins is $M = 10$ for all the investigations carried out in this work. For both the ECE and the reliability diagram, the true positives (TP) and false positives (FP) must be determined. The basic process is shown in Figure 1 and consists of four steps. In the first step, all ground truths (GT) of an image are compared with each prediction, consisting of a label, a confidence score and the bounding box. Initially, all matches are considered that the correct label predicted and where the calculated IoU (Intersection over Union) value between GT and prediction is above an IoU threshold. After that, they are sorted in descending order based on their score (second step). The matches with the highest score are selected first and evaluated as TP, whereby each prediction may be assigned to a maximum of one ground truth and each ground truth to a maximum of one prediction. In this way, the matches with the highest scores are preferred and those with a lower score are evaluated later. This is the third step. The remaining matches and predictions are categorized as false positives (step four). As a result, the true positives with the best scores and the false positives can be determined, which are essential for calculating the ECE (see equation 7). In this work, the two thresholds $IoU\text{-}Threshold \in \{0.5, 0.75\}$ are considered, which are common threshold values for determining the mAP (see [6]).

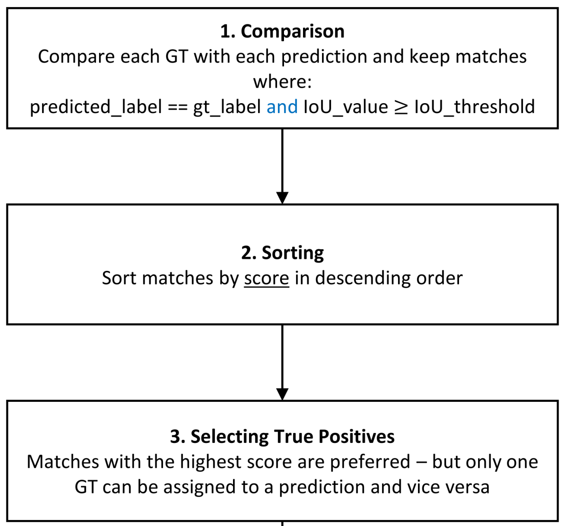


Figure 1: Process for determining the True Positives and False Positives

Subsequently, two calibration data sets, Pascal VOC [13] and Open Images V7 [14, 15], are used to determine the optimal temperature value. Of the Pascal VOC data set, only the validation data set is used, which includes 20 object classes and 5823 images. Since the Open Images validation dataset contains over 40,000 images, only a subset (8000 images) is used. These can be extracted in a deterministic way using this Python code, whereby the total number of object classes (601 in total) is not changed:

```
import fiftyone as fo
fo.zoo.load_zoo_dataset(
    "open-images-v7",
    split="validation",
    label_types=["detections"],
    max_samples=8000)
```

In accordance with [12], it is determined that T can take on the values $T \in \{0.5 + 0.05 \cdot n \mid n \in \mathbb{N}, 0 \leq n \leq 30\}$. The distance of 0.05 was chosen for reasons of time and resources. For each of these T-values, the ECE of the two calibration data sets can then be determined and displayed graphically, with the scaling according to equation 3. The minima (minimum T-values) obtained in this way are then evaluated by the LVIS minival data set. This makes it possible to see whether the found minima also reduce the ECE of the LVIS minival evaluation data set and whether there is a correlation between the minima of the data sets. By comparing the ECE values and the reliability diagrams before and after scaling with T, the effectiveness of the *temperature scaling* procedure with regard to the calibration of the YOLO-World model can be derived.

4 RESULTS

4.1 Before Applying Temperature Scaling

As described in Section 3, the calibration of the YOLO-World model is first examined using the LVIS minival evaluation dataset. The results in Table 1 show the values of the ECE and the mAP for different IoU thresholds before *temperature scaling*. The ECE has a size of 6.78% and 7.09% at IoU thresholds of 0.5 and 0.75, respectively.

Table 1: ECE and mAP for various IoU thresholds before applying temperature scaling.

Metric	Result [%]
mAP@[IoU=0.5:0.95]	34.6
mAP@[IoU=0.5]	45.5
mAP@[IoU=0.75]	37.8
ECE@[IoU=0.5]	6.78
ECE@[IoU=0.75]	7.09

Looking at the reliability diagram and the confidence diagram (see Figure 2), it can be seen that about 75% of the predictions lie in the confidence range of $(0;0.1]$. The majority of the predictions therefore have only a very low score (confidence value). Furthermore, there is a deviation from the diagonal in all bins. The accuracy ($acc(B_m)$), shown in blue, differs significantly from the calculated confidence ($conf(B_m)$), which can be seen from the red gap. This indicates that the model tends to be overconfident in its predictions in all bins.

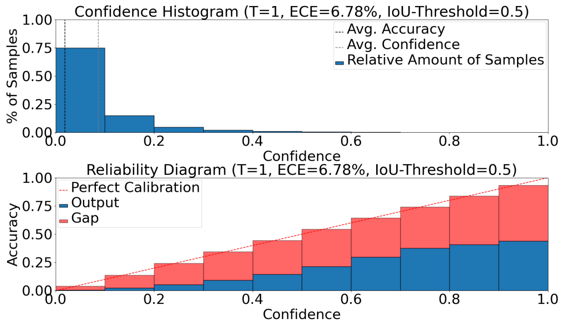


Figure 2: Confidence histogram and reliability diagram for an IoU threshold of 0.5 and $T = 1$.

4.2 After Applying Temperature Scaling

To determine the optimal T-value, the *temperature scaling* methode is applied to the calibration data sets as described in section 3. At this point, it should be mentioned that calibration methods such as temperature scaling do not model either the data or the model uncertainty of a model. [10] The results of this investigation are illustrated in Figure 3. As it can be seen, there are a total of 4 optimal T-values. The temperature values of the two IoU thresholds (0.5 and 0.75, respectively) of the Open Images calibration dataset are relatively close to each other ($T = 0.65$ and $T = 0.6$, respectively), while there are larger deviations for the Pascal VOC dataset ($T = 0.85$ and $T = 1.2$, respectively).

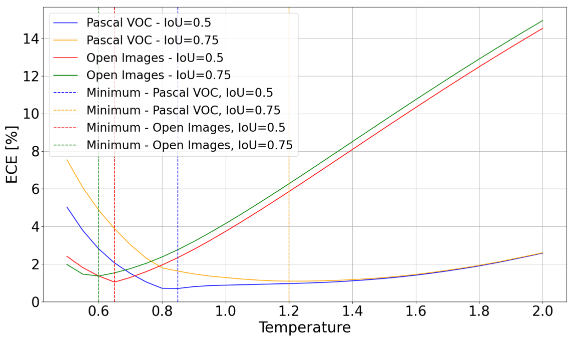


Figure 3: A plot of the ECE as a function of T for the two calibration data sets, Pascal VOC and Open Images, to determine the minima for the IoU thresholds 0.5 and 0.75.

The T-scores obtained in this way are then applied to the LVIS minival evaluation dataset. The accuracy and performance of the model are retained after application, i.e. the mAP remains unchanged for all T-values and IoU thresholds, as listed in Table 1. However, Table 2 shows that the ECE is significantly reduced, which assumes the most minimal value of $ECE = 2.31\%$ and $ECE = 2.66\%$ at $T = 0.6$, respectively. In general, the temperature values determined by Open Images result in the lowest ECE. By using the Pascal VOC, a T-value ($T = 0.85$) could be determined, which also reduces the ECE, but also a T-value ($T = 1.2$), which leads to an increase in ECE. From this, it can be deduced that the Open Images dataset is best suited for calibration using temperature scaling under the given experimental conditions. One possible reason could be the large number of 601 object classes, whereas Pascal VOC comprises just 20, which are more or less covered by the pre-training dataset *Objects365* of the YOLO-World model. A temperature value of $T = 1$ represents the results without applying the calibration method. Thus, temperature scaling could reduce the calibration error ECE by 4.47% and 4.43% for an IoU threshold of 0.5 and 0.75, respectively, in the best case ($T = 0.6$).

Table 2: Determined ECE values for various IoU thresholds after applying temperature scaling.

T value	IoU threshold	ECE [%]
0.6	**0.5**	**2.31**
0.6	**0.75**	**2.66**
0.65	0.5	2.70
0.65	0.75	3.03
0.85	0.5	4.84
0.85	0.75	5.16
1	0.5	6.78
1	0.75	7.09
1.2	0.5	9.45
1.2	0.75	9.76

However, looking at the confidence or reliability diagram in Figure 4 again after calibration, the first thing to note is that now about 90% of the samples lie in the confidence interval of $(0; 0.1]$. The two diagrams are shown here for $T = 0.6$ and an IoU threshold of 0.5. After applying temperature scaling, a significant improvement in calibration within bin B_1 can be seen, as well as a slight improvement in bins B_2 to B_4. However, the model still tends to be overconfident in its predictions, especially in bins B_3 to B_{10}, despite a reduction in ECE. It should be noted that the ECE is calculated from a weighted sum (see equation 7). Since about 90% of the samples are in the first bin, this also has the greatest influence on the calculation of the ECE. In summary, it can be deduced from this study that the proposed method could reduce the ECE, but without reducing the overconfidence of the model in all bins.

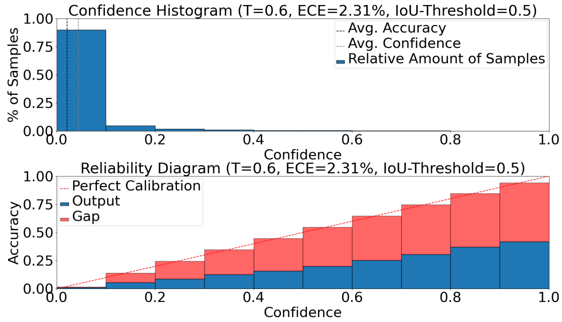


Figure 4: Confidence histogram and reliability diagram for an IoU threshold of 0.5 and $T = 0.6$.

5 CONCLUSION

In this paper, the influence of the calibration method *temperature scaling* on the OVD model YOLO-World was investigated. The optimal T-value could be determined using two calibration data sets, with the Open Images data set proving to be more suitable. It was shown that the model without the application of temperature scaling has a calibration error (ECE) of 6.78% and 7.0% for an IoU threshold of 0.5 and 0.75, respectively, and is overconfident in its predictions. After application, the calibration error ECE is reduced to a value of 2.31% (or 2.66%), while maintaining the accuracy and performance of the model. However, the overconfidence of the model could only be reduced in a few bins in this way. Additional or different calibration approaches are necessary to further improve the calibration of the model and thus minimize overconfidence in the predictions. An accurate modeling of the model uncertainty could also improve the calibration.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of the Research, Transfer and Start-Up Center (Forschungs-, Transfer- und Gründerzentrum), the Anhalt University of Applied Sciences and the state of Saxony-Anhalt for our research.

REFERENCES

[1] S. Twieg and R. Menghani, "Analysis and implementation of an efficient traffic sign recognition based on YOLO and SIFT for TurtleBot3 robot," 2023, [Online]. Available: http://dx.doi.org/10.25673/112993.

[2] 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), 2016, pp. 779-788.

[3] 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.

[4] C. Zhu and L. Chen, "A survey on open-vocabulary detection and segmentation: Past, present, and future," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 46, no. 12, pp. 8954-8975, 2024.

[5] A. Radford et al., "Learning transferable visual models from natural language supervision," 2021, [Online]. Available: https://arxiv.org/abs/2103.00020.

[6] T. Cheng et al., "YOLO-World: Real-time open-vocabulary object detection," in 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024, pp. 16901-16911.

[7] L. Yao et al., "DetCLIP: Dictionary-enriched visual-concept paralleled pretraining for open-world detection," in Advances in Neural Information Processing Systems, vol. 35, 2022, pp. 9125-9138, [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2022/file/3ba960559212691be13fa81d9e5e0047-Paper-Conference.pdf.

[8] S. Liu et al., "Grounding DINO: Marrying DINO with grounded pre-training for open-set object detection," 2024, [Online]. Available: https://arxiv.org/abs/2303.05499.

[9] F. Mumuni and A. Mumuni, "Segment anything model for automated image data annotation: Empirical studies using text prompts from Grounding DINO," 2024, [Online]. Available: https://arxiv.org/abs/2406.19057.

[10] J. Gawlikowski et al., "A survey of uncertainty in deep neural networks," 2022, [Online]. Available: https://arxiv.org/abs/2107.03342.

[11] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," 2017, [Online]. Available: https://arxiv.org/abs/1706.04599.

[12] W. LeVine et al., "Enabling calibration in the zero-shot inference of large vision-language models," 2023, [Online]. Available: https://arxiv.org/abs/2303.12748.

[13] M. Everingham et al., "The PASCAL Visual Object Classes challenge: A retrospective," International Journal of Computer Vision, vol. 111, no. 1, pp. 98-136, 2015.

[14] A. Kuznetsova et al., "The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale," arXiv:1811.00982, 2018.

[15] I. Krasin et al., "OpenImages: A public dataset for large-scale multi-label and multi-class image classification," 2017, [Online]. Available: https://storage.googleapis.com/openimages/web/index.html. [Accessed: 2-Jan-2025].

[16] A. Kamath et al., "MDETR - Modulated detection for end-to-end multi-modal understanding," in 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 1760-1770.

[17] Tencent AI Lab, "YOLO-World GitHub Repository," [Online]. Available: https://github.com/AILab-CVC/YOLO-World. [Accessed: 2-Jan-2025].

[18] T. Hirsch and B. Hofer, "The map metric in information retrieval fault localization," in 2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE), 2023, pp. 1480-1491.

[19] P. Zhang and W. Su, "Statistical inference on recall, precision and average precision under random selection," in 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery, 2012, pp. 1348-1352.

[20] S. Shao et al., "Objects365: A large-scale, high-quality dataset for object detection," in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 8429-8438.

[21] A. Gupta, P. Dollar, and R. Girshick, "LVIS: A dataset for large vocabulary instance segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.

VSC Adaptive Nonuniform Sampling Approach for Resource Optimization in Networked Controlled System

Andrej Sarjaš and Dušan Gleich
University of Maribor, Faculty of Electrical Engineering and Computer Science, Slomškov Str. 15, 2000 Maribor, Slovenia
andrej.sarjas@um.si, dusan.glecih@um.si

Keywords: Network Control System, Variable Control System, Sliding Mode Control, Nonuniform Sampling, Adaptive Event-Triggering.

Abstract: This paper presents an Adaptive Event-Triggered nonlinear Control method suitable for Networked Control Systems (NCS) with optimized resource exploitation. The work aims to ensure Networked Control Systems performance with lowered network usage and adaptive nonuniform controller execution. The nonlinear control strategy utilizes a predefined sliding variable, defined by system states and a nonlinear switching function, to maintain system stability within a specified boundary. This stability boundary is governed by an adaptive triggering condition, which balances system performance against network data-rate constraints and resource optimization. The adaptive triggering condition dynamically determines triggering law based on the state's ℓ^2 metrics and auxiliary internal dynamics system. Internal dynamic systems guarantee the stability property of the NCS and leverage resource optimization for nonuniform control algorithm execution. The minimum inter-event time for a nonuniform approach is derived to address network limitations and controller computation burden. The effectiveness of the proposed NCS method is validated through experimental results on a real system with UDP communication protocol.

1 INTRODUCTION

Sampled data systems have been a subject of research for many decades. When sampling and controller update occurs periodically, a well-established theory exists for analyzing stability and designing control strategies for such systems [1]-[4]. This mature theoretical foundation and stability analysis of the time-delayed system has significantly influenced the development of Networked Control Systems (NCS) [5]-[7]. However, in practical applications, the intervals between consecutive sampling instants in an NCS are typically time-varying rather than fixed. For instance, when an NCS experiences packet dropouts or denial-of-service (DoS) attacks, it can be modeled as a sampled data system with nonuniform sampling, also referred to as aperiodic or variable sampling. To efficiently manage limited network resources, the update frequency of sensors or controller units should be deliberately reduced. Many NCS components such as sensors, actuators, and embedded systems are battery-powered, and periodic activity of signal sampling with minimal variance from the last acquisition and controller updates can lead to unnecessary energy consumption and network resource usage. In such cases, nonuniform controller paradigms emerge as a promising strategy to optimize resource usage by activating sampling only when significant changes occur. The nonuniform sampling controller design for NCS is a viable solution for reliable operation. In the last two decades, many scholars have dealt with the nonuniform sampling approach for linear systems, where the stability is analyzed based on the time-delayed system [8],[9]. The analysis estimates the NCS's robustness and optimal behavior in network imperfection, where the controller execution remains under time triggering policy. The viable of a nonuniform sampling system is an Event-triggering approach (ET). ET is prevalently employed to act when the controller update policy is fulfilled [10]. The ET does not sample the system at uniform time intervals but instead executes actions based on a predefined triggering rule [11],[12]. While constant periodicity is disregarded, it allows for computational relaxation. However, it must still ensure the closed-loop properties of stability, state convergence, and time performance.

The presented research deals with designing adaptive Variable Structure Control (VSC) in the form of ET paradigms. A well-known approach of VSC is Sliding Mode Control (SMC), which ensures robustness and fast response by forcing system trajectories onto a predefined specified surface [13]. The main future of an SMC is a discontinuous control action concerning the preselected sliding variable, where the control alternates between different structures to maintain the system's trajectory on the sliding surface. Fast switching behavior makes sliding mode control well-suited for complex and uncertain systems [14]. While SMC excels at managing uncertainties and disturbances, its main drawback is the nonlinear output, characterized by high-frequency oscillations around the sliding manifold [15]. These oscillations are undesirable as they can lead to increased wear on the physical system, unwanted vibrations, excessive heat dissipation, and, in NCS, excessive network usage. SMC with ET approach relaxes latter conditions and improves controller performance. The primary benefit of this approach is incorporating an adaptive triggering rule that leverages the characteristics of relative triggering policies [16]. Adaptation of the triggering rule is based on the predefined desired value, where the stability of NCS-SMC systems is ensured with absolute convergence to the sliding surface with an adaptive triggering rule.

In the SMC mode, when the state trajectory is far from the sliding surface, the switching function remains unchanged until the trajectory crosses the sliding surface. In such a case, theoretically, no update is needed. When the trajectory is in the vicinity of the sliding variable, the practical sliding mode is introduced to ensure system stability [17],[18]. The practical sliding mode ensures the absolute boundness of the state trajectory, where the bandwidth is proportional to the triggering condition. With an adaptive triggering rule, the execution and network usage can efficiently be alleviated. The efficiency of the proposed approach is examined in the real NCS with the positioning system with a servo drive.

The structure of the paper is as follows: Section 2 presents the problem formulation and state transformation with error variables introduced into the system. The SMC design with an adaptive triggering approach is presented in Section 3. Section 4 introduces the ET approach for two previously designed SMC controllers. Two different triggering rules are suggested and derived lower nonzero TI values are presented. Section 5 presents the results and comparisons of the TT and ET strategies. Section 6. is the conclusion of the paper.

2 PRELIMINARIES

The single-input, single-output nonlinear system is given in the following class,

$$\begin{aligned} \dot{x} &= f(x)x + g_1(x)u + g_2(x)d \\ y &= h(x)x \end{aligned} \tag{1}$$

The functions $f(x)$, $g_1(x)$, $g_2(x)$, $h(x)$ are Lipschitz with respect to its arguments. $x(t)=\left[x_1(t)\ x_2(t)\right]^T \in {}^2$ is a state vector and $u(t)\in$ is the input variable. Due to the nonlinearity of (1), the parameters $f(x)\colon {}^2 \to$ and $g_1(x), g_2(x)\colon {}^2 \to$ depend on the operation point of (1). The matched disturbance is defined as $d\colon \ \to$. All the parameters are assumed to be bounded such as, $|f(x)| \le A$, $B_{i,\min} \le |g_i(x)| \le B_{i,\max}$, $i=1,2$, and $|d| \le D$, where $A, B_{i,\min}, B_{i,\max}$ and D are known positive constants. The state transformation is made by introducing new error variables,

$$e = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} x_d - x_1 \\ \dot{x}_d - x_2 \end{bmatrix}, \tag{2}$$

where x_d and $\dot{x}_d$ are the desired value and its time derivative, respectively and holds $e=\left[e_1\ e_2\right]^T \in {}^2$. The new transformed system is,

$$\begin{aligned} \dot{e} &= f(x)e - g_1(x)u + \tilde{d} + m(x)x_d' \\ y &= h(x)e \end{aligned}, \tag{3}$$

where $x_d' = \left[\dot{x}_d\ \ddot{x}_d\right]^T \in {}^2$ is the time derivative of the desired value x_d. The disturbance $\tilde{d} = -g_2(x)d$ is assumed to be bounded $|\tilde{d}| \le \Delta_d$. The boundary $\varDelta_d$ is a positive value and holds $D < \varDelta_d$.

3 VSC CONTROLLER DESIGN WITH ADAPTIVE TRIGGERING MECHANISM

The VSC controller under SMC paradigms is designed for the given system in (1). The reaching phase of the SMC is developed regarding the preselected sliding function given as,

$$S \quad \{s \in \ ^2 : s = ce = 0\}, \tag{4}$$

where is $c = [c_1 \ 1]$. The derivative of the function (4) with respect to time is,

$$\begin{aligned} \dot{s} &= c_1\dot{e}_1 + \dot{e}_2, \\ \dot{s} &= c\dot{e}, \\ \dot{s} &= -f(x)e_2 - g_1(x)u + c_1e_2(t) + \tilde{d}(t) + \ddot{x}_d + f(x)\dot{x}_d, \end{aligned} \tag{5}$$

The SMC controller is designed to lead the variable s to (4), with respect to disturbance and uncertainty boundary Δ_d.

$$\begin{aligned} u &= g_1(x)^{-1}\begin{pmatrix} k_1e_1 + (c_1 - a + k_2)e_2 + \rho\operatorname{sgn}(s) \\ +\ddot{x}_d + a\dot{x}_d \end{pmatrix}, \\ \rho &\geq \Delta_d \end{aligned} \tag{6}$$

where $k = [k_1 \ k_2], k \in \ ^2$ are linear feedback gains and holds $k > 0$. The gain values will be discussed in relation to ET. The controller u for the system in (3) is designed with an additional feedback loop in term k_1e_1, which does not directly influence the system's stability but has a significant effect on the tracking capabilities of the NCS in ET mode. For the static feedback in (6), the nominal values of the functions are used $f(\tilde{x}) = a,\ g_i(\tilde{x}) = b_i,\ i = 1,2$.

3.1 Event-Triggered VSC

In the latter section, the VSC-SMC controller was derived for the continuous system, as given in (3). The VSC controller will be implemented in discrete form, where the Backward-Euler and Backward-Differencing methods are mainly used in real-time applications. For the controller (6), instead of classic fixed update time T_t, the ET approach introduces interevent time T_i, which is defined by triggering rules and system dynamics. The T_i is defined with two successive updates defined as $T_i = \{t_{i+1} - t_i\}_{i=0}^{\infty}$. When the controller is updated at t_i, the last output value $u(t_i)$ is held until the new update is required t_{i+1} and holds for all $t \in [t_i, t_{i+1})$. The induced error between the last update and the current value due to the discrete implementation is defined as $\xi(t) = e(t) - e(t_i)$, $\xi(t) = [\xi_1(t) \ \xi_2(t)]^T$. At the time of update, the error is $\xi(t) = 0$, where $\xi(t) = e(t_i) - e(t_i) = 0$ holds. The error variable is crucial by determining the triggering condition of ET. The error variable is crucial by determining the triggering condition of ET. The ideal sliding mode is possible only in theory, where the manifold $s = 0$ is ensured with continuous operation of the $sgn(s)$ function. In practice, this cannot be achieved due to the discrete operation of the SMC controller. The variable remains bounded, depending on the selected sampling time T_s. It is similar; in ET, the system trajectory remains bounded by the preselected triggering conditions, which define a practical sliding mode [19]. The practical sliding mode occurs if a finite time $t_1 \in [t_i, \infty)$ exists for any given constant μ when the sliding variable s reaches the vicinity of $s = 0$ and remains there for all time $t > t_1$. The sliding variable is bounded with $|s| \leq \Omega,\ \Omega \in \ _+ / \ _\infty$. The triggering rule and nonzero, minimum positive inter-event time can be defined throughout the variable evolution $\xi(t)$ between two successive updates. The controller in (6) with nominal parameters at the given update time $t = t_i$ is given as,

$$u(t_i) = b_1^{-1}\begin{pmatrix} k_1e_1(t_i) + (c_1 - a + k_2)e_2(t_i) + \rho\operatorname{sgn}(s(t_i)) \\ +\ddot{x}_d(t_i) + a\dot{x}_d(t_i) \end{pmatrix}. \tag{7}$$

Theorem 1: *Consider system (3) with the sliding manifold (4) and controller (6). The parameter γ is given such that,*

$$\left\| \left[(c_1 - a) + k_2 \quad k_2c_1 \right] \right\| \left\| \xi(t) \right\| < \gamma, \tag{8}$$

for all $t > 0$, where $\gamma \in$, $\gamma > 0$, and $k_2 > 0$, $k_1 = c_1k_2$. The event triggering is established if the controller gain is selected as,

$$\rho > \gamma + \Delta_d, \tag{9}$$

Proof: The stability analysis is performed with the Laypunov function $V(t) = \frac{1}{2}s(t)^2$. Substituting (5) to

$$\dot{V} = s\dot{s} = s\left((c_1 - a)e_2 - b_1u + \tilde{d} + \ddot{x}_d + a\dot{x}_d\right), \tag{10}$$

and respect to the time derivative (10) gives,

$$\begin{aligned} \dot{V} &= s(t)\left((c_1 - a)e_2(t) - g_1u(t_i) + \tilde{d}(t) + \ddot{x}_d(t) + a\dot{x}_d(t)\right), \\ &= s(t)\begin{pmatrix} (c_1 - a)e_2(t) - b_1\left(b_1^{-1}\begin{pmatrix} k_1e_1(t_i) + (c_1 - a + k_2)e_2(t_i) \\ +\rho\operatorname{sgn}(s(t_i)) + \ddot{x}_d(t_i) + a\dot{x}_d(t_i) \end{pmatrix}\right) \\ +\tilde{d}(t) + \ddot{x}_d(t) + a\dot{x}_d(t) \end{pmatrix}, \end{aligned}$$

$$
\begin{aligned}
&= s(t)\left((c_1-a)\left(e_2(t)-e_2(t_i)\right)-k_1e_1(t_i)-k_2e_2(t_i)-\rho\,\mathrm{sgn}\left(s(t_i)\right)+\tilde{d}(t)\right),\\
&= s(t)\begin{pmatrix}(c_1-a)\xi_2(t)-\begin{pmatrix}e_2(t_i)+c_1e_1(t_i)+(k_2-1)\\ \cdot\left(e_2(t_i)+\dfrac{k_1-c_1}{k_2-1}e_1(t_i)\right)\end{pmatrix}\\ -\rho\,\mathrm{sgn}\left(s(t_i)\right)+\tilde{d}(t)\end{pmatrix},\quad k_1=c_1k_2,\\
&= s(t)\begin{pmatrix}(c_1-a)\xi_2(t)-\begin{pmatrix}e_2(t_i)+c_1e_1(t_i)+(k_2-1)\\ \cdot\left(e_2(t_i)+c_1e_1(t_i)\right)\end{pmatrix}\\ -\rho\,\mathrm{sgn}\left(s(t_i)\right)+\tilde{d}(t)\end{pmatrix},\\
&= s(t)\left((c_1-a)\xi_2(t)-k_2\left(e_2(t_i)+c_1e_1(t_i)\right)-\rho\,\mathrm{sgn}\left(s(t_i)\right)+\tilde{d}(t)\right),
\end{aligned}
$$

With the introduced sliding variable error $\xi_s(t)=s(t)-s(t_i)$, the term $e_2(t_i)+c_1e_1(t_i)=s(t_i)$ gives $e_2(t_i)+c_1e_1(t_i)=s(t)-\xi_s(t)$. Substituting the relation into a stability analysis gives,

$$
\begin{aligned}
\dot{V} &= s(t)\begin{pmatrix}(c_1-a)\xi_2(t)-k_2\left(s(t)-\xi_s(t)\right)\\ -\rho\,\mathrm{sgn}\left(s(t_i)\right)+\tilde{d}(t)\end{pmatrix},\\
&= s(t)\begin{pmatrix}(c_1-a)\xi_2(t)-k_2s(t)+k_2\xi_s(t)\\ -\rho\,\mathrm{sgn}\left(s(t_i)\right)+\tilde{d}(t)\end{pmatrix},\\
&= s(t)\begin{pmatrix}(c_1-a)\xi_2(t)-k_2s(t)+k_2\left(\xi_2(t)+c_1\xi_1(t)\right)\\ -\rho\,\mathrm{sgn}\left(s(t_i)\right)+\tilde{d}(t)\end{pmatrix},\\
&= s(t)\begin{pmatrix}(c_1-a+k_2)\xi_2(t)+k_2c_1\xi_1(t)-k_2s(t)\\ -\rho\,\mathrm{sgn}\left(s(t_i)\right)+\tilde{d}(t)\end{pmatrix},\\
&\le |s|\gamma-k_2s^2-|s|\rho+|s|\Delta_d,\\
&\le -|s|\left(\rho-\gamma-\Delta_d\right)-k_2s^2<0,
\end{aligned}
$$

where $\rho>\gamma+\Delta_d$ stability s is guaranteed for a long time $t\in\left[t_i,t_{i+1}\right)$. The stability condition $\dot{V}<0$ is ensured if $\mathrm{sgn}\left(s(t_i)\right)=\mathrm{sgn}\left(s(t)\right)$, otherwise boundary is defined as,

$$
\begin{aligned}
&\left|s(t)-s(t_i)\right|=\left|ce(t)-ce(t_i)\right|,\\
&\qquad<\|c\|\left\|\left[(c_1-a)+k_2 \quad k_2c_1\right]\right\|^{-1}\gamma=\tilde{k}_l\gamma,
\end{aligned}
\tag{11}
$$

where $\Omega_l=\left\{e\in\ ,|s|=|ce|<\tilde{k}_l\gamma\right\}$ and the triggering rule (8) is defined as $\|\xi(t)\|<\left\|\left[(c_1-a)+k_2 \quad k_2c_1\right]\right\|^{-1}\gamma$. The stability in the region Ω_l is defined by the Lyapunov function $V(t)=\frac{1}{2}e_1(t)^2$, where holds $e_2=s-c_1e_1$. The trajectory is confined to the region $\|e_1\|<\|c_1\|^{-1}\tilde{k}_l\gamma$.

3.2 Adaptive ET approach

The stability analysis and the static triggering law are defined in (8). The adaptation algorithm for parameter γ can serve as an internal mechanism to adjust parameters γ in specific regions defined as $\gamma\in\left[\underline{\gamma},\bar{\gamma}\right]$, where holds $\gamma>0$, and $\underline{\gamma}<\bar{\gamma}$. Adaptive triggering is introduced with triggering law,

$$
\|e\|\ge\eta+\gamma,\qquad \gamma>0,\eta>0,
\tag{12}
$$

where η is an internal parameter and can be defined as,

$$
\dot{\eta}=-\alpha\eta+\Theta\left(\gamma-\|e\|\right),
\tag{13}
$$

and holds $\eta(0)=\eta_0,\ \eta,\eta_0,\Theta,\alpha\in {}_0^+$. System in (13) is a strictly positive function. The adaptation algorithm acts as a mechanic that converges to zero if the static law γ is employed. Take, for example, the condition when triggering has not occurred.

$$
\|e\|<\eta+\beta,
\tag{14}
$$

and holds $-\eta\le\beta-\|e\|$. If we rewite the system given in (13), we get,

$$
\begin{aligned}
\dot{\eta}&=-\alpha\eta+\Theta\left(\beta-\|e\|\right)\\
&=-\alpha\eta-\Theta\eta\\
&=-\left(\alpha+\Theta\right)\eta
\end{aligned}
,
\tag{15}
$$

where it holds that the derivative of the (13) is a strict negative function. The stability analysis of ET with an adaptation system (13) and (10) is given with the Lyapunov function,

$$
V_A=V+\eta,
\tag{16}
$$

and the derivative of V_A is equal to,

$$
\begin{aligned}
\dot{V}_A&=\dot{V}+\dot{\eta}\\
\dot{V}_A&=-|s|\left(\rho-\gamma-\Delta_d\right)\\
&\quad-k_2s^2-\left(\alpha+\Theta\right)\eta<0
\end{aligned}
.
\tag{17}
$$

The system's stability with the adaptive system (13) is ensured.

4 NETWORKED CONTROL SYSTEM

The structure of the networked control system is illustrated in Figure 1. The controller algorithm runs on a networked computer, while the triggering rule (12) is evaluated on the plant. It is assumed that the plant is equipped with a real-time system that has computational capabilities and communication interfaces based on the LwIP stack. The real-time system handles simple tasks such as evaluating triggering conditions, signal conditioning, and managing communication. The User Datagram Protocol (UDP) is used for the given adaptive ET implementation for data transmission. Data packets are transmitted across multiple network hops, where delays and packet loss may occur. Packet loss in the network is modeled as a loss delay, with the maximum allowable Round Trip Time (RTT) serving as a threshold for packet loss detection. On the plant side, a dedicated packet-loss timer is employed. If the watchdog timer expires, the system requests new data from the server to maintain reliable communication. On the server side, the Python server is employed.

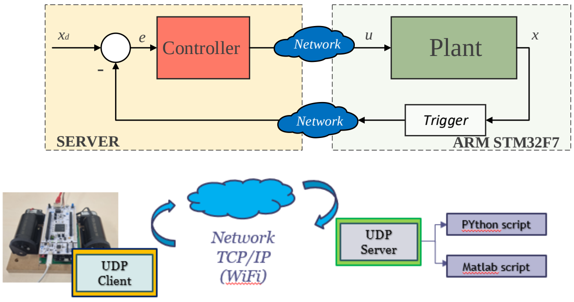


Figure 1: NCS configuration based on adaptive triggering approach.

Regarding the NCS operation and discrete implementation of the ET controller, ensuring that the inter-event time will not tend to the Zeno phenomena is necessary [20],[21]. The inter-event time has to be a lower positive bound, where the lower bound does not violate network performance. The estimated network performance with the RTT parameter is presented in Figure 2. The average RTT is $RTT \approx 2.7ms$, with maximal RTT deviance $\Delta_{RTT} = 0.212ms$.

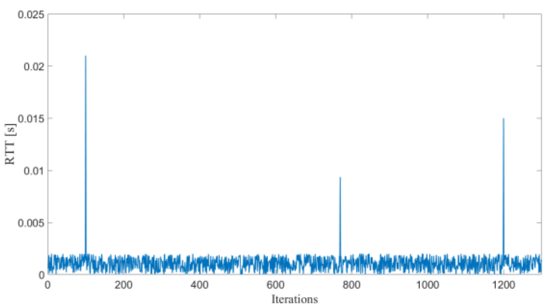


Figure 2: Network performance with RTT parameter.

The inter-event time T_i is determined based on the error analysis between two consecutive sampled states given as,

$$\frac{d}{dt}\|e(t)\| \le \left\|\frac{d}{dt}e(t)\right\| = \left\|\frac{d}{dt}\begin{bmatrix}\xi_1(t)-\xi_1(t_n)\\ \xi_2(t)-\xi_2(t_n)\end{bmatrix}\right\| = \left\|\frac{d}{dt}\begin{bmatrix}\xi_1(t)\\ \xi_2(t)\end{bmatrix}\right\|. \tag{18}$$

where is $\xi(t_n) = 0$, according to the last update, which follows,

$$\frac{d}{dt}\|\xi(t)\| \le \left\| \begin{matrix} \begin{bmatrix}0 & 1\\ 0 & -a\end{bmatrix}\begin{bmatrix}e_1(t)\\ e_2(t)\end{bmatrix} + \begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}\tilde{d}(t) - \\ \begin{bmatrix}0\\ 1\end{bmatrix}\begin{pmatrix} k_1 e_1(t_i) + (c_1 - a + k_2)e_2(t_i) \\ +\rho\,\mathrm{sgn}(s(t_i)) + \ddot{x}_d(t_i) + a\dot{x}_d(t_i)\end{pmatrix} \end{matrix} \right\|, \quad M_l = -\begin{bmatrix}k_1 & 0\\ 0 & (c_1 - a) + k_2\end{bmatrix}$$

$$\begin{aligned} &\le \|f\|\|e(t)\| + \|M_l\|\|e(t_i)\| + \rho + \|P\|x'_d \\ &= \|f\|\|\xi(t)\| + (\|f\| + \|M_l\|)\|e(t_i)\| + \rho + \|P\|\|x'_d\|. \end{aligned}$$

The solution of the differential equation is,

$$\|\xi(t)\| \le \|f\|^{-1}\begin{pmatrix}(\|f\| + \|M_l\|)\|e(t_i)\| + \\ \rho + \|P\|\|x'_d\|\end{pmatrix}\left(e^{\|f\|(T_l)} - 1\right).$$

Regarding (8) holds,

$$\left\|\begin{bmatrix}(c_1 - a) + k_2 & k_2 c_1\end{bmatrix}\right\|^{-1}\eta = \|f\|^{-1}\begin{pmatrix}(\|f\| + \|M_l\|)\|e(t_i)\| + \rho \\ +\|P\|\|x'_d\|\end{pmatrix}\left(e^{\|f\|T_l} - 1\right)$$

The inter-event time is equal to,

$$T_l \ge \|f\|^{-1}\ln\left(\frac{\|f\|\eta}{\left\|\begin{bmatrix}(c_1 - a) + k_2 & k_2 c_1\end{bmatrix}\right\|\begin{pmatrix}(\|f\| + \|M_l\|)\|e(t_i)\| \\ +\rho + \|P\|\|x'_d\|\end{pmatrix}} + 1\right). \tag{19}$$

From the derived condition (19), it is obvious that the parameter T_i is positively lower bound and depends on the adaptation system η, triggering condition γ, and controller parameters ρ, k_1, k_2. The selection of the parameters depends on possible network imperfection, DoS, and package drops. All the network uncertainty is modeled as a delayed system [5],[9],[21], where selection of the triggering law can be determined in the way to compensate such delays and preserve NCS performance [21].

5 EXPERIMENTAL RESULTS

For the experimental system, the positioning system with parameters $a = 0.32$, $b_{1,2} = 1.21$ and state variables x_1-angle $[\deg]$, x_2-velocity $[RPM]$ is presented in Figure 3. The real-time experiments were performed on the ARM® Cortex®-M4 based STM32F7xx MCU with Digital-Signal Processing and Floating-Point Unit (DSP and FPU), operating frequency of 180MHz and implemented *LwIP* stack. The motor was driven with an NXP-MC33926 H-controller and pulse-width modulation-PWM technique, with a carrier frequency of 10kHz. The preselected sampling time of the TT implementation is 1ms. System parameters are; $k = [8.56\ 3.35]$ $c_1 = 9.1$, $\rho = 22.5$, $\gamma = 0.15$, $\alpha = 2.93$, $\Theta = 0.89$ and $\eta_0 = |\max(x_d) - x_0| / 3$.

The NCS performance is evaluated by given performance indices,

$$RMS_n = \sqrt{\frac{1}{n_s}\sum_{k=1}^{n_s} n^2},\ n \in \{x_1, s, u\}, \tag{20}$$

where n_s is the number of evaluated samples. Efficiency is measured by comparing three types of VSC controllers. The first controller is implemented based on the classic time-triggering technique, and its states are marked as TT. The second version is implemented based on a fixed ET technique, and the system states are marked as ET. The third version is implemented with an adaptive ET approach and is marked as AET.

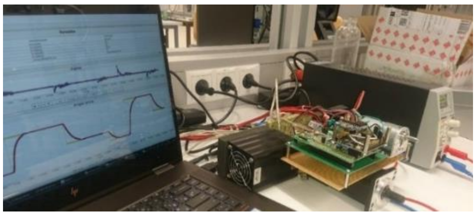


Figure 3: Real-time NCS system.

The results of adaptive ET in NCS structure are presented in the following Figure 4 – Figure 7.

Performance indices are presented in Table 1.

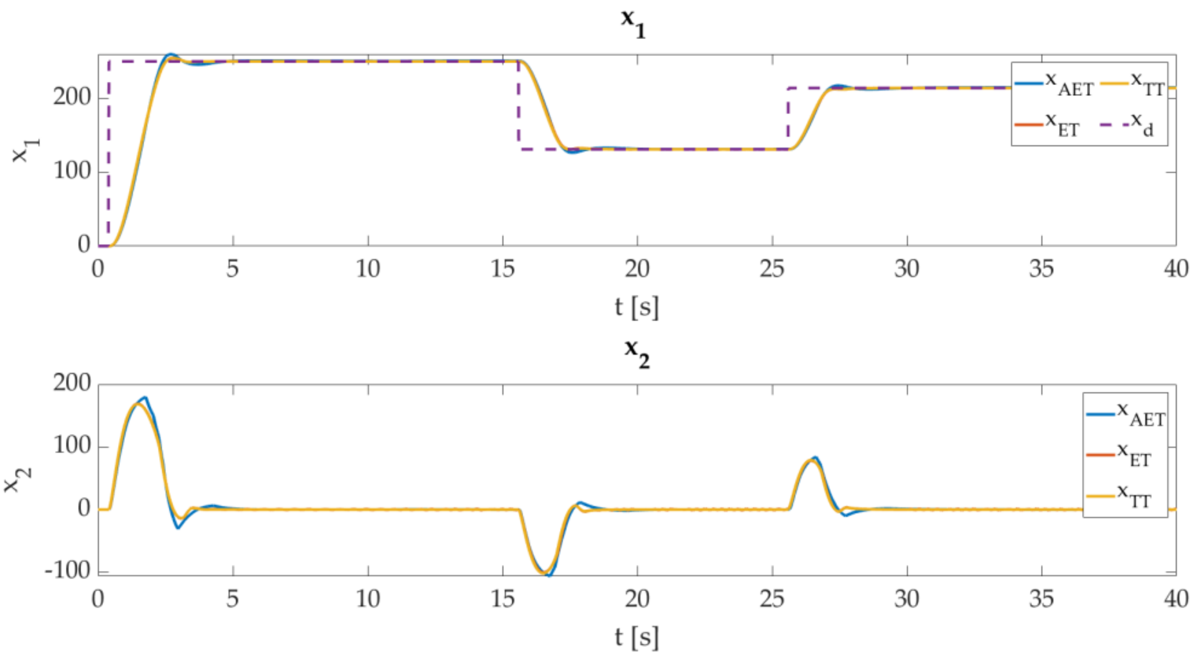


Figure 4: State variables; $x_1[\deg]$, $x_2[RPM]$, $x_d[\deg]$ position, velocity, and desired value, respectively.

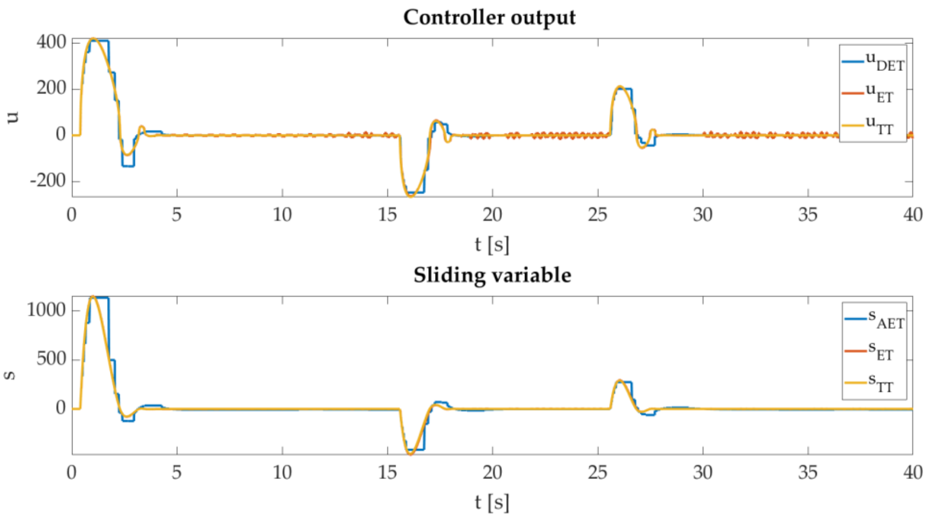


Figure 5: Controller output and sliding variable.

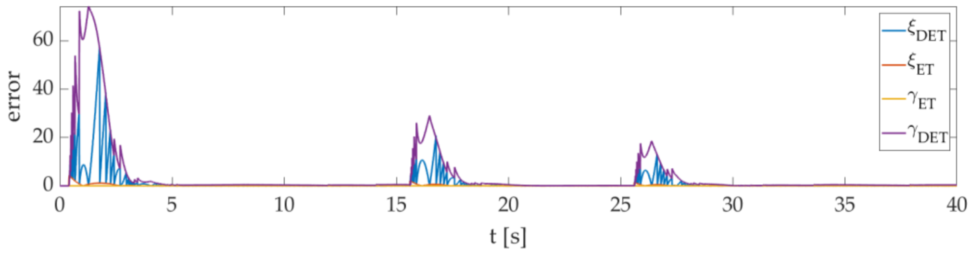


Figure 6: Triggering boundary γ and update error ξ.

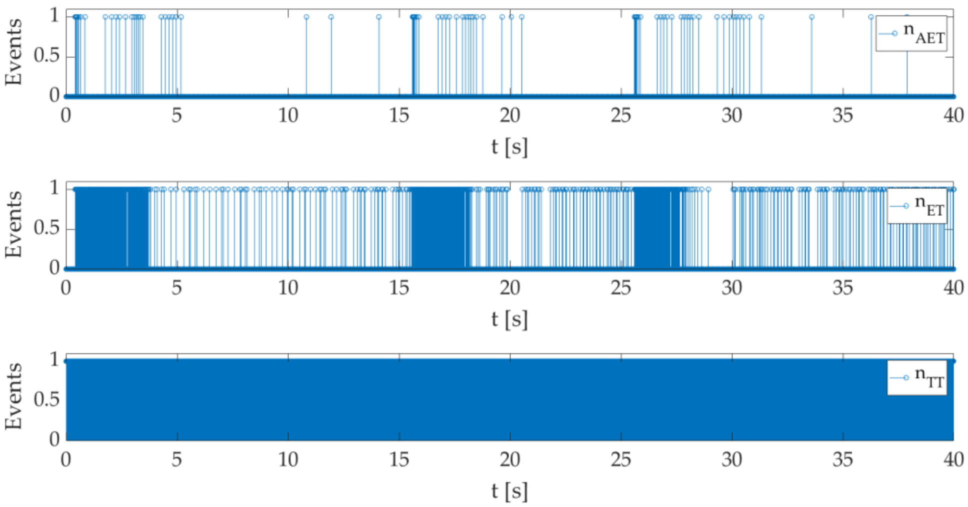


Figure 7: Controller update flags.

Table 1: Performance indices of different controller implementations.

Cont	RMSx	RMSu	RMSs	minT_i	maxT_i	meanT_i	Flag
TT	206.7	88.07	182.9	0.001	0.001	0.001	4e4
ET	206.9	88.23	184.5	0.001	0.84	0.09	993
DET	207.1	90.1	201.3	0.025	5.56	1.17	83

6 CONCLUSIONS

This paper introduces the implementation of an adaptive event-triggered nonlinear controller for networked control systems. The proposed approach offers a viable alternative to time-triggered execution and ET with a fixed triggering bound, making it particularly advantageous for NCS applications with data rate limitations. This method enhances usage efficiency in resource-constrained environments by incorporating network uncertainty directly into the controller design, with a proper selection of the controller parameters, triggering rule, and adaptation system. Furthermore, this work is a valuable foundation for research in different NCS configurations, such as multi-agent systems, distributed control, and task scheduling in embedded systems. The adaptation system can be managed in different scenarios based on the desired values' properties and the state trajectory's vicinity to the sliding manifold. Such an approach will be beneficial for the system with fast varying desired values.

REFERENCES

[1] D. Nesić and A. R. Teel, "Sampled-data control of nonlinear systems: An overview of recent results," in Perspectives in Robust Control, vol. 268, London: Springer, 2001.

[2] C. L. Phillips, H. T. Nagle, and A. Chakrabortty, Digital Control System Analysis and Design, Pearson, 2015.

[3] J. G. Proakis and D. G. Manolakis, Digital Signal Processing: Principles, Algorithms, and Applications, Prentice Hall, 2007.

[4] C. Fiter, L. H. Perruquetti, and W. J. P. Richard, "A state-dependent sampling for linear state feedback," Automatica, vol. 48, no. 8, pp. 1860-1867, 2012.

[5] F. Forni, S. Galeani, D. Nešić, and L. Zaccarian, "Lazy sensors for the scheduling of measurement samples transmission in linear closed loops over networks," in 49th IEEE Conference on Decision and Control (CDC), 2010, pp. 6469-6474.

[6] E. Fridman, "A refined input delay approach to sampled-data control," Automatica, vol. 46, no. 2, pp. 421-427, 2010.

[7] Z. Chen, Q.-L. Han, Y. Yan, and Z. Wu, "How often should one update control and estimation: a review of networked triggering techniques," Science China Information Sciences, vol. 63, 2020.

[8] H. Gao, T. Chen, and J. Lam, "A new delay system approach to network-based control," Automatica, vol. 44, no. 1, pp. 39-52, 2008.

[9] E. Fridman, "A refined input delay approach to sampled-data control," Automatica, vol. 46, pp. 421-427, 2012.

[10] B. Khashooei, D. Antunes, and W. Heemels, "Output-based event-triggered control with performance guarantees," IEEE Transactions on Automatic Control, vol. 62, no. 7, pp. 3646-3652, 2017.

[11] X. Ge, Q.-L. Han, X.-M. Zhang, L. Ding, and F. Yang, "Distributed Event-Triggered Estimation Over Sensor Networks: A Survey," IEEE Transactions on Cybernetics, vol. 50, pp. 1306-1320, 2020.

[12] H. Ning, D. Jiawei, X. Zhongxian, and C. Jiaxuan, "Event-triggered MPC for linear systems with bounded disturbances: An accumulated error based approach," IET Control Theory & Applications, vol. 16, no. 8, pp. 816-827, 2022.

[13] V. I. Utkin and A. S. Poznyak, "Adaptive sliding mode control with application to super-twist algorithm: Equivalent control method," Automatica, vol. 49, no. 1, pp. 39-47, 2013.

[14] J. A. Moreno and M. A. Osorio, "Lyapunov approach to second-order sliding mode controllers and observers," in 47th IEEE Conference on Decision and Control, Cancun, Mexico, 2008, pp. 2856-2861.

[15] Y. Shtessel, C. Edwards, L. Fridman, and A. Levant, Sliding Mode Control and Observation, Birkhäuser, 2014.

[16] P. Tabuada, "Event-triggered real-time scheduling of stabilizing control tasks," IEEE Transactions on Automatic Control, vol. 52, no. 9, pp. 1680-1685, 2007.

[17] A. Behera, K. Bandyopadhyay, and X. Yu, "Periodic event-triggered sliding mode," Automatica, vol. 96, pp. 1916-1931, 2018.

[18] M. Cucuzzella and A. Ferrara, "Practical second-order sliding modes in single-loop networked control of nonlinear systems," Automatica, vol. 89, pp. 235-240, 2018.

[19] A. Behera and K. Bandyopadhyay, "Self-triggering based sliding-mode control for linear systems," IET Control Theory & Applications, vol. 9, no. 17, pp. 2541-2547, 2015.

[20] M. R. Mojallizadeh and B. Brogliato, "Time-discretizations of differentiators: Design of implicit algorithms and comparative analysis," International Journal of Robust and Nonlinear Control, vol. 31, no. 16, pp. 7679-7723, 2021.

[21] V. Utkin, "Discussion aspects of high-order sliding mode control," IEEE Transactions on Automatic Control, vol. 61, pp. 829-833, 2016.

Methods and Algorithms for Decision-Making in Agro-Industrial Environmental Management

Olga Ivashchuk, Vyacheslav Fedorov, Zhanat Kenzhebayeva, Oleg Ivashchuk, Moldir Allaniyazova, Kaiyrbek Makulov and Gulnar Kaziyeva
Yessenov University, 32 District, 13000 Aktau, Kazakhstan
olga.ivashuk@mail.ru, fedorov_v@bsu.edu.ru, zhanat.kenzhebayeva@yu.edu.kz, ooleg.ivashuk@gmail.com, moldir.allaniyazova@yu.edu.kz, kaiyrbek.makulov@yu.edu.kz, gulnara.kaziyeva.@yu.edu.kz

Keywords: Comprehensive Assessment, Ecological Monitoring, Fuzzy Logic, IoT, GIS, Situational Modelling, Agro-Industrial Complex.

Abstract: This paper presents a novel methodology for comprehensive assessment and decision-making in managing the ecological state of agro-industrial territories. The study introduces an intelligent situational modeling approach that integrates fuzzy logic, GIS-based analysis, and IoT-driven environmental monitoring to evaluate both current and forecasted conditions. The proposed cyber-physical system utilizes real-time sensor networks and UAV-based hyperspectral imaging to collect, process, and analyze environmental parameters, including soil pollution index (SPI), air quality index (AQI), and vegetation health index (VHI). Field experiments conducted in the Belgorod Region, Russia, on 50-hectare test sites demonstrated a 27% improvement in forecasting accuracy compared to conventional methods. Key findings reveal that implementing optimized land-use scenarios resulted in: 19% reduction in pollutant accumulation in soil, 27% increase in agricultural productivity, 25% decrease in public health risks. The proposed framework facilitates adaptive management by providing science-based recommendations for establishing protective forest strips, reducing pollutant exposure, and optimizing land-use planning. The findings confirm the necessity of integrating intelligent environmental monitoring into territorial management systems to enhance sustainable agro-industrial development and mitigate ecological risks.

1 INTRODUCTION

The implementation of scientifically grounded solutions in the field of agricultural production greening and sustainable territorial development requires reliable information about the current and forecasted environmental situation. This necessitates a spatial-temporal analysis of environmental dynamics, considering both natural and anthropogenic factors [1].

For the agro-industrial complex (AIC), as an intersectoral system, the collection and processing of large volumes of heterogeneous data pose significant challenges. These data encompass the characteristics of technological processes involving living organisms, as well as a wide range of controlled parameters that are highly dispersed and subject to stochastic variations.

Currently, various national and international environmental monitoring systems have been developed and implemented, including territorial and industrial solutions. These systems rely on stationary and mobile units for automated data collection, transmission, processing, storage, and analysis. They provide decision-makers with access to environmental information while also informing the public through web-based resources and data visualization systems [2].

Among the most widely used environmental monitoring approaches in AIC are:

- Artificial Intelligence (AI) – applied in plant disease detection, water and soil management, weather forecasting, animal behavior monitoring, and crop optimization [3].
- Internet of Things (IoT) – used for tracking agricultural machinery and automating production cycles [4].

- Blockchain systems – ensure supply chain tracking, data security, and payment control [5].
- Big Data technologies – applied for analyzing and optimizing agricultural production [6].
- Remote sensing and unmanned aerial vehicles (UAVs) – employed for land mapping, soil analysis, and vegetation monitoring [7].
- High-precision and energy-efficient sensors – improve environmental state measurement accuracy while minimizing energy consumption [8].
- Advanced environmental monitoring systems – enhance ecological assessment accuracy through the use of intelligent technologies [9].

Despite the widespread adoption of these technologies, existing monitoring systems have several limitations:

- AI and Big Data-based methods require substantial computational resources and large datasets, which may not be feasible in rural areas with limited infrastructure.
- IoT technologies effectively collect data, but their integration with situational models remains a challenge.
- Blockchain solutions are useful for logistics and supply chains, but do not provide predictive environmental analysis.

The proposed approach combines fuzzy logic, situational modeling, and intelligent data analysis, enabling:

- the consideration of multiple heterogeneous parameters,
- the generation of alternative management scenarios,
- adaptation to changing conditions,
- and improved decision-making efficiency in environmental monitoring for AIC.

Thus, the developed methodology is designed to enhance forecasting accuracy, improve adaptability to external changes, and increase the efficiency of ecological state management in agro-industrial objects and processes.

2 SETTING THE TASK OF SCIENTIFIC RESEARCH

The authors propose models and approaches to the creation and organization of the functioning of unique cyberphysical systems for monitoring and managing the environmental condition of agricultural facilities and processes, providing the possibility of organizing agricultural production based on the principles of biospheric compatibility, presented in [2]. According to the requirements and principles of functioning of the proposed cyberphysical system, the task is to develop a method for a comprehensive adequate assessment of agricultural facilities and processes.

The key condition under study in this case is the environmental situation in the territory under consideration, which is located in the zone of influence of specific studied objects and/or agro-industrial complex processes. We introduce the designation of the corresponding linguistic variable – , which (taking into account and developing the approach proposed by D.A. Pospelov [10]) can be defined as a complex spatial and temporal assessment (carried out on the basis of analysis and generalization) of the totality of environmental characteristics of components of the natural environment and living objects of agricultural production, their relationships with the parameters of the agro-industrial complex and the external environment which form a certain level of environmental safety, as well as the level of impact on food security.

Based on the general principles of situational modeling [11], we introduce the concepts of: the current environmental situation (we will denote when implementing the modeling process as EL^{act}), which is determined at a given time in the territory under consideration, located in the zone of influence of objects and/or processes of the agro-industrial complex; the complete environmental situation (we will denote when implementing the modeling process as EL^{full}), Including: status EL^{act}, knowledge about the state of the studied objects and/or agro-industrial complex processes, the state of the territory's infrastructure, external influences at a given time; knowledge about management mechanisms and technologies, about cause-and-effect relationships that determine the conditions of dynamics and the possibility of optimizing their parameters. An elementary act of managing the environmental situation in the territory under consideration can be presented on the basis of a logical transformation rule:

$$EL_i^{full} : EL_i^{act} \xrightarrow[U_k]{} EL_j^{act} \quad (1)$$

where, U_k – k is the control effect on any parameter of the studied object and/or the agro-industrial complex process, infrastructure of the territory under consideration, which determines the spatial and temporal dynamics of the environmental situation.

When choosing a rational management scenario from the many alternatives provided, it is necessary to consider the process of forming an environmental situation based on an integrated approach. Thus, changes in the same technological parameters of business processes can affect to varying degrees the level of change in chemical and/or physical effects on individual components of the natural environment. The need to maintain an ecological and economic balance should also be taken into account.

The information received from sensors, instruments and measuring devices is expressed accurately, i.e. in specific numbers. However, there is no instrumentation that allows for an accurate integrated assessment of the environmental situation as a result of the impact of a set of objects and processes on various components of the natural environment, taking into account the participation of living objects. Such an assessment can only be carried out qualitatively and expressed using natural language and a linguistic variable. At the same time, such a qualitative assessment of the state of environmental safety using linguistic meanings is based on the knowledge of expert specialists in the subject area and does not require technical measuring devices.

So, the model *EL* should be synthesized on the basis of individual elements of knowledge extracted during an experimental or model assessment of specific quality indicators of components of the natural environment, objects of wildlife. The formation of management scenarios will be based precisely on the results of such a synthesized assessment.

2.1 The Method of Comprehensive Assessment of the Environmental Situation

Based on the set-theoretic description, the linguistic variable *EL* can be represented by $\{EL,T,Q,G,H\}$, where *EL* – the name of the variable in question entered above; *T* – this is the main term set for the basic values of the linguistic variable *EL; Q* – some numerical set on which fuzzy variables are defined; *G* – a set of syntactic rules for the formation of new meanings *EL*, not included in the main term set; *H*– the corresponding mathematical rules.

The minimum accuracy of the complex is determined by two main terms: T_1 = «favorable», T' = «unfavorable». Depending on the result of the cumulative state of each of the components of the natural environment selected for analysis and/or the state of living objects of agricultural production in the study area, a differentiation of a comprehensive assessment is carried out «unfavorable» (that is, its accuracy increases): $T = \{T_1, T'\} = \{T_1, T_2, T_3, T_4, T_5\}$, T_2 = «relatively unfavorable»; T_3 = «dangerous»; T_4 = «very dangerous»; T_5 = «critical».

To form an output *EL* when deploying a certain cyberphysical system in the study area, where specific objects/processes of the agro-industrial complex operate/are implemented, necessary to construct a set of N_{EL} logical rules. For numerical evaluation, it is recommended to use points based on updating the Sugeno fuzzy inference algorithm of the 0th order [11]. At the same time, the environmental situation is characterized by a conditional 5-point scale in accordance with the characteristics of the terms of their set *T*.

In this paper, it is proposed to use the following approach: a comprehensive assessment is based on the analysis of two main subsets *EL:*

$$EL = \{E, L\}, \tag{2}$$

where *E* – comprehensive assessment of the state of the components of the natural environment; *L* – comprehensive assessment of the state of components of living agricultural production facilities that affect the formation of environmental and food security.

In turn *E* and *L* they are also compound variables: $E = \{E_1, E_2, E_3, E_4\}$, where E_1 – comprehensive assessment of the state of the atmosphere, E_2 – comprehensive assessment of the state of water resources, E_3 – comprehensive assessment of the soil condition and E_4 – comprehensive assessment of the state of physical impact (noise, thermal, radiation, etc.). At the same time, we introduce a comprehensive assessment of each component according to the aggregate state of various parameters, for example, we assess the state of atmospheric air according to the cumulative effect of various pollutants. $L = \{L_1, L_2\}$, where L_1 – comprehensive assessment of the condition of plant objects, L_2 – comprehensive assessment of the condition of animal objects.

2.2 An Algorithm that Implements a Method for a Comprehensive Assessment of the Current and Forecast Environmental Condition of Agricultural Facilities and Processes

As part of the study, an algorithm was developed for the proposed method of comprehensive adequate assessment, schematically presented in Figure 1.

At the first step, the components of the natural environment are identified, which are subjected to intense negative effects during the functioning of the studied object / process of the agroindustrial complex. Ecomonitoring parameters are being determined. To do this, digital business models are being investigated, an example of which is shown in Figure 2 and Figure 3.

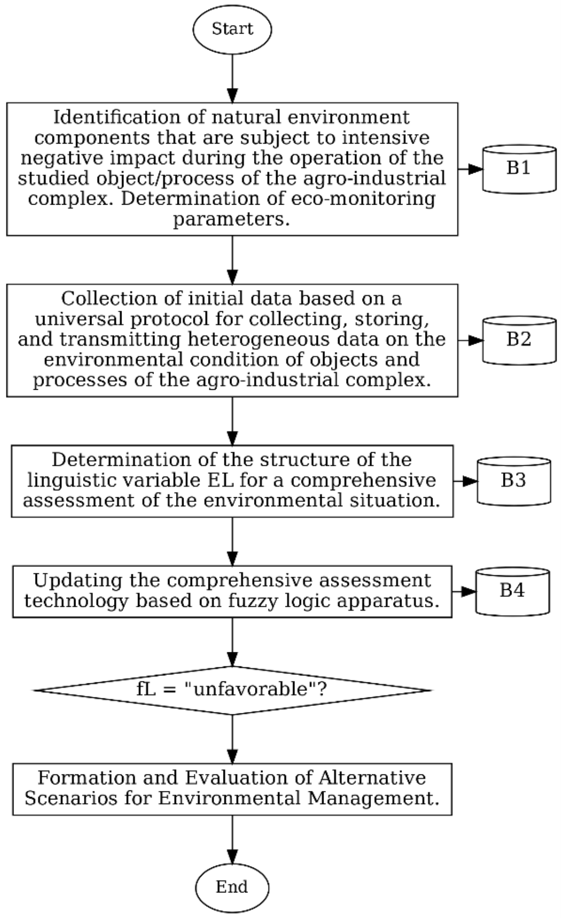


Figure 1: Block diagram of the algorithm implementing the method of comprehensive adequate assessment.

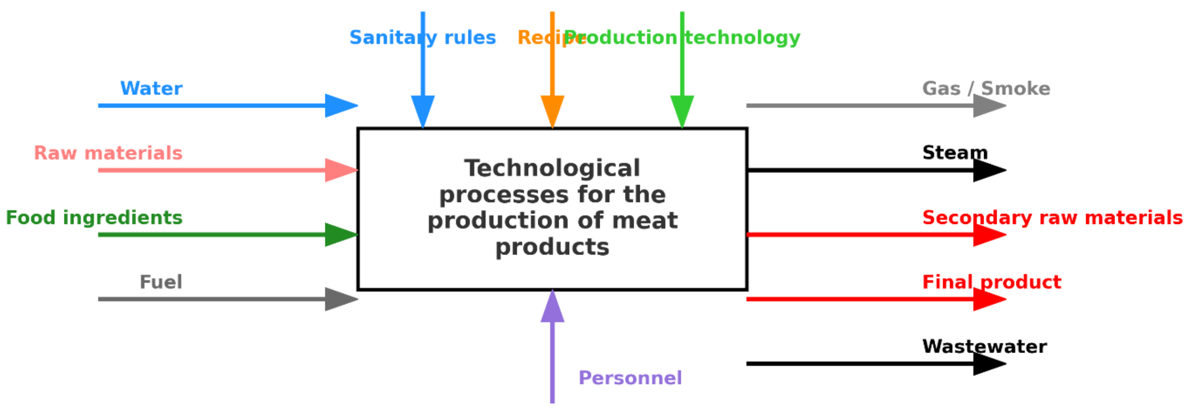


Figure 2: Contextual diagram of the model of the production organization process at the meat processing plant.

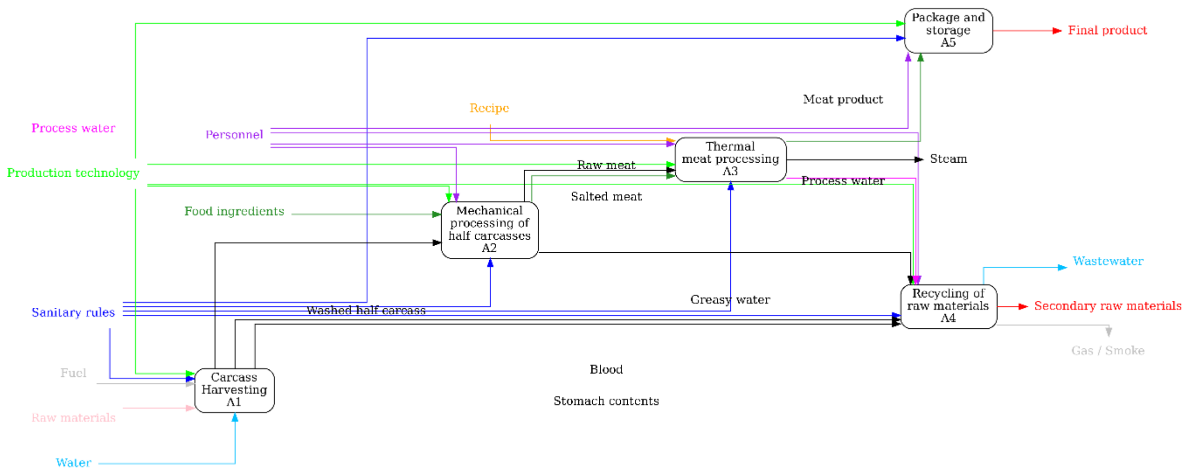


Figure 3: Decomposition of the context diagram.

At the second step, the method of collecting eco-data developed by the author's team is implemented on the basis of a universal protocol for collecting, storing and transmitting heterogeneous data on the environmental condition of objects and processes of the agro-industrial complex [12]. At the third stage of the proposed algorithm, according to the analysis of the studied business processes of the agro-industrial complex and the identified components, which have a negative technogenic impact, and the parameters of ecomonitoring, the structure of the linguistic variable EL is formed. The next stage updates the proposed integrated assessment technology, namely, for each component of EL, the corresponding membership functions and fuzzy inference rules are formed to form the values of T. The process is hierarchical: a comprehensive assessment of each component of E and L is implemented, and then a comprehensive assessment of the resulting EL is formed. To carry out a comprehensive assessment at each step of the above hierarchy, it is necessary to have a variety of mathematical, situational and simulation models.

If, after updating the integrated assessment technology, we get the value of the variable EL = «favorable», then the situation does not require intervention and the implementation of any environmental measures. Otherwise, when receiving an EL = «unfavorable» assessment, it is necessary to detail this assessment, identify the most effective control actions based on updating the method of formation and evaluation of forms.

2.3 Method and Algorithm for Evaluating Alternative Scenarios for Environmental Management

Let's imagine a variety of possible solutions to regulate the level of man-made impact of agricultural facilities/processes on environmental and food security as

$$D = \{d_1, d_3, \cdots d_{s-1}, d_s\} \tag{3}$$

A set of parameters for assessing the state of components of the natural environment and/or components of living agricultural production facilities:

$$C = \{c_1, c_2, c_3, \cdots c_{n-1}, c_n\} \tag{4}$$

and the set of values of each parameter:

$$\begin{aligned} R_1 &= \{r_{11}, r_{12}, r_{13}, \cdots r_{1j-1}, r_{1j}\} \\ R_2 &= \{r_{21}, r_{22}, r_{23}, \cdots r_{2k-1}, r_{2k}\} \\ R_3 &= \{r_{31}, r_{32}, r_{33}, \cdots r_{3l-1}, r_{3l}\} \\ R_{n-1} &= \{r_{n-11}, r_{n-12}, r_{n-13}, \cdots r_{n-1m-1}, r_{n-1m}\} \\ R_n &= \{r_{n1}, r_{n2}, r_{n3}, \cdots r_{ni-1}, r_{ni}\} \end{aligned}$$

Then the products of the formation of effective measures based on the conclusion of a comprehensive assessment:

$$p = \langle S; c_n = r_{ni} \to d_s; Q \rangle, \tag{5}$$

where p – product name, S – description of the class of situations in which the product is triggered; $c_n = r_{ni} \to d_s$ – the core of the product (if the value of the state parameter of the selected component is equal to $r_{ni} \in R$, then the assessment of his condition is equal to $d_s \in D$);

Thus, we can obtain sets of various control actions that form certain management scenarios to reduce the negative impact on the components of the natural environment and improve the quality of the condition of living agricultural production facilities that affect the formation of the level of food security. The product base is dynamic and is supplemented on the basis of new knowledge, as well as depending on the specifics of the infrastructure of the territories, the studied objects and agro-industrial complex processes.

An algorithm has been developed that implements this method, shown in Figure 4.

2.4 Practical Implementation

Figure 5 shows examples of the implementation of the developed method and algorithm.

Modeling was conducted in the Belgorod Region, Russia, using situational analysis to evaluate management interventions. The study included both simulation and field experiments conducted on test plots covering 50 hectares, representing various land-use conditions.

Field measurements utilized smart environmental monitoring sensors, including soil moisture sensors, air quality sensors, and detectors for heavy metals and organic pollutants (zinc, copper, lead, nickel, chromium, arsenic, benzene, benzopyrene). Data from the sensors were transmitted in real time via an IoT-based system [13].

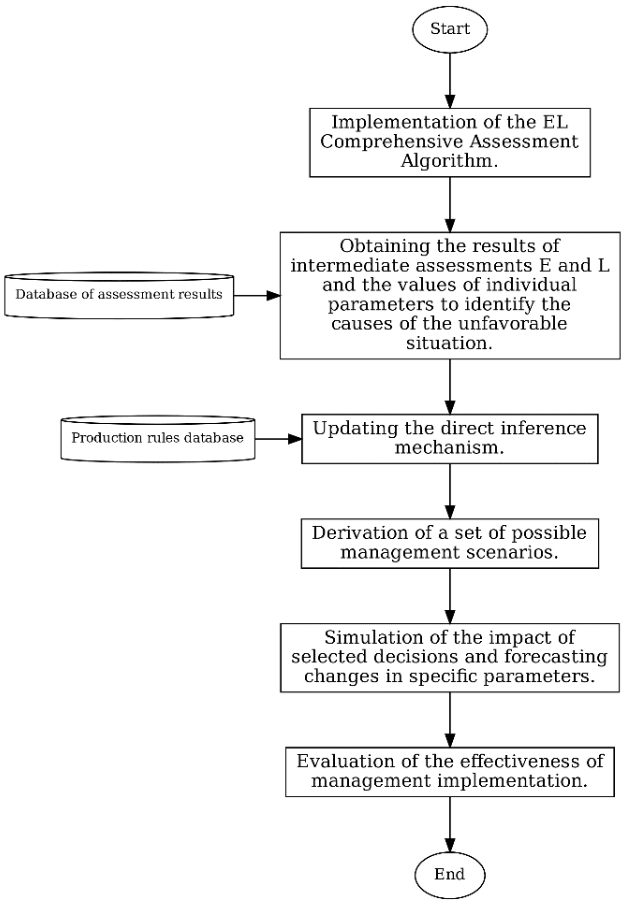


Figure 4: Algorithm for evaluating alternative scenarios for environmental management.

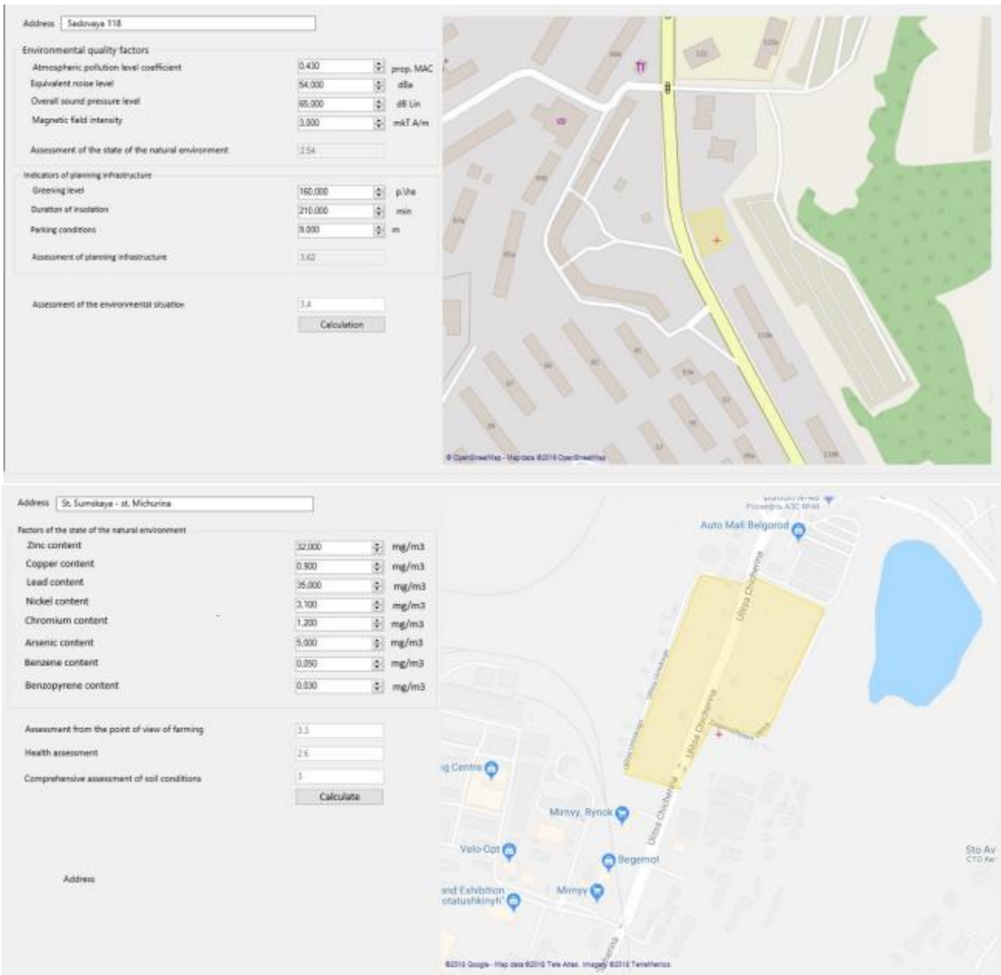


Figure 5: The results of situational modeling for the assessment of control actions.

Additionally, unmanned aerial vehicles (UAVs) equipped with hyperspectral cameras were used to assess vegetation cover and pollutant distribution. Data analysis was performed using a GIS-based system, incorporating historical environmental records.

The developed model analyzed key environmental indicators, such as the Soil Pollution Index (SPI), Air Quality Index (AQI), and Vegetation Health Index (VHI). Comparative analysis with historical data demonstrated an 27% improvement in forecasting accuracy, enabling real-time adaptive decision-making.

The implementation of the proposed method led to the following outcomes:

- 19% reduction in pollutant accumulation in the soil;
- 27% increase in agricultural productivity;
- 25% decrease in public health risks associated with environmental degradation.

Based on the findings, recommendations were developed for land-use modifications, including the establishment of protective forest strips, restriction of pollutant exposure, and adjustments to territorial planning structures.

The study results confirm the necessity of integrating intelligent ecological monitoring into territorial planning to enhance sustainable agricultural development and mitigate environmental risks.

3 CONCLUSIONS

The study developed an advanced methodology for comprehensive ecological assessment and intelligent decision-making in agro-industrial environmental management. The proposed cyber-physical system integrates fuzzy logic, IoT-based environmental monitoring, GIS analytics, and situational modeling, enabling real-time adaptive management of agricultural territories.

Field and simulation experiments conducted in the Belgorod Region, Russia (50-hectare test sites) demonstrated the model's effectiveness. The 27% increase in forecasting accuracy compared to traditional monitoring techniques highlights the system's capability to enhance decision-making processes. Implementation of optimized land-use scenarios resulted in:

- 19% reduction in soil pollutant accumulation,
- 27% increase in agricultural productivity,
- 25% decrease in health risks related to environmental pollution.

The proposed system provides actionable insights for policymakers and agricultural enterprises, ensuring biospheric compatibility and sustainable development. Future research will focus on refining machine learning algorithms for improved predictive modeling and expanding the cyber-physical system's integration with blockchain-based environmental data security frameworks.

These findings confirm that intelligent ecological monitoring should be embedded into territorial planning strategies, supporting a resilient and environmentally sustainable agro-industrial sector.

ACKNOWLEDGMENTS

This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP23489215).

REFERENCES

[1] O. A. Ivashchuk, I. S. Konstantinov, V. I. Fedorov, N. V. Shcherbinina, and Yu. N. Maslakov, "Modeling of the cyber-physical system of intelligent monitoring and management of environmental safety of the agro-industrial complex," Scientific and Technical Bulletin of the Volga Region, no. 11, pp. 52-55, 2021 (in Russian).

[2] X. Wang, Y. Liu, Z. Li, Z. Yang, and H. Zhao, "Cyber-physical system architecture for intelligent environmental monitoring in smart farms," Computers and Electronics in Agriculture, vol. 194, Art. no. 106783, Oct. 2022, [Online]. Available: https://doi.org/10.1016/j.compag.2022.106783.

[3] P. G. Ritzel, A. M. Ramos, L. J. Esquerdo, N. T. P. da Silva, and L. P. Magalhães, "Geospatial information systems to optimize crop production: a review," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 173, pp. 24-44, Jan. 2021, [Online]. Available: https://doi.org/10.1016/j.isprsjprs.2020.12.006.

[4] R. Bongiovanni and J. Lowenberg-Deboer, "Precision agriculture and sustainability," Precision Agriculture, vol. 21, no. 6, pp. 1068-1094, Dec. 2020, [Online]. Available: https://link.springer.com/article/10.1007/s11119-020-09741-3.

[5] A. N. Khosla, M. R. Abhishek, S. Vidyarthi, and P. K. Roy, "Digital agriculture: technologies and global adoption trends," IEEE Access, vol. 9, pp. 115957-115980, Aug. 2021, [Online]. Available: http://dx.doi.org/10.1109/ACCESS.2021.3071766.

[6] A. G. Stewart, L. H. Sklar, and D. R. Jensen, "Satellite-based assessment of soil moisture dynamics in arid agro-ecosystems," Remote Sensing of Environment, vol. 247, Art. no. 111937, Apr. 2020, [Online]. Available: https://doi.org/10.1016/j.rse.2020.111937.

[7] Y.-P. Lin, J. R. Petway, J. Anthony, H. Mukhtar, S.-W. Liao, C.-F. Chou, and Y.-F. Ho, "Blockchain: the evolutionary next step for ICT e-agriculture," Environments, vol. 4, Art. no. 50, Nov. 2017, [Online]. Available: https://doi.org/10.3390/environments4040050.

[8] Y.-P. Lin, T.-K. Chang, C. Fan, J. Anthony, J. R. Petway, W.-Y. Lien, C.-P. Liang, and Y.-F. Ho, "Applications of information and communication technology for improvements of water and soil monitoring and assessments in agricultural areas: a case study in the Taoyuan irrigation district," Environments, vol. 4, Art. no. 6, Jan. 2017, [Online]. Available: https://doi.org/10.3390/environments4010006.

[9] Y. Jiang, T. Wang, H. Zhao, X. Shao, W. Cui, K. Huang, and L. Li, "Development of an intelligent agricultural automation system using AI-based decision support," Applied Engineering in Agriculture, vol. 35, no. 2, pp. 147-162, 2019, [Online]. Available: https://doi.org/10.3390/agronomy13010190.

[10] N. Guimarães, L. Pádua, P. Marques, N. Silva, E. Peres, and J. J. Sousa, "Forestry remote sensing from unmanned aerial vehicles: a review focusing on the data, processing, and potentialities," Remote Sensing, vol. 12, no. 6, Art. no. 1046, Mar. 2020, [Online]. Available: https://doi.org/10.3390/rs12061046.

[11] Y.-H. Tehrani and S.-M. Atarodi, "Design and implementation of a high-precision and high-dynamic-range power-consumption measurement system for smart energy IoT applications," Measurement, vol. 146, pp. 458-466, Dec. 2019, [Online]. Available: https://doi.org/10.1016/j.measurement.2019.07.008.

[12] A. Berger, T. Hölzl, L. B. Hörmann, H.-P. Bernhard, A. Springer, and P. Priller, "An environmentally powered wireless sensor node for high-precision temperature measurements," in Proc. IEEE Sensors Applications Symp., Glassboro, NJ, USA, Mar. 2017, pp. 1-6, [Online]. Available: https://doi.org/10.1109/SAS.2017.7894065.

[13] S. L. Ullo and G. R. Sinha, "Advances in smart environment monitoring systems using IoT and sensors," Sensors, vol. 20, no. 11, Art. no. 3113, Jun. 2020, [Online]. Available: https://doi.org/10.3390/s20113113.

Scalable Video Processing and Frame Analysis System for Automated Monitoring of Chicken Behavior Based on Artificial Intelligence Technologies

Svitlana Antoshchuk[1], Oleksii Danchuk[2] and Tetiana Kunup[1]
[1]*National University Odesa Polytechnic, Shevchenko Avenue 1, 65044 Odesa, Ukraine*
[2]*Institute of Climate-Smart Agriculture of the National Academy of Agrarian Sciences of Ukraine, Mayatska Doroha Str. 24, Khlibodarske, Odesa Region, 67667 Ukraine*
asg@op.edu.ua, olexdanchuk@gmail.com, kunup.t.v @op.edu.ua

Keywords: Artificial Intelligence (AI), LLM, OpenAI, GPT, Agentic Object Detection, Poultry, Automated System, Video Analysis.

Abstract: Modern global crises have significantly impacted the agribusiness sector, particularly poultry farming, which has experienced issues with logistics, labor shortages, and disruptions in the supply of feed and veterinary drugs. Under these conditions, the implementation of artificial intelligence (AI) has become a necessity to ensure production stability. The use of AI in poultry farming allows for the automation of monitoring processes, improves the management of poultry health, and reduces reliance on human resources, which is especially important amid the pandemic and quarantine measures. This article examines the benefits of automated systems for monitoring the condition of chickens, including methods for tracking both natural and sick behavior that allow for the timely detection of diseases and minimization of production risks. The use of digital technologies and AI helps adapt to changes in market demand and ensures higher resilience of enterprises in crisis situations. The article highlights the advantages of using automated systems to monitor chicken health under conditions of limited human resources. A software solution has been developed for monitoring chicken health, which enables video uploads from local sources or video services such as YouTube, selection of the communication language, and the LLM model. It provides a user-friendly interface for interacting with AI. The program consists of two components: the frontend and the server side. Interaction with the server side occurs via an API, allowing seamless integration into any interface. The software architecture ensures convenient scalability and functionality expansion through the addition of agents and services. Such innovative solutions hold great potential for the development of the agribusiness sector, contributing to increased efficiency and resilience to crisis situations.

1 INTRODUCTION

Poultry farming is the largest branch of terrestrial animal production, with approximately 70 billion birds slaughtered annually, and the demand for chicken meat continues to grow, especially in developing countries [1]. To boost productivity, production systems have been intensified – often at the expense of animal health and welfare. In large industrial operations, the low economic value of each bird complicates individual care, yet farmers have a moral obligation to provide proper housing conditions [2].

Modern digital technologies are opening new opportunities for the automated monitoring of agricultural animal behavior, particularly in chickens, which is a crucial aspect of ensuring their welfare and productivity. One promising direction is the application of computer vision and deep learning techniques for real-time video stream analysis. This approach enables the detection of behavioral anomalies, monitoring of activity levels, and evaluation of stress factors without the need for manual observation.

Ethology of chickens studies their behavior, social interactions, and adaptation to housing conditions. Chickens are social birds with a clear hierarchy, known as the "pecking order," which determines access to food, resting areas, and resources. They exhibit a wide range of behavioral responses, including exploratory, feeding, nesting, and communicative activities. Chickens are capable of recognizing the faces of other individuals, distinguishing them by status, and producing various vocalizations to express fear, satisfaction, or danger. Under natural conditions, chickens spend a significant amount of time foraging, scratching the ground with their feet and beaks. Instinctive behavioral patterns include dust bathing for feather care, nest building before laying eggs, and maternal care. Restrictions on natural behavior in industrial conditions may lead to stress, aggression, or the development of abnormal reactions, such as feather pecking. Understanding chicken ethology is key to improving their welfare and production efficiency. In this context, the development of a scalable video processing and frame analysis system for automated monitoring of chicken behavior in production environments represents a highly relevant challenge in modern animal husbandry. The combination of neural networks, image processing algorithms, and specialized tracking methods will ensure high accuracy in recognizing behavioral patterns. [1, 3, 4]. The use of such solutions contributes to enhanced management efficiency on farms, reduced reliance on human factors, and improved conditions for poultry housing [3, 5, 6].

Optimization of Chicken Welfare Monitoring. Traditional observation methods require significant human resources and do not always provide accurate assessments. Automated analysis enables the detection of signs of stress, illness, or aggressive behavior continuously, 24/7.

Enhancement of Poultry Production Efficiency. Monitoring activity levels, feeding behavior, and other key indicators allows for timely adjustments to housing conditions. Reduction of Stress Factors. Lowering the impact of stress contributes to improved poultry health and increased production efficiency.

Early Diagnosis and Disease Prevention. The system can automatically recognize behavioral changes that may indicate the early stages of diseases, allowing for prompt intervention.

Reduction of Farm Management Costs. Automation reduces the need for personnel to monitor, analyze, and process data. Early problem detection leads to decreased veterinary expenses.

Scalability and Adaptability. The system can be integrated into both large-scale poultry farms and small family-run operations. Its adaptability to various housing conditions (cage, floor, free-range) allows for widespread application of the technology.

Environmental Aspect. Behavior monitoring helps reduce the use of antibiotics through disease prevention. Improving animal welfare meets the requirements of sustainable development and ethical agricultural practice.

2 PROBLEM STATEMENT

Amid the growing demand for high-quality products and the need to optimize production processes, artificial intelligence is becoming an essential tool for the modernization and development of agribusiness sectors, including modern poultry farming. The application of artificial intelligence, machine learning technologies, and computer vision in poultry farming allows for significant improvements in farm management, the automation of monitoring poultry health, the optimization of feeding processes, and the timely detection of diseases. The use of digital technologies contributes not only to increased productivity but also to reduced costs for maintaining and caring for poultry. Identifying birds with issues on poultry farms is crucial for ensuring their welfare and preventing disease outbreaks.

This work demonstrates the relevance of implementing artificial intelligence in poultry farming and presents a scalable server system for video processing and frame analysis using artificial intelligence technologies such as FastAPI, OpenCV, Streamlit, and others. The described system enables video processing and frame analysis for studying chicken behavior based on video data.

3 SIGNS OF CHICKEN BEHAVIOR FOR HEALTH MONITORING

To develop a video analysis system for poultry farming, it is necessary to identify the main behavioral signs of chickens that must be monitored to address this task. Chickens exhibit natural behavior characterized by certain habits that indicate their health. Even if a chicken behaves "sickly," it does not always indicate the presence of pathogenic diseases - the cause may be injuries, nutrient deficiencies, or the impact of adverse environmental factors, such as heat stress. To determine the health status of the bird, it is important to understand chicken ethology [7, 8, 9]. The natural behavioral signs of chickens include:

1) Preening. Using its beak, the chicken distributes oil among its feathers, which not only helps maintain feather condition but also signals to other flock members that it is in good health and ready for reproduction.

2) Dust Bathing. This is a hygienic process during which the chicken flaps its wings to raise dust and then carefully shakes it off, cleaning its body of parasites.
3) Foraging. In addition to feeding from the feeder, chickens often search for additional food sources, which is an integral part of their social interaction.
4) Perching at a Height. For protection against predators and to maintain the flock's hierarchy, chickens choose perching spots that are slightly elevated above the ground.
5) Play. Chickens engage in playful activities both individually and in groups, such as sparring, chasing, and lively jumps, which also serve as training for defense.
6) Aggression, Fear, and Anxiety. These are manifested through behaviors like pecking, jumping, and displays of strength (such as spreading their wings and vocalizing) to establish social hierarchy.
7) Reproduction. Reproduction begins with the onset of sexual maturity (usually at 4–5 months), when roosters display their traits (spreading their wings, loud vocalizations, dancing), and hens form nests for laying eggs and exhibit maternal instincts while raising their offspring. [1, 3, 10].

Thus, by observing these signs, it is possible to assess whether the behavior of chickens corresponds to their natural, physiological state. The assessment of the physiological state of chickens in production is carried out using a set of methods that include both visual observations and instrumental studies. The main assessment indicators are:

1) Appearance and behavior [11]:
   - Activity. Healthy chicken is active and shows interest in food and its surroundings.
   - Feather condition. Feathers should be smooth, clean and without any signs of plucking (an indicator of comfort and the absence of cannibalism).
   - Appearance of the comb and wattle. They should be bright red, well-moisturized and have no signs of pallor or swelling.
   - Movement coordination. Gait disturbance may indicate a deficiency of trace elements or joint disease.
2) Body condition and weight:
   - Regular weighing allows you to monitor growth rates and compliance with performance standards. Palpation of the pectoral muscle is used to assess muscle development and potential exhaustion.
   - Feed intake. A sharp decrease in appetite signals possible health problems.
   - Water intake. Excessive or insufficient water intake may indicate a violation of the water-salt balance or infection.
3) Litter condition. The color, consistency, presence of undigested food residues or mucus in the litter may indicate infectious or metabolic diseases.
4) Physiological parameters:
   - Body temperature. Normally 40.6–41.7°C; measured if pathology is suspected.
   - Respiratory rate. Rapid breathing may be a sign of respiratory disease or heat stress.
   - Heart rate. Normal is 250–350 beats per minute; abnormal heart rate may indicate stress or cardiovascular disease.
5) Reproductive assessment (for layers):
   - Egg production monitoring. A sharp decline in production may be due to stress, illness, or malnutrition.
   - Egg shell thickness analysis. Thinning of the shell indicates calcium deficiency or metabolic disorders.
6) Hematology and biochemical studies:
   - Blood tests are used to determine glucose, protein, liver enzymes, and electrolytes. Hematocrit and hemoglobin levels are used to detect anemia or inflammation.
   - The combination of these methods allows for effective monitoring of the physiological state of the hens [12, 13].

It should be noted that such approaches for assessing the physiological signs of chicken health and sickness are not only high-cost in terms of human resources but also rather subjective, which makes them impractical for large-scale production.

Therefore, a system is proposed that enables video processing and frame analysis to study chicken behavior based on video.

4 APPLICATION OF A VIDEO PROCESSING AND FRAME ANALYSIS SYSTEM

4.1 General Structure of the System

In today's world, automation of data collection and analysis processes is a key component of machine learning and computer vision technologies. The developed system provides video processing and frame analysis for studying chicken behavior based on video.

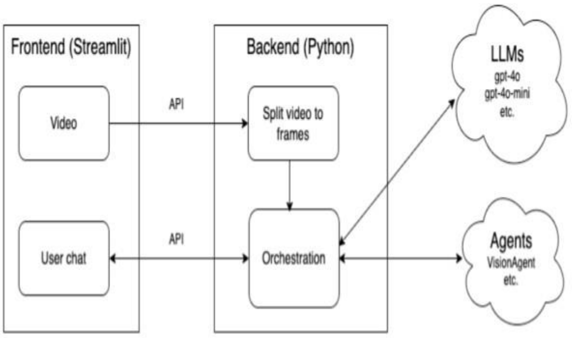


Figure 1: Program structure.

Figure 1 illustrates the general structure of the system, which consists of three main elements: Frontend (Streamlit), Backend (Python) and external services in the form of LLMs (large language models) and Agents (specialized modules).

Description:

1) Frontend (Streamlit): interacts with the backend via API: sends requests for video processing and receives results or answers to questions.
2) Backend (Python): consists of two main components:
   - Split video into frames: a module that splits the video into frames at given intervals.
   - Orchestration: a central logical unit that coordinates the entire system. It receives requests from the Frontend, calls the appropriate services (video processing modules, LLMs, agents, etc.) and compiles the response to send back to the Frontend. It makes calls to external services (LLMs and agents) to solve various tasks: image analysis, text response generation and anomaly detection.
3) LLM (Large Language Models): These models can perform tasks such as text generation, description and image analysis, and answering user questions. They are called by the backend to process text queries and images extracted from videos.
4) Agents (VisionAgent): Specialized modules or microservices that perform specific functions related to image or video processing and analysis. They can use computer vision, machine learning, or other algorithms to detect anomalies, classify objects, and evaluate chicken behavior.

The diagram shows how the frontend (Frontend), data processing logic (Backend), and external intelligent services (LLM, Agents) interact with each other to effectively analyze videos and chicken behavior.

The system allows users to upload videos from video services or local files, choose a communication language, and utilize large language models. Additionally, it provides the option to configure video segmentation parameters by specifying the interval at which frames (images extracted from the video) are generated – that is, setting the number of seconds between frames.

Left Panel – "Upload Options" (see Figure 2):

- Choose upload method: The user can select the source for uploading the video (either a local file or a YouTube link).
- Select frame interval (seconds): Specify the interval in seconds at which frames will be generated from the video.
- Choose language: Allows the user to choose the language for communication or analysis.
- Select model: Enables selection of a large language model (LLM) to be used for further processing.

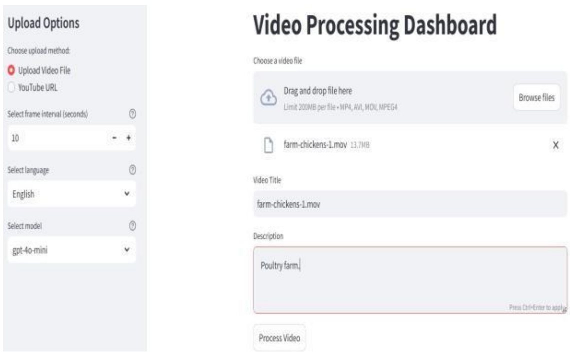


Figure 2: Screenshot of the main page of the program.

Right Panel – "Video Processing Dashboard" (see Figure 2):

- Drag and drop file here / Browse files: An area where the file can be dragged and dropped for uploading or selected from local storage.
- Video Title: A field displaying the title of the uploaded video (e.g., “farm-chickens-1.mov”) and its file size.
- Description: A text field for additional information or a description of the video (e.g.,

"Poultry farm").

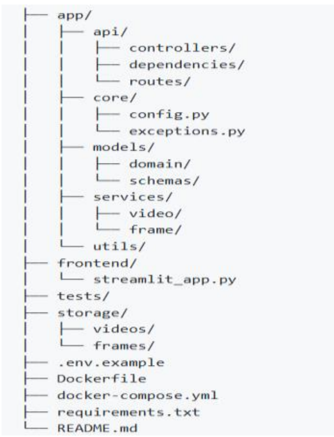
- Process Video: A button that initiates video processing (splitting the video into frames, passing the frames to the selected LLM).

This interface is built on the Streamlit framework, which allows for the rapid and convenient development of web applications for processing and analysis video. The backend is implemented using the Python programming language, and communication between the interface and the server is carried out via a RESTful API. The system development uses several powerful technologies[1]:

- FastAPI is a modern web platform for creating high-performance RESTful APIs. Thanks to its asynchronous capabilities, FastAPI provides fast request processing and easy scaling.
- OpenCV is a library for image and video processing. It allows you to extract frames from videos, resize them, and apply filters and algorithms for analysis.
- Streamlit is a library for rapid development of interactive web applications for scientific research and data analysis. It provides a convenient interface for interacting with the processing results.

The system has a modular architecture that includes the following components:

- API (FastAPI) – the API (FastAPI) is responsible for processing requests and interacting with the user via the REST API. This allows you to load videos, select the frame extraction interval, and access the processing results.
- The video processing service (OpenCV) is a module that performs basic video operations, such as extracting frames at specified intervals and saving the resulting images.



```
├── app/
│   ├── api/
│   │   ├── controllers/
│   │   ├── dependencies/
│   │   └── routes/
│   ├── core/
│   │   ├── config.py
│   │   └── exceptions.py
│   ├── models/
│   │   ├── domain/
│   │   └── schemas/
│   ├── services/
│   │   ├── video/
│   │   └── frame/
│   └── utils/
├── frontend/
│   └── streamlit_app.py
├── tests/
├── storage/
│   ├── videos/
│   └── frames/
├── .env.example
├── Dockerfile
├── docker-compose.yml
├── requirements.txt
└── README.md
```

Figure 3: Screenshot of the system structure.

The interface (Streamlit) creates a user interface where users can upload videos, select intervals for analysis, and view the results in real time. Figure 3 shows a screenshot of the system structure.

4.2 Video Processing and Frame Analysis Methods

The video processing and frame analysis pipeline consists of three key stages designed to systematically analyze chicken behavior, beginning with data acquisition and culminating in automated anomaly detection[2]:

1) Video Upload. The user can upload videos from a local computer or paste a YouTube link.
2) Frame Extraction. The video processing service extracts frames from the video at a specified time interval (from 1 to 20 seconds). This is accomplished by calculating the necessary number of frames based on the video's duration and the specified interval.
3) Frame Analysis. Additional processing methods, such as filtering, color analysis, or the use of machine learning algorithms, can be applied to the extracted frames to detect anomalies in chicken behavior. See Figure 4 and Figure 5.
4) Interactive Interface. Through the interface, the user can view the processed frames, ask questions, clarify details, and analyze individual frames.

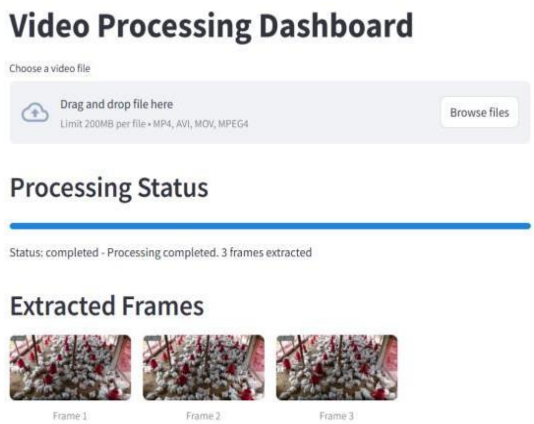


Figure 4: Screenshot of selecting frames from a video at a specific time interval.

Experimental studies of the developed scalable server system for video processing and frame analysis were conducted on test data sets. The studies showed that when working in real time, the accuracy

[1] https://www.renesas.com/eu/en/document/oth/um-wi-023-da16200-threadx-evaluation-kit-user-manual

[2] https://nypost.com/2025/02/15/us-news/us-egg-farmers-concerned-bird-flu-unwinnable-as-experts-consider-vaccinations/?utm_source=chatgpt.com

of counting chickens in a frame is at least 90%, which is comparable to analogues. However, the developed system has more flexible capabilities.

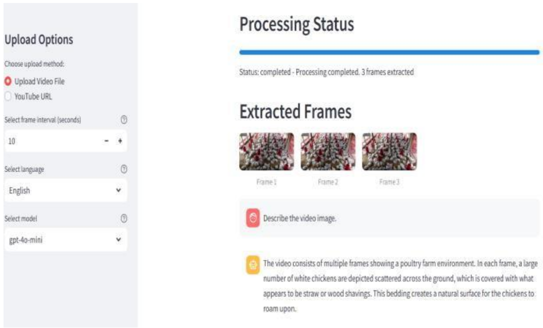


Figure 5: Screenshot of the request processing shown in the video.

5 CONCLUSIONS

A scalable server system for video processing and frame analysis using artificial intelligence technologies FastAPI, OpenCV, Streamlit, etc. has been developed. Its application to study the behavior of chickens based on video is shown.

During the study, the developed system demonstrated high efficiency in computer vision and animal behavior analysis across various video materials. The system does not rely on traditional model training methods (such as collecting visual data, image processing and analysis, or machine learning). Instead, it features an intuitive interface that allows users to optimally configure processing parameters and rapidly obtain results. The proposed system is an effective tool for automated video processing and the analysis of chicken behavior. The application of artificial intelligence methods contributes to the creation of a high- performance and user-friendly solution that does not require lengthy neural network training. This is particularly relevant for agribusiness research, where automated monitoring of animal health is crucial. Improvements to the system may include integration with other agents and services for image and video analysis, which will provide a more detailed examination of chicken behavior and health status. The system enables communication with the user in various formats, ranging from text chat to visual frame review. Its architecture allows for the rapid integration of new agents, language models, or image and video processing services. Configurable analysis parameters and an open RESTful API simplify integration with third-party tools and services. Although the system is designed for use in the agribusiness sector (specifically poultry farming), it can be adapted for analyzing the behavior of other animal species or any objects in video. It can serve as a foundation for scientific research, automated health monitoring, and ensuring animal welfare. Thanks to these features, the system provides a comprehensive solution for efficient and scalable video analysis, significantly simplifying the process of collecting and interpreting data on the behavior of objects-particularly chickens-in real time.

REFERENCES

[1] J. Mench, I. C. de Jong, L. Berg, M. Raj, and A. Butterworth, "Broiler Chickens: Welfare in Practice," Essex, UK, vol. 5, pp. 21-34, 2024, [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/00439339.2023.2264824.

[2] S. Azarpajouh, S. Weimer, J. Calderón Díaz, and H. Taheri, "Smart Farming: A Review of Animal-Based Measuring Technologies for Broiler Welfare Assessment," CABI Reviews, Oct. 2022, [Online]. Available: https://doi.org/10.1079/CABIREVIEWS202217033.

[3] J. Bao and Q. Xie, "Artificial intelligence in animal farming: A systematic literature review," Journal of Cleaner Production, vol. 331, 2022, [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0959652621041251.

[4] H. Patel, A. Samad, M. Hamza, A. Muazzam, and M. Harahap, "Role of artificial intelligence in livestock and poultry farming," Sinkron: jurnal dan penelitian teknik informatika, vol. 6, no. 4, pp. 2425-2429, 2022.

[5] H. Patel and A. Sana, "Role of Computer Science (Artificial Intelligence) In Poultry Management," Devotion: Journal of Research and Community Service, vol. 3, no. 12, pp. 2068-2088, 2022, [Online]. Available: https://www.researchgate.net/publication/368285922_Role_of_Computer_Science_Artificial_Intelligence_In_Poultry_Management.

[6] R. Khanal, W. Wenqin, and J. Lee, "Automated Dead Chicken Detection in Poultry Farms Using Knowledge Distillation and Vision Transformers," Applied Sciences, vol. 15, no. 1, pp. 124-136, 2024.

[7] G. Havenstein, "Performance and Efficiency of Broilers as Influenced by Genetics, Management, and Nutrition," Poultry Science, vol. 82, no. 10, pp. 1500-1508, 2003, [Online]. Available: https://www.scirp.org/reference/referencespapers?referenceid=1291166.

[8] T. Knowles, B. Hughes, and M. Appleby, "The Welfare of Broiler Chickens: a review," Poultry Science, vol. 86, no. 5, pp. 938-949, 2007, [Online]. Available: https://food.ec.europa.eu/system/files/2020-12/sci-com_scah_out39_en.pdf.

[9] L. Patreva and O. Koval, "Technology of poultry production: lecture course," Mykolaiv: MNAU, pp. 146-156, 2018.

[10] G. Damerow, The Chicken Health Handbook, 2nd Edition: A Complete Guide to Maximizing Flock Health and Dealing with Disease, pp. 36-67, 2015.

[11] M. Reiter, "Social Behavior in Chickens," Journal of Poultry Science, vol. 43, no. 3, pp. 143-150, 2006.

[12] C. Nicol, S. Abeyesinghe, and Y. Chang, "An analysis of the welfare of fast-growing and slower-growing strains of broiler chicken," Frontiers in Animal Science, vol. 5, 2024, [Online]. Available: https://www.frontiersin.org/journals/animal-science/articles/10.3389/fanim.2024.1374609/full.

[13] L. Schlegel, A. Kleine, M. Doherr, and F. Tenhagen, "How to see stress in chickens On the way to a Stressed Chicken Scale," Poultry Science, vol. 103, no. 8, 2024, [Online]. Available: https://www.vetmed.fu-berlin.de/forschung/paper-of-the-month/preistraeger/202407/Schlegel_et_al_Stressed_Chicken_Scale.pdf.

System for Simultaneous Control of Multiple Unmanned Aerial Vehicles

Yehor Zheliazkov
SBH Nordost GmbH, Dresden Chip Academy, Hermann-Reichelt Str. 3A, 01109 Dresden, Germany
yzheliakov2@gmail.com

Keywords: Unmanned Aerial Vehicles, Swarm Drone, Software-Defined Radio, Over-The-Air.

Abstract: Nowadays, airspace is a strategically important resource for both national security and economic development. The state is responsible for its efficient use and safety. One of the key challenges of today is countering unmanned aerial vehicles, which includes the use of electronic warfare, laser weapons and other technologies. The popularity of drones is growing rapidly not only in the consumer sector, but also in industry, agriculture, energy, geodesy, and military. Military conflicts demonstrate a significant increase in the use of unmanned aerial vehicles, in particular in the form of drone swarms, which creates new challenges for countermeasures. One of the most promising developments is the creation of a platform for managing heterogeneous swarms of drones, including those that were previously hostile but can be integrated into their own system. It provides new opportunities for warfare and an advantage in the management of military operations. This paper presents solutions that provide simultaneous control over heterogeneous unmanned aerial vehicles, which makes it difficult to use them effectively in modern warfare and defense strategies. An approach is realized as real prototype electronics circuit, which can create a flexible platform capable of intercepting, analyzing and integrating enemy unmanned aerial vehicles into its own network through electronic interference, hacking communication protocols and reprogramming. The results of proposed approach case study showed the possibility to take under the control “enemie’s” unmanned aerial vehicles to the control of main operator.

1 INTRODUCTION

Globally, airspace is an important resource for various sectors of national security and the economic sector. It is the duty of the state to ensure the efficient use of airspace and its security [1].

Today, new ways of countering unmanned aerial vehicles (UAVs) are being explored, including the use of electronic suppression means - various electronic warfare (EW) devices, as well as directed energy emission means - laser weapons [1, 2].

Until recently, there has been a rapid increase in the popularity of drones, and it was based on everyone's interest in flying. Photo and video recording from the air, the ability to shoot in previously inaccessible places, new angles - all this made the technology in demand. Drones topped the charts of the most desired gifts in the consumer market. But UAVs are not just a hobby or a selfie, they are becoming assistants in various industries: agriculture, mining, topography, geodesy, and energy. All of these are industrial and economic sectors [1].

Dominance in space gives an advantage over the enemy in controlling troops and weapons. This is facilitated by electronic warfare (EW) and electronic warfare (EW) means, which are one of the leading elements of current wars and armed conflicts and have the highest development dynamics among all modern weapons [3].

The experience of current military conflicts and the war in Ukraine shows the growing role of drones and the variety of their models (air, surface, ground) on the battlefield for various functional purposes. At the same time, there is an active use of not only single UAVs, but also a large number of them simultaneously. The nearest modern prospect is the transition to swarms of drones [4].

Military and civilian life are undergoing a profound transformation based on advances in two critical technologies: machine autonomy and artificial intelligence (AI). The current global trend is the creation and use of various sets of drones - drone swarms - which is already revolutionizing the way we look at military art [4].

At the moment there is no platform that can simultaneously control a swarm of unmanned aerial vehicles (UAVs) of different models and different purposes, including those that belonged to “enemies” and are becoming “our own”.

The main aim is to develop a new drone control platform capable of controlling a swarm of unmanned aerial vehicles with the ability to simultaneously control different models of drones, as well as the ability to “intercept” enemy UAVs at close range.

2 THE PRINCIPLE OF CONTROLLING A SWARM OF UNMANNED AERIAL VEHICLES

2.1 The Principle of Controlling a Swarm of Unmanned Aerial Vehicles

In a group of drones interacting based on swarm intelligence, each of the devices interacts only with some of the devices closest to it now [5]. At the same time, energy consumption for information transmission is relatively low. UAVs make decisions about their current behavior based on self-collected environmental data as well as data transmitted by neighboring vehicles. Energy-consuming communication with the central control device is carried out only to receive information about the tasks assigned to the group and to transmit a report with information about the group's status during the implementation of the completed tasks [5]. The swarm algorithm is an effective solution to ensure optimal control of a UAV swarm [6].

According to [5], a simplified general scheme of swarm intelligence algorithms includes the main stages:

1) Initialization of the population - this stage initializes (often randomly) the population of agents that are given directions to search for targets;
2) Migration of agents (finding targets, exchange of information between agents, changing the target if other agents have found a higher priority target);
3) Completion of the search (occurs if all the found targets have been achieved).

The swarm algorithm in the form of a flowchart is shown in Figure 1.

This algorithm is the basis for the further creation of any other UAV swarms related to management and communication in the group. The authors in [5] also provide bee swarm algorithms for controlling a UAV swarm and responding to emergencies separately, which is also an additional basis for building a UAV control interception algorithm.

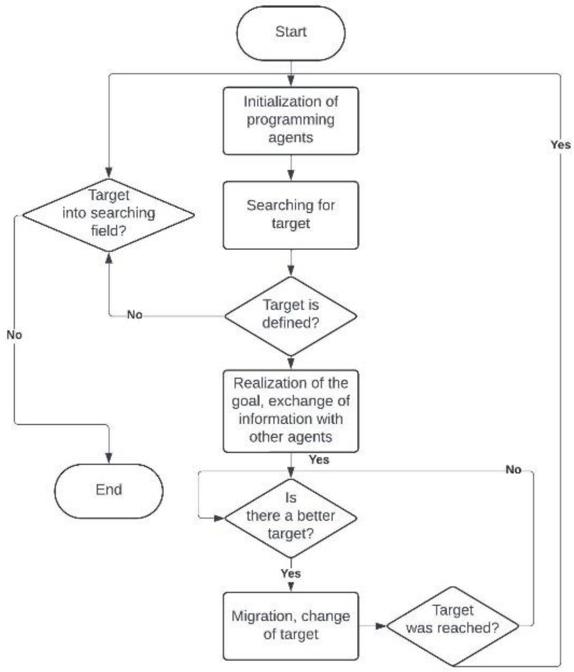


Figure 1: The general swarm algorithm.

In order to implement the algorithm, you need to know all the components for its construction. The main aspects:

1) The ability to remotely flash the device.
2) Compliance of the UAV module for remote flashing.
3) The ability to identify the drone within range

2.2 Wiring Diagram of the Device for Intercepting and Repurposing an Enemy Drone

The description of the connection scheme of a device for intercepting and repurposing an enemy drone consists of three important components: a signal interception circuit, a reflashing circuit, and a signal jamming circuit.

General flowchart for intercepting and repurposing an enemy drone is shown in Figure 2. It was suggested similar as in [7].

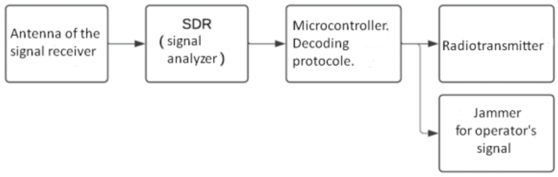


Figure 2: Flowchart for intercepting and repurposing an enemy drone.

The following steps should be taken into account in order to realize this:

1) To receive a signal, connect the antenna to a radio receiver for analyzing Software-Defined Radio (SDR) signals, in particular, to analyze the frequencies at which the drone operates. SDR is designed to analyze the frequency spectrum and data transfer protocols between the drone and its operator.
2) Protocol decoding. It is necessary and sufficient to study the structure of protocol packets (MAVLink, DSMX or other protocol).
3) Generation of a replacement signal. An RF transmitter connected to the microcontroller is used to generate commands.
4) Blocking the operator signal. The jammer generates intense noise on the control channel.

Instead of the conventional “jammer”, another gun is implemented, which sends a pulse that the drone “understands”. The software is flashed with the help of “Over-The-Air” (OTA); in particular, “migration programming” is used [8]. The “gun” itself must have the appropriate software for reprogramming.

There must be a corresponding module for each remote flashing. One of the components of any UAV is a flight controller, or control module. A flight controller is a UAV control board that processes signals from its own sensors (accelerometer, gyroscope, etc.), devices connected to it, pilot commands, and calculates the speed to be set for the motors [9].

The controller has possible corresponding interfaces (UART, I2C, PWM, etc.) for connecting ESC, VTX, RX, camera, GPS, servomotors, and other devices. The mentioned interfaces for remote flashing must be compatible with OTA updates [9].

2.3 Scheme for Signal Interception

To implement the hardware to intercept and repurpose an enemy drone, you need a circuit that includes elements for analysis, jamming, device connectivity, and reflashing. Below are the main components and their roles in the circuitry implementation.

The main components of the hardware circuit include: a radio receiver for signal analysis - Software-Defined Radio (SDR); a transmitter for signal replacement, a signal jammer, an interface for connecting to the drone, microcontrollers, sensors for debugging and testing signals [10].

Components:

1) Receiving antenna (e.g., Yagi type directional antenna);
2) SDR module (HackRF, RTL-SDR) for analyzing control signals;
3) Microcontroller (STM32/Arduino) for signal processing;
4) Radio transmitter (nRF24L01, CC2500) for transmitting new commands to the drone.

The antenna is connected to the SDR via an SMA connector. The SDR is connected to a computer (PC) via USB. The PC is used to analyze the frequency and signal protocols using software (e.g. GQRX, SDR#). After analyzing the signal, the PC transmits the protocol information to the microcontroller via UART or USB.

The microcontroller generates data packets and transmits them to the RF transceiver via SPI.

3 STRUCTURE SCHEME

To take control of the drone, the system first intercepts the signal, then analyzes the protocol, and then transmits commands, labeling the operator's signals, via the RF module.

A “cannon” circuit for implementing the principle of intercepting a third-party UAV is proposed, as shown in Figure 3. The power supply provides stable DC and AC voltages for all system components - 3.3 V, 5 V, and 230 V. The microcontroller (MCU) is the main component of the system, responsible for signal processing, coordination of other modules, and execution of the interception algorithm. The MCU exchanges signals via digital pins to communicate with the radio frequency module (RF module) and the signal jammer.

In this paper, we consider a new device with the ability to intercept and control a swarm of UAVs using jamming and spoofing capabilities.

A signal jammer is a device that interferes with a drone's signal, causing the UAV to either lose its signal and fall from the sky or return to the operator.

By hacking the UAV system, it becomes possible to take control and fly in any direction.

Based on the received signals, the microcontroller performs calculations to determine the type of protocol and drone control parameters, and connects to the drone to gain control.

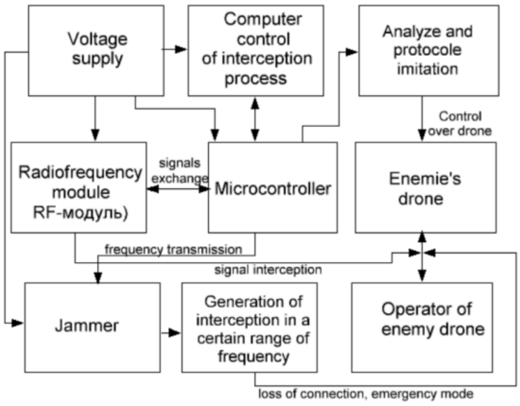


Figure 3: Block diagram of the system for intercepting an unauthorized UAV.

The RF module is responsible for intercepting signals [11] used in communication between the enemy drone and its operator. Sequence of the module operation:

- The microcontroller initializes the module via the SPI interface; the microcontroller transmits frequency data to the signal jammer;
- the jammer generates interference in a certain frequency range;
- The UAV loses communication with the operator and goes into emergency mode.

3.1 Stage of Signal Jamming

The jamming stage is the key to blocking communication between the enemy drone and its operator. Here is a sequential description of the process:

1) Determining the control frequency:
   - Before jamming, it is important to determine the operating frequency that the drone uses for communication. Common frequencies used for FPV are 900 MHz, 1.2 GHz, 1.3 GHz, 2.4 GHz, 3.3 GHz, and 5.8 GHz. The device must be capable of working with the appropriate frequencies [12].
   - Signal intensity (RSSI) is used to determine the exact channel on which communication is taking place [13]. Since the signal strength changes throughout the flight, communication degradation can be assessed in advance, more smoothly and before “problems” in control occur - this is an advantage of RSSI. The signal strength can be determined on both analog and digital receivers.
2) Setting up the jammer. To jam the signal, you need to set the exact frequency of the channel where the UAV operates. Then adjust the interference width to cover the drone's signal [14].
3) Generating interference. The jammer transmits a powerful signal of the same frequency as the operator's signal, but with high intensity (usually > 10 dB stronger). This creates “noise” that prevents the drone from receiving commands. The jamming method used is broadband jamming, which creates noise in the entire frequency range [15].

3.2 Signal Interception Stage

The signal interception phase involves several steps, each of which is critical to successfully seizing control of the drone.

Sequential description of the process:

A) Setting up the RF module to listen for frequencies. The RF module is configured on the microcontroller via the SPI interface [16]. The main parameters are: a) frequency range (in most drones it is 2.4 GHz), the RF module starts scanning channels in the range; b) data rate (common rates are 250 kbps, 1 Mbps, 2 Mbps); c) channel width (1, 2 or 5 MHz).

B) Scanning the frequency range. The RF module switches to the “receive” mode (RX-mode) to listen for signals within 2.4 GHz. The module checks each channel and records the activity of the signals (for example, RSSI level - radio signal intensity) [17].

C) Intercepting a data packet. Once a signal is found, the RF module stores the data packet in its buffer, and the microcontroller reads the packet via the SPI interface for further analysis. The packet includes the drone's ID, transmitter address, and commands (e.g., “takeoff”, “turn”). If activity is detected on a particular channel, the microcontroller stops scanning and starts analyzing the signal.

D) Determination of communication parameters. The microcontroller analyzes the received packet and extracts important parameters: 1) the transmitter address - a unique ID of the operator to which the drone is connected; 2) control commands that are decoded depending on the protocol; 3) cyclically redundant CRC code - checking the correctness of data to avoid errors [18]. The collected parameters are stored in the microcontroller's memory: 1) signal frequency; 2) drone address; 3) communication protocol.

The microcontroller uses these parameters to connect to the drone and send its commands.

3.3 The Stage of Transferring Control to the Enemy UAV

The main requirements for successful UAV interception are accurate collection of drone communication parameters (ID, frequency, packet structure), quick setup of the RF module (interception and connection takes fractions of a second), and stable operation of the jammer (to maintain control advantage) [19].

The process of transferring control to an enemy UAV involves several stages that are performed sequentially:

1) Interception of the signal from the drone. The RF module operates in a given frequency range, it is configured to listen to signals coming to the UAV from its operator. When the microcontroller receives these signals through the RF module, it:
   - Analyzes the signal structure (communication protocol, data rate, transmitter address).
   - Records the parameters used for communication (drone ID, command codes).
2) Protocol analysis and simulation. After intercepting the signal, the microcontroller performs analysis [20]:
   - Determines the type of communication protocol (e.g., DSMX, SBUS, or other).
   - Decodes the structure of the data packet (commands transmitted by the drone operator, for example: movement, ascent, landing).
   - Generates new data packets that simulate the operator's signal.
3) The RF module is set to the same communication parameters as those used by the drone operator. These parameters include signal transmission frequency, recipient address (drone ID), and data rate (bits/s).
4) Blocking the real operator's signal. To prevent the UAV from receiving signals from its original operator, the jammer generates radio interference on the frequency used by the operator [21]. This blocks signals from the control panel to the drone, leaving your device as the only source of commands. The microcontroller fully controls the UAV through the RF module, including transmitting commands (e.g., "Raise the UAV to a certain height," "Land at a given point," "Turn off the engines"). The commands are formed in accordance with the structure of the protocol that the UAV is waiting for.
5) Support for stable control. The microcontroller constantly communicates with the UAV, periodically sending confirmation or new commands. The signal jammer continues to block the signals of the real operator.

4 CALCULATING THE RANGE OF ACTION

During the operation of the radio module, there is a permanent exchange of messages between the receiver and the transmitter. Under normal operating conditions, the receiver responds to the transmitter with a response about the correct reception of the information packet. Thus, the calculation of the potential range can be reduced to the calculation of the range of the radio line with an active response. Figure 4 and Figure 5 show the structural diagrams of the radio-controlled system (RCS) and the control panel in a simplified form. According to fig. 4, the radio-controlled system has an antenna, a module for radio control, which can be performed by the certain chip [16].

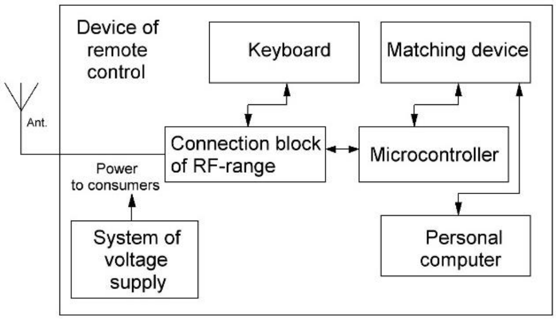


Figure 4: Radiocontrol system for signal receiver.

Diagram of the remote control. Ours is a remote control, which is implemented on the basis of the NRF24L01 chip. In addition to the chip itself, the remote control may contain additional amplifiers and an antenna that also provides amplification.

In this case, it is important to connect with main drone among the swarm of unmanned aerial vehicles. It is difficult technology with different parameters and internal algorithms for drone detection, especially when they different distance and different construction have, which may difficult to determine.

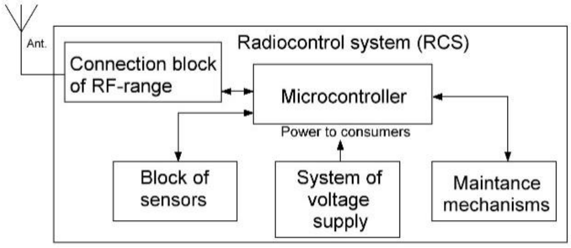


Figure 5: Signal transmitter as device of remote control.

In the most of situation people use the chip NRF24L01. In addition to the chip itself, the RCS communication unit may contain additional amplifiers and an antenna that also provides amplification [16].

To calculate the potential range of the RFS use the theory of the potential range of an active response radio line. The maximal possible range in free space, in the absence of active and passive interference, do not taking into account attenuation in the atmosphere, is determined by the equation:

$$R_{\max} = \sqrt{\frac{P_{Tx} G_{pkc} G_{n\delta y} G_{nrf} \lambda^2}{(4\pi)^2 P_{\min}}} \qquad (1)$$

P_{Tx} is maximum transmitter power of the radio module (W), $G_{n\delta y}$ is the gain of the remote control antenna, G_{pkc} is the antenna gain of the RCS, P_{min} receiver sensitivity (W), R_{max} is maximal possible range (m), G_{nrf} is additional gain (taking into account possible additional amplifying).

Taking into account the fact that the calculations are made for a ground-moving platform, it is necessary to pay attention on the line-of-sight range, since this frequency range does not provide for operation without line-of-sight [17-18]. There was used the line-of-sight range equation:

$$D_{\max} = 4.12\left(\sqrt{H_a} + \sqrt{h_b}\right) \qquad (2)$$

H_a is height of the remote control antenna, h_b is height of the RCS antenna.

A comparison of the possible range of the radio control system (R_{max}) with the distance to the radio horizon (D_{max}), which is determined by the design of the control station's antenna mounting and elevation system, shows a way to increase the range of the control system [19-20].

There are two ways to increase the efficiency of the system: to introduce additional signal amplifiers into the system, or to build an antenna system with a higher gain.

Taking into account for NRF24L01 technical characteristic (frequency – *2.4 GHz*, control station antenna height – *1.5 m*; control station antenna gain *20 dB*; the height of the antenna on the mobile platform is *0.1 m*; the pin antenna gain on the mobile platform – *5 dB*) and parameters defined by programmatic agencies (NRf24 receiver sensitivity – *-85 dBm*; data transfer rate – *1 Mbps*; transmitter output power – *0 dBm*), were received from (1) and (2) distance parameters such as R_{max}*≈3.12 km* and D_{max}*≈6.35 km* [21].

5 PRACTICAL REALIZATION OF DEVICE

Distance parameters such as *Rmax≈3.12 km* and *Dmax≈6.35 km*. It is enough to maintain of Swarm Drone, because the plural average of drones are flying near each other with a distance less than 100 m. As the result, it was built the circuit in Figure. 6.

Device in Figure 6 takes into account two PLL-synthesizers ADF4351. The ADF4351 synthesizer is used as the basis and signal source for jamming. The output signal of the RF interference is formed by mixing the signal of the local oscillator and the signal of the intermediate frequency noise. The intermediate frequency signal is generated by amplifying and filtering the base noise. The local oscillator signal is generated from the ADF4351 signal generator. Since the frequency of the local oscillator can be changed using the ADF4351, the center frequency of the RF signal can be controlled by program [22].

The second ADF4351 has a built-in voltage controlled oscillator (VCO) with a fundamental output frequency in the range of 2200 MHz to 4400 MHz. In addition, the 1/-2/-4/-8/-16/-32/-64 divide-and-convert circuits allow the user to generate RF output frequencies up to 35 MHz. All on-chip registers are controlled via a simple 3-wire interface [22]. The device operates from a power supply in the range of 3.0 V to 3.6 V and can be turned off when not in use.

It takes into account: PLL-synthesizer (ADF4351), flashing device (ST-Link V2), microcontroller (STM32), RF-Transceiver intercepting (NRF24L01) and another additional components. Detailed scheme of device is represented in Figure 6.

Radio frequency (RF) enables communication between a drone and its ground transmitter or remote control, typically operating in the 2.4 GHz to 5.8 GHz range – with 2.4 GHz and 5.8 GHz being the most commonly used frequencies for remote drone control.

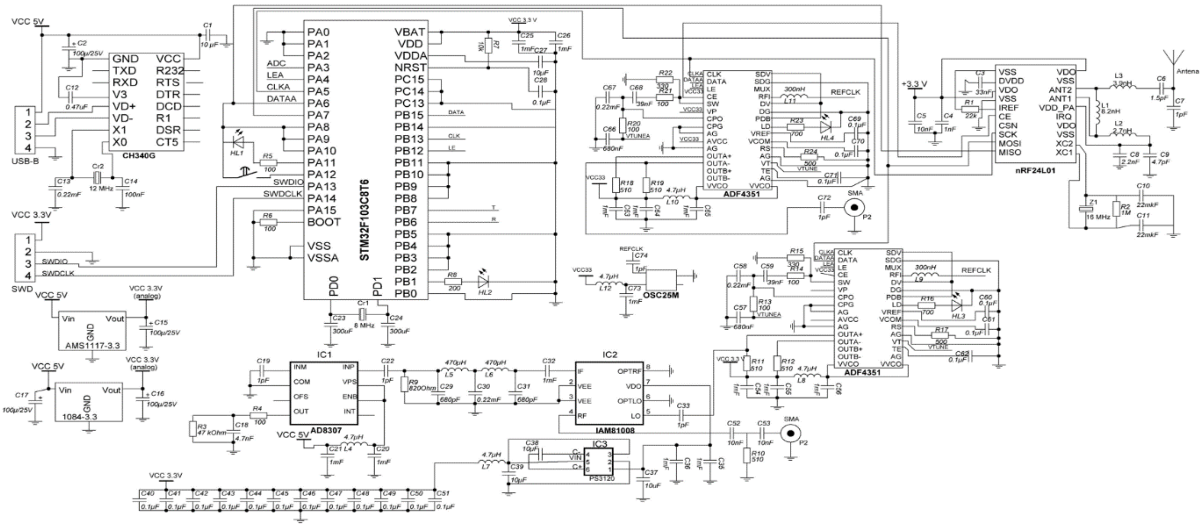


Figure 6: Electronic circuit diagram for UAVs intercepting.

Designing a loop filter is a key part of PLL design. In order to obtain a stable VCO control voltage, a loop filter should be used to filter out high-frequency interference in the line voltage [23]. In this work, a passive low-pass notch filter of two orders is used. System for simultaneous control of multiple unmanned aerial vehicles means that there is main drone among the swarm of drones and one drone can arrange synchronized work for other nearest UAVs. After the jamming antenna receive the signal with a following processing and then sends the feedback signal with remotely update firmware using official applications and OTA technologies. Code takes into account programming for ADF4351 and STM32, which are connected to each other:

```
   for (uint32_t freq = 100000000; freq
<= 1000000000; freq += 1000000) {
   ADF4351_set_frequency(freq);//freque
ncy
   HAL_Delay(10); // wait for Stabil PLL
   uint16_tadc_value                   =
HAL_ADC_GetValue(&hadc1);
   float              voltage          =
convert_adc_to_voltage(adc_value);}
```

In an RF jamming attack, the attacker first identifies the drone's active communication channel and then tunes into that frequency. RF jammers emit stronger signals on the same frequency as the target, overwhelming the receiver and preventing it from decoding legitimate signals [24]. Figure 7 illustrates a RF jamming attack scenario. Swarm drones are working through the protocols [9] and this system hacks not only main drone, but also auxiliary drones among the swarm.

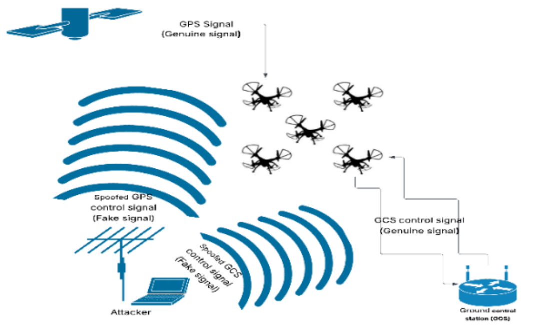


Figure 7: Schematic realization of jamming and spoofing attacks.

CONCLUSIONS

A new drone control platform has been developed which differs from famous by capability of controlling a swarm of unmanned aerial vehicles with the ability to simultaneously control different models of drones. During the research it was hacked 5 swarm drones, which take into account 1 main and 4 auxiliary.

The description of the wiring diagram of a device for intercepting and repurposing an enemy drone consists of three important components: a signal interception circuit, a reflashing circuit, and a signal jamming circuit. It is enough to use possible range in free space of 3.12 km and range of 6.35 km for hacking the swarm of drones.

For most modern drones from well-known manufacturers (e.g. DJI), it is possible to remotely update the firmware using official applications and

OTA technologies. However, the drone must be connected to the Internet via a controller or mobile device for a successful reprogramming, and care should be taken to avoid the risk of data loss or damage to the device during the update.

REFERENCES

[1] O. Lutskyi, "Tactics of using unmanned aerial vehicles in the protection of the state border: a textbook," Khmelnytskyi: NADPSU Publishing House, 2023, 164 p.

[2] D. Shamanov and A. Sorokin, "Analysis of modern methods of electronic warfare," Kharkiv National University of Radio Electronics, Control, Navigation and Communication Systems, no. 1, pp. 211-214, Kharkiv, Ukraine, 2024, doi: 10.26906/SUNZ.2024.1.211.

[3] I. Mayboroda, K. Vlasov, and M. Glushchenko, "Application and Prospects for the Development of Mobile Electronic Intelligence Means of the Tactical Link of the Security and Defense Forces of Ukraine," National Academy of the National Guard of Ukraine, Journal "Honor and Law", vol. 2, no. 89, 2024, doi: 10.33405/2078-7480/2024/2/89/309207.

[4] S. Mosov, "Swarming of Military Drones: Realities and Prospects," Institute of Public Administration and Civil Defense Research, no. 1(80), pp. 77-86, 2024, [Online]. Available: https://doi.org/10.33099/2304-2745/2024-1-80/77-86.

[5] O. Karatanov, O. Ustymenko, M. Jena, E. Bova, and V. Kalashnikova, "The use of swarm intelligence algorithms in the design of control systems for groups of unmanned aerial vehicles," Young scientist, vol. 10, no. 98, pp. 98-103, 2021, [Online]. Available: https://doi.org/10.32839/2304-5809/2021-10-98-24.

[6] K. Klyarskyi, "Quadrocopter control system: bachelor's thesis," 151 Automation and computer-integrated technologies, Kyiv, 2024, 100 p.

[7] G. Howell, J. M. Franklin, V. Sritapan, M. P. Souppaya, and K. Scarfone, "Guidelines for Managing the Security of Mobile Devices in the Enterprise," National Institute of Standards and Technology, Gaithersburg, MD, NIST Special Publication (SP) NIST SP 800-124r2, 51 p., May 2023, [Online]. Available: https://doi.org/10.6028/NIST.SP.800-124r2.

[8] F. Mahfoudhi, A. K. Sultania, and J. Famaey, "Over-the-Air Firmware Updates for Constrained NB-IoT Devices," Sensors, vol. 22, no. 19, p. 7572, 2022, doi: 10.3390/s22197572.

[9] P. Hou et al., "Distributed DRL-Based Intelligent Over-the-Air Computation in Unmanned Aerial Vehicle Swarm-Assisted Intelligent Transportation System," IEEE Internet of Things Journal, vol. 11, no. 21, pp. 34048-34054, 2024, doi: 10.1109/JIOT.2024.3418882.

[10] M. Sadiku, N. Akujuobi, and O. Cajetan, "Software-defined radio: A brief overview," Potentials, IEEE, vol. 23, pp. 14-15, 2004, doi: 10.1109/MP.2004.1343223.

[11] B. R. Mahafza, S. C. Winton, and A. Z. Elsherbeni, "Handbook of Radar Signal Analysis," 1st ed., Chapman and Hall/CRC, 2021, 706 p., doi: 10.1201/9781315161402.

[12] Planeta hobby, "Tips for managing FPV frequencies," [Online]. Available: https://modelistam.com.ua/ua/sovety-upravleniyu-chastotami-a-300/.

[13] "RSSI verstehen: Eine Schlüsselmetrik für FPV-Systeme mit großer Reichweite DJI-Systeme im Vergleich zu analogen Systemen," [Online]. Available: https://fpvracingdrone.de/fpv/rssi-sender-empfaenger-grosse-reichweite-fpv/.

[14] "How to avoid intermodulation distortion and ensure a stable FPV video signal," [Online]. Available: https://fpvua.org/threads/jak-uniknuti-intermoduljacijnix-spotvoren-i-zabezpechiti-stabilnij-videosignal-fpv.361/.

[15] "The role of satellite interference in failures in modern communications," [Online]. Available: https://www.szmidjammer.com/uk/blog/satellite-jammer/.

[16] V. Chmeliov, "Radio control systems. Home control work, Study guide for students majoring in 172 'Electronic Communications and Radio Engineering'," Igor Sikorsky Kyiv Polytechnic Institute, Kyiv: Igor Sikorsky Kyiv Polytechnic Institute, 2023, 24 p.

[17] V. Savko, "Development of an ARI stand for machine-to-machine interaction: diploma thesis," Bachelor's degree: 121 Software Engineering, Kyiv, 2021, 75 p.

[18] P. Holodiaieva, "Algorithm for estimating the UAV motion parameters based on the Kalman filter: diploma thesis," Bachelor's degree: 172 Telecommunications and Radio Engineering, Kyiv, 2023, 78 p.

[19] M. Yatsymon, "Methods for increasing the protection of the control radio channel from interception by unmanned aerial vehicles: graduation qualification work: 125 'Cybersecurity'," National University "Chernihiv Polytechnic", Department of Cybersecurity and Mathematical Modeling, Chernihiv, 2023, 71 p.

[20] V. Pavlenko, "Investigation of the influence of intentional interference on UAV communication channels: diploma thesis," Bachelor's degree: 172 Telecommunications and Radio Engineering, 2023, 51 p.

[21] X. Jiang, L. Jiang, Y. Xiao, and C. Xiao, "Design of Wireless Communication System Based on nRF24L01," Advanced Materials Research, vol. 945-949, pp. 1756-1759, 2014, doi: 10.4028/www.scientific.net/AMR.945-949.1756.

[22] H. Wang, C. Li, and F. Wu, "Design and Implementation of Broad-Band Jamming Signal Source for GPS," Proceedings of the 2016 International Conference on Computer and Information Technology Applications, pp. 68-73, 2016, ISSN 2352-538X, doi: 10.2991/iccita-16.2016.13.

[23] J. Li, "Working principle and application analysis of phase-locked loop," Applied and Computational Engineering, vol. 11, pp. 174-180, 2023, doi: 10.54254/2755-2721/11/20230228.

[24] D. Debojyoti, N. R. Sivaraaj, and M. Abdul, "Exploring the Landscape of Phase-Locked Loop Architectures: A Comprehensive Review," IEEE Access, pp. 1-1, 2024, doi: 10.1109/ACCESS.2024.3446393.

Integrated Machine Learning Models for Bakery Product Defect Prediction

Nataliia Zaiets[1], Nataliia Lutska[2] and Lidiia Vlasenko[1]
[1]*Department of Automation and Robotic Systems, National University of Life and Environmental Sciences of Ukraine, Heroiv Oborony 12v, 03041 Kyiv, Ukraine*
[2]*Department of Automation and Computer Technologies of Control Systems, National University of Food Technologies, Volodymyrska 68, 01033 Kyiv, Ukraine*
z-n@ nubip.edu.ua, lutskanm2017@gmail.com, vlasenko.lidia1@gmail.com

Keywords: Defect Prediction, Machine Learning Models, Correction, Bakery Products.

Abstract: The paper discusses the development of a model for predicting the probability of occurrence of defects in bakery products using a set of input variables at different stages of the technological process. The model is based on the analysis of data including control variables, such as oven temperature and humidity, as well as disturbance variables characterizing the properties of flour, the dough preparation process and baking of products. Based on the results of the study, a GMM-based model was selected, which demonstrated the highest accuracy, with the achieved Precision and Recall values equal to 1.0 for the class of defective products, which indicates high correctness of forecasts. In terms of Log-Likelihood, the model demonstrated a large difference between the classes, which confirms its ability to accurately classify both defective and non-defective products. The proposed model is an effective tool for predicting defects and optimizing process parameters. It allows you to adjust control variables, such as temperature and humidity, to reduce the amount of defects, ensuring stability of product quality. The article also proposes different methods for adjusting the values of control variables based on historical data. This allows for optimization of the technological process and improvement of the quality of bakery products in real-time production conditions.

1 INTRODUCTION

The development of artificial intelligence, digitalization of all stages of production, implementation of digital twins, organization of processes according to the concept of intelligent production, requires the development and use of models of varying complexity and type. There is also a tendency to increase the complexity of emerging engineering and technical problems. Increasingly, for their accurate and successful solution, it is necessary to combine several different models. Therefore, integrated models are often used, providing solutions to complex problems based on a systems approach by combining several individual models into a single structure, taking into account the relationships between different components of the system [1].

The modern food industry is faced with a constant increase in product quality requirements, especially in the bakery industry. One of the key tasks is to minimize the defects of bakery products that occurs due to imperfections in the technological process, in particular at the baking stage. This stage is critically important, since the structure, taste, texture and appearance of the finished product depend on it. Real-time defect prediction and adaptive control of baking parameters are promising approaches to improving product quality and reducing defects. For this purpose, it is advisable to use defect models that allow you to analyze, predict, prevent the occurrence in time and reduce the total number of defects and/or discrepancies in production processes [2].

The combination of defect models in integrated models allows for local analysis of defects taking into account the interrelations existing in the existing production system. In particular, to identify the impact of changes in the quality indicators of raw materials, the operation of process equipment or working conditions on the total number of defects [3] or to analyze the systemic causes of defects through a multifactorial approach that takes into account technologies and external influences [4, 5]. The combination of these two approaches ensures effective quality control of final products and semifinished products, reduces losses, downtime, unplanned repairs and increases productivity.

The object of the study is the technological processes of manufacturing bakery products at the enterprise, including key variables of dough preparation, dough piece aging and baking, which affect the defects of the final product.

The subject of the study is methods for optimizing the technological process of preparing bakery products using machine learning models that allow predicting and monitoring product defects, automating the selection of technological parameters and adapting the process to changing production conditions.

The purpose of the study is to develop and substantiate approaches to optimizing the bakery production process using machine learning models, which ensures the minimization of defective products at the baking stage and increases the efficiency of production processes in general.

This paper discusses the development of a control system that uses machine learning to predict bakery product defects based on process characteristics and to correct control actions in order to minimize defects.

2 LITERATURE REVIEW

There are various models, methods and tools that help productions to identify and manage defects during their technological processes [6]. Combining several separate adequate models or approaches into a single structure provides a significant improvement in solving complex problems where it is necessary to take into account many different factors and relationships. This is confirmed by the results of research by scientists in various subject areas. This is due to the main characteristics of integrated models: multi-modeling, interdisciplinarity, adaptability, system approach, the ability to conduct various types of testing and simulations in complex systems.

Thus, Integrated Assessment Models (IAMs) determine hydrogen demand through the analysis of technology competition for energy services [7], but their low spatial and temporal resolution limits their application in global hydrogen trade. The presented framework combines IAMs with high-resolution models, optimizing the production, storage and transmission of hydrogen and electricity, allowing the exploration of future trade scenarios.

Integrated models incorporating multiple pharmacodynamic variables are used to support drug development, in particular, dosing optimization and study strategy in different patient subgroups. The authors [8] provide current examples of such models, key aspects and benefits of their application, and prospects for the development of individualized development of anticancer drugs. The use of integrated models in industry is actively developing. The study [9] focuses on an integrated planning model for electrical wiring manufacturers, which improves the synchronization of production processes and product inventories, ensuring uninterrupted supply for the automotive sector. The use of a model based on integer programming showed a significant reduction in lead time, which proves its effectiveness and the importance of collaborative planning in supply chains.

Integrated models are also used to analyze and determine optimal digital transformation strategies (Industry 4.0) in the automotive industry, taking into account technological, environmental and business aspects. They provide a comprehensive approach to assessing companies' readiness, developing new business models and taking into account sustainability in the process of implementing digital technologies [10]. Integrated models can be used to create a sustainable agri-food supply chain, combining cultivation, processing and distribution solutions to minimize costs and environmental impact. Using multivariate analysis, geospatial data and mathematical modeling, the study identifies optimal location, production and logistics strategies, ensuring profit and employment maximization while adapting to uncertain conditions [11].

Therefore, to improve the efficiency of defect tracking and processing, it is advisable to use integrated models that take into account data on various defects, their life cycle, historical data, the influence of external factors, available resources for working with defects, etc. Such modeling is used to determine the causes and evaluate defects, isolate failures and analyse their impact.

Defect modeling is used to identify failure causes and estimate defects by isolating faults and analyzing their impact. This approach significantly improves the accuracy of estimates, reduces the number of possible defects and the area of interest, allowing for effective problem detection even under limited data conditions [12].

Recent research in defect detection highlights the limitations of traditional image processing under complex textures and varying lighting conditions. As a result, deep learning methods are increasingly applied due to their ability to automatically extract features and handle complex data [13], [14]. These methods are effective in various quality control scenarios, including surface and X-ray image analysis [14]. However, challenges such as data imbalance, limited sample size, and the need for real-time processing remain relevant [13], [15]. Overall, the integration of non-contact technologies with ML-based methods represents a promising direction for developing intelligent inspection systems [15].

One of the powerful areas of work with defects is the use of anomaly detection methods [16]. Anomaly detection is a method and technique for detecting and identifying atypical or deviant events or observation data that do not correspond to the expected process. Such methods are widely used in data analysis, cybersecurity, predicting system failures, quality control, etc.

The main methods for detecting anomalies can be distinguished: statistical methods based on the analysis of data samples and the identification of outliers and deviations from the normal distribution [17]; methods based on time series for data that change over time [18]; methods based on distances [19] and density [20] for working with data in multidimensional space; Bayesian methods that estimate the probability of a point being abnormal based on cause-and-effect relationships in the data [21]; when working with network nodes, it is convenient to use graph-based methods [22]; when analyzing complex data structures, machine learning methods have proven to be the best for determining various deviations [23-25].

Therefore, the advantages of using integrated models include their ability to take into account the maximum number of interdependencies between the components of the system and the processes occurring in it; they also make more accurate forecasts and provide a better basis for decision-making. This also causes certain difficulties in working with these models: the need to use high-quality data for their development; high computational requirements; difficulties may arise when integrating heterogeneous models.

3 PREPARATION AND PRELIMINARY DATA ANALYSIS

In modern bakery production, a significant proportion of product defects are detected at the final stage - during baking. This is due to the fact that visual defects, such as uneven rise, excessive density or undesirable crust color, become noticeable only at the stage of thermal treatment. In such cases, the production process is forced to end without the possibility of correction, which leads to a loss of resources such as raw materials, energy and time. In addition, timely identification of defective products requires qualified specialists or automated visual inspection systems.

The article proposes an integrated approach to predicting defects at early stages of production, based on the characteristics of flour, dough and proofing process parameters. Using machine learning models will allow these characteristics to be analyzed to determine the probabilities of a defect even before the baking stage, allowing for early process termination or adjustments. This will not only improve production efficiency, but also minimize the costs associated with recycling or disposal of defective products.

3.1 Data Collection and Pre-Processing

To build a forecasting model, data were collected, including all measured technological product change portions at the stages of dough preparation, dough piece standing and baking. These data were obtained from automated production line control systems. At the stage of preliminary data processing, noise is removed, and values are normalized for subsequent use in machine learning models. For this, standard data processing methods are used, in particular, standardization and filling in missing values.

In addition to automated approaches to identifying emissions, each of the calculated emissions was additionally examined manually, taking into account the features of the technological process and the specifics of production. This made it possible to take into account the possible causes of emissions and correctly make decisions on their removal, correction or preservation in the sample for further analysis.

When selecting features for building the model, an individual approach was applied to each variable, regardless of the level of correlation with other features. Both statistical characteristics and the technological significance of each feature for predicting defects were taken into account. This made it possible to form a set of variables reflecting real physical and chemical processes at all stages of production.

The total dataset contains 321 samples, of which 48 are defective (class label $y = 1$) and 273 are non-defective (class label $y = 0$). This indicates a class imbalance, which was taken into account during model selection and evaluation.

Table 1 shows the selected variables, and Figure 1 shows their correlation matrix. Note that $x_1, \dots, x_4$ determine the initial quality of the raw material, which has a significant impact on the formation of the dough structure. For example, flour strength (x_3) correlates with the sugar-forming capacity of the dough during fermentation, and gas-forming capacity (x_2) is closely related to the rise of the dough. The group of variables $x_5, \ldots, x_7$ describes the main technological variables of dough mixing and fermentation. In particular, water consumption (x_5) affects the consistency of the dough, while titran acidity (x_7) reflects the activity of enzymatic processes and determines the taste. At the stage of proofing of dough pieces, two variables were selected: humidity and weight of the dough piece (x_8, x_9), which have a direct impact on the plasticity and readiness of the dough for baking. The high correlation of these variables with the quality of the finished product allows them to be used as predictors. Baking variables such as temperature and humidity in the humidification zone (x_{10}, x_{11}), have a significant impact on crust formation, baking and bread color. Deviations in these variables often result in defects.

Table 1: Input and output variables of the model.

Raw material or process	Variable notation	Variable decryption
Properties of flour	x_1 (z_1)	Quantity and quality of gluten
	x_2 (z_2)	Gas-forming capacity
	x_3 (z_3)	Flour strength
	x_4 (z_4)	Sugar-forming capacity
Preparing the dough	x_5 (z_5)	Water consumption
	x_6 (z_6)	Dough fermentation time
	x_7 (z_7)	Titric acidity of the dough
Proofing the dough	x_8 (z_8)	Dough piece moisture content
	x_9 (z_9)	Dough piece weight
Baking	x_{10} (u_1)	Humidity in the oven and in the humidification zone
	x_{11} (u_2)	Baking temperature
	y	Defect [0, 1] (0 – no defect, 1 – defect)

The mathematical model is represented by the following (1):

$$y = h(\overline{x}), \overline{x} = [x_1, x_2, \ldots, x_{11}], \quad (1)$$

where $h()$ – model (hypothesis); y, x_i – the inputs and output of the model, respectively (Table 1).

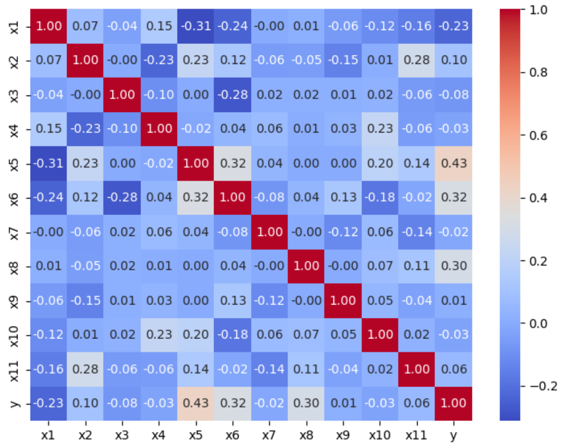


Figure 1: Visualization of process variable correlations.

3.2 Selection of Model Construction Methods

Due to the small amount of data with defective products ($y = 1$), the model is based on data only for non-defective products ($y = 0$), focusing on identifying patterns in this data. Consider the parallel coordinates plot (Fig. 2) showing the change in input variables for defective and non-defective products. It is clear that the values of some variables have stable levels, while others vary significantly, but no patterns are observed. Figure 3 shows that the non-defective product data is distributed compactly and without clearly visible clusters. This indicates that: dimensionality reduction methods (PCA) can be used to eliminate correlated variables.

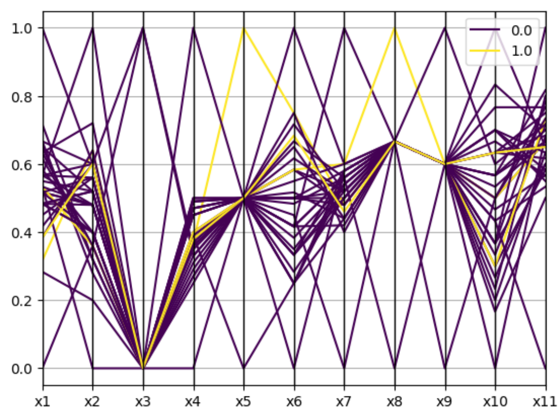


Figure 2: Parallel coordinates graph.

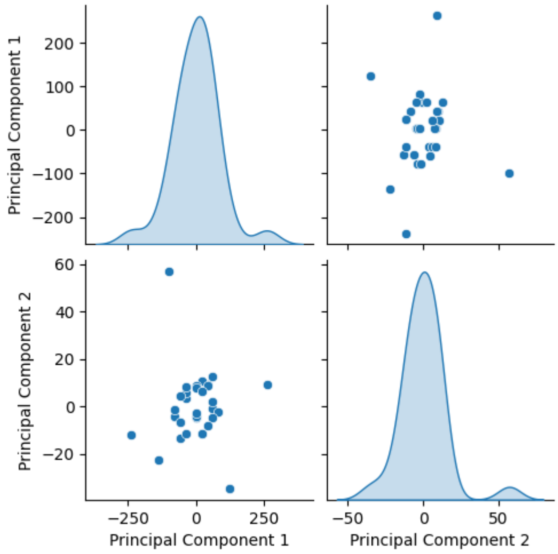


Figure 3: Principal component distribution densities for $y = 0$.

Thus, several methods can be proposed for constructing a model that do not require clear clusters or explicit boundaries for separation:

1) Gaussian Mixture Model (GMM) –is one of the best methods for detecting anomalies when data can be represented as a mixture of several normal distributions.
2) Isolation Forest – It is an effective algorithm for anomaly detection that works particularly well in high-dimensional spaces and for identifying anomalies even when data clusters are not obvious.
3) One-Class SVM (Support Vector Machine) – The method constructs a hyperplane separating "normal" data from anomalous data.
4) Local Outlier Factor (LOF) – the method calculates the relative density of a point relative to its neighbors. If a point has a much lower density than its neighbors, it will be considered an anomaly. This works well when the anomalies are local.
5) k-Nearest Neighbors (k-NN) can be used to detect anomalies by comparing the distances between points. If a point is too far from its nearest neighbors, it can be considered an anomaly.
6) Autoencoders – is a type of neural network that learns to reconstruct input data. It is suitable for working with multidimensional data, when there are complex relationships between features.

The Gaussian Mixture Model (GMM) is an optimal choice for this task because it can effectively model complex data distributions where there are no clear clusters or boundaries between normal and abnormal points. By being able to describe data as a mixture of several normal distributions, GMM can detect abnormal points that deviate from the underlying distribution. This is especially useful in situations where the number of defective points is limited and where traditional methods based on clear separations or clustering may be less effective. GMM can work with small amounts of data and accurately detect anomalies, making it the most suitable method for this task. In addition, the choice of GMM for this task is determined by the specifics of the process and the nature of the data. In particular, GMM is able to effectively work with multidimensional data that have a complex structure with possible latent groups. At the preliminary analysis stage, the distribution of input features showed the presence of mixed distributions, which further confirmed the feasibility of using GMM. For comparison, other methods were also tested - One-Class SVM, Isolation Forest and ensemble methods (for example, Random Forest for classification). GMM demonstrated high accuracy and stability with a small amount of defective data, which is important for production conditions, where defects are rare events.

To make a rational choice, metrics such as the difference in logarithmic likelihoods between individual Log-Likelihood classes and precision/recall on test data will be used to evaluate the quality of forecasting and hyperparameters of the GMM model.

After building and testing the model in laboratory conditions, the system will be tested in real bakery enterprises. Testing includes collecting additional data and adjusting the models to ensure their effective application in real production conditions.

4 MODELING OF ANOMALIES DETECTION AND COMPARISON OF RESULTS

The modeling process was performed using the Python programming language, using the pandas, matplotlib, seaborn and sklearn libraries. Colaboratory (Google) was chosen as the programming environment.

The process of tuning the GMM hyperparameters (the number of components n_components and the type of covariance matrix) was performed based on cross-validation with optimization according to the criterion of minimizing the logarithmic likelihood

(Log-Likelihood) for non-rejected data. Analysis of the test models showed that for 4 components with a full covariance matrix (covariance_type='full') the best balance between sensitivity and specificity of the model is achieved. This choice is consistent with the physical nature of the process, where each component can correspond to a separate technological state or a group of production conditions.

Based on the preliminary analysis, as well as through modeling, the following models were selected:

- Model 1: GMM by 2 principal components (PCA) with the number of Gaussian components n_components=5 and with the type of covariance matrix covariance_type= 'full';
- Model 2: GMM by 2 principal components (PCA) with the number of Gaussian components n_components=7 and with the type of covariance matrix covariance_type= 'full';
- Model 3: GMM for all features with the number of Gaussian components n_components=3 and with the type of covariance matrix covariance_type= 'full';
- Model 4: GMM for all features with the number of Gaussian components n_components=4 and with the type of covariance matrix covariance_type= 'tied';
- Model 5: GMM for all features with the number of Gaussian components n_components=4 and with the type of covariance matrix covariance_type= 'full'.

For each model, an optimal wear threshold was selected based on the logarithmic probability estimate of training and test points. This allows to determine the most optimal threshold for detecting anomalies in unknown data.

Table 2 shows the simulation results for different model configurations in terms of Precision and Recall metrics, as well as the minimum Log-Likelihood values. Model 1 shows a medium balance between prediction accuracy and anomaly detection ability. Precision is 0.8, indicating that most of the predicted anomalies are correct, but Recall at 0.34 indicates that most of the anomalies were not identified. The suitable Log-Likelihood range shows a small gap between the classes, which explains the low Recall score. Models 2 and 3 show ideal Precision values (1.0), but Recall for Model 2 remains unchanged (0.34), while Recall increases to 0.67 for Model 3. This indicates an improvement in anomaly detection while maintaining accuracy. It is worth noting that Model 3 has a significantly lower minimum Log-Likelihood value for class y=0 than Model 2, which contributes to better anomaly classification. Model 4 maintains the results of Model 3 for Precision and Recall, but has an even lower minimum Log-Likelihood value for class y=0, which may indicate an improved ability of the model to separate classes. Model 5 shows the best results with Precision and Recall at 1, which means that all the model's predictions were correct and all anomalies were detected. It is important to note that this model has a lower Log-Likelihood value for class y=1 than the previous models, which contributed to its high accuracy.

The analysis shows that the key factors for successful modeling are the optimal choice of parameters to ensure the maximum gap between the Log-Likelihood values for the classes. Model 5 is the best, but its behavior should be further assessed on a larger data sample.

Table 2: Simulation results.

Model	Precision/ Recall	Normalized Log-Likelihood (min(y=0) / max(y=1))
1	0.8 / 0.34	0.12 / 0.15
2	1 /0.34	0.13 / 0.15
3	1 / 0.67	0.06 / 0.15
4	1 / 0.67	-0.14 / 0.01
5	1 / 1	0.17 / 0.05

Figure 4a shows a scatter plot with normalized probabilities based on two principal components (PCA), showing how normal and abnormal points are classified based on the GMM method. Normal points are preferentially located in the same region of the principal component space, and their probabilities (estimates based on the GMM) are normalized and represented by a color scale. Abnormal points have significantly lower log-probability values, corresponding to their separation from normal points. They differ from normal points in the principal components, indicating their deviation from the expected behavior.

Figure 4b shows the histogram of the log-likelihood distribution for the two classes. Normal points have log-likelihood values between 0 and -5, with most values concentrated near values close to zero (indicating a high probability of belonging to the normal class). Abnormal points have significantly lower log-likelihood values, confirming their deviation from normal behavior. This distribution clearly shows that the model separates abnormal points from normal ones, since their log-likelihoods are very different.

Figure 4c shows the test results, which are the log-likelihoods for each point for two classes: y=0 (normal) and $y = 1$ (abnormal). For the $y = 0$ class,

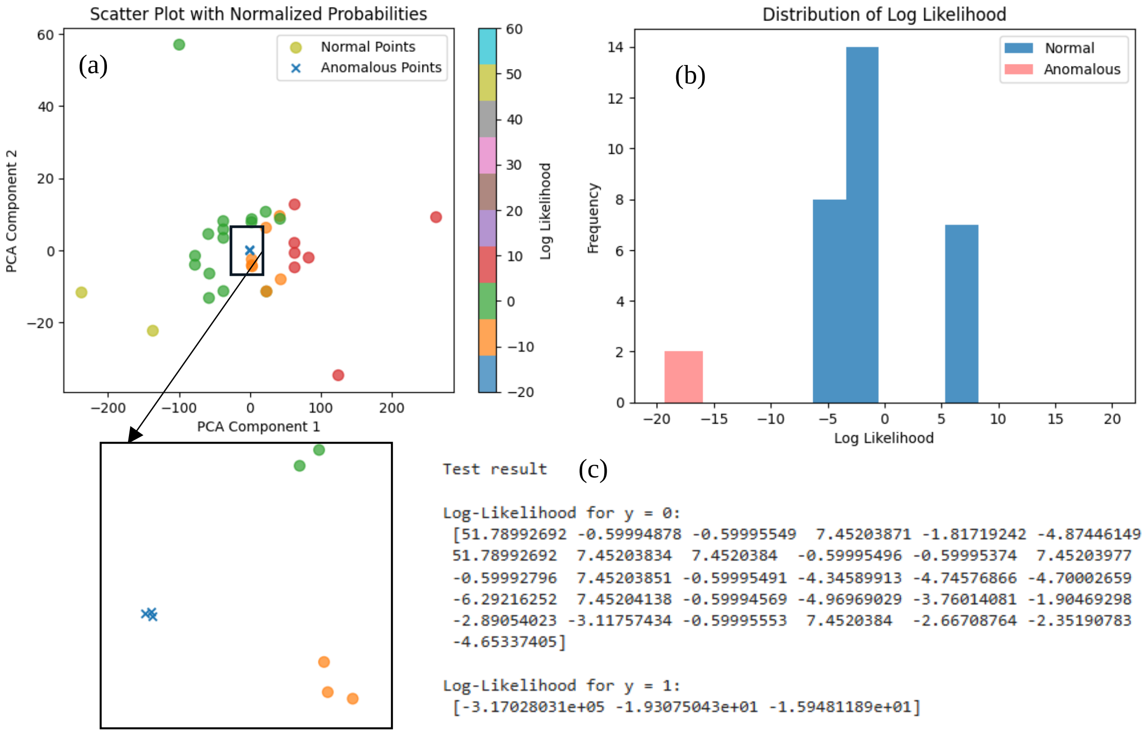


Figure 4: Simulation results for Model 5 on test data: (a) Scatter plot with log-likelihoods; (b) Log-likelihood distribution (nearest scale); (c) Test results.

the log-likelihood values range from -6 to 0. This indicates that normal points have probability values that are not very low, which correlates with a high probability of their belonging to the normal category. For the $y = 1$ class, the log-likelihood value is significantly lower, indicating that these points have a low probability of being part of the normal population, confirming their abnormality.

5 STRATEGY FOR DEFECT REDUCTION

Based on the constructed model for predicting defects in bakery products, a control strategy is developed that, by changing the set values of the oven temperature and humidity regulators, allows reducing the amount of defects (Fig. 5). The developed model based on the current values of the input data $\overline{x}$ (control variables u_1, u_2 and disturbance variables $z_1, \ldots, z_9$), before the baking stage, provides the state of the original product (not defective/defective) and if y=1 is predicted, then the control variables u_1, u_2 are corrected. If such values are not found, then a message is generated about the need for changes at the previous stages of the technological process. In particular, at the stages of dough preparation and baking by changing the process mode variables z_5, z_6, z_7 and z_8, z_9, respectively.

The correction of the control variables u_1, u_2 can be implemented through various approaches depending on the available resources. One option is to use historical data to find cases with similar values of the disturbance variables $z_1, \ldots, z_9$, where y=0 was predicted, and apply the corresponding values of u_1, u_2 for correction according to the nearest neighbor principle. Another approach is to build a regression model studying the relationship between u_1, u_2 and the disturbance variables at y=0, and use this model to predict the optimal values of the control variables. Numerical optimization can also be used to find the values of u_1, u_2 that minimize the risk of defects by modeling and selecting the control variables.

Another approach is to use inverse modeling based on an existing GMM, where feasible values of u_1, u_2 are selected by gradually varying the input variables until the forecast y=0 is achieved. Alternatively, iterative simulation can be performed using the main model, gradually changing the value of u_1, u_2 until the forecast y=0 is achieved. Heuristic rules developed based on expert experience or data analysis can be effective, especially in complex or dynamic settings.

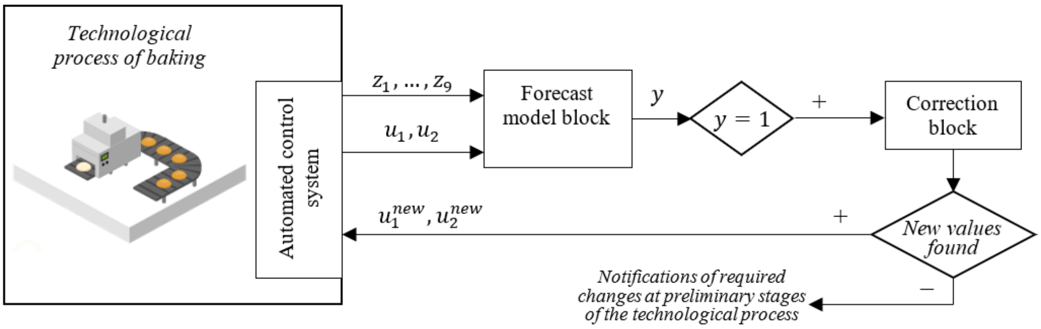


Figure 5: Structural diagram of the baking process correction strategy.

The application of both approaches – defect prediction and control variable correction – is carried out through an integrated system that can operate in real time. Machine learning models predict the probability of defects based on current values of process parameters, and the correction system changes the baking conditions to minimize defects.

6 CONCLUSIONS

As a result of the conducted research, a model for predicting defects in bakery products was developed, which uses a set of input variables to predict the probability of defects at different stages of the technological process. The accuracy of the model was 1.0 for Precision and Recall, which confirms the correctness of all defect predictions and the absence of missed defects, and the model is able to correctly identify all defective products. In terms of Log-Likelihood, the largest difference between the two classes is obtained, indicating a high probability of accurate predictions for both classes and confirming the adequacy of the model. Therefore, given the results obtained, Model 5 is the most effective in predicting defects, able to correctly classify both missing and defective products, making it suitable for application in manufacturing process conditions to reduce defects.

The model is part of a system that allows for the correction of control variables, such as oven temperature and humidity, based on current process disturbance data. Predicting defects and optimizing control variables help reduce the amount of defective products and ensure the stability of bakery product quality. The paper also considered various approaches to correcting control variables, in particular, numerical optimization methods, heuristic rules, and the use of nearest neighbor search algorithms to predict and correct values based on historical data. Thus, the proposed management strategy is an effective tool for improving product quality and optimizing the technological process in bakeries. The results of the study can be useful for implementation in industrial conditions in order to minimize defects and ensure high quality of the final product.

REFERENCES

[1] H. J. Bungartz, S. Zimmer, H. Buchholz, and D. Pfluger, "Modeling and simulation," Springer Undergraduate Texts in Mathematics and Technology, vol. 10, 2014, [Online]. Available: https://doi.org/10.1007/978-3-642-39524-6.

[2] F. Beaudoin, K. Sanchez, and P. Perdu, "Dynamic laser stimulation techniques for advanced failure analysis and design debug applications," Microelectronics Reliability, vol. 47, no. 9-11, pp. 1517-1522, 2007, [Online]. Available: https://doi.org/10.1016/j.microrel.2007.07.054.

[3] F. Lin, T. Jia, R. Y. Fung, and P. Wu, "Impacts of inspection rate on integrated inventory models with defective items considering capacity utilization: Rework-versus delivery-priority," Computers & Industrial Engineering, vol. 156, p. 107245, 2021, [Online]. Available: https://doi.org/10.1016/j.cie.2021.107245.

[4] A. Mondal, A. K. Halder, S. Nayak, R. R. Kumar, A. Chakraborty, S. K. Gudimetla, and M. Dubois, "Root cause analysis of indentation mark defect on Zn coated steel sheet in continuous galvanizing line," Engineering Failure Analysis, vol. 157, p. 107883, 2024, [Online]. Available: https://doi.org/10.1016/j.engfailanal.2023.107883.

[5] C. Chen, C. Liu, T. Wang, A. Zhang, W. Wu, and L. Cheng, "Compound fault diagnosis for industrial robots based on dual-transformer networks," Journal of Manufacturing Systems, vol. 66, pp. 163-178, 2023, [Online]. Available: https://doi.org/10.1016/j.jmsy.2022.12.006.

[6] C. Gao, B. Cai, Y. Zhang, X. Shao, C. Yang, and L. Gao, "A life cycle reliability testing and assessment method for deepwater oil and gas equipment systems," Ocean Engineering, vol. 311, part 1, p. 118928, 2024, [Online]. Available: https://doi.org/10.1016/j.oceaneng.2024.118928.

[7] D. Mathews, M. Brinkerink, and P. Deane, "A framework for high resolution coupled global electricity & hydrogen models based on integrated assessment model scenarios," International Journal of Hydrogen Energy, vol. 97, pp. 516-531, 2025, [Online]. Available: https://doi.org/10.1016/j.ijhydene.2024.11.077.

[8] H. Liu, E. I. K. Ibrahim, M. Centanni, C. Sarr, K. Venkatakrishnan, and L. E. Friberg, "Integrated modeling of biomarkers, survival and safety in clinical oncology drug development," Advanced Drug Delivery Reviews, vol. 216, p. 115476, 2025, [Online]. Available: https://doi.org/10.1016/j.addr.2024.115476.

[9] I. Safra, K. Ghachem, F. Benabdallah, H. Albalawi, and L. Kolsi, "Integrated operations planning model for the automotive wiring industry," Heliyon, vol. 10, no. 11, p. e31820, 2024, [Online]. Available: https://doi.org/10.1016/j.heliyon.2024.e31820.

[10] Ö. F. Görçün, A. R. Mishra, A. Aytekin, V. Simic, and S. Korucuk, "Evaluation of Industry 4.0 strategies for digital transformation in the automotive manufacturing industry using an integrated fuzzy decision-making model," Journal of Manufacturing Systems, vol. 74, pp. 922-948, 2024, [Online]. Available: https://doi.org/10.1016/j.jmsy.2024.05.005.

[11] D. Aghajani, H. Seraji, H. Kaur, and J. Vilko, "A sustainable integrated model for multi-objective planning of an agri-food supply chain under uncertain parameters: A case study," Computers & Chemical Engineering, vol. 188, p. 108766, 2024, [Online]. Available: https://doi.org/10.1016/j.compchemeng.2024.108766.

[12] A. Li, H. Bian, B. Liu, Y. Li, and Y. Zhao, "A general defect modelling and simulation-assisted approach for fault isolation in failure analysis," Microelectronics Reliability, vol. 139, p. 114805, 2022, [Online]. Available: https://doi.org/10.1016/j.microrel.2022.114805.

[13] Y. Lin, J. Ma, Q. Wang, and D. W. Sun, "Applications of machine learning techniques for enhancing nondestructive food quality and safety detection," Critical Reviews in Food Science and Nutrition, vol. 63, no. 12, pp. 1649-1669, 2023.

[14] A. Saberironaghi, J. Ren, and M. El-Gindy, "Defect detection methods for industrial products using deep learning techniques: A review," Algorithms, vol. 16, no. 2, p. 95, 2023.

[15] Z. Jia, M. Wang, and S. Zhao, "A review of deep learning-based approaches for defect detection in smart manufacturing," Journal of Optics, vol. 53, no. 2, pp. 1345-1351, 2024.

[16] I. Elía and M. Pagola, "Anomaly detection in Smart-manufacturing era: A review," Engineering Applications of Artificial Intelligence, vol. 139, Part B, p. 109578, 2025, [Online]. Available: https://doi.org/10.1016/j.engappai.2024.109578.

[17] Z.-H. Jiao and X. Shan, "Consecutive statistical evaluation framework for earthquake forecasting: Evaluating satellite surface temperature anomaly detection methods," Journal of Asian Earth Sciences: X, vol. 7, p. 100096, 2022, [Online]. Available: https://doi.org/10.1016/j.jaesx.2022.100096.

[18] N. Lutska, N. Zaiets, and L. Vlasenko, "Development of Diagnostic System for the State of Electric Drives of Food Enterprise," Przegląd elektrotechniczny, vol. 2024, no. 2, pp. 164-167, 2024, [Online]. Available: https://doi.org/10.15199/48.2024.02.33.

[19] J. Pang, D. Liu, Y. Peng, and X. Peng, "Temporal dependence Mahalanobis distance for anomaly detection in multivariate spacecraft telemetry series," ISA Transactions, vol. 140, pp. 354-367, 2023, [Online]. Available: https://doi.org/10.1016/j.isatra.2023.06.002.

[20] A. Oluwasegun and J.-C. Jung, "A multivariate Gaussian mixture model for anomaly detection in transient current signature of control element drive mechanism," Nuclear Engineering and Design, vol. 402, p. 112098, 2023, [Online]. Available: https://doi.org/10.1016/j.nucengdes.2022.112098.

[21] K. Wang, C. Liu, and Y. Lu, "Ensemble Bayesian Network for root cause analysis of product defects via learning from historical production data," Journal of Manufacturing Systems, vol. 75, pp. 102-115, 2024, [Online]. Available: https://doi.org/10.1016/j.jmsy.2024.06.001.

[22] C. Song, L. Niu, and M. Lei, "A Brief Survey on Graph Anomaly Detection," Procedia Computer Science, vol. 242, pp. 1263-1270, 2024, [Online]. Available: https://doi.org/10.1016/j.procs.2024.08.145.

[23] G. Czibula, I.-G. Chelaru, I. G. Czibula, and A.-J. Molnar, "An unsupervised learning-based methodology for uncovering behavioural patterns for specific types of software defects," Procedia Computer Science, vol. 225, pp. 2644-2653, 2023, [Online]. Available: https://doi.org/10.1016/j.procs.2023.10.256.

[24] N. Zaiets, N. Lutska, L. Vlasenko, and A. Zhyltsov, "Forecasting Breakdowns of Electric Motors of a Sugar Factory Using Machine Learning Methods," 2023 IEEE 4th KhPI Week on Advanced Technology (KhPIWeek), pp. 1-6, Nov. 2023, [Online]. Available: https://doi.org/10.1109/KhPIWeek61412.2023.10312974.

[25] A. Syamsuddin, A. C. Adhi, A. Kusumawardhani, T. Prahasto, and A. Widodo, "Predictive Maintenance Based on Anomaly Detection in Photovoltaic System Using SCADA Data and Machine Learning," Results in Engineering, p. 103589, 2024, [Online]. Available: https://doi.org/10.1016/j.rineng.2024.103589.

Application of X-bar R Control Charts for Process Efficiency Monitoring: A Data-Driven Approach in Quality Management

Aleksy Kwilinski[1,2], Maciej Kardas[2] and Nataliia Trushkina[3]
[1]*The London Academy of Science and Business, Mary At Hill Str. 6-7, EC3R 8EE London, England*
[2]*WSB University, 1C Ciplaka Streeet, 41-300 Dabrowa Gornicza, Poland*
[3]*Research Center for Industrial Problems of Development of the National Academy of Sciences of Ukraine, Inzhenernyi Lane 1a, 61165 Kharkiv, Ukraine*
a.kwilinski@london-asb.co.uk, kardas.maciej@wp.pl, Corresponding author: nata_tru@ukr.net

Keywords: Lean Six Sigma, Lean Manufacturing, Six Sigma, Continuous Improvement, Statistical Process Control (SPC), Control Charts, X-bar R, Variation, Adhesion, Thermoplastic Polyurethane (TPU) Foil.

Abstract: Ensuring process efficiency and product quality remains a critical challenge in modern manufacturing, necessitating the implementation of robust methodologies for process monitoring and optimisation. Lean Six Sigma (LSS), which integrates Lean Manufacturing and Six Sigma principles, is widely adopted to enhance productivity while minimising process variability and waste. A key component of LSS is Statistical Process Control (SPC), which employs control charts to assess process stability and compliance in real time. Despite extensive research on SPC applications, existing studies often fail to systematically differentiate common cause variation from special cause variation and to identify their critical sources in industrial processes. Addressing this gap, the present study evaluates the effectiveness of X-bar R control charts as a data-driven methodology for identifying process inefficiencies. Using Minitab Statistical Software, the study analyses the adhesion parameter of Thermoplastic Polyurethane (TPU) film, a material widely used for electronic screen protection. The methodology involves constructing X-bar R control charts to monitor variability patterns, establish stability thresholds, and pinpoint critical sources of process deviations. The findings demonstrate that X-bar R control charts provide a robust framework for differentiating process variations, facilitating targeted corrective actions to enhance process stability. This research highlights the importance of statistical modelling in industrial decision-making, particularly within automated manufacturing environments. A key contribution of the study lies in its demonstration of the practical applicability of control charts in quality management and the integration of data-driven techniques for process control. Future research should investigate advanced machine learning-based SPC approaches to refine real-time decision-making and expand the applicability of control charts to dynamic and complex production systems. By reinforcing the role of statistical tools in quality engineering and operational excellence, this study contributes to the broader discourse on digital transformation in industrial process optimisation.

1 INTRODUCTION

Quality is an integral aspect in the proper and efficient functioning of production processes. In recent decades, a significant increase in its relevance has been observed [1], and multifaceted quality control has become a key element in maximising quality levels [2]. The described approach leads to building a market advantage for companies implementing activities within the described policy.

Of paramount importance is the need to combine activities aimed at improving quality, with a parallel increase in process efficiency [3]. Aiming to continuously improve productivity in the realised processes and increase the quality of manufactured products through continuous improvement, requires the implementation of specific methodologies such as Lean Six Sigma (LSS), which is an integration of Lean Manufacturing and Six Sigma. The general premise of the above-mentioned strategy is the elimination of waste, aimed at increasing productivity while reducing production costs [4-7].

It should be emphasised that in order to realise the assumptions that constitute the essence of successful LSS implementation, statistical process control (SPC), which involves the use of a wide range of indicator-based statistical methods to monitor and control process quality, becomes an extremely important aspect [8,9]. A fundamental tool within SPC are control charts, which enable the assessment of process compliance in terms of their stability and real-time performance [10-12]. The creator of control charts is considered to be Walter Shewhart [11]. Their primary purpose is to detect at an early stage undesirable trends, behaviour in the process area, which could be an initiating factor for defects or losses. The essence of the operation of control charts needs to be emphasised that they are not based on mathematical models, and the prediction of the further course of the process (future results), is based solely on process data obtained within the realised area.

The essential elements of the control chart are the boundary lines, which define the ranges of acceptable process variability. A distinction is made between: The central line, which represents the process average, as well as the upper (UCL) and lower control limits (LCL). Through simplicity of interpretation, when points on the control line fall outside the limits or when characteristic trends or other specific signals – common cause or special cause variation – are observed, it should indicate that the process is not performing as intended, and the consequence is to take intervention measures aimed at identifying and eliminating the factors creating the problem.

If, on the other hand, all measurements are within the control limits, such a situation should be interpreted as the fact that there is a common cause variation in the process, which is consistent, stable and predictable within a certain range of data, and actions aimed at reducing the variability of the process should focus on the area of identifying the sources of variability, by, among other things, understanding the cause-effect sequence of the analysed process.

Thus, for individual data, a single point outside the control limits on a Moving Range (MR) chart is classified as special cause variation, whereas for subgroup data, a point outside the control limits on a Range (R) chart indicates special cause variation. If all points remain within the control limits without significant trends or patterns, the process is considered to exhibit only common cause variation. These identification criteria are consistent with widely accepted control chart rules used to detect special cause signals.

Based on the classification according to quality characteristics, two types of control charts should be distinguished – those based on a numerical evaluation; for example: product sizes, their weight, etc. and control charts based on an alternative evaluation; for example, the number of defective pieces in a batch. An important point to note is that the most commonly used tools focused on monitoring variables when evaluating numerically are: X-means chart (sample averages are plotted), R chart (range values), S chart (standard deviation values) or S^2 chart (variance values). An example of a control chart is shown in Figure 1. UCL and LCL are included, which, based on process data, define the limits of process variability around the mean (CL).

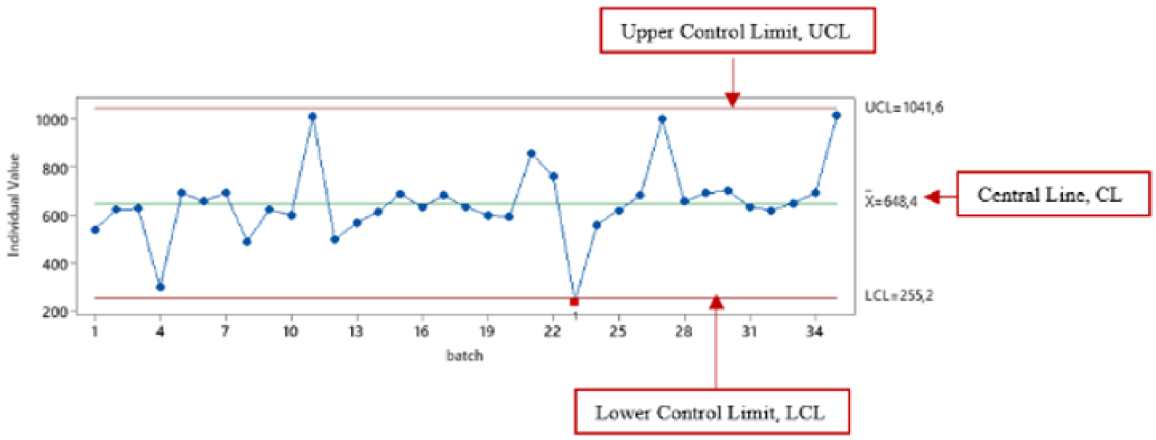


Figure 1: An example of a control chart the distribution of LCL, UCL and CL.

The ULC, LCL values determined on the control chart represent the limits of variation defined as 3 Sigma [13, 14]. The indicated value is primarily due to the fact that the use of 3 Sigma limits enables an effective distinction between common and special cause variation. In addition, it should be noted that this setting of control limits represents a balance between the risk of Type I and Type II error, i.e. over- and under-reaction to process variations. A key element is that the Sigma value is calculated differently from the classic standard deviation value, as this figure represents only its estimate. In addition, depending on the type of control chart, the Lower and Upper Control Limit values are calculated differently.

2 LITERATURE REVIEW

A literature review in the field of scientific publications shows many sources highlighting the importance of the use of control charts as tools used in the area of quality improvement, by minimising process variance [15,16].

In the article [17], the authors point out that statistical process control is a method of monitoring and improving a process through statistical analysis, with control charts as a key tool to detect changes that require intervention in the area of production processes. The publication [18] defines control charts as a basic statistical tool, and their simplicity in the area of design, using the solutions proposed by Shewhart, based on a statistical or economic criterion, significantly facilitate the interpretation of the process and the taking of specific corrective actions when we have specific signals to do so. The author [19], on the other hand, emphasises the versatility of the use of control charts by monitoring each production stage. Furthermore, the article defines control charts as the most important method of quality control.

In the article [20], the authors address issues related to the proper selection of control charts so that process control is as effective as possible; the proper identification of variability within an area of activity. The publication further provides an overview of the effectiveness of different types of control charts, such as Shewhart type and exponentially weighted moving range.

The publication [21] draws attention to the fact that the implementation of conventional Shewhart control charts can generate risks in the form of relatively longer time needed to identify small or medium process disturbances. An alternative solution to improve efficiency by modifying the control charts is therefore proposed. The importance of control charts as an essential tool in the area of statistical process control is also addressed in a publication [22]. The authors define their role in the maintenance layer and in improving the quality level of the process, and the correct use of the data is an essential element in terms of reducing variability.

Due to their significant added value, control charts are widely used in many industries [19,23]: medical [24-27], pharmaceutical [28], automotive [29], chemical [30,31] and others [32,33].

The analysis of the literature clearly indicates the validity of the use of control charts in terms of process monitoring. It should be emphasised that there is a research gap with regard to how to identify the type of variability (common cause or special cause variation) and the sources that create it. Based on the above, the research area of this publication is aimed at verifying the hypothesis that control charts X-bar R are an excellent tool for distinguishing variability, as well as defining the critical areas that are its source (common cause variation). An additional aim of the publication is to identify recommendations for strategies to deal with correctly defining types of variability and identifying their potential substrate. The activities are comprehensive and based on an analysis of the process from a cause and effect perspective.

3 METHODOLOGY

The study carried out concerned the presentation of X-bar R check sheets as a tool to identify the optimal approach, in order to identify the type of variability (common cause or special cause variation), as well as to define the potential areas constituting its source. This study included both quantitative and qualitative aspects. For the analysis of the collected process data, the statistical software Minitab 21.4.1 was used, which is an advanced statistical tool that enables a multidimensional yet in-depth analysis. It should be noted that Minitab effectively presents the aspect of diversity in terms of statistical techniques and their application in the experimental studies carried out. The software is centred around Automated Machine Learning for both binary and continuous responses, leveraging a Predictive Analytics Module [34]. Tailored for business-focused operations, it provides users with convenient methods to input statistical data, manipulate it, identify patterns and trends, and ultimately analyse the data to address real-world problems. Minitab streamlines data analysis, making it particularly suitable for statistical interpretation at the business level. It offers a variety of visual tools like histograms, boxplots, and scatterplots, enabling professionals to conduct statistical analysis more efficiently and derive insights from their data. Furthermore, it empowers users to compute descriptive statistics for their datasets.

Based on the research hypothesis, a case study focused on the supply control process of TPU (thermoplastic polyurethane) films was conducted. The adhesion parameter - the strength of the adhesive with which the film is coated - was investigated. This parameter is a key value for TPU films protecting the screens of electronic devices, so its constant control is very important.

The limitations of the inference in terms of the case study carried out may be due to several independent factors. Crucially, when investigating process variability, two lab technicians were used at the data collection stage and samples from four different material deliveries were analysed. Two samples were taken from each delivery, which were subjected to three independent measurements (each). The sampling tree created at the planning stage, together with a combination of X-bar R control charts, allows the sources of variability to be identified, as well as the area in terms of variability that most builds up the global variability of the process analysed.

In order to test the adhesion of the films, proper sample preparation becomes a key element. The material to be tested is cut in the form of strips 15 cm. long and 2.5 cm. wide. The film samples prepared in this way are glued onto 100 x 50 x 4 mm. laboratory float glass. (one batch of glass was used in the study). The adhesion force was measured using an Axis force meter, model FB on an automatic tripod, model STAH. The speed of the automatic gantry was a constant value - 20 mm./s for each run. Ambient conditions - 23.1°C ± 0.1; humidity - 60% ± 1.

4 RESEARCH RESULTS

The research carried out within the scope of this case study aims to find an answer in the area of the hypothesis, which was defined on the basis of the identified research gap. The area of results represents a broad spectrum of statistical analysis, based directly on quantitative analysis, enriched by elements of qualitative analysis. The numerical data on which the statistics focus are primary in nature.

The histogram presented in Figure 2 shows the total distribution of adhesion measurement data over three months. It should be clearly emphasised that the data included in the analysis includes the total variability of the process, which is a critical element for correct inference based on the results obtained. The lower limit of the adhesive force specification is 12N, while the value of the upper limit was set at 18N.

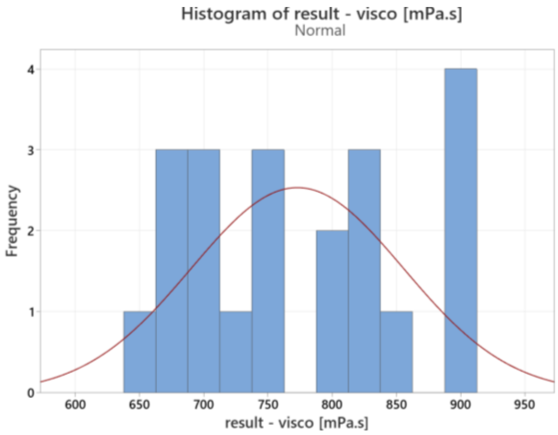


Figure 2: Histogram of the analysed values (three months).

When analysing Figure 2, it should be noted that some of the measurements obtained during the process are outside the accepted specification limits, which may pose a problem in terms of ensuring the expected quality. In addition, the distribution of the data presented in this chart reveals some differences in the context of a classical normal distribution, which should be particularly taken into account when determining further research directions and the tools to be used.

Figure 3; Time Series Plot, in contrast to Figure 2, corresponds to the distribution of the process data, taking into account their actual sequence. The data included in the Time Series Plot represent the absence of characteristic waveforms or strictly defined trends, and their arrangement is completely random, confirming that no strong external factor acted on the process, which could cause a serious problem within erroneous inference.

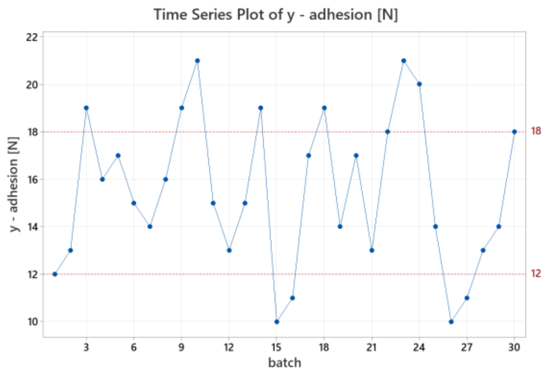


Figure 3: Time Series Plot of the analysed values (three months).

Based on the assumptions of the research study carried out, a sampling tree was prepared - Figure 4 - before proceeding to the data collection stage, which was analysed directly to define the type of variability dominating the process, as well as to identify its sources.

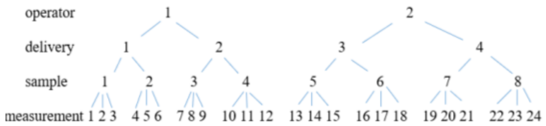


Figure 4: Sampling Tree; case study.

The sampling tree was developed with reference to the initial assumptions made - two operators in the process, four deliveries of material from which two samples were taken for testing, and each sample was measured three times.

The data acquired at the sampling stage reflects the total variability of the process and can therefore be used for further analysis. Based on the collected results, an X-bar R - Figure 5 control chart was prepared using Minitab, which enables an extensive statistical analysis to be carried out in the area of defining the parameters that constitute the purpose of this publication - determining the type of variability, as well as defining the source of variability dominating the process (operator to operator, delivery to delivery, sample to sample, measurement to measurement), taking into account the percentage of global variability. The interpretation of control charts for distinguishing the type of variation was based solely on identifying single points outside the control limits on the Range (R) chart in the analyzed case.

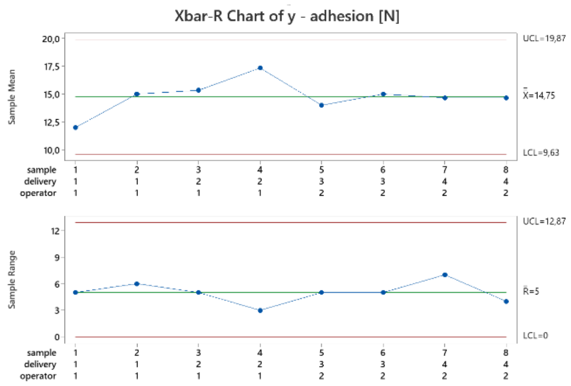


Figure 5: X-bar R control chart; adhesion.

Average of the range:

$$\bar{R} = \frac{\sum_{i=1}^{k} R_i}{k} = \frac{5+6+5+3+5+5+7+4}{8} = 5, \qquad (1)$$

where R_i is the range of the i-th subgroup and k is the number of subgroups.

Average of averages:

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{k} \bar{x}_i}{k} = \frac{12.00+15.00+15.33+17.33+14.00+15.00+14.67+14.67}{8} = 14.75\,, \qquad (2)$$

where $\bar{x}_i$ is the average of the i-th subgroup and k is the number of subgroups.

Upper Control Limit (UCL) and Lower Control Limit (LCL) for R chart:

$$UCL_R = D_4 \cdot \bar{R} = 2.575 \cdot 5 = 12.87, \qquad (3)$$

$$n < 7,\ LCL_R = 0, \qquad (4)$$

$$n \geq 7,\ LCL_R = D_3 \cdot \bar{R}. \qquad (5)$$

The value of the constant D_4 for the subgroup size k=3, is 2.575.

Upper Control Limit (UCL) and Lower Control Limit (LCL) for X-bar chart:

$$UCL_{\bar{x}} = \bar{\bar{x}} + A_2 \cdot \bar{R} = 14.75 + 1.023 \cdot 5 = 19.87, \qquad (6)$$

$$LCL_{\bar{x}} = \bar{\bar{x}} - A_2 \cdot \bar{R} = 14.75 - 1.023 \cdot 5 = 9.63. \qquad (7)$$

The value of the A_2 constant for the subgroup size k=3, is 1.023.

Determining the percentage distribution of individual sources of variability in terms of global variability makes it easy to identify which source of variability is the dominant source in the process and which one should be addressed first:

$$100\% = \hat{\sigma}^2{}_{total} = \hat{\sigma}^2{}_{measurement\ to\ measurement} + \hat{\sigma}^2{}_{sample\ to\ sample} + \hat{\sigma}^2{}_{delivery\ to\ delivery} + \hat{\sigma}^2{}_{operator\ to\ operator}. \quad (8)$$

Variability: measurement to measurement:

$$\hat{\sigma}^2{}_{measurement\ to\ measurement} = \left(\frac{\bar{R}}{d_2}\right)^2 = = \left(\frac{5}{1.693}\right)^2 \approx 8.72 \quad (9)$$

where d_2 is a constant value, which for a subgroup of 3 is 1.693. The value of $\bar{R}$ (5), is read from the R chart created by "wrapping" the sampling tree and calculating the new parameters.

Variability: sample to sample:

$$\hat{\sigma}^2{}_{sample\ to\ sample} = \left(\frac{\bar{R}}{d_2}\right)^2 - \frac{\hat{\sigma}^2{}_{measurement\ to\ measurement}}{3} = \left(\frac{1.5}{1.128}\right)^2 - \frac{8.72}{3} \approx -1.14 \quad (10)$$

where d_2 is a constant value, which for a subgroup of 2 is 1.128. The value of $\bar{R}$ (1.5), is read from the R chart created by "wrapping" the sampling tree and calculating the new parameters.

Since the obtained sample-to-sample variability score according to the above calculation reached a negative value, it is necessary to assume that it is 0.

Variability: delivery to delivery:

$$\hat{\sigma}^2{}_{delivery\ to\ delivery} = \left(\frac{\bar{R}}{d_2}\right)^2 - \frac{\hat{\sigma}^2{}_{sample\ to\ sample}}{2} - \frac{\hat{\sigma}^2{}_{measurement\ to\ measurement}}{3 \cdot 2} = \left(\frac{1.5}{1.128}\right)^2 - 0 - \frac{8.72}{6} \approx 0.32\ , \quad (11)$$

where d_2 is a constant value, which for a subgroup of 2 is 1.128. The value of $\bar{R}$ (1.5), is read from the R chart created by "wrapping" the sampling tree and calculating the new parameters.

Variability: operator to operator:

$$\hat{\sigma}^2{}_{operator\ to\ operator} = \left(\frac{\bar{R}}{d_2}\right)^2 - \frac{\hat{\sigma}^2{}_{delivery\ to\ delivery}}{2} - \frac{\hat{\sigma}^2{}_{sample\ to\ sample}}{2 \cdot 2} - \frac{\hat{\sigma}^2{}_{measurement\ to\ measurement}}{3 \cdot 2 \cdot 2} = \left(\frac{0.33}{1.128}\right)^2 - \frac{0.32}{2} - 0 - \frac{8.72}{12} = 0.086 - 0.16 - 0 - 0.73 \approx -0.80 \quad (12)$$

where d_2 is a constant value, which for a subgroup of 2 is 1.128. The value of $\bar{R}$ (0.33), is read from the R chart created by "wrapping" the sampling tree and calculating the new parameters.

Since the obtained sample-to-sample variability score according to the above calculation reached a negative value, it is necessary to assume that it is 0.

Based on calculations of the significance of individual types of sub-variability in terms of global variability, the dominant source is measurement to measurement variability - 96.46%, a small fraction is delivery to delivery variability - 3.54%, while the remaining sub-variabilities are not statistically significant.

5 DISCUSSION

The case study provides a valuable resource in terms of the use of X-bar R control charts and the direct benefits of their implementation. The statistical analysis provides important data on the approach to identifying the type of variability and which factor is key to shaping the holistic variability of the process. Distinguishing the types of variability has a fundamental impact on defining the right course of action in the context of variance reduction, while identifying the critical area that is most responsible for building total process variability is a fundamental aspect as an element that can lead to holistic variability reduction. It should be noted that any case study involves certain limitations that may result in incomplete objectivity in adequately representing the area of knowledge studied.

Publications [35-38] point out that the primary objective in terms of implementing control charts is to minimise the costs resulting from process control, as well as to increase the quality of products and services. On the other hand, the article [40] points out that X-bar and R chart are effective tools in detecting signals that may indicate that a process is out of control.

In the article [38], the authors emphasise that the X-bar R control charts is widely used in quality management systems, but attention should be paid to the likelihood of false alarms due to its direct limitations. Publications [35-38] indicate that X-bar, which are used to control the process mean, and R charts, which are responsible for controlling variance, do not have an adequate level of sensitivity to respond to small changes in process parameters. Instead, the authors [35] propose the use of the adaptive non-central chi-square statistic chart, which has a significantly higher level of performance compared to standard control charts.

A number of literature sources indicate a trend towards the development of control charts, aiming to develop robust tools that eliminate the risk of detecting signals lacking statistical significance. In [41], the authors highlight the fact that X-bar and R chart are extremely useful for controlling and detecting causes generating adverse effects in terms of process variability, pointing out that in order to increase the sensitivity of a standard control chart, it is necessary to include an auxiliary variable in the analysis. The content of the article [42] focused on the aspect of improving the effectiveness of signalling variance increases with respect to the classic R chart (Shewhart), and the values obtained showed that an R chart with variable parameters is more sensitive to variance increases.

Moreover, the findings of this study, particularly regarding the use of statistical tools for monitoring and controlling variations in industrial processes, align with the broader methodological approaches applied in other domains. For instance, research on electronic administration systems has highlighted the role of digitalisation in improving process efficiency and ensuring regulatory compliance, which is conceptually related to process control and quality assurance mechanisms in industrial settings [43]. Similarly, studies on anti-money laundering strategies in digital economies have demonstrated the importance of systematic risk detection and mitigation, paralleling the approach used in identifying variability sources within production processes [44]. In addition, cybersecurity research underscores the necessity of continuous monitoring and early detection of anomalies, reinforcing the importance of real-time statistical control in quality management [45; 46]. Furthermore, methodological advancements in experimental design and factorial analysis for quality management provide additional insights into improving the precision of variance detection [47]. Lastly, research on artificial intelligence in process optimisation suggests that machine learning-driven SPC techniques could enhance the predictive capabilities of control charts, providing a future direction for expanding the applicability of X-bar R charts in automated manufacturing environments [48].

The literature sources covered in this chapter highlight the importance of control charts, pointing to the potential for problems in terms of the proper detection of factors that generate small changes in process parameters (X-bar R). The analysis of the literature indicates that there is a lack of detailed description of how to proceed in the interpretation of X-bar R control charts in the context of defining the type of variation, identifying critical areas corresponding to potential sources of variation, as well as indicating the percentage distribution of partial variation on a global basis.

6 CONCLUSIONS

The statistical analysis carried out within the scope of this case study provides a broad spectrum of information in terms of the hypothesis set up based on the identified research gap. It should be mentioned that the analysis carried out indicates the importance of distinguishing common cause and special cause variation, as an overarching element that initiates actions aimed at minimising variability, and also includes an aspect related to the identification of the source or sources of partial variation, which are critical in terms of global process variability.

The article confirms the research hypothesis, highlighting the importance of using the control charts X-bar R. The R chart makes it possible to assess stability in the area of process variability. Furthermore, it provides important information by indicating the type of variability present; common cause or special cause variation. Correctly distinguishing the type of variation defines the subsequent workflow related to minimising the total variation. If special cause variation is identified in the process, it becomes necessary to take immediate, short-term action. In the case of common cause variation, on the other hand, it is necessary to first understand the cause-effect sequence of the process, an important aspect being then to define the areas from which the greatest variability arises. In identifying the area of variation, the X-bar plays an invaluable role, being responsible for monitoring changes at a central level, thus making it possible to pinpoint the origin of the source of variation, which in turn significantly facilitates the identification of the process factors that can have a decisive influence on its creation. X-bar and R-control charts are particularly useful in batch and continuous production, where continuous quality monitoring is key to maintaining standards and minimising waste.

Future research directions should focus on enhancing the effectiveness of statistical process control by integrating artificial intelligence and machine learning algorithms. The development of predictive SPC models based on historical data and real-time sensor inputs could provide more accurate anomaly detection and facilitate proactive decision-making. Additionally, expanding the application of X-bar R control charts to dynamic and adaptive manufacturing environments, such as Industry 4.0 and smart factories, would allow for a deeper understanding of process behaviour under varying conditions. Further interdisciplinary studies can explore the correlation between statistical process control and digital transformation trends, including blockchain-based traceability in supply chains and the use of cyber-physical systems for real-time quality monitoring. Moreover, addressing the limitations of traditional SPC methods by developing hybrid approaches that combine statistical techniques with computational intelligence can open new avenues for improving production efficiency and reducing defects. Lastly, applying control charts in non-industrial sectors, such as healthcare, finance, and environmental monitoring, could provide valuable insights into variability management and contribute to a broader adoption of SPC methodologies in diverse domains.

REFERENCES

[1] F. Blanco-Encomienda, E. Rosillo-Díaz, and J. Muñoz-Rosas, "Importance of Quality Control Implementation in the Production Process of a Company," European Journal of Economics and Business Studies, vol. 4, pp. 240-244, 2018.

[2] M. Zacharias, "The Importance of Quality Control for The Success of A Company," Asian Journal of Logistics and Management, vol. 1, no. 2, pp. 99-106, 2022.

[3] F. Psarommatis, S. Prouvost, G. May, and D. Kiritsis, "Product quality improvement policies in Industry 4.0: characteristics, enabling factors, barriers, and evolution toward zero defect manufacturing," Frontiers in Computer Science, vol. 2, p. 26, 2020.

[4] A. H. Gomaa, "A Systematic Review of Lean Six Sigma in Manufacturing Domain," Engineering Research Journal (Shoubra), vol. 52, no. 4, pp. 139-148, 2023.

[5] F. Olanrewaju, A. Chima Uzorh, and I. Nnanna, "Lean Six Sigma Methodology and Its Application in the Manufacturing Industry–A Review," American Journal of Mechanical and Industrial Engineering, vol. 4, no. 3, p. 40, 2019.

[6] N. Habidin, M. Salleh, N. Latip, M. Azman, and N. Fuzi, "Lean Six Sigma performance improvement tool for automotive suppliers," Journal of Industrial and Production Engineering, vol. 33, pp. 215-235, 2016.

[7] L. Minh, V. Ni, and D. Hien, "Continuous Improvement of Productivity and Quality with Lean Six-Sigma: A Case Study," Applied Mechanics and Materials, vol. 889, pp. 557-566, 2019.

[8] R. Khan, "Statistical Process Control," pp. 261-312, 2013.

[9] E. Bottani, R. Montanari, A. Volpi, L. Tebaldi, and G. Di Maria, "Statistical Process Control of assembly lines in a manufacturing plant: Process Capability assessment," Procedia Computer Science, vol. 180, pp. 1024-1033, 2021.

[10] P. Qiu, "Statistical process control charts as a tool for analyzing big data," Big and Complex Data Analysis: Methodology and Applications, pp. 123-138, 2017.

[11] S. Shah, P. Shridhar, and D. Gohil, "Control chart: A statistical process control tool in pharmacy," Asian Journal of Pharmacy, vol. 4, p. 184, 2010.

[12] S. Jamadar, "Statistical Process Control," in Operations Management in Healthcare, vol. 2, Springer, p. 88, 2020.

[13] H. Schneider, W. Kasperski, T. Ledford, and W. Kraushaar, "Control Charts for Skewed and Censored Data," Quality Engineering, vol. 8, pp. 263-274, 1995.

[14] R. Radhakrishnan and P. Balamurugan, "Six Sigma Based Control Chart for Fraction Defectives," Journal of Testing and Evaluation, vol. 39, pp. 678-681, 2011.

[15] J. Ng, "Statistical Process Control Chart as a Project Management Tool," IEEE Engineering Management Review, vol. 46, pp. 26-28, 2018.

[16] D. Montgomery and W. Woodall, "A Discussion on Statistically-Based Process Monitoring and Control," Journal of Quality Technology, vol. 29, p. 121, 1997.

[17] H. Nazir, M. Riaz, and R. Does, "Robust CUSUM Control Charting for Process Dispersion," Quality and Reliability Engineering International, vol. 31, pp. 369-379, 2015.

[18] E. Saniga, "Economic Statistical Control-Chart Designs With an Application to and R Charts," Technometrics, vol. 31, pp. 313-320, 1989.

[19] J. Li, "Application of Statistical Process Control in Engineering Quality Management," IOP Conference Series: Earth and Environmental Science, vol. 831, 2021.

[20] A. Ali, T. Mahmood, H. Nazir, I. Sana, N. Akhtar, S. Qamar, and M. Iqbal, "Control Charts for Process Dispersion Parameter under Contaminated Normal Environments," Quality and Reliability Engineering International, vol. 32, pp. 2481-2490, 2016.

[21] S. Psarakis, "Adaptive Control Charts: Recent Developments and Extensions," Quality and Reliability Engineering International, vol. 31, pp. 1265-1280, 2015.

[22] J. Oakland and R. Oakland, Statistical Process Control, 3rd ed. Routledge, pp. 3-4, 2018.

[23] R. G. Aykroyd, V. Leiva, and F. Ruggeri, "Recent developments of control charts, identification of big data sources and future trends of current research," Technological Forecasting and Social Change, vol. 144, pp. 221-232, 2019.

[24] W. Woodall, "The Use of Control Charts in Health-Care and Public-Health Surveillance," Journal of Quality Technology, vol. 38, pp. 104-89, 2006.

[25] Y. K. Seoh, V. H. Wong, and M. Z. Sirdari, "A study on the application of control chart in healthcare," ITM Web of Conferences, vol. 36, p. 1001, 2021.

[26] A. Sachlas, S. Bersimis, and S. Psarakis, "Risk-Adjusted Control Charts: Theory, Methods, and Applications in Health," Statistics in Biosciences, vol. 11, pp. 630-658, 2019.

[27] M. Thien, Q. Cordier, J. Lifante, M. Carty, and A. Duclos, "Control Charts Usage for Monitoring Performance in Surgery: A Systematic Review," Journal of Patient Safety, vol. 19, pp. 110-116, 2023.

[28] F. Zhu, "Application of Shewhart Control Charts in the inspection of pharmaceutical manufacturing process," China Pharmacy, pp. 358-362, 2018.

[29] İ. Kaya, E. Devrim, and H. Baraçli, "Design of Attributes Control Charts for Defects Based on Type-2 Fuzzy Sets with Real Case Studies from Automotive Industry," Journal of Multiple-Valued Logic and Soft Computing, vol. 41, 2023.

[30] D. Scott and D. Firth, "Using control charts early in the quantitative analysis laboratory curriculum," Journal of Chemical Education, vol. 96, no. 5, pp. 1037-1041, 2019.

[31] B. Mendes de Jesus, I. S. Mesquita dos Santos, J. V. Marinho de Albernaz, R. Palandi Cardoso, and N. A. de Souza Sampaio, "Using the attribute control chart to monitor a chemical manufacturing process," GeSec: Revista de Gestao e Secretariado, vol. 14, no. 11, 2023.

[32] A. Vries and J. Reneau, "Application of statistical process control charts to monitor changes in animal production systems," Journal of Animal Science, vol. 88, no. 13, pp. 11-24, 2010.

[33] S. Cheng and K. Thaga, "Single Variables Control Charts: An Overview," Quality and Reliability Engineering International, vol. 22, 2006.

[34] Minitab 21.4.1 Statistical Software, Minitab, 2023, [Online]. Available: https://www.minitab.com/content/dam/www/en/uploadedfiles/documents/readme/MinitabReadMe_en.pdf.

[35] A. F. Costa and M. S. De Magalhaes, "An adaptive chart for monitoring the process mean and variance," Quality and Reliability Engineering International, vol. 23, no. 7, pp. 821-831, 2007.

[36] J. Zhang, Z. Li, and Z. Wang, "A new adaptive control chart for monitoring process mean and variability," International Journal of Advanced Manufacturing Technology, vol. 60, pp. 1031-1038, 2012.

[37] A. Ganguly and S. Patel, "Joint economic design of X-bar and R charts using simulated annealing," International Journal of Process Management and Benchmarking, vol. 4, pp. 305-323, 2014.

[38] B. Elahi, A. Deheshvar, M. Franchetti, and S. Tokaldany, "Optimum evaluation performance of adaptive X-bar control charts," Journal of Industrial and Production Engineering, vol. 39, pp. 644-653, 2021.

[39] M. Wan-yu, "Application of 3D Control Chart," Equipment for Electronic Products Manufacturing, vol. 2, 2009.

[40] M. Riaz and F. Muhammad, "An application of control charts in manufacturing industry," Journal of Statistics and Econometric Methods, vol. 1, pp. 77-92, 2012.

[41] M. Amin, S. Tariq, and M. Hanif, "Control charts for simultaneously monitoring of process mean and coefficient of variation with and without auxiliary information," Quality and Reliability Engineering International, vol. 35, pp. 2639-2656, 2019.

[42] P. Lee, "Adaptive R charts with variable parameters," Computational Statistics and Data Analysis, vol. 55, pp. 2003-2010, 2011.

[43] J. Ayala-Mora, A. Infante-Moro, and J. C. Infante-Moro, "Electronic administration in the Council of the Judiciary of Ecuador [La administración electrónica en el Consejo de la Judicatura de Ecuador]," RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação, no. E60, pp. 511-523, 2023.

[44] A. Boyko, A. Zimbroff, S. Mynenko, and Y. Chen, "Anti-money laundering in the digital economy: Institutional, financial, and educational channels," Virtual Economics, vol. 7, no. 3, pp. 59-80, 2024, [Online]. Available: https://doi.org/10.34021/ve.2024.07.03(4).

[45] A. Infante-Moro, J. C. Infante-Moro, and J. Gallardo-Pérez, "Key factors to raise cybersecurity awareness among employees [Factores claves para concienciar la ciberseguridad en los empleados]," Revista de Pensamiento Estratégico y Seguridad CISDE, vol. 7, no. 1, pp. 69-79, 2022.

[46] A. Kwilinski and N. Trushkina, "Impact of Cyber Risks and Threats on the Critical Infrastructure Development: Visualization of Scientific Research," Proceedings of the 12th International Conference on Applied Innovation in IT (ICAIIT), vol. 12, iss. 2, pp. 107-119, 2024, [Online]. Available: http://dx.doi.org/10.25673/118123.

[47] A. Kwilinski and M. Kardas, "Efficiency of full and fractional factorial designs in quality management: A case study on identifying key adhesion factors," Forum Scientiae Oeconomia, vol. 12, no. 3, pp. 138-155, 2024, [Online]. Available: https://doi.org/10.23762/FSO_VOL12_NO3_7.

[48] O. Lyulyov, T. Pimonenko, A. Infante-Moro, and A. Kwilinski, "Perception of artificial intelligence: GSR analysis and face detection," Virtual Economics, vol. 7, no. 2, pp. 7-30, 2024, [Online]. Available: https://doi.org/10.34021/ve.2024.07.02(1).

Sensor-Based Gait Analysis: A Comparative Study of Ultrasonic and Laser Sensors for Gait Monitoring in Rollator-Assisted Walking

Ivan Omeko, Stefan Twieg and Martin Obert
[1]*Department of Electrical, Mechanical and Industrial Engineering, Anhalt University of Applied Sciences, Bernburger Str. 55, Köthen, Germany*
Ivan.Omeko@student.hs-anhalt.de, Stefan.Twieg@hs-anhalt.de, Martin.Obert@hs-anhalt.de

Keywords: Sensor, Signal Processing, Gait Analysis, Ultrasonic Sensor, Laser Sensor, Machine Learning, Algorithms, Elderly Support, Rollator.

Abstract: The AktiMuW project aims to enhance mobility assistance for elderly individuals by developing a smart rollator equipped with advanced posture monitoring. A crucial aspect of this system is the detection and correction of the user's posture of legs. This is based on measuring the distance between the user and the rollator, among other methods. The study evaluates three different distance sensors – HC-SR04, HC-SR04-P, and TFmini-S to determine the most reliable and suitable option. To achieve this, a series of use case centered experiments were conducted, where each sensor's performance was tested. The HC-SR04 demonstrated relatively low measurement error, with Root Mean Square Error (RMSE) values ranging from 0.64 cm to 0.89 cm but required a 5V power supply and additional voltage conversion components, complicating integration to single board computers (SBC). The HC-SR04-P, an updated model, operates reliably at 3.3V – compatible with Raspberry Pi boards – and maintained comparable measurement precision, with RMSE values of 0.57 cm and 0.78 cm. In contrast, the TFmini-S LiDAR sensor exhibited higher RMSE values of 5.42 cm and 2.89 cm, particularly struggling at shorter distances, making it unsuitable for this application. Further gait analysis tests confirmed that the HC-SR04-P could effectively monitor the user's position, despite occasional signal reflections. The study concludes that the HC-SR04-P is the optimal choice for the rollator Machine Learning algorithms due to its balance of accuracy, compatibility, and cost-effectiveness. These findings contribute to the theoretical understanding of sensor-based posture monitoring and hold practical significance for the development of assistive mobility devices and further algorithms.

1 INTRODUCTION

1.1 Motivation

As people age, they often require more assistance [1]. According to [2], mobility declines significantly with age, as majority of people over 85 experience some difficulty walking, and mobility disability is linked to increased risks of social isolation, falls, and depression. Robotics can help provide essential support to enhance their independence and quality of life [3]. The goal of the AktiMuW [4] project is to develop a smart rollator for elderly individuals, providing them with enhanced assistance [5] in daily tasks, whether navigating indoors or walking to the nearest grocery store (for instance, with the help of road signs detection [6]).

One key area of development in the project is posture monitoring, which relies on data from multiple sensors. To achieve accurate posture assessment, it is crucial to identify the most suitable sensor for measuring the distance between the AktiMuW Rollator and the user for gait identification [7], [8]. The following tests aim to determine which sensor best fulfills this role.

1.2 Problem Statement

Ultrasonic sensors HC-SR04 [9] are widely used for various applications. These ultrasonic sensors have been implemented to measure the distance between the rollator and the user's legs but have demonstrated inconsistent performance. This leads to reliability issues for addressed use case of gait analysis. To address this challenge, a study based on a series of tests was conducted to evaluate sensor technology and determine the most suitable option for the project.

The main technical focus is set to compare ultrasonic versus laser sensor. By analyzing the performance of pre-selected sensors, the goal is to identify the most effective solution for accurately measuring the user's distance from the rollator. The findings from these tests will lead to further improvements in the smart rollator's design, particularly in enhancing its posture detection machine learning algorithms [10], [11], [12], [13].

2 SENSORS REVIEW

The evaluation process involved comparing three different sensor types:

- HC-SR04 (5V),
- HC-SR04-P (3.3V),
- TFmini-S (5V).

The current study setup for distance reading consists of Raspberry Pi Zero W with, for instance, two ultrasonic sensors HC-SR04, like shown in Figure 1.

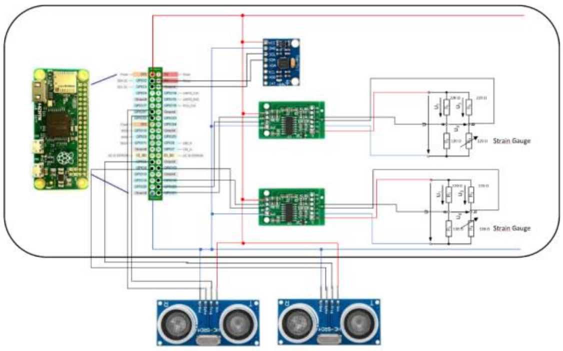


Figure 1: Raspberry Pi Zero W Schematic of Sensor Connections.

2.1 HC-SR04

HC-SR04, shown in Figure 2, is the ultrasonic sensor that was used for first proof of concept implementation on rollator.

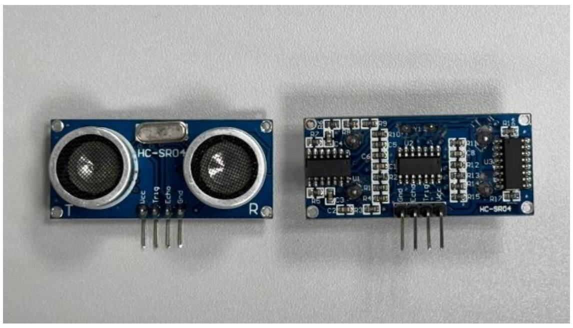


Figure 2: HC-SR04 ultrasonic sensor.

Characteristics are:

- Working Voltage: DC 5V;
- Working Current: 15 mA;
- Working Frequency: 40 Hz;
- Maximum Range: 4 m;
- Minimum Range: 2 cm;
- Measuring Angle: 15 degrees;
- Dimensions: 45x20x15 mm;
- Price: approx. 3€.

Reviewing the device specifications and connection schematics reveals that the HC-SR04 sensor needs to be powered by 5V. GPIO of Raspberry Pi Zero board requires 3.3V, but supports 5V power supply. This mismatch causes issues with the Raspberry Pi.

However, the main problem is that sensor does not perform voltage conversion on its ECHO pin, while the Raspberry Pi's GPIO can only handle 3.3V. As a solution, a voltage divider consisting of two resistors can be used. With all this in mind, electrical connection scheme is shown in Figure 3.

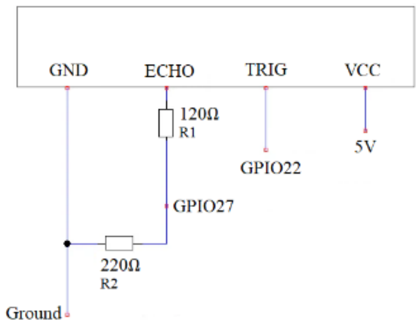


Figure 3: Scheme of HC-SR04 connected to Raspberry Pi powered by 5V and voltage divider at ECHO pin.

2.2 HC-SR04-P

Figure 4 displays HC-SR04-P [14], which is a version of HC-SR04 sensor that can operate both at 5V and 3.3V.

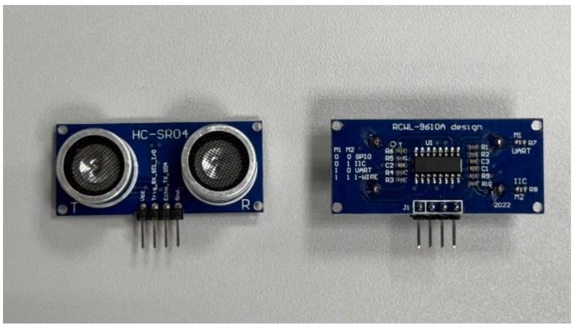


Figure 4: HC-SR04-P ultrasonic sensor.

Characteristics are:
- Working Voltage: DC 3.3–5V;
- Working Current: <2 mA;
- Working Frequency: 40 Hz;
- Maximum Range: 500 cm;
- Minimum Range: 2 cm;
- Measuring Angle: 15 degrees;
- Dimensions: 45x20x15 mm;
- Price: approx. 3€.

Since the HC-SR04-P is merely a modification of the HC-SR04, the main advantage is, that this sensor will be directly connected with 3.3V power and GPIO. Within following evaluation, it is analyzed if performance differs between 5V and 3.3V version.

2.3 TFmini-S

The TFmini-S [15] is a single-point LiDAR sensor. It was included in the tests to introduce a different type of sensor, distinct from ultrasonic sensors, and to compare their performance. Figure 5 illustrates the visual appearance of the sensor.

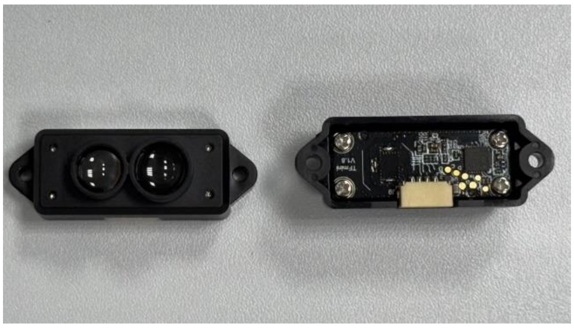


Figure 5: TFmini-S single-point LiDAR sensor.

Its characteristics are as follows:
- Operating range: 0.1m–12m;
- Accuracy: ±6 cm at 0.1–6m, ±1% at 6m–12m;
- Measurement unit: cm;
- Range resolution: 1cm;
- FOV: 2 degrees;
- Frame rate: 1~1000Hz;
- Supply voltage: 5V±0.1V;
- Average current: ≤140mA;
- Peak current: 200mA;
- Average power: 700mW;
- Communication level: LVTTL (3.3V);
- Price: approx. 42€.

3 PRE-EVALUATION ANALYSIS

3.1 Sensor Test Setup

Each sensor will be tested individually by detecting an obstacle at four distances: 5 cm, 30 cm, 60 cm, and 100 cm. The boundaries of 5 cm and 100 cm were chosen because rollator users are likely to operate within this range. The goal is to evaluate the accuracy of each sensor type and compare their performance, like shown in Figure 6.

To provide a compact measure of each sensor's performance, tables, such as Table 2, present the average distance measurements recorded by each sensor at four test distances. These averages offer a simplified view of each sensor's typical response at each range.

The bottom row of the table reports the Root Mean Square Error (RMSE) [16] for each sensor, calculated relative to the true distances. RMSE is computed from the average readings and provides a single-value summary of overall deviation. It is defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(measured_i - actual_i)^2} \quad (1)$$

where n is the number of test distances.

The test setup consists of:
- Raspberry Pi Zero W,
- Laptop with SSH connection to the Raspberry Pi,
- USB-A to Micro-USB cable,
- a breadboard,
- jumper cables,
- 120Ω and 220Ω resistors,
- Measuring tape fixed to a desk,
- Centered cardboard box as an obstacle.

Figure 6: Example of the test setup. HC-SR04 tested at 30 cm.

3.2 HC-SR04 Test

As mentioned before, HC-SR04 will be powered by 5V with connected and voltage divider at ECHO pin, consisting of 120Ω and 220Ω resistors.

Connections between the HC-SR04 and the Raspberry Pi Zero W are shown in Table 1.

Table 1: Connections between two HC-SR04 or HC-SR04-P and the Raspberry Pi Zero W.

HC-SR04 (-P) Left	Raspberry Pi Zero W
VCC	5V/ 3V3
TRIG	GPIO22
ECHO	GPIO27
GND	Ground

HC-SR04 (-P) Right	Raspberry Pi Zero W
VCC	5V/ 3V3
TRIG	GPIO5
ECHO	GPIO6
GND	Ground

Figures 7 - 10 show the results of conducted tests with HC-SR04 sensors.

Table 2: Average distance readings for each HC-SR04 unit.

Test Distance, cm	Average readings, cm			
	HC-SR04 1	HC-SR04 2	HC-SR04 3	HC-SR04 4
5	5.39	5.20	5.54	5.42
30	30.52	30.12	30.88	30.30
60	59.95	59.55	59.70	60.15
100	98.73	98.40	98.97	98.87
RMSE, cm	0.74	0.89	0.79	0.64

As shown in Figures 7–10 and Table 2, all four sensor units generally provide consistent distance measurements to the object in front of them. However, occasional outliers to random distances occur. Notably, these outliers appear on different sensors at various distances, suggesting they may be caused by external interference.

The Root Mean Square Error for individual sensors ranges from 0.64 cm to 0.89 cm, reflecting relatively small overall deviations from the true values.

3.3 HC-SR04-P Test

Since HC-SR04-P is just a modification of HC-SR04 that can operate at both 5V and 3.3V, in this test for variability it will be powered by 3.3V without voltage divider at ECHO pin. Additionally, the current Raspberry Pi setup on the rollator already powers the ultrasonic sensors via 3.3V. Connections are shown in Table 1. Figures 11–14 show the results of conducted tests with HC-SR04-P.

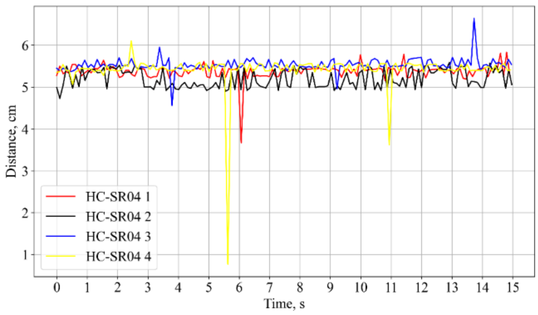


Figure 7: Comparison of four HC-SR04 units at 5 cm.

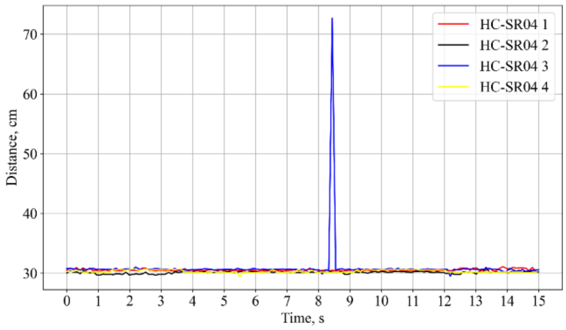


Figure 8: Comparison of four HC-SR04 units at 30 cm.

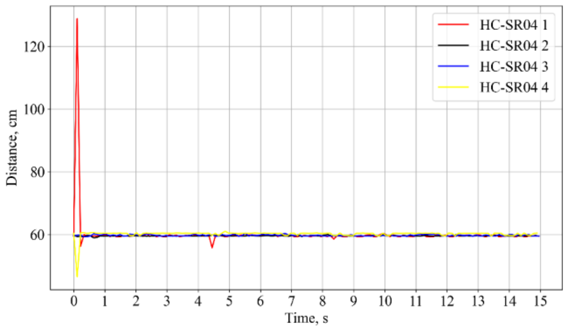


Figure 9: Comparison of four HC-SR04 units at 60 cm.

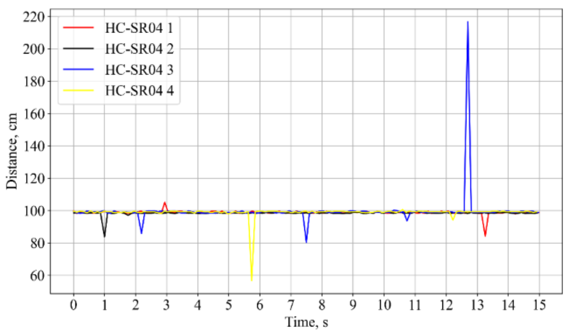


Figure 10: Comparison of four HC-SR04 units at 100 cm.

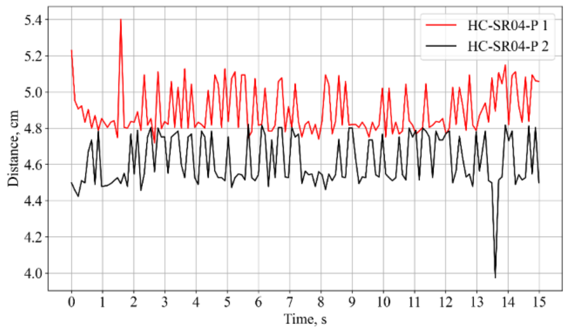


Figure 11: Comparison of two HC-SR04-P units at 5 cm.

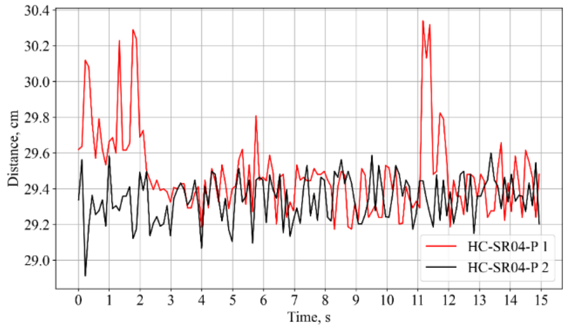


Figure 12: Comparison of two HC-SR04-P units at 30 cm.

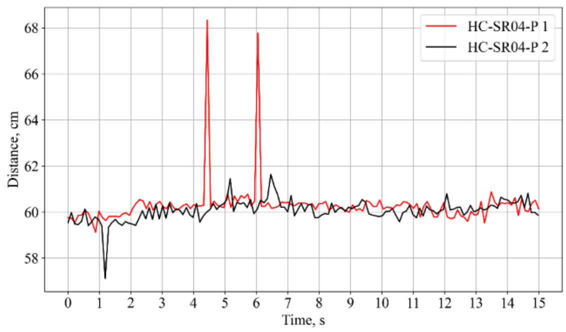


Figure 13: Comparison of two HC-SR04-P units at 60 cm.

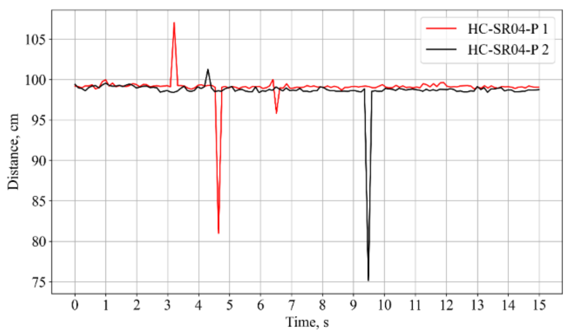


Figure 14: Comparison of two HC-SR04-P units at 100 cm.

Table 3: Average distance readings for each HC-SR04-P unit.

Test Distance, cm	Average readings, cm	
	HC-SR04-P 1	HC-SR04-P 2
5	4.90	4.62
30	29.47	29.35
60	60.30	60.07
100	99.05	98.63
RMSE, cm	0.57	0.78

It can be observed from Table 3 that the HC-SR04-P is very similar to the HC-SR04 in terms of accuracy. However, it appears to be less prone to sudden outliers in readings.

The Root Mean Square Error is 0.57 cm and 0.78 cm for the first and second units, respectively, indicating high accuracy relatively to the actual test distances.

3.4 TFmini-S Test

Connections between the TFmini-S and the Raspberry Pi are shown below in Table 4.

Table 4: Connections between the TFmini-S and the Raspberry Pi Zero W.

TFmini-S	Raspberry Pi
5V(RED)	5V
GND(BLACK)	Ground
RX(WHITE)	TXD0
TX(GREEN)	RXD0

Results of testing are shown in Figures 15 to 18.

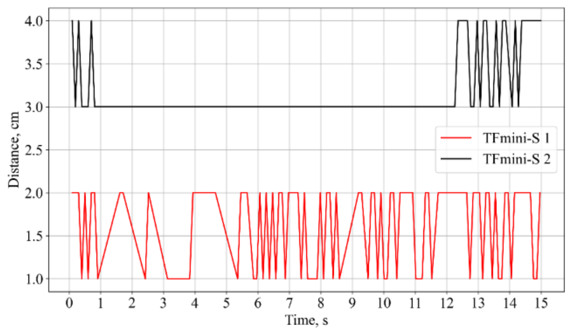


Figure 15: Comparison of two TFmini-S units at 5 cm.

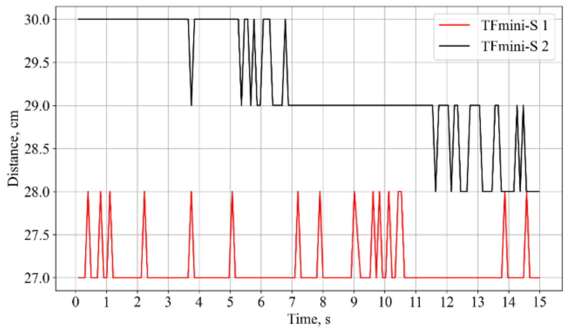


Figure 16: Comparison of two TFmini-S units at 30 cm.

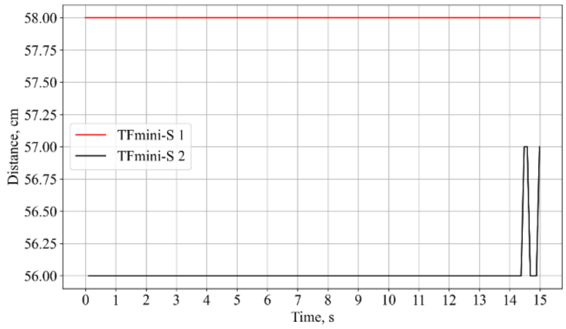


Figure 17: Comparison of two TFmini-S units at 60 cm.

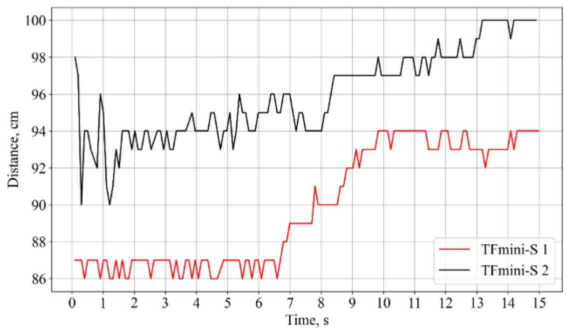


Figure 18: Comparison of two TFmini-S units at 100 cm.

Table 5: Average distance readings for each HC-SR04-P unit.

Test Distance, cm	Average readings, cm	
	TFmini-S 1	TFmini-S 2
5	1.67	3.18
30	27.12	29.15
60	58.00	56.08
100	90.30	96.25
RMSE, cm	5.42	2.89

Results, shown in Table 5, indicate that the TFmini-S exhibits larger deviations at all tested distances compared to HC-SR04 and HC-SR04-P sensors.

The Root Mean Square Errors for individual TFmini-S sensors are 5.42 cm and 2.89 cm, indicating a notable decrease in accuracy compared to the results from ultrasonic sensors.

Examining its specifications provides a possible explanation. The claimed operating range is 0.1 m to 12 m – three times greater than that of the HC-SR04 and HC-SR04-P. Additionally, its range resolution is 1 cm, which prevents more precise distance readings.

This suggests that the TFmini-S may be more appropriate for applications where extended range is prioritized over fine measurement resolution.

4 ROLLATOR GAIT ANALYSIS

Based on the results of previous tests, the HC-SR04-P appears to be the most suitable sensor for use with the rollator.

To assess the performance of the sensor for use in Machine Learning applications, it is important to evaluate how well it captures data under controlled conditions that simulate human gait.

Therefore, a series of tests imitating a person's gait will be conducted. Instead of walking with the device, the subject's legs will be positioned at various relative distances to the sensors, installed on the rollator. This approach ensures a controlled and repeatable testing environment, allowing for consistent data collection and more reliable insights into sensor performance.

4.1 Rollator System Setup

Figure 19 illustrates the sensor mounting configuration on the rollator. The sensors are positioned beneath the rollator's seat and attached to the housing that carries the PCB with the Raspberry Pi Zero W, auxiliary sensors, and the Jetson Nano 2 GB.

Figure 19: Prototype Rollator Integration of Sensors and Processing Hardware.

The Raspberry Pi is responsible for collecting data from all connected sensors and transmitting it to the Jetson Nano, which functions as the MQTT broker.

Sensor data is accessed via a laptop that connects to the MQTT broker, and subsequently stored for future analysis.

4.1.1 Nyquist Theorem

Before proceeding with gait tests, it is needed to check whether the sampling rate, in our case the step frequency, satisfies the Nyquist Theorem [17].

If a signal contains frequency components up to a maximum frequency f_{max}, then the minimum sampling rate required to avoid loss of information is:

$$f_s \geq 2f_{max} \quad (2)$$

where f_s is the sampling frequency (or sampling rate).

During previous test with sleep time of 0.1 seconds between cycles, samples were taken with Raspberry Pi Zero at a rate of $f_s = 8.7\, Hz$, with average sampling period of $T_s = 0.115\, s$. This implies that, to satisfy the Nyquist Theorem, the maximum step frequency f_{max} would need to be:

$$f_{max} \leq \frac{f_s}{2} = \frac{8{,}7\, Hz}{2} = 4{,}35\, Hz. \quad (3)$$

This corresponds to a minimum step period of $T_{min} \geq 0.23\, s$. In other words, to violate the theorem, a single leg would need to step faster than every 0.23 seconds, or make approximately 5 steps per second. The step frequency is estimated to be around 1 step per second, or 1 Hz.

This confirms that the Nyquist Theorem is satisfied and tests can be continued.

4.1.2 HC-SR04-P Maximum Sampling Rate

While still on the subject of sampling rate, it is worth examining the maximum achievable sampling rate of the HC-SR04-P when used with the Raspberry Pi Zero W.

The sampling rate is dependent on the distance from the sensor to the obstacle. For example, if one object is located 5 cm from the sensor and another at 100 cm, the signal’s flight time will be 20 times shorter in the first case than in the second. As a result, the sensor can proceed with the next reading much faster in the first scenario, leading to a higher sampling rate.

For the following calculations, HC-SR04-P’s maximum specified operating distance of 400 cm will be used as the reference distance.

Knowing the distance to the obstacle, which is 4 m, and speed of sound, which is 343 m/s, maximum time of flight t_{max} can be calculated:

$$t_{max} = \frac{4\, m \times 2}{343\, m/s} \approx 23.3\, ms. \quad (4)$$

The sensor requires a short delay (~1 ms) before a new measurement starts, so assume a minimum cycle time of $t_{max} \approx 24ms$.

Having maximum time of flight t_{max}, maximum theoretical sampling frequency can be found:

$$f_{max} = \frac{1}{0.024\, s} \approx 41.7\, Hz. \quad (5)$$

Given that on the rollator two sensors are being used sequentially, the theoretical maximum sampling rate would be:

$$\frac{41.7}{2} \approx 20.8\, Hz. \quad (6)$$

On practice, maximum sampling rate of $f_{max} = 15.9\, Hz$ was achieved.

4.1.3 Maximum Sampling Rate with Raspberry Pi

The previous result was achieved using a test script focused solely on collecting data from the ultrasonic sensors. However, in the actual project, the aktimuwGetData.py script will be used. This script gathers information from five additional sensors, which increases the runtime of each cycle and consequently reduces the sampling rate.

Using the previously mentioned 400 cm distance to an obstacle and the current script setup with a sleep time of 0.1 seconds between cycles, the sampling rate with aktimuwGetData.py is 7.3 Hz.

Referring back to the Nyquist Theorem, it can be observed that the condition is still satisfied.

By minimizing the sleep time between cycles, a maximum sampling rate of 10.3 Hz was achieved.

4.2 Gait Test Methodology

Gait analysis is essential for tracking a person's posture, helping to improve their overall health and well-being [18], [19], especially for elderly [20].

The following test will be conducted as follows (an example is shown in Figure 20):

- Start with the right leg in front and the left leg in the back; hold this position for 5 seconds.
- Move both legs to the middle position; hold for 5 seconds.
- Switch to the right leg in the back and the left leg in front; hold for 5 seconds.
- Repeat the cycle from the beginning.

Test will be repeated with a reduced holding time of 2 seconds per position:

- Start with the right leg in front and the left leg in the back; hold this position for 2 seconds.
- Move both legs to the middle position; hold for 2 seconds.
- Switch to the right leg in the back and the left leg in front; hold for 2 seconds.
- Repeat the cycle for 20 seconds.

The "middle position" refers to a stance in which a straight leg is positioned in front of the rollator, with an approximate distance of 30–35 cm between the leg and the sensors.

The test will be performed with two different distances between the back leg's knee and the middle position – 30 cm and 15 cm – to simulate larger and smaller steps.

Figure 20: Example of a stance with right leg in the front (the middle position) and the left leg in the back.

Additionally, a moving average filter with a window size of 5 will be applied to produce smoothed results alongside the raw data.

4.3 Results

In Figure 21, the results of the first test are shown. Each stance change – which happens approximately every 5 seconds – is represented by vertical dashed orange line.

The test begins with the left leg positioned back, 30 cm from its knee to the middle position, while the right leg is in the middle position. Around the 5-second mark, the right leg is moved forward to the middle position which is indicated by vertical dashed orange line. At this point, a large spike in the readings is observed, jumping to approximately 200 cm. This is most likely caused by the ultrasonic signal being reflected off the folds in the fabric of the trousers. Additionally, having the back leg positioned 30 cm behind creates a rather acute angle between the leg and the sensor (Fig.20), further increasing the likelihood of signal reflection and inaccurate readings.

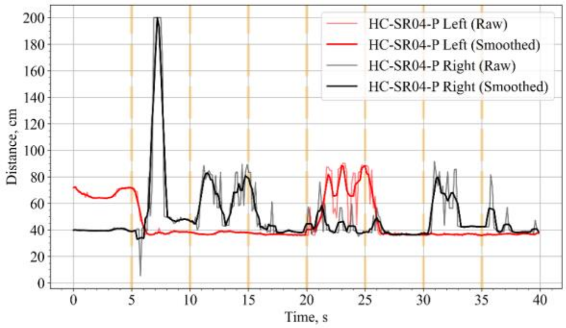


Figure 21: Step test with 30 cm amplitude and 5 s cycle time.

After this spike, the readings stabilize. Around the 10-second mark, the right leg is moved back 30 cm and held in place for 5 seconds. During this phase, a noticeable dip in the distance readings is observed. Following this, the right leg is returned to the middle position, and the cycle is repeated once more.

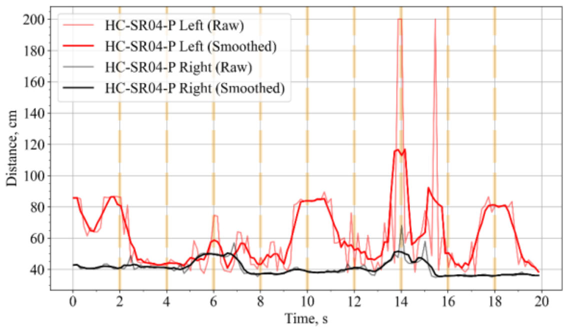


Figure 22: Step test with 30 cm amplitude and 2 s cycle time.

Figure 22 shows similar behavior to the previous test; however, each position is now held for a shorter duration – only 2 seconds. In this test stance change happens every 2 seconds, as shown by vertical dashed orange line.

It is worth noting that, since the holding period is shorter and transitions between positions are not instantaneous, there is an increased potential for confusion in the sensor readings. This is particularly visible between the 12- to 16-second marks.

Additionally, occasional spikes to 200 cm continue to occur, likely due to signal reflections.

Despite these irregularities, the different leg positions are still recognizable in the plot.

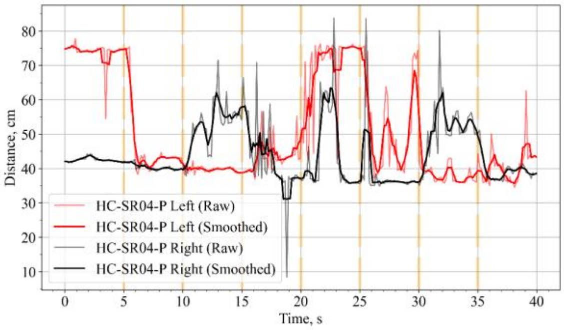


Figure 23: Step test with 15 cm amplitude and 5 s cycle time.

Moving to the next test, Figure 23 displays the results of the run with the back leg positioned with the knee approximately 15 cm behind the middle position, with each stance held for 5 seconds.

Overall, the different positions can be distinguished; however, the back position of the left leg is measured as 35 cm away from the middle, indicating some inaccuracy. On the right leg, the sensor readings are closer to the actual distance between the sensor and the leg.

It is also notable that the spikes up to 200 cm observed in previous tests are now absent. This supports the assumption that reducing the back leg's distance decreases the angle between the leg and the sensor, thereby lowering the likelihood of signal reflections and missed measurements.

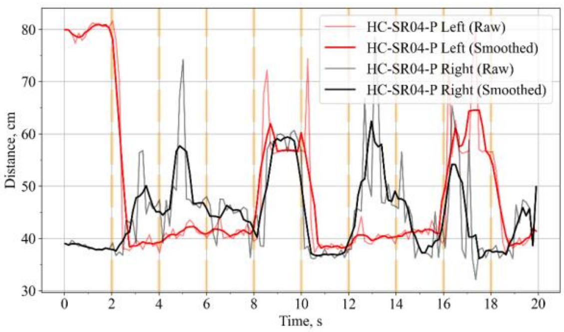


Figure 24: Step test with 15 cm amplitude and 2 s cycle time.

Figure 24 shows the results of a test similar to the previous one, but with each position held for 2 seconds. Distance detection across various stances is once again quite satisfactory, especially considering the potential sources of inaccuracy previously discussed: signal reflections from fabric folds, the shorter holding time for each position, and the transition period between stances.

Additionally, the slight inaccuracy observed on the left sensor appears to be somewhat reduced in this test. This suggests that the issue is likely not with the sensor itself but rather influenced by external factors.

Overall, the HC-SR04-P sensors demonstrate reliable performance in position and distance detection, though they are not without flaws. However, it is important to consider that the AktiMuW project is intended to assist elderly individuals with mobility, and they are unlikely to take steps larger or faster than those observed during the tests.

Taking this into account – along with the previous analysis of the different sensor types' performance in the current rollator Raspberry Pi setup, as well as factors like price and reliability – the HC-SR04-P remains the most suitable choice for the AktiMuW project, it gives reliable information for further gait analysis and processing by machine learning algorithms.

5 CONCLUSIONS

After completing all the tests and analyzing the sensors, a decision can be made regarding which sensor should be used.

The HC-SR04 demonstrated very good accuracy with the Root Mean Square Error at 0.64–0.89 cm; however, to use it on the rollator, a new PCB with 5V power traces for the sensors and voltage dividers on the ECHO pin would need to be created.

The HC-SR04 exhibited low Root Mean Square Error values between 0.64 cm and 0.89 cm, indicating a relatively precise measurement capability. However, integrating it into the rollator would require the development of a new PCB with 5V power traces and voltage dividers on the ECHO pin.

The HC-SR04-P resolves the power issue of the HC-SR04, as it can operate directly with 3.3V of Raspberry Pi Zero. It also demonstrated comparable precision, with RMSE values of 0.57 cm and 0.78 cm across two units.

The TFmini-S, a LiDAR-based sensor operating on a different measurement principle, was considered as an alternative. However, test results indicated lower measurement accuracy, with RMSEs of 5.42 cm and 2.89 cm. This reduced precision can be explained by the sensor's measurement principle and focused measurement spot. Furthermore, it is at least 10 times more expensive and requires 5V power.

The final gait test of the HC-SR04-P on the rollator confirms that the sensors are a suitable solution for the project and further processing by machine learning algorithms to identify leg position, usage and other states of use of rollator.

ACKNOWLEDGEMENTS

This 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," October 2022, [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health.

[2] E. Jaul and J. Barron, "Age-Related Diseases and Clinical and Public Health Implications for the 85 Years Old and Over Population," Front Public Health, vol. 5, p. 335, 2017.

[3] V. Kurumbaparambil, S. Rajanayagam and S. Twieg, "Modular Robotic Reinforcement Learning Platform for Object Manipulation," 2024, Universitäts- und Landesbibliothek Sachsen-Anhalt, [Online]. Available: https://doi.org/10.25673/115703.

[4] Translationsregion für digitalisierte Gesundheitsversorgung, "Aktiv im Alter durch Multisensorische Umfeldwahrnehmung – pflege- und versorgungswissenschaftlich evaluative Begleitung der co-kreativen Entwicklung," [Online]. Available: https://inno-tdg.de/projekte/aktimuw/.

[5] F. Mubarak and R. Suomi, "Elderly Forgotten? Digital Exclusion in the Information Age and the Rising Grey Digital Divide," INQUIRY, 2022, [Online]. Available: https://doi.org/10.1177/00469580221096272.

[6] S. Twieg and R. Menghani, "Analysis and Implementation of an Efficient Traffic Sign Recognition," in 11th International Conference on Applied Innovations in IT (ICAIIT), Köthen, 2023.

[7] Z. Wu, Y. Huang, L. Wang, X. Wang and T. Tan, "A comprehensive study on cross-view gait based human identification with deep cnns," IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 2, 2016.

[8] E. J. Harris, I. Khoo and E. Demircan, "A Survey of Human Gait-Based Artificial Intelligence Applications," Front. Robot. AI, 03 January 2022, Sec. Biomedical Robotics, Volume 8 – 2021, [Online]. Available: https://doi.org/10.3389/frobt.2021.749274.

[9] Elecfreaks.com, "Ultrasonic Ranging Module HC - SR04," [Online]. Available: https://www.elecfreaks.com/download/HC-SR04.pdf.

[10] E. Blasch et al., "Machine Learning/Artificial Intelligence for Sensor Data Fusion–Opportunities and Challenges," in IEEE Aerospace and Electronic Systems Magazine, vol. 36, no. 7, pp. 80-93, 1 July 2021, [Online]. Available: https://doi.org/10.1109/MAES.2020.3049030.

[11] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," Nature, vol. 521, 2015.

[12] R. S. Sutton and A. G. Barto, "Reinforcement Learning: An Introduction," IEEE Transactions on Neural Networks, vol. 9, no. 5, pp. 1054-1054, 1998, [Online]. Available: https://doi.org/10.1109/TNN.1998.712192.

[13] S. M. Alfayeed and B. S. Saini, "Human Gait Analysis Using Machine Learning: A Review," 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), Dubai, United Arab Emirates, 2021, pp. 550-554, [Online]. Available: https://doi.org/10.1109/ICCIKE51210.2021.9410678.

[14] Roboter-Bausatz.de, "Datenblatt - HC-SR04-P Ultraschallsensor Entfernungsmesser," [Online]. Available: https://www.roboter-bausatz.de/datasheet/RBS11813.pdf.

[15] Benewake (Beijing) Co. Ltd., "Product Manual of TFmini-S," [Online]. Available: https://en.benewake.com/uploadfiles/2025/04/20250430175136263.pdf.

[16] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," Geosci. Model Dev., vol. 15, pp. 5481–5487, 2022, [Online]. Available: https://doi.org/10.5194/gmd-15-5481-2022.

[17] C. E. Shannon, "Communication in the Presence of Noise," in Proceedings of the IRE, vol. 37, no. 1, pp. 10-21, Jan. 1949, [Online]. Available: https://doi.org/10.1109/JRPROC.1949.232969.

[18] Y. Wahab and N. A. Bakar, "Gait analysis measurement for sport application based on ultrasonic system," 2011 IEEE 15th International Symposium on Consumer Electronics (ISCE), Singapore, 2011, pp. 20-24, [Online]. Available: https://doi.org/10.1109/ISCE.2011.5973775.

[19] R. B. Huitema, A. L. Hof and K. Postema, "Ultrasonic motion analysis system–measurement of temporal and spatial gait parameters," Journal of Biomechanics, vol. 35, no. 6, pp. 837-842, 2002, [Online]. Available: https://doi.org/10.1016/S0021-9290(02)00032-5.

[20] 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, [Online]. Available: https://doi.org/10.1016/j.gaitpost.2023.03.004.

Intelligent Learning Support System

Eugene Karashevych, Svitlana Sulima and Mariia Skulysh
Department of Information technologies in telecommunications, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Industrialnii Lane 2, 03056 Kyiv, Ukraine
{its30316, itssulima, mskulysh}@gmail.com

Keywords: Automated System, Artificial Intelligence, Web Application, Adaptive Learning, Task Analysis, Test Generation, Educational Process Support, Individualisation of Learning, Educational Technologies.

Abstract: Modern educational processes require automation to improve learning efficiency. The use of artificial intelligence (AI) allows to optimize the management of the learning process, increase the personalization of learning, and automate assessment. In the context of digitalization of education and the growing role of distance learning, it is important to create adaptive systems that meet the needs of students and teachers. Thus, the aim of the work is to develop and implement a web service that will support the learning process by automating the generation of test tasks, checking answers, and integrating with learning systems. A machine learning module has been implemented to automatically analyze student work (grading, checking for uniqueness). Natural language processing (NLP) was used to analyze student responses and create adaptive content. Automatic generation of test tasks based on learning materials is implemented, which increases the personalization of learning. Standard assessment systems (e.g., Moodle testing) are often limited to multiple choice, while the use of semantic analysis allows you to evaluate creative tasks and open-ended answers without manual verification by the teacher. Most LMS systems (Moodle, Google Classroom) provide standard content for all students, while the developed system adapts to the needs of a particular user, increasing the efficiency of learning. Instead of creating another isolated LMS system, the platform is designed with a flexible API that allows it to be easily integrated into existing educational solutions (for example, university portals).

1 INTRODUCTION

The evolution of educational technology has fundamentally transformed teaching and learning methodologies, significantly impacting the efficiency of educational processes. Early educational tools, such as chalkboards and projectors, laid the groundwork for better communication between educators and students. The introduction of computers in classrooms during the 1980s marked a pivotal shift, allowing for interactive learning experiences through specialized software programs [1]. As the Internet became mainstream in the 1990s, it opened new avenues for education, enabling the widespread availability of online courses and resources, thereby increasing accessibility. In recent years, there has been a growing emphasis on incorporating various instructional formats, including videos, podcasts, and interactive simulations, to address diverse learning styles among students. This multi-format approach has shown to enhance engagement and effectiveness, accommodating visual, auditory, and kinesthetic learners. As educators strive to create inclusive learning environments, the integration of technology into teaching practices has become essential. However, educational challenges persist, including the digital divide that limits access to technology in underprivileged areas, as well as the need for comprehensive teacher training to utilize these technological resources effectively. Additionally, maintaining student engagement in online learning environments has proven to be challenging, leading to higher dropout rates. To further improve the efficiency of teaching, there is a movement towards implementing automated systems for generating test tasks. Such systems aim to alleviate the time-consuming aspects of traditional testing methods, which often rely on human expertise and are prone to errors. Traditional testing approaches can be labor-intensive, involving meticulous creation and execution of test cases, and they may not adequately support the needs of all learners [2],[3]. Automated testing systems promise to streamline these processes, allowing educators to focus more on content delivery and student interaction rather than administrative tasks. Moreover, the use of adaptive assessment systems and data-driven approaches can enhance learning outcomes by providing tailored experiences that cater to individual student needs, thus promoting greater engagement and retention. As educational technology continues to advance, the integration of automated systems for test generation represents a significant step forward in addressing current educational challenges while improving overall teaching efficiency.

Unlike most existing automated learning systems, which rely on traditional methods such as multiple-choice assessments, predefined rule-based grading, and simple keyword matching, our approach incorporates AI to enhance the learning process. Many current systems do not utilize AI, focusing instead on static evaluation techniques. In contrast, our system is designed to assist teachers in common tasks like checking assignments and generating learning materials, making the educational process more efficient and adaptable.

However, the implementation of such systems is not without its challenges and controversies. Ethical concerns regarding data privacy, the potential for bias in AI algorithms, and the need for comprehensive teacher training to effectively use these technologies are critical issues that must be addressed. Moreover, ensuring fairness and equity in automated assessments is essential to prevent disparities in educational outcomes among different student demographics [4-6]. Overall, the development and integration of automated systems for generating test tasks signify a promising advancement in educational technology. These systems aim to empower educators by reducing administrative burdens, fostering innovative teaching practices, and ultimately improving learning outcomes in diverse educational contexts [7],[8].

2 SYSTEM DEVELOPMENT

The development of an automated system for generating test tasks is a multifaceted process that requires careful planning and execution. It involves the integration of various technologies and methodologies to ensure that the system is not only efficient but also adaptable to the diverse needs of educators and learners. Central to this process is the architecture of the system, which must support scalability, resilience, and real-time access to data, allowing for quick adjustments to meet educational demands [9].

The AI module utilizes the OpenAI API with GPT-4 for grading and evaluating uniqueness. This approach was implemented as a prototype to demonstrate the system’s capabilities. However, for educational purposes, pretrained models can also be integrated, allowing for more customizable and domain-specific assessment solutions. Additionally, NLP techniques such as named entity recognition (NER) and sentiment analysis can be applied to enhance semantic evaluation of open-ended answers, ensuring a more adaptive and intelligent learning support system.

The main components of the system include a chat module with artificial intelligence, task verification based on uploaded materials, and test generation from training materials. The system uses Next.js, Firebase, and OpenAI API technologies that ensure fast query processing and real-time response generation. Google authentication is integrated to identify users, which increases data security.

The project is built using Tailwind CSS [10] for styling, which simplifies the process of creating an adaptive interface, and TypeScript ensures code security through static typing. The settings include tsconfig to optimize the work with TypeScript and prettier [11] and eslint [12] settings for automatic formatting and maintaining a single code style.

2.1 Description of the Basic Project Structure

The project is built as a modular system with the main structure of the /app folder (Fig. 1), which stores individual pages and components for chat, task checking, and test generation.

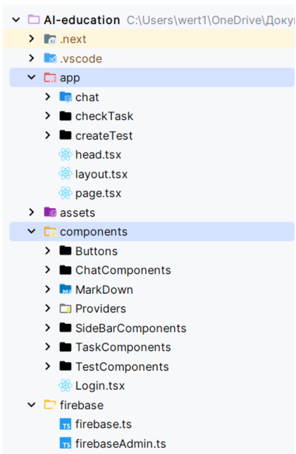


Figure 1: Basic project structure.

2.2 Designing the System Architecture: Modules, Connections

The system consists of the following main modules:
1) authentication and authorization module;
2) chatbot module with AI;
3) module for checking tasks according to the material;
4) test generation module;
5) module for saving and exporting data;
6) administration and user management module.

2.2.1 Authentication and Authorization Module

This module is responsible for system security and controls user access to various functions and resources. Its main functions are:
- user registration: creating a new account using Google authentication;
- user authentication: verification of user credentials to log in to the system;
- authorization: determining user access rights to various functions and resources of the system (students, teachers, administrators);
- security: the use of encryption protocols (e.g., TLS/SSL) to protect personal data and server requests.

Connections. The authentication module interacts with the database to store and retrieve information about users, as well as with other system modules to control access to their functionality.

2.2.2 Chatbot Module with AI

The main task of this module is to provide interactive communication between the user and the system via text chat. The chatbot integrates with external artificial intelligence APIs for natural language processing (e.g., GPT-3.5).

Module functions:
- acceptance of text requests from the user;
- processing and transferring requests to the artificial intelligence API;
- displaying the received answers in the format of a text message;
- ability to learn from previous chat sessions to personalize responses.

Connections. The chatbot module is connected to the AI API for natural language processing and the database for storing chat history. It also interacts with the authentication module to personalize chats depending on the user's access level.

2.2.3 Module for Checking Assignments by Material

This module processes the task materials uploaded by users to check their completion. It uses OCR technologies to recognize text from PDF or Word files.

Module functions:
- uploading materials by the user through the web application interface;
- use of algorithms to process and recognize text from files;
- automatically checking the results against the tasks and providing feedback.

Connections. The module interacts with the Material Processing API and the AI API, as well as with the Authentication module to verify user rights to perform tasks. The results of the checks are saved to a downloadable final file.

2.2.4 Test Generation Module

The module is responsible for the automatic creation of tests based on the training materials entered by the user. It uses text processing algorithms to analyze and structure the materials.

Module functions:
- uploading text materials or entering information through a form;
- create a template for export to Google Forms based on text analysis;
- generate tests with different types of questions (multiple choice, matching, short answers);
- ability to edit the created materials before exporting them.

Connections. The module uses internal text processing algorithms and interacts with the data saving module to export tests to various formats.

2.2.5 Module for Saving and Exporting Data

This module is responsible for saving user data and system results, as well as exporting information in user-friendly formats.

Module functions:
- saving chat history, results of task checks, and generated materials;
- providing users with access to the saved data through their personal account;
- exporting tests and reports to PDF, Word, or other convenient formats.

Connections. The module interacts with the database to store user information and with other modules to provide access to this data.

2.2.6 User Administration and Management Module

The module allows administrators to manage system users, their roles, and access rights. It provides control over system operation and configuration.

Module functions:

- adding, editing, and deleting users;
- assign roles and access rights (student, teacher, administrator);
- monitoring user activity and system status;
- configuring system parameters (for example, integration with new APIs or security settings).

Connections. The administration module integrates with the authentication module to manage roles and access rights, as well as with the database to store user information.

2.3 Selected Technologies and Tools

To develop an automated learning support system, it is important to choose effective and convenient tools and technologies that will allow you to achieve the maximum level of performance and functionality (Table 1).

2.4 UI

Main page for AI queries is shown on Figure 2.

The module for checking assignments based on uploaded materials is designed to greatly simplify the process of assessing student responses. Using the power of artificial intelligence and automation, this module allows teachers to quickly upload student answers (Fig. 3) and get an assessment of the results based on pre-provided templates of correct answers. This approach makes it easy to check even large volumes of tasks, which saves time and ensures objective assessment.

Figure 2: Main page for AI queries.

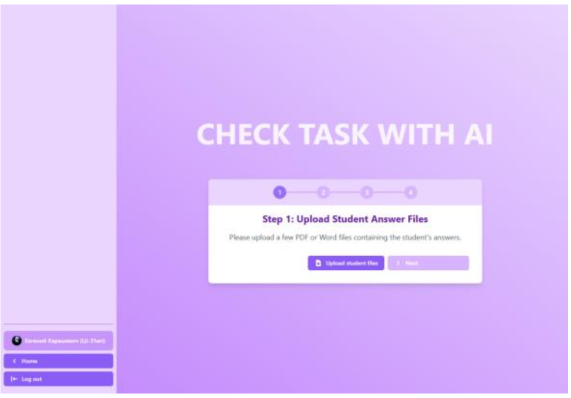


Figure 3: Main page for checking AI tasks.

The automatic test generation module is designed to help teachers quickly create tests based on educational materials (Fig. 4). Example of a test generation request with a template is shown in Figure 5, where request to get an answer from AI, generateQuiz.ts utility, is demonstrated.

Table 1: Selected technologies: purpose and advantages.

Technology	Purpose	Advantages
Figma	Design and prototyping	Real-time collaboration and commenting
WebStorm	Development environment	Tools for productive development and integration with Git
React	Interface creation	Dynamic components, high performance
Next.js	Server rendering	SEO optimization, fast page generation
Firebase	Authentication and data storage	Real-time, easy integration
OpenAI API	AI integration	Extension of system functions
Framer Motion	Animations	Smooth transitions, improved UX
Mammoth.js	Text extraction from Word files	Automatic conversion to text format
PDF-Parse	Working with PDF	Extending support for documents

Figure 4: Generated test using AI.

2.5 Evaluation and Testing of the System

Various analysis methods were used to evaluate the effectiveness of the developed automated learning support system. The evaluation was conducted in two main areas:

- quantitative indicators (productivity and efficiency metrics);
- qualitative indicators (user satisfaction, interaction analysis).

2.5.1 Quantitative Performance Indicators

The following parameters were selected:

1) Time to check assignments. 00The average time required for a teacher to check assignments manually was compared with the time taken by the automated system. The use of automated checking reduced grading time by 63%.

   Data processing time shows how quickly the system is able to perform operations such as loading, processing files, and generating results. Data for this metric can be obtained by measuring the duration of various operations in the system. For example, below is a Table 2 with execution time data for three main functions of the system: test generation, task evaluation, and file loading.

```
export default async function handler(
  req: NextApiRequest,
  res: NextApiResponse,
) {
  const { text, countQuestion } = req.body;

  const prompt = `!Return response only as json format for Google
forms!
  Create a ${countQuestion ?? "5"} questions (!not more!, !not less!)
quiz based on the following material
  Material:  \n\n${text}

  !!!Template should be like this only!!!:
  {
    title: "",
    description: "",
    questions: [
      {
        type: "multipleChoice",
        question: "",
        options: ["", "", ""],
        correctAnswer: "",
      },
      {
        type: "shortAnswer",
        question:"",
        answer: "",
      },
      ...other questions
    ],
  }`;

  try {
    const response = await openai.chat.completions.create({
      model: "gpt-4",
      messages: [{ role: "user", content: prompt }],
      max_tokens: 1500,
    });
    const result = response.choices[0].message?.content;
    res.status(200).json({ quiz: result });
  } catch (error) {
    res.status(500).json({ error: "Failed to generate quiz" });
  }
}
```

Figure 5: Test generation request with a template.

Table 2: Data processing time in the developed system.

Operation	Min. processing time, sec	Average processing time, sec	Max. processing time, sec
Upload 1 file	1.5	2	3.3
Upload 10 files	3	5	7.1
Generate a test based on 1 file	7.3	10.5	15.0
Grade tasks for 1 file	3.5	4.6	6.9
Grade tasks for 10 files	7.2	8.8	12.5

2) Accuracy of grading answers. The correspondence between the system's assigned grades and those given by teachers was analyzed. The system correctly evaluated 92% of test tasks and 85% of open-ended responses (considering semantic analysis).

   To assess the accuracy, a comparative analysis of the results calculated by the system and the control data can be performed. An example of accuracy data is given in Table 3.

Table 3: Data processing time in the developed system.

Test file	Number of questions	System (correct answers – 3 attempts)	Assessment accuracy, %
English test	15	14-15	94%
Ukrainian test	20	18-19	93%
Grade 4 math test	15	15	100%
Grade 9 math test	15	13-15	95%
Higher math test	10	7-8	70-80%
Programming test	10	8-9	80-90%

3) Test generation speed. The system generated test tasks from uploaded educational materials within 5-10 seconds, significantly reducing the time needed compared to manual preparation. The number of saved tests reflects the efficiency of automation, showing how much faster the system generates tests compared to a manual process. Creating one test manually takes an average of 1 to 3 hours, depending on the volume and complexity of the training materials. The system is able to generate tests in a much shorter time - from a few seconds to a few minutes, depending on the size of the files and the number of questions. This provides significant time savings, as shown in the Table 4 below.

Table 4: Time saving when creating tests automatically.

Volume of educational materials	Average time to prepare a test manually	Average test generation time by the system	Time savings, %
Small (5-10 pages)	1 hour	10-20 seconds	98%
Medium (10-20 pages)	2 hours	30-60 seconds	96%
Large (20-50 pages)	3 hours	1-2 minutes	96%

To ensure the validity of these results, a controlled experimental setup was used. The evaluation process included a comparison of system performance across different educational disciplines and teaching styles. Statistical significance was assessed using t-tests to determine the reliability of the improvements observed.

2.5.2 Qualitative Performance Indicators

To assess the qualitative aspects, a survey was conducted among teachers.

1) Teacher satisfaction:
   - 82% of teachers noted that the system reduces their workload for checking assignments;
   - 75% of teachers rated the quality of automated grading as sufficiently accurate for use in the educational process.
2) Analysis of user interaction. The system monitored teacher activity on the platform, revealing a 40% increase in time spent using the system. This suggests that the interface and functionalities are practical and convenient.

3 CONCLUSIONS

The developed system effectively improves learning, personalizes the educational process and significantly reduces the workload of teachers, which makes it promising for widespread implementation. The introduction of AI has increased the accuracy of assessment and made the learning process more personalized. Student performance has improved due to adaptive learning and automatically generated tests. Feedback from students and teachers indicates high satisfaction with the system and its effectiveness in the learning process.

The developed automated learning support system unlike traditional LMS actively uses AI for adaptive learning, automatic task checking, and test generation, unlike platforms where the teacher has to check the work manually, the system can automatically evaluate answers and even analyze text assignments using NLP, the system can create unique tests based on training materials, which avoids templates and improves the quality of assessment.

Future developments could incorporate adaptive learning pathways based on student performance analytics, as well as tools to support teachers in curriculum planning and assessment. By tracking learning patterns and difficulties, the system could provide personalized recommendations for both students and educators, enhancing the overall learning experience. Additionally, improving AI models to specialize in specific educational fields and increasing the accuracy of analysis and results will further enhance the system’s effectiveness.

REFERENCES

[1] "AI Agents Revolutionize Personalized Learning in 2025," Rapid Innovation, Feb. 2025, [Online]. Available: https://www.rapidinnovation.io/post/ai-agents-for-personalized-learning-paths. [Accessed: 17-Feb-2025].

[2] Devops Automation, "AI Testing vs. Traditional Testing," Medium, Jan. 2024, [Online]. Available: https://medium.com/@devopsautomation93/ai-testing-vs-traditional-testing-374b658d8d1a. [Accessed: 17-Feb-2025].

[3] "Test Automation Strategy Guide: Best Practices & Checklist," TestRail, [Online]. Available: https://www.testrail.com/blog/test-automation-strategy-guide/. [Accessed: 17-Feb-2025].

[4] "Benefits and Examples of Automation in Education," Salient Process, [Online]. Available: https://salientprocess.com/blog/benefits-and-examples-of-automation-in-education/. [Accessed: 17-Feb-2025].

[5] "Automated Test Generation: Streamline Assessments for Smarter Learning," Coursebox, [Online]. Available: https://www.coursebox.ai/blog/automated-test-generation. [Accessed: 17-Feb-2025].

[6] "22 Top AI Quiz and Exam Generators to Test Your Students," Thinkific, [Online]. Available: https://www.thinkific.com/blog/ai-quiz-and-exam-generators/. [Accessed: 17-Feb-2025].

[7] "Exploring Automated Test Generation Techniques," Machinet, [Online]. Available: https://blog.machinet.net/post/exploring-automated-test-generation-techniques. [Accessed: 17-Feb-2025].

[8] N. A. Kumar and A. S. Lan, "Using Large Language Models for Student-Code Guided Test Case Generation in Computer Science Education," in Proc. 2024 AAAI Conf. on Artificial Intelligence, Proceedings of Machine Learning Research, vol. 257, pp. 170-179, 2024, [Online]. Available: https://proceedings.mlr.press/v257/kumar24.html.

[9] "Automation Challenges Factors: How to Overcome Them?" InventionGen, [Online]. Available: https://www.inventiongen.com/automation-challenges/. [Accessed: 17-Feb-2025].

[10] "Tailwind CSS," Tailwind Labs, [Online]. Available: https://tailwindcss.com. [Accessed: 17-Feb-2025].

[11] Prettier, "Prettier - Code formatter," [Online]. Available: https://www.npmjs.com/package/prettier. [Accessed: 17-Feb-2025].

[12] ESLint, "ESLint - Find and fix problems in your JavaScript code," [Online]. Available: https://www.npmjs.com/package/eslint. [Accessed: 17-Feb-2025].

The Relationship Between Artificial Intelligence and Critical Infrastructure Development: Bibliometric Analysis

Nataliia Trushkina
Research Center for Industrial Problems of Development of the National Academy of Sciences of Ukraine, Inzhenernyi Lane 1a, 61165 Kharkiv, Ukraine
nata_tru@ukr.net

Keywords: Critical Infrastructure, Critical Infrastructure Development, Artificial Intelligence, Cyber Threat, Cybersecurity, Information Security, Digital Space, Strategic Management, Risk Management, Digital Transformation, Information System, Information Technology, Bibliometric Analysis.

Abstract: Currently, the priority direction of the national economy in the world's countries is the formation of an appropriate system of protection and security of critical infrastructure, which is able to quickly respond to exogenous and endogenous threats and risks and flexibly adapt to new operating conditions using digital technologies and information systems. At the same time, in the last decade, the role of artificial intelligence as a powerful tool in the management system for the development of critical infrastructure will grow. Therefore, the purpose of this article is to identify the relationship between artificial intelligence and the development of critical infrastructure by characterizing the evolution of key patterns of scientific publications on this problem. To achieve this goal, a relevant sample of scientific articles was formed based on identifying periods of publication activity and bibliometric analysis of keyword matches to identify promising areas of research in this area. The formed sample of publications for the study includes 606 documents indexed by the international scientometric database Scopus for the period 2011-2025. Bibliometric analysis and visualization of its results were carried out using the VOSviewer software product. Based on the visualization maps, seven clusters were identified and characterized by the content coincidence of keywords in the publications and five stages of the evolutionary development of artificial intelligence technologies for the effective functioning of critical infrastructure. Based on the analysis of empirical data, the exponential growth of the number of publications on the selected issues was confirmed (the annual growth rate of the number of scientific papers on this topic is 23.5% for 2011-2024). The results of the analysis can be used in further research to substantiate and develop an algorithm of actions for rapid response to crisis phenomena and adaptation of the functioning of critical infrastructure facilities to new global challenges.

1 INTRODUCTION

In today's dynamic environment, the development of critical infrastructure is becoming an integral part of the national economy of the world's countries. This is due to the fact that disruption of the operation of critical infrastructure facilities (energy, agro-industrial, transport, telecommunications, information networks, energy and water supply systems, etc.) can lead to significant economic losses. It should be noted that these systems are very vulnerable to various threats – from climatic and geological hazards to industrial accidents, terrorist attacks, cyber threats, armed conflicts and military actions, which can cause a cascading negative impact on different levels of management.

In this regard, there is an extremely urgent need to search and apply innovative and management approaches, smart technologies, information and intellectual systems, qualitatively new methods, tools and mechanisms for ensuring the development of critical infrastructure facilities. One of these mechanisms is digital transformation, which becomes the basis for achieving the outlined goals of sustainable development of critical infrastructure in the context of global changes [1; 2; 3]. It is worth emphasizing that the rethinking of the principles of functioning of critical infrastructure facilities and fundamental changes in their activities occur through the creation of an appropriate ecosystem of artificial intelligence. According to Statista [4], the volume of the global artificial intelligence market increased from 93.3 to 184 billion dollars in 2020-2024, or by 97.2%. It is predicted that this growth will continue – the value of this indicator will be 243.7 billion dollars in 2025, 415.6 billion dollars in 2027, and 826.7 billion dollars in 2030 [4]. That is, the growth rate of the global artificial intelligence market will be 24.4% in 2020-2030.

According to AIPRM experts [5], the volume of public investment in artificial intelligence in the United States amounted to 328.5 billion dollars in 2019-2023. This is 195.8 billion dollars more than in China (132.7 billion), which took second place in the same period. The United Kingdom takes third place in the ranking with a value of this indicator of 25.5 billion dollars, which is 92.2% less than in the United States. India takes fourth place (16.1 billion), and Canada takes fifth place (12.5 billion dollars). In terms of "share of investment in artificial intelligence as a percentage of GDP," Singapore ranks first (15%), followed by Sweden (14.1%), the United States (12.9%), Estonia (10.9%), and the United Kingdom (8.3%) [5].

A study conducted in October 2024 by consulting firm PwC [6] found that 49% of global technology leaders indicated that artificial intelligence was fully integrated into their companies' core business strategy. The survey found that 63% of the most effective companies are increasing their cloud budgets to leverage Generative Artificial Intelligence (GenAI). And 34% of companies say that sustainability considerations are driving the expected budget increase. At the same time, 67% of successful companies are already realizing the value of using GenAI for innovative products and services. In addition, 73% of executives emphasize that they plan to use generative artificial intelligence in the future to make changes to their company's business model [6].

According to S. Ghimire [7], in the ever-changing landscape of critical infrastructure development, artificial intelligence is a transformative force that helps to achieve unprecedented efficiency, resilience, and innovation. By continuously monitoring the health of infrastructure components, artificial intelligence systems can predict maintenance needs, optimize repair schedules, and extend the life of critical assets. This proactive approach not only reduces costs, but also increases the reliability and security of infrastructure networks, making the development of critical infrastructure more sustainable.

As noted in the 2024 Workshop Report "Securing Critical Infrastructure in the Age of AI" by the Center for Security and Emerging Technology [8], the introduction of artificial intelligence can lead to the creation of more efficient systems, improved business operations, and better tools for detecting and responding to cyber threats. At the same time, artificial intelligence systems can cause new cyber threats [9], which suppliers of critical infrastructure facilities must deal with. And, first of all, it is necessary to assess the potential risks associated with the use of artificial intelligence in the involved sectors of critical infrastructure, which will allow monitoring and diagnosing the level of vulnerability of the system to critical failures, physical attacks and cyberattacks [8].

According to D. M. Gerstein & E. N. Leidy [10], artificial intelligence must be used to ensure national security and monitor critical infrastructure systems. However, a study conducted by NATO to monitor the ability of AI to protect critical infrastructure from cyberattacks found that AI can operate without human intervention, help identify patterns of cyberattacks on critical infrastructure and network activity, and detect malicious software to improve decision-making on defensive responses, while warning that many government agencies are neglecting critical infrastructure security, failing to implement most of the recommendations for its protection since 2010 [11].

Therefore, it is particularly relevant to substantiate the directions for ensuring cybersecurity and increasing the resilience of critical infrastructure facilities, as well as to study the relationship between artificial intelligence and the critical infrastructure development.

2 LITERATURE REVIEW

Various aspects of the development of critical infrastructure, substantiation of the conceptual foundations of modernization and theoretical and methodological approaches to increasing the efficiency of functioning in various sectors of the economy are among the scientific interests of many leading foreign scientists (C. Baudrit et al. [12]; S. Bruno, M. De Fino & F. Fatiguso [13]; D. Buhalis et al. [14]; C. Cath [15]; K. Dick et al. [16];

R. Doshi, N. Apthorpe & N. Feamster [17]; S. Feng et al. [18]; F. Filgueiras [19]; F. Santoso & A. Finn [20]; A. C. Serban & M. D. Lytras [21]; F. van der Vlist, A. Helmond & F. Ferrari [22] and others).

At the same time, paying tribute to the scientific achievements of scientists in the study of the selected issues, it should be noted that some issues of the development of critical infrastructure require further development and finding ways to solve them. And especially the solution of this problem is becoming more relevant due to the emergence of new challenges associated with the digital transformation of strategic sectors of the economy and the intensive use of artificial intelligence technologies.

Thus, this problem determined the purpose of this article, which is to identify the relationship between artificial intelligence and the development of critical infrastructure based on the characteristics of the evolution of key patterns of scientific publications on the selected topic.

3 METHODOLOGY

The theoretical and methodological basis of the study is the provisions of institutional theory, in particular the paradigm of evolutionary development; theories of systems, globalization, transaction costs, infrastructure; concepts of strategic management, national, information and cybersecurity, sustainable development. The study is based on systemic, structural-functional, linguistic, synergistic and logical-semantic approaches.

The information base of the study is analytical materials of AIPRM, Center for Security and Emerging Technology, CSO, Frost & Sullivan Institute, Homeland Security Operational Analysis Center, PwC, Statista, which highlight the results of surveys and statistical analysis on the problems of the impact of artificial intelligence on the management of the development of critical infrastructure.

The following general scientific methods were used in the research process: dialectical, historical, formal-logical, axiomatic, hypothetical-deductive, analysis and synthesis, induction and deduction, expert survey, bibliometric analysis, comparative analysis, analogy, classification, structural-logical generalization.

The study selected bibliometric analysis as a method that reveals the connection between artificial intelligence and the development of critical infrastructure. This type of analysis is based on the mathematical theory of graphs, clustering methods and scientific visualization, which makes it widely applicable in various fields of science [1].

Based on the structuring of a large volume of metadata of scientific publications, bibliometric analysis allows us to identify the essence of the subject area and its conceptual foundations and to substantiate the evolution of the research area [1]. The research methodology includes such main stages as data collection and analysis, selection of a visualization tool, graphical representation of the identified connections and interpretation of the results obtained.

This was implemented using the software product VOSviewer v.1.6.19. The functionality of this program involves creating keyword visualization maps based on compatibility data, maps of authors or countries based on the number of citations, etc. [3]. In addition, network visualization maps in VOSviewer are displayed in several ways (for example, by content criterion, by publication period) [1].

An important stage of bibliometric analysis, which ensures its quality, is the selection of a data source and the form of a relevant sample of publications. The Scopus database was chosen as a data source due to the breadth of its coverage of such subject areas as computer science; engineering; social sciences; energy; decision sciences; environmental sciences; business, management and accounting; economics, econometrics and finance, etc. To form the data sample, the chronology of the release of 642 publications was first determined for the search query “Artificial Intelligence” (or AI) and “Critical Infrastructure Development” for the years 1990-2025, and then the period of publication activity of the topic under study, namely from 2011 to 2025. Secondly, the sample was limited by the stage of publication, that is, only published works were taken. The key categories for selecting scientific publications were the title, abstract and keywords for them. Therefore, the studied sample of publications included 606 works that met the above criteria, published in the period 2011-2025 and indexed by the Scopus database.

4 RESULTS

Quantitative analysis of the formed sample of scientific works showed an exponential growth of research in the context of the relationship between artificial intelligence and the development of critical infrastructure during the period 2011-2025 (Fig. 1). On average, the growth rate of the number of publications was 23.5% for 2011-2024. At the same time, it can be assumed that by the end of 2025 there will also be a trend of significant growth in the number of publications on the specified topic compared to previous years. According to preliminary estimates, the annual growth rate of the number of publications for 2011-2025 will be 10.7%.

As the analysis shows, the following keywords are mostly used in the publications: Artificial Intelligence (364 documents), Machine Learning (92), Internet of Things (58), Decision-Making (54), Decision Support Systems (52), Sustainable Development (49), Critical Infrastructures (41), Cybersecurity (40), Network Security (33), Risk Assessment (28), Learning Systems (28), Big Data (24), Climate Change (23), Blockchain (23), Automation (23), Sustainability (22), Smart City (22), Information Management (21), Learning Algorithms (20), Investments (20), 5G Mobile Communication Systems (20), Risk Management (19), Digital Storage (19), AI (19), Cost Effectiveness (17), Infrastructure (15 documents) etc.

In addition, it is advisable to study the geographical structure of scientific works on the development of critical infrastructure using artificial intelligence technologies (Fig. 2). According to the geographical structure, the leaders are countries such as United States (134 documents), India (79), China (61), United Kingdom (52), Italy (36), Germany (26), South Africa (24), Spain (23), Saudi Arabia (22), Canada (21), France (17), Malaysia (17), Netherlands (15 documents).

However, the overwhelming number of publications by researchers from India, Canada, China, and Italy is inferior in terms of citation level to the USA, Great Britain, and France. Thus, each article by scientists from the USA, Great Britain, and France is cited on average 20 times. This is more than 4 times the value of this parameter for representatives from India (on average 5 times), 1.5 times for scientists from Canada (on average 13 times), 1.3 times for China and Italy (on average 15 times).

In our opinion, it is important to determine the most cited works devoted to the study of the impact of artificial intelligence on the development of critical infrastructure (Table 1). According to the data in Table 1, the work of D. Buhalis et al. [14], published in 2019, has the largest number of citations. And to date, this work has 636 citations. This work is devoted to the study of technological changes (Internet of Things, autonomous devices, advanced analytical capabilities (artificial intelligence), virtual and augmented reality), which create a smart environment that transforms industry structures and processes of critical infrastructure facilities.

The second place is taken by the publication of the authors R. Doshi, N. Apthorpe, N. Feamster [17], which considers new methods for automatic detection of consumer traffic of IoT attacks on critical information infrastructure.

In third place is the publication by C. Cath [15], which has been cited 279 times. In this article, the author pays attention to the ethical, legal and technical possibilities of artificial intelligence, as well as the risks and challenges of its application in various areas of economic activity.

Table 2 shows the Top 5 journals with the largest number of publications on artificial intelligence and critical infrastructure development, indexed in the international scientometric database Scopus. The largest number of articles was published in the journal “Lecture Notes in Computer Science including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics”. The first article on the topic was published in 2008 and is devoted to the study of the issue of developing adaptive software to be used in high-level reliability systems, such as critical infrastructure management and protection systems [23].

Among universities and scientific institutions whose researchers have studied the impact of artificial intelligence on the development of critical infrastructure, the leading position is occupied by University of Johannesburg (10 documents), Southeast University (6), Tsinghua University (6), Virginia Polytechnic Institute and State University (6), Stanford University (5), National Technical University Kharkiv Polytechnic Institute (5), ETH Zürich (5), University of Cambridge (5), Virginia Tech College of Engineering (5 documents).

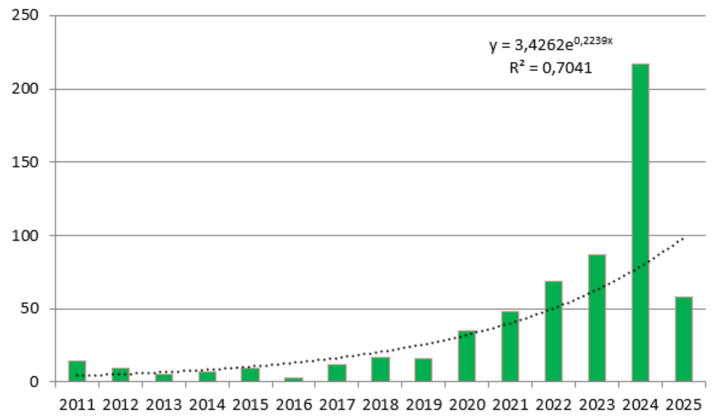


Figure 1: Dynamics of publications on artificial intelligence and critical infrastructure development for 2011-2025.

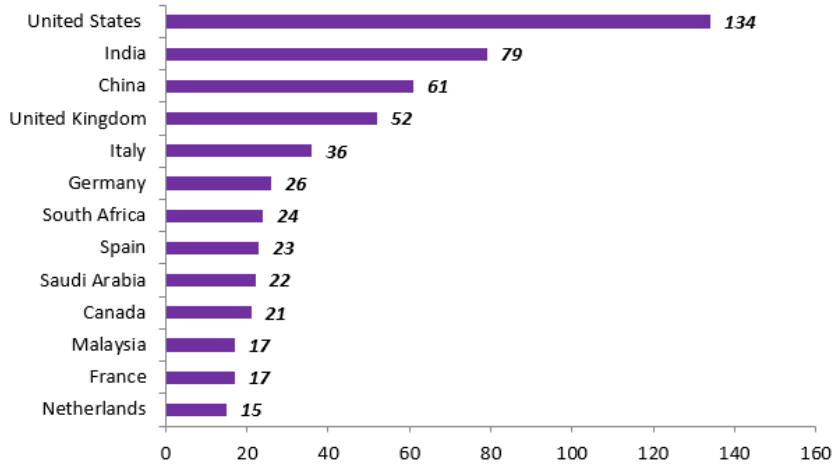


Figure 2: Geographic structure of scientific publications search query "artificial intelligence" and "critical infrastructure development".

Table 1: Ranking of scientific papers by number of citations [13; 14; 15; 17; 18].

Title	Author(s)	Year	Source	Number of citations
Technological disruptions in services: lessons from tourism and hospitality	D. Buhalis et al.	2019	Journal of Service Management	636
Machine learning DDoS detection for consumer internet of things devices	R. Doshi, N. Apthorpe, N. Feamster	2018	2018 IEEE Symposium on Security and Privacy Workshops	602
Governing artificial intelligence: Ethical, legal and technical opportunities and challenges	C. Cath	2018	Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences	279
Dense reinforcement learning for safety validation of autonomous vehicles	S. Feng et al.	2023	Nature	240
Historic Building Information Modelling: performance assessment for diagnosis-aided information modelling and management	S. Bruno, M. De Fino, F. Fatiguso	2018	Automation in Construction	224

Table 2: Top 5 journals with the largest number of publications on the selected research topic, indexed in the scientometric database Scopus.

Journal Title	Indexing Period	Publisher	Field of Knowledge	Cite Score 2023	SJR 2023	SNIP 2023	Number of Articles
Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	from 1973 to the present	Springer Nature	Computer Science: General Computer Science; Mathematics: Theoretical Computer Science	2.6	0.606	0.590	16
Lecture Notes in Networks and Systems	from 2016 to the present	Springer Nature	Computer Science: Signal Processing; Engineering: Control and Systems Engineering; Computer Science: Computer Networks and Communications	0.9	0.171	0.282	11
Sustainability (Switzerland)	from 2009 to the present	Multidisciplinary Digital Publishing Institute (MDPI)	Social Sciences: Geography, Planning and Development; Computer Science: Computer Networks and Communications, Hardware and Architecture; Environmental Science: Management, Monitoring, Policy and Law; Energy: Renewable Energy, Sustainability and the Environment etc.	6.8	0.672	1.086	10
ACM International Conference Proceeding Series	1993, from 1996 to 1997, from 1999 to 2023	-	Computer Science: Computer Networks and Communications, Computer Vision and Pattern Recognition, Software, Human-Computer Interaction	1.5	0.253	0.233	9
IEEE Access	from 2013 to the present	IEEE	Engineering: General Engineering; Computer Science: General Computer Science; Materials Science: General Materials Science	9.8	0.960	1.440	8

The institutions that most fund research by scientists on critical infrastructure development problems using artificial intelligence tools include the following: European Commission (22 documents), National Natural Science Foundation of China (20), UK Research and Innovation (17), National Science Foundation (15), Ministry of Science and Technology of the People's Republic of China (14), Horizon 2020 Framework Programme (12), U.S. Department of Defense (7), Horizon 2020 (6), U.S. Department of Energy (6 documents).

The ranking of scientific papers by document type is given in Table 3. As we can see, scientists mostly test the obtained research results during conferences of various levels and highlight them in scientific articles.

Table 3: Number and share of scientific publications by document type.

Type of publication	Number of publications	Share of publications, %
Conference Paper	248	38.6
Article	224	34.9
Review	74	11.5
Book Chapter	60	9.3
Conference Review	15	2.3
Book	13	2.0

Most scientific papers devoted to identifying the impact of artificial intelligence on the development of critical infrastructure are published in the following fields of knowledge: Computer Science (342 documents), Engineering (296), Social Sciences (115), Energy (81), Mathematics (80), Environmental Science (72), Medicine (52), Decision Sciences (49), Business, Management and Accounting (46 documents) (Table 4). All this indicates the multifaceted and multidisciplinary nature of the chosen research topic.

Table 4: Share of scientific publications by key fields of knowledge.

Field of knowledge	Share of scientific publications, %
Computer Science	24.8
Engineering	21.4
Social Sciences	8.3
Energy	5.9
Mathematics	5.8
Environmental Science	5.2
Medicine	3.8
Decision Sciences	3.6
Business, Management and Accounting	3.2

Based on the results of the analysis of the coincidences and closeness of the relationship between the keywords of the selected sample of publications, network visualization maps were constructed (Fig. 3 and Fig. 4), and 7 clusters on the studied topic were identified and characterized (Fig. 3).

The first cluster (red colour, Fig. 3) contains the largest number of terms (namely 48 positions), among which the following can be mentioned: “artificial intelligence”, “decision support system”, “information technology”, “decision-making”, “risk management”, “information systems”, “infrastructure development”, “disasters”, “project management”, “critical challenges”, “critical success factor”, “telecommunication networks”. It is worth noting the keyword “Artificial Intelligence”, the frequency of its co-use in the studied sample is 364, and the strength of the connection is 1945.

The second cluster (46 concepts, green colour. Fig. 3) combines such terms as “AI”, “climate change”, “cost effectiveness”, “digital technologies’, “energy policy”, “energy management”, “Industry 4.0”, “IoT”, “machine learning”, “renewable energy”, “smart city”, “smart grid”, “strategic planning”, “sustainable development”. In this cluster, the keyword “Internet of Things” has the highest frequency – the ratio is 58, while the strength of the association is 393.

The third cluster (45 elements, blue colour, Fig. 3) describes the connection between artificial intelligence and the critical infrastructure development with the following terms: “5G mobile communication system”, “artificial intelligence technologies”, “blockchain”, “cybersecurity”, “national security”, “network security”, “security systems”. The main keyword in this cluster is “Cybersecurity”, the frequency of its co-use in the studied sample of scientific publications is 40, and the strength of the association is 290.

The fourth cluster (41 elements, yellow colour, Fig. 3) is associated with the following categories: “algorithms”, “Big Data”, “data analysis”, “diagnosis”, “industrialization”, “infrastructure”. The keyword “Human” has the highest frequency of co-use – 49, while the strength of the association is 424.

The fifth cluster (41 elements, purple colour, Fig. 3) covers the following categories: “anomaly detection”, “cloud computing”, “cloud platforms”, “deep learning”, “digital storage”, “digital transformation”, “digitalization”, “intelligent systems”, “life cycle”, “predictive analytics”, “real time systems”, “transport infrastructure”. The frequency of co-use of the main keyword “Machine Learning” in this cluster is 92, and the strength of association is 604.

The sixth cluster (38 elements, turquoise colour, Fig. 3) is associated with the concepts of critical infrastructure, critical infrastructure protection, risk analysis, digital twin, generative AI, risk assessment, AI systems. The seventh cluster (16 categories, orange colour, Fig. 3) includes such concepts as automation, computer crime, cyber physical system, damage detection etc.

Therefore, based on the constructed terminological map of categories and the most significant keywords identified related to artificial intelligence issues in critical infrastructure development management, it can be stated that the areas under study are multidimensional and interdependent, since there are numerous connections between terms, as well as their high prevalence in research.

According to the results of bibliometric analysis in the evolutionary and temporal dimensions, it can be stated that in the development of scientific research related to the development of critical infrastructure using artificial intelligence tools, five most significant stages can be distinguished (Fig. 4).

The first stage of development was observed until 2016, when most publications considered general issues of forming networks and decision support systems using data analysis technologies.

In the second stage, which lasted from 2016 to 2018, the focus of the scientists' research shifted to the terms “critical infrastructure”, “information management”, “decision-making”.

From 2018 to 2020, that is, in the third of the selected stages, the dominant key terms were “artificial intelligence”, “robotics”, and “metadata”. The fourth stage (from 2020 to 2022) is characterized by the predominance of the terms – “Big Data”, “human”, “intrusion detection systems”.

The last, fifth stage began from 2022 to 2024. According to the results of its analysis, it can be stated that the main terms in the study were: “cybersecurity”, “cyber threats”, “smart grid”, “machine learning”, “IoT”, “blockchain”, “denial-of-service attack” and others.

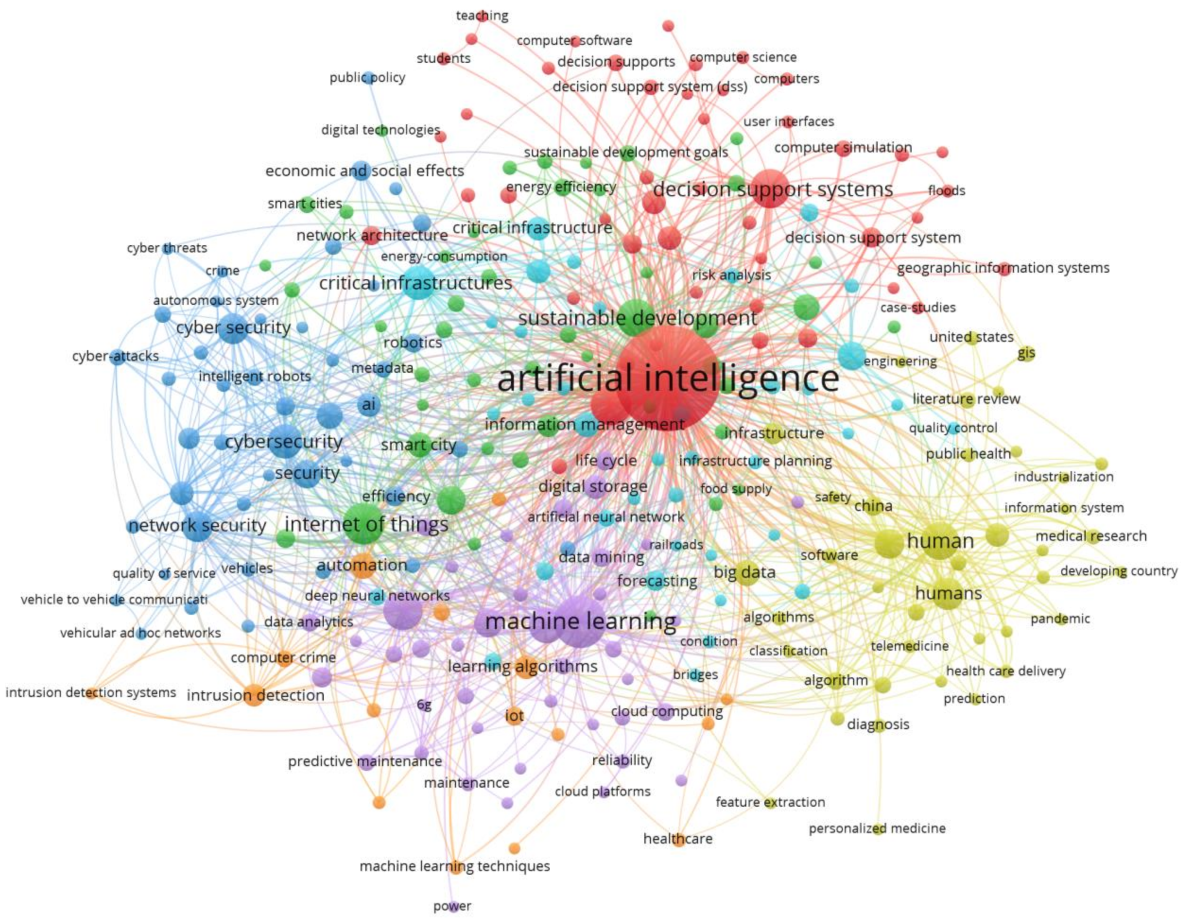


Figure 3: Visualization map of bibliometric analysis of publications that highlight the impact of artificial intelligence on the development of critical infrastructure (content aspect).

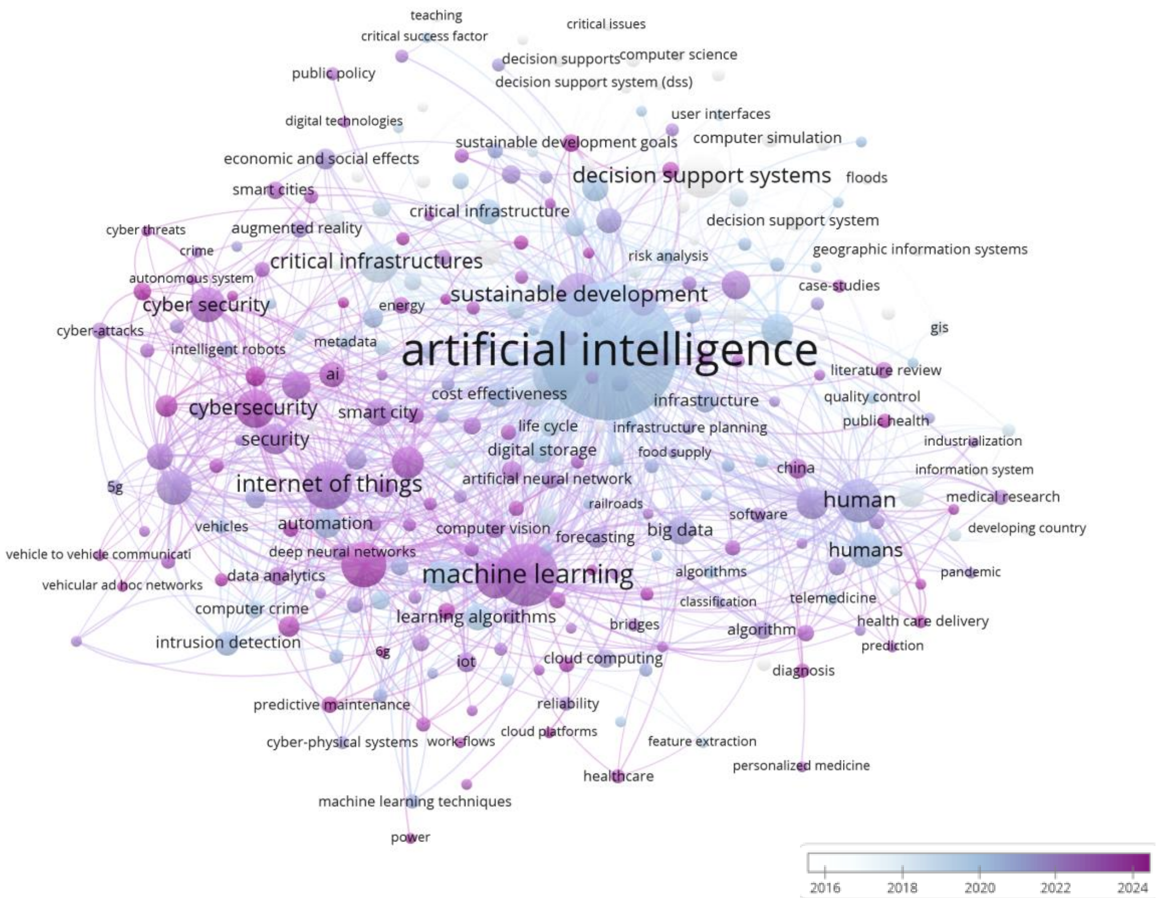


Figure 4: Visualization map of bibliometric analysis of publications that highlight the impact of artificial intelligence on the development of critical infrastructure (evolutionary and temporal aspects).

Thus, summing up the above, we can trace a change in emphasis in scientific publications, caused by the development and improvement of information technologies (from the already traditional ones: “Internet”, “software”, “information systems”, common at the first – third stages (Fig. 4) to the increasing importance of artificial intelligence tools at the fourth – fifth stages. The widespread use of modern technologies has led to the emergence of new phenomena and areas of scientific research, such as “information security management”, “critical infrastructure development management”, “cybersecurity”, “artificial intelligence”, “machine learning”, etc., which determines the importance of simultaneous study of the subject area in the context of digital transformations.

5 CONCLUSIONS

Based on the purpose and results of the study, we can conclude that there is a high level of closeness between the concepts of “Artificial Intelligence” and “Critical Infrastructure Development”. Thus, empirical data showed an exponential growth of research in the context of the relationship between artificial intelligence and the development of critical infrastructure (the annual growth rate of the number of scientific papers on this topic is 23.5% for 2011-2024, and 10.7% for 2011-2025). Therefore, it can be assumed that, in accordance with the dialectical law, the accumulated quantitative changes will turn into qualitative ones, which will lead to the complementarity and mutual stimulation of these factors.

This article provides a bibliometric analysis of scientific publications indexed in the international scientometric database Scopus, which highlight aspects of the application of artificial intelligence technologies to improve the efficiency of the development of critical infrastructure. This analysis made it possible to identify current trends in publication activity on the selected research topic. Using the VOSviewer software, network visualization maps of keyword matches of publications indexed by the international scientometric database Scopus from 2011 to 2025 were created.

Based on the semantic correspondence of the keywords of the studied sample, seven clusters were identified and described, and the presence of five most significant stages of the development of scientific research dedicated to identifying the relationship between artificial intelligence and the development of critical infrastructure was established.

This, in turn, provides an opportunity to expand theoretical knowledge on managing the development of critical infrastructure as an object of economic research. In addition, it also gives impetus to the development of a modern methodology for the formation and implementation of a strategy for the protection and security of critical infrastructure objects using artificial intelligence technologies and information systems.

Prospects for further research are to substantiate the feasibility of applying an integrated approach to managing the risks of critical infrastructure development in a changing security environment.

REFERENCES

[1] A. Kwilinski, "The relationship between sustainable development and digital transformation: Bibliometric analysis," Virtual Econ., vol. 6, no. 3, pp. 56–69, 2023, [Online]. Available: https://doi.org/10.34021/ve.2023.06.03(4).

[2] D. Pudryk et al., "Towards achieving sustainable development: Interactions between migration and education," Forum Sci. Oecon., vol. 11, pp. 113–132, 2023, [Online]. Available: https://doi.org/10.23762/FSO_VOL11_NO1_6.

[3] V. Khaustova, M. Kyzym et al., "Digital transformation of energy infrastructure in the conditions of global changes: Bibliometric analysis," in Proc. 12th Int. Conf. Appl. Innov. IT (ICAIIT), Koethen, Germany, Mar. 7, 2024, vol. 12, no. 1, pp. 135–142, [Online]. Available: http://dx.doi.org/10.25673/115664.

[4] Artificial intelligence (AI) market size worldwide from 2020 to 2030 (in billion U.S. dollars), Statista, Nov. 28, 2024, [Online]. Available: https://www.statista.com/forecasts/1474143/global-ai-market-size.

[5] AI Statistics 2024, AIPRM, [Online]. Available: https://www.aiprm.com/ai-statistics/.

[6] 2025 AI Business Predictions, PwC, 2025, [Online]. Available: https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-predictions.html.

[7] S. Ghimire, "The role of AI in infrastructure development of developed nations," Frost & Sullivan Institute, Jan. 22, 2025, [Online]. Available: https://frostandsullivaninstitute.org/the-role-of-ai-in-infrastructure-development-of-developed-nations/.

[8] K. Crichton et al., Securing Critical Infrastructure in the Age of AI. Workshop Report, Center for Security and Emerging Technology, Oct. 2024, [Online]. Available: https://cset.georgetown.edu/wp-content/uploads/CSET-Securing-Critical-Infrastructure-in-the-Age-of-AI.pdf.

[9] A. Kwilinski and N. Trushkina, "Impact of cyber risks and threats on the critical infrastructure development: Visualization of scientific research," in Proc. 12th Int. Conf. Appl. Innov. IT (ICAIIT), vol. 12, no. 2, pp. 107–119, 2024, [Online]. Available: http://dx.doi.org/10.25673/118123.

[10] D. M. Gerstein and E. N. Leidy, Emerging Technology and Risk Analysis Artificial Intelligence and Critical Infrastructure, Homeland Security Operational Analysis Center, 2024, [Online]. Available: https://www.rand.org/content/dam/rand/pubs/research_reports/RRA2800/RRA2873-1/RAND_RRA2873-1.pdf.

[11] M. Hill, "NATO tests AI's ability to protect critical infrastructure against cyberattacks," CSO, Jan. 5, 2023, [Online]. Available: https://www.csoonline.com/article/574281/nato-tests-ai-s-ability-to-protect-critical-infrastructure-against-cyberattacks.html.

[12] C. Baudrit et al., "Artificial intelligence and non-destructive testing data to assess concrete sustainability of civil engineering infrastructures," Materials, vol. 18, no. 4, p. 826, 2025, [Online]. Available: https://doi.org/10.3390/ma18040826.

[13] S. Bruno, M. De Fino, and F. Fatiguso, "Historic Building Information Modelling: Performance assessment for diagnosis-aided information modelling and management," Autom. Constr., vol. 86, pp. 256–276, 2018, [Online]. Available: https://doi.org/10.1016/j.autcon.2017.11.009.

[14] D. Buhalis et al., "Technological disruptions in services: Lessons from tourism and hospitality," J. Serv. Manag., vol. 30, no. 4, pp. 484–506, 2019, [Online]. Available: https://doi.org/10.1108/JOSM-12-2018-0398.

[15] C. Cath, "Governing artificial intelligence: Ethical, legal and technical opportunities and challenges," Philos. Trans. A Math. Phys. Eng. Sci., vol. 376, no. 2133, p. 20180080, 2018, [Online]. Available: https://doi.org/10.1098/rsta.2018.0080.

[16] K. Dick et al., "Deep learning for critical infrastructure resilience," J. Infrastruct. Syst., vol. 25, no. 2, p. 05019003, 2019, [Online]. Available:

https://doi.org/10.1061/(ASCE)IS.1943-555X.0000477.

[17] R. Doshi, N. Apthorpe, and N. Feamster, “Machine learning DDoS detection for consumer internet of things devices,” in Proc. 2018 IEEE Symp. Secur. Priv. Workshops, pp. 29–35, 2018, [Online]. Available: https://doi.org/10.1109/SPW.2018.00013.

[18] S. Feng et al., “Dense reinforcement learning for safety validation of autonomous vehicles,” Nature, vol. 615, no. 7953, pp. 620–627, 2023, [Online]. Available: https://doi.org/10.1038/s41586-023-05732-2.

[19] F. Filgueiras, “Artificial intelligence governance challenges in Latin America: Infrastructure, decolonization and new dependency,” Reforma y Democracia, no. 87, pp. 44–70, 2023, [Online]. Available: https://doi.org/10.69733/clad.ryd.n87.a3.

[20] F. Santoso and A. Finn, “An in-depth examination of artificial intelligence-enhanced cybersecurity in robotics, autonomous systems, and critical infrastructures,” IEEE Trans. Serv. Comput., vol. 17, no. 3, pp. 1293–1311, 2024, [Online]. Available: https://doi.org/10.1109/TSC.2023.3331083.

[21] A. C. Serban and M. D. Lytras, “Artificial intelligence for smart renewable energy sector in Europe – Smart energy infrastructures for next generation smart cities,” IEEE Access, vol. 8, pp. 77364–77377, 2020, [Online]. Available: https://doi.org/10.1109/ACCESS.2020.2990123.

[22] F. van der Vlist, A. Helmond, and F. Ferrari, “Big AI: Cloud infrastructure dependence and the industrialisation of artificial intelligence,” Big Data Soc., vol. 11, no. 1, 2024, [Online]. Available: https://doi.org/10.1177/20539517241232630.

[23] H. J. Goldsby, B. H. C. Cheng, and J. Zhang, “AMOEBA-RT: Run-time verification of adaptive software,” Lect. Notes Comput. Sci., vol. 5002, pp. 212–224, 2008, [Online]. Available: https://doi.org/10.1007/978-3-540-69073-3_23.

SECTION 4

POWER ENGINEERING

Proceedings of the 13th International Conference on Applied Innovations in IT (ICAIIT)

APRIL 2025

Analysis of Electrical Parameters of a Metal Collector Photovoltaic-Thermal Device

Boysori Yuldoshov[1], Azamat Narbayev[1], Tulkin Buzrukov[2],
Gayrat Raimov[1], Jasur Khaliyarov[1], Sojida Raupova[3] and Sirojiddin Toshpulatov[1]
[1]Termez State University, Barkamol avlod Str. 43, Termez, Uzbekistan
[2]Termiz University of Economics and Service, Farovon Str. 41B, Termez, Uzbekistan
[3]Termiz State Pedagogical Institute, At-Termiziy Str. 1A, Termez, Uzbekistan
b.yuldoshov10@mail.ru, narbaev@ro.ru, tulqin_buzrukov@tues.uz, raimov.gayrat@mail.ru, xjxjasur@mail.ru, sojidaraupova38@gmail.com toshpulatovs@tersu.uz.

Keywords: Renewable Energy, Solar Energy, Photovoltaic-Thermal Device, Metal Heat Collector, Electrical and Thermal Efficiency, Short Circuit Current, Open Circuit Voltage, Electric Power, Energy, Reflector.

Abstract: A photovoltaic-thermal device (PVTD) is developed to reduce energy losses due to high temperatures in photovoltaic batteries (PVBs) and to obtain thermal energy. To enhance heat exchange with the heat-carrying fluid, a copper absorber and a metal heat collector (HC) were used in the PVTD. Additionally, reflectors were integrated to increase the amount of radiation energy incident on the PVB surface. In experiments conducted under natural conditions, the electrical power outputs of PVB and PVTD were compared. Based on experimental results, the key electrical parameters of the PVB and the PVTD were analyzed. The PVTD included a metal heat collector and reflectors to enhance cooling and solar radiation intensity. As a result, the open circuit voltage of the PVTD was 1V higher, and the short circuit current was 1.56 times greater than those of the PVB. Additionally, the average electrical power values for the PVTD and PVB were recorded as 70.3W and 43W, respectively.

1 INTRODUCTION

The utilization of solar energy worldwide is increasing year by year, and this trend is expected to continue in the future [1, 2]. Additionally, the efficiency and technical solutions for using solar energy effectively are advancing annually [3, 4]. The development of science and manufacturing technologies also plays a significant role in the progress of this field. The creation of devices capable of generating both electrical and thermal energy simultaneously is of great importance [5, 6]. During the operation of PVBs, cooling and reflectors are widely used to prevent the decline in electrical efficiency due to high temperatures and insufficient solar radiation [7]. Extensive research is being conducted globally to optimize the shape, material, and other physical properties of heat collectors attached to the back of PVBs to enhance cooling efficiency [8, 9]. Studies also focus on improving system energy efficiency by increasing the density of solar energy incident on the PVB surface through reflectors [10, 11]. Cooling of PVBs is typically achieved using air, liquid, or nanofluid-based systems [12-16]. The geometric and material selection of heat collectors is also crucial [17-19] since the flat backside of PVBs necessitates maximum contact area with the heat collector to improve heat exchange. The choice of material with high thermal conductivity further enhances system efficiency. Ensuring effective heat exchange between components leads to simultaneous improvements in electrical and thermal efficiency. When the system is designed to extract thermal energy, the heat collector must be optimized in all aspects [20].

2 METHODS AND MATERIALS

2.1 Experimental Device

A 60W PVB was selected for the experiment and equipped with a metal heat collector. A thermal paste was used to attach the HC to the PVB. The electrical and additional parameters of the PVB are provided in Table 1, and the characteristics of the thermal paste are listed in Table 2.

Table-1: Physical parameters of photoelectric battery.

Parameters	Size
Geometric size	
PVB surface, S_{PVB}	0.33m^2
PVB frame width, d	2.5sm
Back cover thickness, d_q	4mm
Electrical parameters	
Electric power, P_{max}	60W
η	18,36%
Salt circuit voltage, $U_{o.c.}$	21,6V
Short circuit current, $I_{s.c.}$	3,53A

Thermal paste was used to improve heat transfer between parts and create good thermal contact.

Table-2: Physical parameters of thermal paste.

Parameter	Numerical value
Working temperature range	from -50°C to 180°C
Thermal conductivity	3–4.5 W/m·K
Average density	2.5 g/cm³
Validity period	4-5 years

The components of the PVTD are shown in Figure 1. The heat collector was made from steel metal.

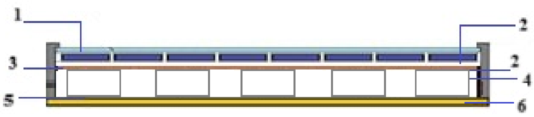


Figure 1: The structure of PVTD. 1 – PVB, 2 – thermal paste, 3 – copper absorber, 4 – metal HC, 5 – foil insulation, 6 – back cover.

The copper absorber serves to ensure uniform heat exchange between the PVB and HC. The absorber thickness is 0.5mm. The general structure of the HC is shown in Figure 2. Five profiles are used, allowing water to move from the bottom to the top. The profile dimensions are 5×2.5cm, with an inter-profile distance of 3.5cm.

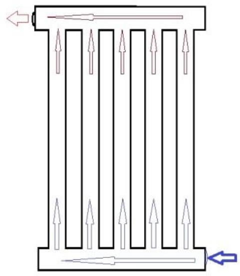


Figure 2: General view of the HC.

The system includes two reflectors, which can be adjusted via brackets as needed. The total reflector surface area is twice that of the PVB surface, with a reflection coefficient of ~ 0.75. The PVB used in the system has a power rating of 60W and consists of 32 solar cells.

2.2 Experimental Method and Basic Measuring Devices

During the experiment, short circuit current, open circuit voltage, solar radiation, inlet and outlet water temperatures, and cooling water mass flow rate were measured. The results were used to create graphs and tables, leading to conclusions.

Short circuit current and open circuit voltage values were measured using UNI-T (UT52 and UT89X) multimeters, while solar radiation was determined using a DT-1307 (Solar Power Meter) device. The parameters of the measurement device are listed in Table 3.

Table 3: DT-1307 Solar Power Meter.

Specifications	
Display	3-1/2 digits LCD with maxmum reading 1999.
Range	1999W/m^2,634BTU/(ft^2·h)
Resolution	1W/m^2;1BTU/(ft^2·h).
Accuracy	typically within ±10W/m^2 [±3BTU/(ft^2·h)]or ±5%
Sampling time	Approx 0.25 second.

Variable resistors (potentiometers) are used to obtain volt-ampere (VAC) and volt-watt characteristics (VWC).

3 RESULTS AND DISCUSSION

The experiments were conducted in September 2024 under natural conditions at the heliopolygon of the Physics-Technical Institute, with measurements taken at 20-minute intervals. The general setup of the PVB and PVTD during the experiment is shown in Figure 3.

Figure 3: General view of PVB and PVTD during the experiment.

The short circuit current time dependence of PVB and PVTD is shown in Figure 4. From the graph, it can be seen that the increase in solar radiation falling on the PVTD surface due to reflectors led to a 1.5-1.7-fold increase in short circuit current. Since the short circuit current is proportional to the solar radiation falling on the surface, the reflectors increased the average current by 56.64%. Despite the fact that the short circuit current of the 60W FEB is 3.53A, the average short circuit current of the PVTD is 4.28A, and for the PVB this indicator is 2.73A.

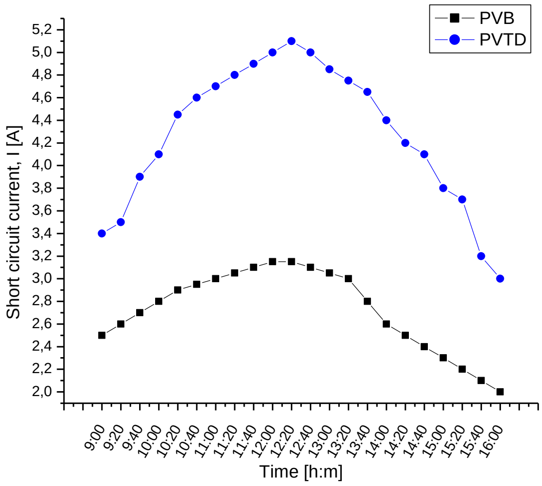


Figure 4: Graphs of short circuit currents of PVB and PVTD over time.

Open circuit voltage versus time is shown in Figure 5. Cooling reduced electrical losses, and additional thermal energy was obtained from the temperature difference between incoming and outgoing water.

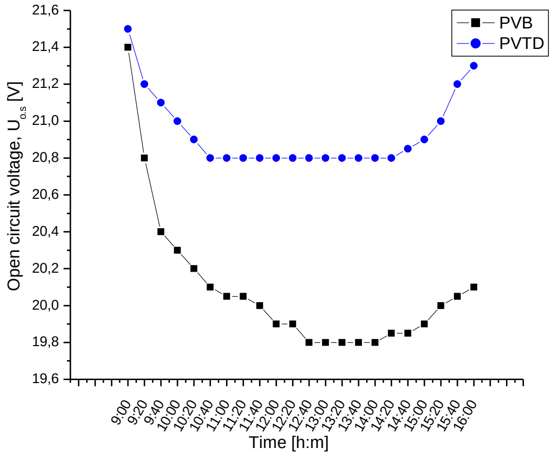


Figure 5: Graphs of the time dependence of the PVB and PVTD's open circuit voltage.

As a result of cooling, the PVTD open circuit voltage was a minimum of 20.8V, while in PVB this value was 19.8V. The difference in open circuit voltage, and electrical power is clearly visible in the volt-ampere characteristic (VAC) and volt-watt characteristic (VWC) of the devices (Fig. 6). Two multimeters (UNI-T brand UT89X and UT52) and a reostat were used to measure VAX and VVX. Figures 6 and 7 depict the VAC and VWC of PVB and PVTD measured over the same period of time under natural conditions.

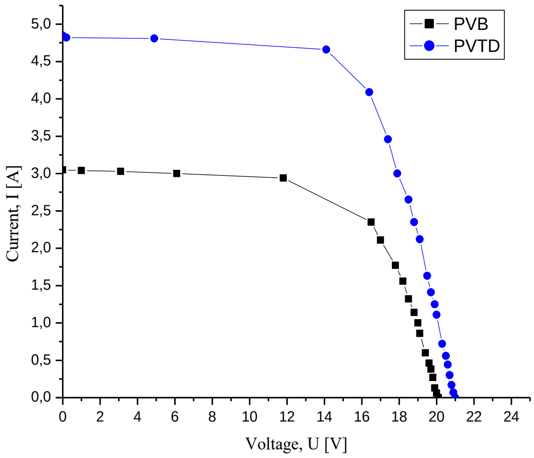


Figure 6: PVB and PVTD VAC.

Due to the cooling and reflectors of the PVTD, the voltage was improved by 1V and the current by 1.6A. As a result, the electrical power was 38.78W

(900W/m^2) for the FEB and 67.1W (1500W/m^2) for the PVTD.

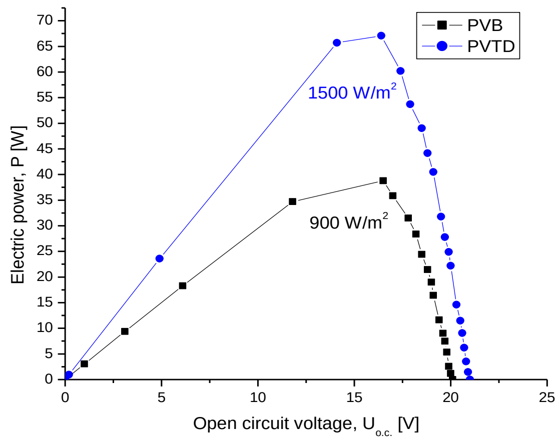


Figure 7: PVB and PVTD VWC.

At the same time, when the PVB power reached 37W, the PVTD power was around 68W. The PVB and PVTD electrical power indicators are displayed in a single graph for the purpose of analyzing the efficiency of reflectors and cooling (Figure 8). The maximum power in PVTD was 83.38W, and in PVB it was 49.27W. The average power values showed values of 70.3W and 43W, respectively.

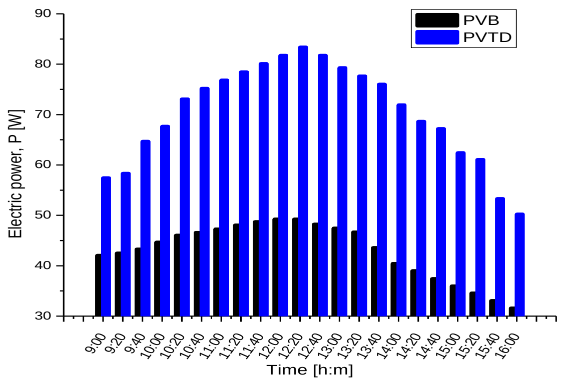


Figure 8: Time dependence of PVB and PVTD electric powers.

In the experiment, the simultaneous use of cooling and reflectors led to an increase in system efficiency. With an inlet water temperature of 18^{0}C, the outlet water temperature was 30^{0}C (water consumption 0.4 liters/min). Experimental results have shown that the average electrical efficiency in FEIQ is 15-16%, and the thermal efficiency is over 50%. The average daily electricity output power for PVTD was 63% higher than for PVB.

4 CONCLUSIONS

In PVTD, the results of cooling and reflector efficiency are reflected in open circuit voltage, short circuit current, and electrical power. In addition, it was possible to extract thermal energy from the temperature difference between the incoming and outgoing water. The average power values for PVTD and PVB were 70.3W and 43W, respectively. This showed that the average daily electricity for PVTD production was 63% higher. Such low-power FEB-based devices can be used as the main power supply for charging and operating smartphones, notebooks, and other low-power devices in extreme situations. Additionally, it can be used for certain purposes in hot water supply.

5 ACKNOWLEDGMENTS

We would like to express our gratitude to Termiz State University, which greatly supports the field of renewable energy sources and scientific work.

REFERENCES

[1] IRENA, Renewable Capacity Statistics 2024, International Renewable Energy Agency, 2024.

[2] Energy Institute, Statistical Review of World Energy, 2024, [Online]. Available: https://ourworldindata.org/grapher/installed-solar-pv-capacity.

[3] A. O. M. Maka and J. M. Alabid, "Solar energy technology and its roles in sustainable development," Clean Energy, vol. 6, no. 3, pp. 476-483, Jun. 2022, [Online]. Available: https://doi.org/10.1093/ce/zkac023.

[4] A. Zubeer and O. Mhohammed, "Performance analysis and electrical production of photovoltaic modules using active cooling system and reflectors," Ain Shams Eng. J., pp. 1-8, 2020, ISSN: 2090-4479.

[5] M. Rosa-Clot, P. Rosa-Clot, G. Tina, and C. Ventura, "Experimental photovoltaic-thermal power plants based on TESPI panel," Sol. Energy, vol. 133, pp. 305-314, 2016.

[6] A. Kabeel, M. Abdelgaied, and R. Sathyamurthy, "A comprehensive investigation of the optimization cooling technique for improving the performance of PV module with reflectors under Egyptian conditions," Sol. Energy, vol. 186, pp. 257-263, 2019.

[7] R. Muminov, M. Tursunov, Kh. Sabirov, Sh. Abilfayziyev, B. Yuldoshov, and S. Toshpulatov, "Testing of crystalline silicon-based photoelectric and photothermal batteries in real climate conditions and comparison of parameter changes," J. Phys.: Conf. Ser., vol. 2388, p. 012128, 2022, [Online]. Available: https://doi.org/10.1088/1742-6596/2388/1/012128.

[8] N. Aste, C. Del Pero, F. Leonforte, and M. Manfren, "Performance monitoring and modeling of an uncovered photovoltaic-thermal (PVT) water collector," Sol. Energy, vol. 135, pp. 551-568, 2016.

[9] P. Bevilacqua, R. Bruno, and N. Arcuri, "Comparing the performances of different cooling strategies to increase photovoltaic electric performance in different meteorological conditions," Energy, vol. 195, pp. 1-19, 2020.

[10] S. Mandal and S. Ghosh, "Experimental investigation of the performance of a double pass solar water heater with reflector," Renew. Energy, vol. 149, pp. 631-640, 2020.

[11] S. Praveenkumar, E. Agyekum, A. Kumar, and V. Velkin, "Thermo-enviro-economic analysis of solar photovoltaic/thermal system incorporated with u-shaped grid copper pipe, thermal electric generators and nanofluids: An experimental investigation," J. Energy Storage, vol. 60, pp. 1-15, 2023.

[12] I. Hossain, Y. E. Ahmed, and S. Sharma, "The use of both metallic-insert and TiO2/water nanofluid to improve the performance of photovoltaic modules; experimental study," Therm. Sci. Eng. Prog., vol. 42, pp. 1-12, 2023.

[13] R. Baccoli, A. Kumar, A. Frattolillo, C. Mastino, E. Ghiani, and G. Gatto, "Enhancing energy production in a PV collector - Reflector system supervised by an optimization model: Experimental analysis and validation," Energy Convers. Manage., vol. 229, pp. 1-17, 2021.

[14] F. Yazdanifard, E. Ebrahimnia-Bajestan, and M. Ameri, "Investigating the performance of a water-based PV/thermal (PV/T) collector in laminar and turbulent flow regime," Renew. Energy, vol. 99, pp. 295-306, 2016.

[15] M. Alobaid, B. Hughes, J. Calautit, D. O'Connor, and A. Heyes, "A review of solar driven absorption cooling with photovoltaic thermal systems," Renew. Sustain. Energy Rev., vol. 76, pp. 728-742, 2017.

[16] S. Preet, "Water and phase change material based photovoltaic thermal management systems: A review," Renew. Sustain. Energy Rev., vol. 82, pp. 791-807, 2018.

[17] A. H. Al-Waeli, K. Sopian, A. Hussein, and M. Chaichan, "Photovoltaic/Thermal (PV/T) systems: Status and future prospects," Renew. Sustain. Energy Rev., vol. 77, pp. 109-130, 2017.

[18] V. N. Palaskar and S. P. Deshmukh, "Waste heat recovery study of spiral flow heat exchanger used in hybrid solar system with reflectors," Int. J. Energy Sci., vol. 5, no. 1, pp. 6-16, 2015.

[19] B. I. Kazandzhan, "Combined solar collectors," J. Phys.: Conf. Ser., 2020, pp. 1-4.

[20] M. Tursunov, K. Sabirov, T. Axtamov, U. Abdiyev, B. Yuldoshov, J. Khaliyarov, S. Bobomuratov, and S. Toshpulatov, "Analysis of electric and thermal efficiency of crystal silicon small power suppliers," Proc. Int. Conf. Appl. Innov. IT, vol. 11, no. 2, pp. 167-171, 2023.

Experimental Comparative Study on the Use of Photovoltaic Converters for Cooling Photovoltaic Modules

Farkhodjon Turaev[1], Isroil Yuldoshev[1,2], Bozorbek Botirov[1], Fazliddinov Sadulla[1]
[1]Tashkent State Technical University named after Islam Karimov, Universitet Str. 2, 100095 Tashkent, Uzbekistan
[2]Tashkent University of Applied Sciences, Gavkhar Str. 1, 100081 Tashkent, Uzbekistan,
farxodjont@bk.ru, yuldashev.i2004@gmail.com, bbozorbek@bk.ru, Fazliddinovsadullo1@gmail.com

Keywords: Photoelectric Converter, PV Cooling, Gauze, Spray Water.

Abstract The aim of this research is to study evaporative cooling as an effective method for cooling photovoltaic (PV) modules in the climatic conditions of Uzbekistan. The experiment was conducted during the summer period at the heliopolytechnic site of Tashkent State Technical University named after Islam Karimov. Among various cooling options for modules, evaporative cooling was selected as the most cost-effective and technologically simple solution for manufacturing and use. The principle of evaporative cooling of a wet object is based on heating and evaporating moisture using the heat of the cooled object. According to researchers, at an ambient temperature of 35-45°C, the maximum temperature of a photovoltaic module without cooling reached 66.1°C. The maximum temperature of the photovoltaic module with evaporative cooling was 46.8°C. Without cooling, the module's temperature would have reached 66.1°C. The reduction of the module temperature by 33°C demonstrates the effectiveness of evaporative cooling. Two photovoltaic panels with a power output of 290 W were used for the experiment. The difference in average electrical power between the uncooled PV module and the PV module with evaporative cooling was 23 W/h. The results indicate that evaporative cooling is indeed effective and can significantly reduce the power loss of PV modules due to overheating.

1 INTRODUCTION

Interest in non-traditional energy sources, particularly direct electricity generation from solar energy through photovoltaic converters (PVCs), is steadily increasing worldwide. Solar energy is one of the most widespread sources of environmentally friendly energy, reducing environmental impact and being applied in various fields [1-2].

Currently, the most common PVCs are made from silicon. Photovoltaic modules operate in open environments and are continuously influenced by weather conditions such as temperature, solar radiation, and wind [4-5]. The efficiency of most PVCs decreases as the temperature of the panels rises. With an increase in panel temperature, the already modest efficiency of PVCs declines further, raising the cost of generated energy. Published data suggest that a 1°C increase in panel temperature above the standard temperature of 20°C leads to a 0.45-0.50% decrease in efficiency [7-13].

However, various sources indicate that existing PVCs have relatively low efficiency, ranging from 12 to 16% [14]. These efficiency values are given for standard conditions, i.e., when the surface temperature of the solar panel is 20°C and the radiant flux density is 1000 W/m².

In hot climates such as Uzbekistan, where summer air temperatures often exceed 40°C, the efficiency of PVCs is significantly lower than 12-16% [15].

The proposed model involves studying evaporative cooling as an efficient method for cooling photovoltaic modules in Uzbekistan's climatic conditions. Among various cooling options, evaporative cooling was chosen as the most cost-effective and technologically simple method for manufacturing and use [16]. Experiments have shown that in conditions of hot climates, low ambient humidity, and wind, the wetting ability of gauze depends on the number of gauze layers due to capillary forces. Under these conditions, the water rise along the fabric reaches up to four layers of gauze, indicating that wetting is minimal due to the evaporation rate exceeding the water uptake rate via capillary forces. As the number of gauze layers increases, the height of water rise along the fabric also increases due to capillary forces.

The optimal number of gauze layers was found to be 10, providing the maximum water rise height, which enhances the efficiency of solar energy conversion into electrical energy. To maintain form and ease of use, the ten-layer gauze fabric was stitched together.

The proposed model is illustrated by drawings, with Figure 1 showing the general view of the photovoltaic converter and Figure 2 presenting view A, the rear side of the photovoltaic converter. The proposed photovoltaic converter consists of an aluminum frame (1), in which the photovoltaic module (2) is installed. On the rear side, there is an evaporative cooling system including a cotton element made of gauze fabric (4) consisting of ten stitched layers pressed by a metal mesh (5). A water reservoir (3) is located in the upper part.

Of the various cooling options for the modules, the evaporative cooling option was chosen as the cheapest, technologically easiest to manufacture and use [13].

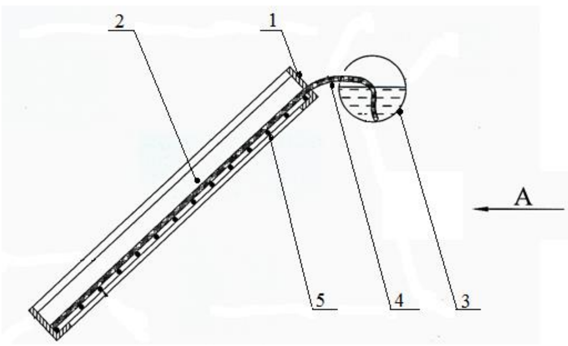


Figure 1: General type of the photoelectric converter.

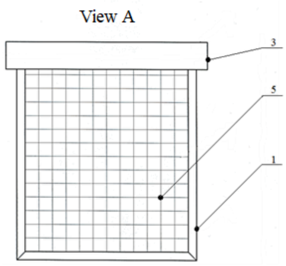


Figure 2: General view of the photoelectric converter, in (Figure 2) – view A.

A method of passive thermal regulation was developed in [16] using automated wetting of gauze fabric via water spraying to regulate the temperature of the photovoltaic module during operation. To verify the accuracy of the proposed theoretical calculations, experiments were conducted with a 290 W photovoltaic module, confirming its functionality. The results demonstrate that the developed cooling system reduced the photovoltaic module temperature by 12% and increased electricity generation by 14%. The main energy balance equation used for evaluating module temperature was applied to the photovoltaic module, showing strong agreement between theoretical and experimental values for both cooled and uncooled conditions.

2 MATERIALS AND METHODS

The photovoltaic converter consists of an aluminum frame housing the photovoltaic module, with an evaporative cooling system on the rear side. This system includes a cotton element made of ten stitched layers of gauze fabric pressed by a metal mesh. A water sprayer, controlled by an automation system, is located at the back of the converter. Water serves as the coolant, entering the water jet and exiting through inlet openings. A small pump inside the water storage tank directs the water to the mesh fabric. A thermal sensor was installed on the rear panel of the photovoltaic converter to measure the module's temperature. Figure 3 presents additional details of the system. The performance of the photovoltaic converter module is compared with that of conventional uncooled photovoltaic modules.

Figure 3: General type of photoelectric converter.

From Figure 4 it can be seen that the basic scheme of the photovoltaic converter consists of: 1 – photovoltaic module; 2 – water spray; 3 –

temperature sensor; 4 - manometer; 5 – water meter; 6 – check valve; 7 - pump; 8 – water tank.

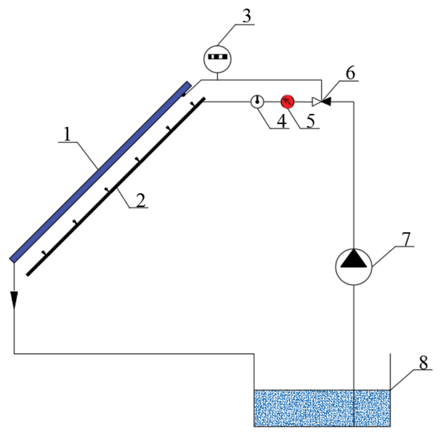


Figure 4: Principal scheme of photovoltaic converter.

The experiments were carried out from 06.05. to 03.09.2024 at the beginning of the summer season. Daily values of solar radiation, ambient temperature, PV temperature, relative humidity outside, etc. were measured.

List of instruments used for measurement;

- 1 FLIR E5 - to measure the temperature on the surface of the radiators;
- 2 anemometer AS856 - for measurement of wind speed and air temperature;
- 3 pyranometer Solar Power Meter Di-LOG SL101 - for solar radiation;
- 4 pressure manometer - for water pressure;
- 5 termometer with output sensor - for temperature measurements.

In particular, a multi-channel digital thermometer of the brand DS18B20 with portable temperature sensors, with a temperature range of -55°C to +125°C, was used to measure the temperature of the photovoltaic converter. The temperature measurement accuracy in the range from --10°C to +85°C is ±0.5°C.

3 RESULTS AND DISCUSSIONS

Increasing solar radiation intensity raises ambient temperature, which in turn increases the surface and rear side temperatures of the photovoltaic converter. The highest temperature was recorded on the surface of the photovoltaic converter due to direct solar exposure. Heat transfer occurs from the hot surface, directly exposed to intense solar radiation, to the cooler surface, leading to an increase in the temperature of the rear panel compared to the ambient temperature. At the start of the experiment, the surface temperature of the photovoltaic converter was 29.2°C, and the rear side (absorbing plate) temperature was 26.7°C. These temperatures gradually increased, reaching a maximum of 46.8°C at 12:20. The rear surface temperature was 33°C, as shown in Figure 5. Water sprayed onto the rear panel during active cooling reduced the module's temperature. The average temperature reduction of approximately 30% for the surface and rear panel of the photovoltaic converter, compared to an uncooled module, was attributed to heat transfer facilitated by the cotton element on the module's rear side.

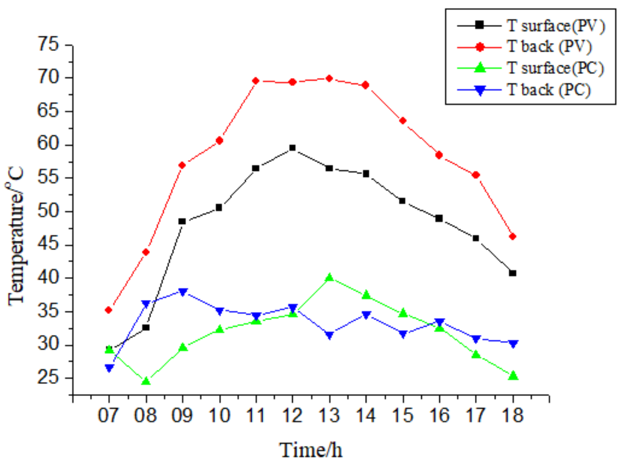


Figure 5: Temperature distribution on PV modules with cooling and without cooling.

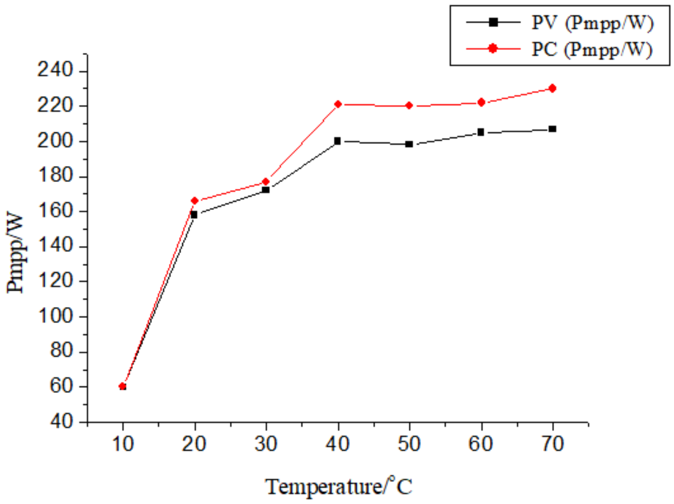


Figure 6: Performance of the photovoltaic module with cooling (PC) and without cooling (PV).

In the study, an infrared thermal imager (FLIR E63900) was used (Fig. 7).

a) b)

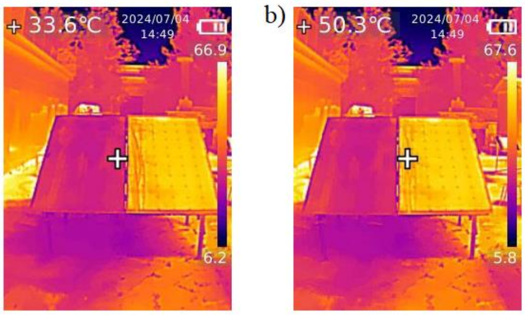


Figure 7: Dynamics of temperature change in of PV module with photovoltaic converter (a) and b without cooling.

4 CONCLUSIONS

The following conclusions can be drawn from the results of the studies:

1) At a photovoltaic layer temperature of 46.8°C, the efficiency reduction of PVCs does not exceed 5%. Throughout the experiments, the average efficiency reduction was no more than 5%.;
2) Experimental data indicate that lower air humidity and higher wind speeds enhance the effectiveness of evaporative cooling.;
3) Photovoltaic modules used without cooling in hot climates significantly degraded in performance.;
4) The applied cooling method is suitable for small-scale photovoltaic installations. Due to its effectiveness, the evaporative cooling method requires further research.

REFERENCES

[1] A. S. Abdelrazik, F. A. Al-Sulaiman, R. Saidur, and R. Ben-Mansour, "A review on recent development for the design and packaging of hybrid photovoltaic/thermal (PV/T) solar systems," Renew. Sustain. Energy Rev., vol. 95, pp. 110-129, 2018, [Online]. Available: https://doi.org/10.1016/j.rser.2018.07.013.

[2] D. H. W. Li, L. Yang, and J. C. Lam, "Zero energy buildings and sustainable development implications—a review," Energy, vol. 54, pp. 1-10, 2013, [Online]. Available: https://doi.org/10.1016/j.energy.2013.01.070.

[3] M. Alktranee and P. Bencs, "Application of nanotechnology in hybrid photovoltaic/thermal systems: A review," J. Appl. Eng. Sci., vol. 19, no. 2, pp. 292-306, 2021, [Online]. Available: https://doi.org/10.5937/jaes0-28760.

[4] A. Saqaff, H. Nikesh, G. Y. Ii, and V. K. Venkiteswaran, "A study on implementation of PV tracking for sites proximate and away from the equator," Process Integr. Optim. Sustain., vol. 3, pp. 375-382, 2019, [Online]. Available: https://doi.org/10.1007/s41660-019-00086-7.

[5] T. T. Chow, G. Pei, K. F. Fong, Z. Lin, A. L. S. Chan, and J. Ji, "Energy and exergy analysis of photovoltaic-thermal collector with and without glass cover," Appl. Energy, vol. 86, pp. 310-316, 2009, [Online]. Available: https://doi.org/10.1016/j.apenergy.2008.04.016.

[6] A. Ndiaye, C. M. F. Kébé, A. Charki, P. A. Ndiaye, V. Sambou, and A. Kobi, "Degradation evaluation of crystalline-silicon photovoltaic modules after a few operation years in a tropical environment," Sol. Energy, vol. 103, pp. 70-77, 2014, [Online]. Available: https://doi.org/10.1016/j.solener.2014.02.006.

[7] V. I. Krutov, Ed., Technical Thermodynamics. Moscow: High School, 1971.

[8] M. Chandrasekar, S. Rajkumar, and D. Valavan, "A review on the thermal regulation techniques for non integrated flat PV modules mounted on building top," Energy Build., vol. 86, pp. 692-697, 2015, [Online]. Available: https://doi.org/10.1016/j.enbuild.2014.10.071.

[9] T. T. Chow, "A review on photovoltaic/thermal hybrid solar technology," Appl. Energy, vol. 87, no. 2, pp. 365-379, 2010, [Online]. Available: https://doi.org/10.1016/j.apenergy.2009.06.037.

[10] G. N. Tiwari, R. K. Mishra, and S. C. Solanki, "Photovoltaic modules and their applications: a review on thermal modeling," Appl. Energy, vol. 88, no. 7, pp. 2287-2304, 2011, [Online]. Available: https://doi.org/10.1016/j.apenergy.2011.01.005.

[11] A. Jamar, Z. A. A. Majid, W. H. Azmi, M. Norhafana, and A. A. Razak, "A review of water heating system for solar energy applications," Int. Commun. Heat Mass Transf., vol. 76, pp. 178-187, 2016, [Online]. Available: https://doi.org/10.1016/j.icheatmasstransfer.2016.05.028.

[12] S. B. Riffat and E. Cuce, "A review on hybrid photovoltaic/thermal collectors and systems," Int. J. Low Carbon Technol., vol. 6, no. 3, pp. 212-241, 2011, [Online]. Available: https://doi.org/10.1093/ijlct/ctr016.

[13] E. Skoplaki and J. A. Palyvos, "On the temperature dependence of photovoltaic module electrical performance: a review of efficiency/power correlations," Sol. Energy, vol. 83, no. 5, pp. 614-624, 2009, [Online]. Available: https://doi.org/10.1016/j.solener.2008.10.008.

[14] Z. I. Zhuraeva, "Overview analysis of the main types of solar cells and identification of ways to increase the efficiency of their operation and application," Tashkent State Tech. Univ., 2020, [Online]. Available: https://7universum.com/ru/tech/archive/item/6460.

[15] R. A. Muminov, S. K. Shoguchkarov, I. A. Yuldashev, and S. S. Makhmudov, "Assessment of temperature conditions and parameters of an autonomous solar photovoltaic plant with a capacity of 3 kW," in *Proc. Republican Conf. Mod. Probl. Semicond. Phys. (SPFP-2019)*, Nukus, 2019.

[16] I. A. Yuldoshev, F. Sh. Turaev, B. M. Botirov, Y. M. Kurbanov, D. R. Kakharova, and J. B. Juraev, "Evaporative cooling photovoltaic panel in the conditions of a hot period," Probl. Energy Sources Saving, vol. 4, pp. 115-123, 2024, [Online]. Available: https://doi.org/10.5281/zenodo.14580468.

Additional Protection Device for Grid Inverter of Photovoltaic Station

Isroil Yuldoshev [1,2], Sanjar Shoguchkarov[1], Shakhnoza Rustamova[1], Bozorbek Botirov [1], Tulkin Jamolov[1] and Zukhra Jurayeva[1]

[1]*Tashkent State Technical University named after Islam Karimov, Universitet, Str. 2, 100095 Tashkent, Uzbekistan*

[2]*Tashkent University of Applied Sciences, Gavkhar 1 Str, 100081 Tashkent, Uzbekistan*

yuldashev.i2004@gmail.com, sanjar4242@inbox.ru, bbozorbek@bk.ru

Keywords: Photovoltaic Battery, Electric Power, Grid-Tie Inverter, Low-Voltage Power Grid, Protection Device, Voltage, Frequency, Power Quality, Integration.

Abstract: The article describes the large-scale implementation of solar photovoltaic installations and systems in the regions of Uzbekistan, and considers the issues of operation of the component equipment of photovoltaic power plants (PVPP) integrated with the low-voltage electric grid. There is carried out the analysis of the influence of low quality of electric power on the operating modes of the grid inverter in PVPP. There are studied the values of voltage, current and frequency of each phase in the low-voltage electric grid using a two-channel flat-panel oscilloscope FNIRSI -1013D. The voltage deviations of each phase in the low-voltage electric grid are determined; their values are $\delta U^{A}_{m(-)} - 7{,}8$ %, $\delta U^{B}_{m(-)} - 20\%$ and $\delta U^{C}_{m(-)} - 20\%$ under winter conditions, respectively. The archives of most inverters in PVPP integrated with the low-voltage distribution grid in the regions of Uzbekistan are studied and analyzed. The reasons for frequent disconnection and connection of grid inverters have been identified, which include voltage instability in phases, losses in the electrical network, incorrect grounding of the inverter, etc. Additional protective equipment has been developed and created for inverters of the solar power plant integrated with the low-voltage electrical network, distinguished by reliability, automation and availability of electrical components on the domestic market of Uzbekistan.

1 INTRODUCTION

In the world, the use of photovoltaic installations and systems based on renewable energy sources (RES) is one of the promising areas for providing consumers with electric energy. In Uzbekistan, the share of renewable energy sources in electricity production will reach 54% by 2030. The President of the country announced this on December 13 at the launch ceremony of new energy capacities. Additional "green capacities" of 19,000 MW will be built during this period, said the head of state [1].

Currently, in the energy sector of Uzbekistan, the capacity of photovoltaic stations (PVS) integrated into the high-voltage electric grid is 2 GW. Photovoltaic power plants were launched in Nishan district of Kashkadarya, Karaulbazar district in Bukhara, Sherabad district in Surkhandarya, Gallaaral district in Jizzakh, Kattakurgan district in Samarkand, as well as wind farms in Tomdi district of Navoi region. In the republic, low-power photovoltaic power plants with a total capacity of 970 MW are connected to a 0,4 kV low-voltage (local) distribution network at 98,246 different facilities, of which 47,5 MW of capacity were installed at higher educational institutions (according to data from the end of 2023) [2].

One of the main factors for the correct operation of the component equipment (mainly inverters for various purposes) of the PVS connected to the local distribution network is the quality of the electricity supplied by the manufacturer. In this regard, there are a number of problems when connecting these sources to the power grid. The main problems are ensuring the requirements for the quality of electricity, limiting power, security measures, protection systems, the synchronization process, reducing the inertia of the system, etc. [3, 4]. Certain values of quality characteristics are set for the electric grid: in terms of nominal frequency, voltage, current, etc., therefore, the required quality of electricity must be ensured for the normal operation of the components of the PVS. [5]. Deviation of the main indicators of electricity quality from the normalized values can lead to operational disruptions of network inverters, and as a result, losses [6]. Connecting photovoltaic arrays to a grid inverter can be cost-effective, but the service life of a grid inverter is often shorter than that of a photovoltaic plant. This will require replacing the inverter at least once during the lifetime of the solar photovoltaic modules, which will lead to additional costs [7].

In order to ensure safety, network inverters are equipped with protective equipment: in case of significant harmonic changes in the network, i.e. with a strong deviation of the voltage and frequency readings from the permissible limit, the circuit breaker turns off and closes the outlet from the mains. The operation of the protective equipment in this form will depend on the setting of the mains inverter and the quality of the power grid. In the negative case, when the mains voltage decreases by 0.5 V compared to the nominal value, and the frequency deviates from the nominal value by 0.5-0.7 Hz, the mains inverter will stop generating electricity to the electrical network in less than 100 ms [8-10]. Despite equipping the network inverters with a protective device, the task of protecting the inverters from a strong deviation of the main parameters of the electrical network still remains unresolved. A sudden increase in voltage can also damage the integrity of the inverter's memory, as well as prematurely cause it to fail.

The purpose of the research is to develop an additional protective device for a network inverter of a photovoltaic plant connected in parallel to a low–voltage distribution network, as well as to study the main indicators of the quality of electrical energy.

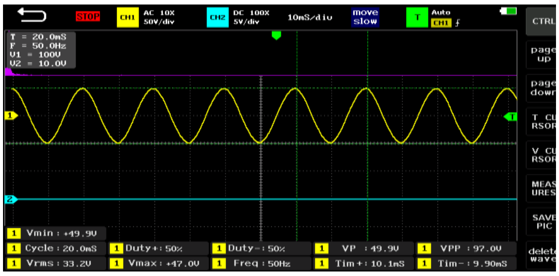
2 MATERIALS AND METHODS

In the course of studying the problems of electricity quality in a low-voltage distribution network, an analysis of waveforms describing voltages, currents and frequency was used. The archive of network inverters of a photovoltaic plant connected in parallel to a low-voltage distribution network during operation has also been analyzed.

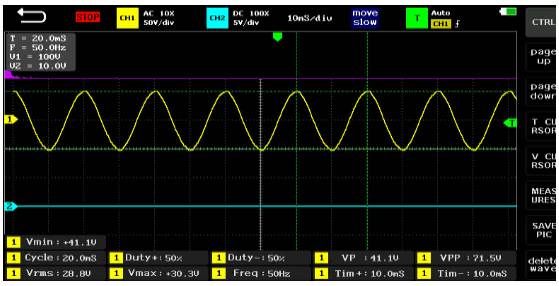
3 RESULTS AND DISCUSSIONS

On January 24-25, 2023, experimental studies were conducted to measure the voltage values of each phase in the low-voltage distribution network of the Faculty of Energy Engineering of Tashkent State Technical University. The voltage, current, and frequency waveforms were recorded using a two-channel FNIRSI -1013D type flat oscilloscope.

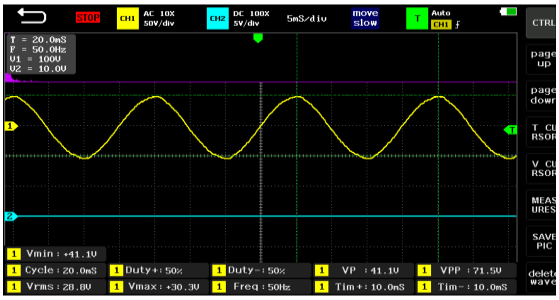
The voltage values for each phase in the low-voltage distribution network, measured and recorded by an oscilloscope at 9:30 a.m. on 01/24/2023 in winter, are shown in Figure 1. The values of the phase voltage in a low-voltage electrical network, respectively, are U_A-202,8 B, U_B – 176 B, U_C- 176 B.



Phase A: oscillation time T=20.0 mS, frequency F=50,0 Hz, U=202.8 B



Phase B: oscillation time T=20.0 mS, frequency F=50,0 Hz, U=176 B



Phase C: oscillation time T=20.0 mS, frequency F=50,0 Hz, U=176 B

Figure 1: Voltage waveforms for each phase in a low-voltage distribution network.

The electric power quality indicators related to slow changes in the power supply voltage are negative $\delta U_{(-)}$ and positive $\delta U_{(+)}$ deviations of the power supply voltage at the point of transmission of electric energy from the nominal agreed value., %:

$$\delta U_{(-)} = \left[1 - \frac{U_{m(-)}}{U_0}\right] \cdot 100\% \qquad (1)$$

$$\delta U_{(+)} = \left[1 - \frac{U_{m(+)}}{U_0}\right] \cdot 100\% \qquad (2)$$

where, $U_{m(-)}, U_{m(+)}$- power supply voltage values lower than U_0 and higher than U_0, respectively, averaged over a time interval of 10 minutes in accordance with requirements of GOST 30804.4.30.

According to the indications of the quality of electricity, the following standards are established: positive and negative voltage deviations at the point of transmission of electric energy should not exceed 10% of the nominal or agreed voltage value [11, 12].

Analyzing Figure 1, it can be concluded that in the conditions of the winter and summer period, with an increase in the loads of consumers with a nonlinear characteristic.

It is the nonlinearity of the load that leads to harmonic distortion of the mains voltage. In most cases, consumers of electric energy have a nonlinear volt-ampere characteristic, which leads to a change in the harmonic composition of the voltage and current of the power supply network [13]. Frequent and unpredictable fluctuations in voltage and frequency at the connection point of the PVS lead to exceeding the tolerances for them [14].

It can be seen that (Figure 1) in severe weather in a low-voltage distribution network, the voltage deviation of each phase is less than U_0. According to the results, the interval $U_{m(-)}$ is respectively $\delta U^{A}_{m(-)}$ −7,8 %, $\delta U^{B}_{m(-)} - 20\%$ и $\delta U^{C}_{m(-)} - 20\%$. A strong phase deviation of the voltage in phase B and C in a low-voltage electrical network when connecting the PVS affects the thermal condition of the heated nodes, the insulation strength and the switching apparatus of the central inverter as a whole. Exceeding this indicator ($\delta U_{(-)}$, $\delta U_{(+)}$) will cause frequent failures or accelerated wear of network inverter parts due to increased nonlinear loads.

Due to the fact that frequent failures lead to malfunctions of the electrically erasable programmable permanent storage device (EEPPSD) in network inverters, which is the main memory contained in the working firmware and configuration of the inverter. Restoration and replacement of EEPPSD in a network inverter does not require high costs, but in operation it leads to a significant decrease in the efficiency of the PVS.

Based on the above, for PVS connected in parallel to a low-voltage distribution network, there is an intermediate range of voltage and frequency values at which the network inverters will turn off and turn on again (Table 1). When analyzing the archive, most inverters in PVS integrated with a low-voltage distribution network in Uzbekistan, the number of disconnections and connections of inverters to local electric network, was 1805 times. This was especially observed in a Huawei 30KTL M3 type network inverter with a capacity of 30 kW in the Urgench city in the period from 11/26/2024 to 01/23/2025. Studies show that when analyzing the inverter archive, the reason for frequent disconnection and connection of the protective device to the low-voltage electrical network is the poor quality of electricity: reduction and increase of voltage in phases, losses in the low-voltage electrical network, improper grounding of the inverter, etc.Table 1

Table 1: Requirements for the operation mode of a PVS connected to a low-voltage electrical network [15]

Mode	Requirements
Shutdown Limits:	Voltage drop protection (U <) <184 V Surge protection (U>)> 253 V Surge protection (U >>)> 264.5 V Frequency drop protection (f <) <47.5 Hz Over-frequency protection (f>)> 51.5 Hz
Reconnection Limits:	Voltage greater than 195.5 V and less than 253 V Frequency greater than 47.5 Hz and less than 50.05 Hz

Excessive switching can eventually lead to overheating and wear of relay contacts in network inverters, which are vital for both normal operation and protective functions. Relay failures can cause interruptions in the energy conversion processes, which will lead to unstable power supply or complete disconnection of the PVS from the low-voltage electrical network.

On October 23, 2019, a PVS 10 kW was installed on the Heliopolygon of the Tashkent State Technical University and connected in parallel to the local electric grid [16]. In December 2020, due to a strong phase deviation in the local electrical network, the inverter prematurely failed under normal operating conditions. Was supplied and replaced 10 kW mains inverter from the company that is the customer of the personnel, but more than 1 year passed before the replacement. During this time, the authors managed to develop and create an additional protective device (Figure 3; 3 - protective device) for PVS grid inverter operating in parallel with a low-voltage distribution network.

Figure 3 shows a general schematic diagram of a solar photovoltaic plant connected in parallel to a local distribution network (0.4 kV), in which additional protective equipment of a network inverter is indicated by the number 3. Figure 3 shows an electrical circuit and components of protective equipment for PVS.

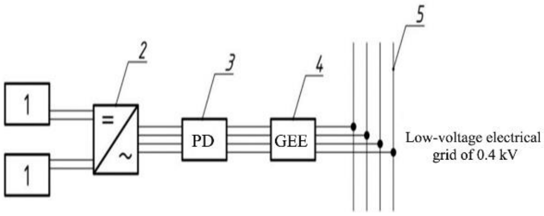


Figure 2. Block diagram of a network photovoltaic plant with a peak power of 10 kW.

In Figure 2: 1 - photovoltaic modules of 20 pieces (total of 40 pieces); 2 - grid inverter; 3 - protective device; 4 - green electricity meter; 5 - low-voltage electrical grid of 0.4 kV.

The solar photovoltaic arrays (1) shown in Figure 3 have an average constant DC voltage of 760 V, a current of 16 A is applied, power is supplied to the mains inverter (2), simultaneously connected to ground (6) and then at a three-phase alternating voltage of ~380 V through the protective equipment (3) of the central inverter through the green electric meter (4) and exported to local electric distribution grid (5). Consumers can import electricity through a green electric meter from the local electric grid [17].

Figure 3 shows in full detail the electrical parts and circuits that make up the protective equipment shown in Figures 2, 3. As noted above, the grid inverter operates in an integrated mode with the local electrical grid, that is, the voltage in the network is synchronized in accordance with the frequency harmonics.

The electrical circuit of the additional protective equipment is switched as follows. The electrical cables are routed from the inverter with three-phase alternating voltage to the input of the SF circuit breaker (3.1), designed for 25A, and are also connected to the neutral (3.5). From the SF circuit breaker, cables are routed to the contactor (3.2) MS and another SF circuit breaker (3.3), rated for a current value of 1A. Current through the SF (3.3) circuit breaker is applied to the input of the RNPP-311 M voltage relay (3.4). In accordance with the requirements of GOST 30804.4.30 the voltage in one or all L1, L2, L3 phases, which is significantly less than or greater than the normalized value of U_0, the smart relay (RNPP-311M – 3.4) disconnects the central inverter from the low-voltage electrical grid and takes control of each phase.

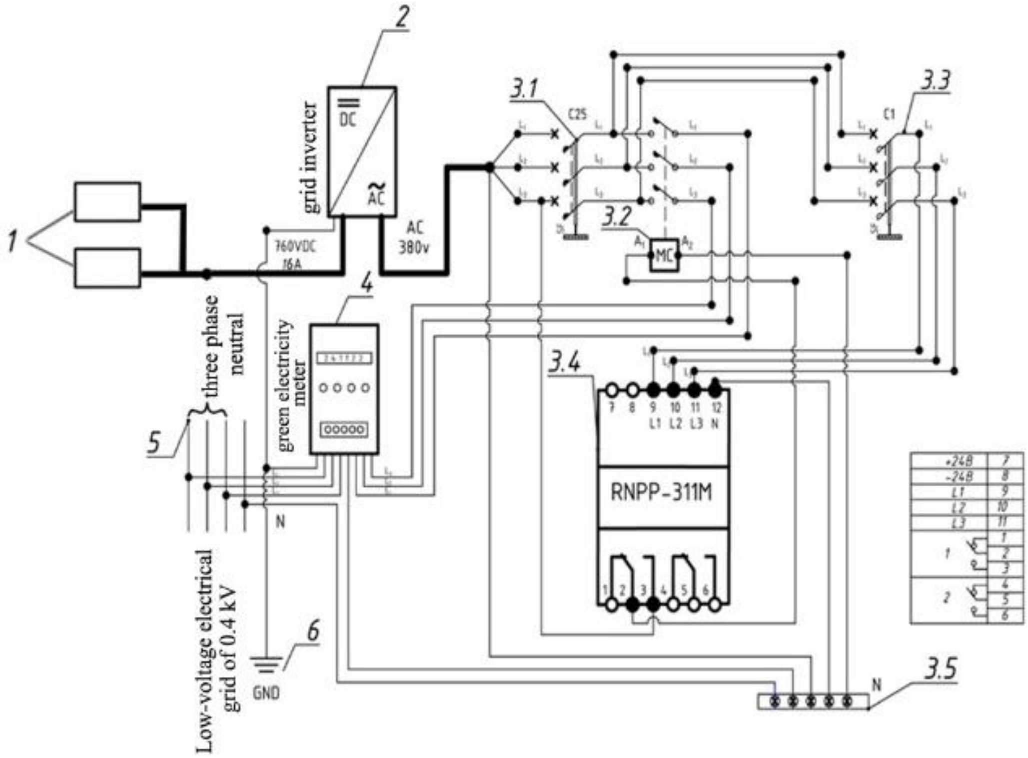


Figure 3: Additional protective equipment for PVS integrated into a low-voltage distribution grid.

In the electrical circuit, the function of the contactor MS (3.2) directly acts as a bridge between the central inverter (2) and the local electrical grid (5), that is, it acts as an automatic switch. The contactor MS (3.2) does not work spontaneously, a signal from the voltage relay RNPP– 311M (3.4) is applied to its operation. It is connected from contact A1 of the contactor MS (3.2) to the voltage relay RNPP– 311M (3.4), and from the other contact A2 it is connected to the neutral (3.5). The RNPP – 311M voltage relay (3.4) issues a signal to disconnect the contacts of the contactor MS (3.2) from the low-voltage distributed network as soon as a defect in the voltage parameters in the electrical grid (5) is detected. Phases L1, L2 and L3 from the MS contactor (3.2), as well as one cable from the neutral (3.5), are connected to the green meter (4) and it is connected in parallel to a low-voltage distributed grid (5).

The following advantages have been identified in the proposed development: it prevents the occurrence of burns and fires caused by overheating of the mains inverter in PVS with low-quality electricity in the local electric grid. The total weight of the development is light, it works without noise, and the cost is very low. There is access to purchase electrical elements and parts in the domestic market of Uzbekistan.

Additional protective equipment is intended for:

- control of the permissible voltage level;
- control of the correct alternation and absence of phase sticking;
- control of the full-phase and symmetry of the mains voltage (phase misalignment);
- disconnecting the load at low-quality mains voltage;
- quality control of the mains voltage after disconnecting the inverter and automatically switching it on after restoring voltage parameters;
- emergency indications in case of an emergency and indication of the presence of voltage in each phase. The development also performs zero control by an indirect method.

4 CONCLUSIONS

In the world, special importance is attached to the issues of electricity generation based on low-power photovoltaic plants and their integration into a low-voltage distribution grid for energy supply to consumers.

The conducted studies allowed us to evaluate the quality of electrical energy in the low-voltage distributed electric grid of Uzbekistan. Based on the results obtained, it can be analyzed that the values of the voltage deviation of each phase in a low-voltage electrical grid are much greater than the nominal value in winter and summer. The voltage deviation of each phase in the low-voltage electrical network and their values, respectively, were $\delta U^{A}_{m(-)} -7{,}8$ %, $\delta U^{B}_{m(-)} - 20\%$ и $\delta U^{C}_{m(-)} - 20\%$ in winter conditions the period in Tashkent. Numerous studies have been conducted on the archives of grid inverters of various types for PVS integrated into the low-voltage electric grid, which are installed in the regions of the country. Data from the memory archive of the grid inverter shows that the reason for frequent disconnection and connection of the inverter to the electrical grid are the following main factors: not phase stability of voltage; losses in the electrical grid; improper grounding of the inverter, etc. Additional protective equipment has been developed and created for PVS inverters integrated with a low-voltage electrical network characterized by reliability, automation and resistance to the thermal condition of heated nodes.

ACKNOWLEDGMENTS

We would like to express our gratitude to the leading scientist of Tashkent State Technical University, Doctor of Technical Sciences, Professor R.A. Sitdikov for his help in discussing the research task.

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] "Uzbekistan intends to increase the share of renewable energy sources in the country's energy balance to 54% by 2030," Presidential Meeting Materials, Dec. 13, 2024, [Online]. Available: https://www.finmarket.ru/news/6323342.

[2] K. Allaev, I. Rakhmonov, and U. Mamadaminov, "Current state and development prospects of Uzbekistan's electric power industry," Probl. Energy Sources Saving, no. 1, pp. 1-7, 2025, [Online]. Available: https://doi.org/10.5281/zenodo.15094235.

[3] M. K. Bobojanov, R. Ch. Karimov, T. H. Qosimov, and S. D. Zh. Dzhuraev, "Development and experimental study of circuits of contactless device for automation of compensation of reactive power of capacitor batteries," E3S Web Conf., vol. 289,

p. 07012, 2021, [Online]. Available: https://doi.org/10.1051/e3sconf/202128907012.

[4] O. Zyukina and D. Ryabova, "Factors of deterioration of electric energy quality and their negative impact on electric energy receivers," Mod. Mater. Equip. Technol., no. 1, pp. 93-96, 2015.

[5] R. Karimov, A. Kuchkarov, M. Xodjalimova, R. Makhamadjanov, and A. Numonov, "Analysis and study of energy efficiency by the operation of a voltage stabilizer," J. Phys.: Conf. Ser., vol. 2094, no. 5, p. 052050, 2021, [Online]. Available: https://doi.org/10.1088/1742-6596/2094/5/052050.

[6] "Solar inverter failures: Causes, consequences and impact on energy output," Enlitia, Jan. 23, 2025, [Online]. Available: https://www.enlitia.com/resources-blog-post/solar-inverter-failures-causes-consequences-and-impact-on-energy-output.

[7] N. Jenkins and J. Ekanayake, Renewable Energy Engineering, 2nd ed. Cambridge: Cambridge Univ. Press, 2024, [Online]. Available: https://www.amazon.com/Renewable-Energy-Engineering-Nick-Jenkins/dp/1009295764.

[8] R. Karimov, "Improving the quality of 0.4 kV electricity in household appliances due to voltage regulation," E3S Web Conf., vol. 384, p. 01056, 2023, [Online]. Available: https://doi.org/10.1051/e3sconf/202338401056.

[9] F. L. Luo and H. E, Advanced DC/AC Inverters: Applications in Renewable Energy. Boca Raton, FL: CRC Press, 2013, [Online]. Available: https://doi.org/10.1201/b13750.

[10] K. Mertens, Photovoltaics Fundamentals, Technology and Practice. Munich: Carl Hanser Verlag, 2014, [Online]. Available: http://ndl.ethernet.edu.et/bitstream/123456789/87792/3/Photovoltaics%20Fundamental%20and.pdf.

[11] *GOST 32144-2013: Electrical Energy. Electromagnetic Compatibility of Technical Equipment. Power Quality Standards in General Purpose Power Supply Systems*, 2014, [Online]. Available: https://meganorm.ru/Data2/1/4293776/4293776477.pdf.

[12] *GOST 30804.4.30-2013 (IEC 61000-4-30:2008): Electrical Energy. Electromagnetic Compatibility of Technical Equipment. Methods of Measuring Power Quality Indicators*, 2013, [Online]. Available: https://docs.cntd.ru/document/1200104665.

[13] S. D. Dzhuraev et al., "Study and analysis of power quality in the electrical networks of the outdoor lighting of the Dushanbe City," in Proc. 2022 Conf. Russ. Young Res. Electr. Electron. Eng. (ElConRus), 2022, pp. 1167-1169, [Online]. Available: https://doi.org/10.1109/ElConRus54750.2022.9755782.

[14] N. Avezova, S. Shoguchkarov, and A. Kudratov, "Methodology for calculating the indicator of availability of network PV power plants," Probl. Energy Sources Saving, no. 1, pp. 106-117, 2025, [Online]. Available: https://doi.org/10.5281/zenodo.15099748.

[15] "Reports by Prof. A. Kiryukhin and others under the Solar Universities project," Solar Universities, 2024, [Online]. Available: https://solar-universities.org/.

[16] I. Yuldoshev et al., "Features of operation of the grid connected photovoltaic power station with a capacity of 10 kW," E3S Web Conf., vol. 216, p. 01172, 2020, [Online]. Available: https://doi.org/10.1051/e3sconf/202021601172.

[17] I. Yuldoshev et al., "Additional protective device of the network inverter of a photovoltaic plant against harmonic changes in the local electric network," Util. Model Patent FAP 2022 0442, 2022.

SECTION 5

INFORMATION TECHNOLOGY IN ECONOMICS

Proceedings of the 13th International Conference on Applied Innovations in IT (ICAIIT)

APRIL 2025

Global Energy Systems Modeling: Structure and Environmental Impacts

Olha Kovalchuk[1], Oksana Tulai[1], Olga Zavytii[1], Ludmila Babala[1], Roman Ivanytskyy[2] and Pavlo Chopyk[2]
[1]*West Ukrainian National University, Lvivska Str. 11, 46009 Ternopil, Ukraine*
[2]*Department of Physics and Mathematics, Ternopil Volodymyr Hnatiuk National Pedagogical University, M. Kryvonosa. Str. 2, 46009 Ternopil, Ukraine*
o.kovalchuk@wunu.edu.ua, o.tulai@wunu.edu.ua, zavitii@ukr.net, ludaduma7@gmail.com, romikiv@ukr.net, chip.ua@ukr.net

Keywords: Geopolitical Challenges, Global Energy Systems, Cluster Model, Primary Energy Consumption, Carbon Emissions, Fuel Balance, Environmental Impact.

Abstract: The global energy landscape undergoes transformative changes characterized by unprecedented challenges in climate and geopolitical dynamics. This study analyzes energy systems across 66 countries using machine learning methods, based on comprehensive energy performance and environmental indicators. Using the K-means clustering method, three distinct country clusters were identified. The baseline cluster includes 62 countries with moderate energy parameters. China forms a unique cluster, characterized by exceptionally high total energy consumption and maximum carbon emissions, yet relatively low per capita consumption. The third cluster comprises the United States, the Russian Federation, and India, which are distinguished by diversified energy policies with high nuclear and natural gas consumption. Statistical analysis reveals robust correlations between primary energy consumption and environmental indicators, with correlation coefficients exceeding 0.9. The research demonstrates that primary energy consumption patterns and fuel balance structure, rather than emission volumes directly, are primary determinants of environmental impact. Coal-based energy generation emerges as the predominant source of anthropogenic environmental challenges globally. These findings underscore the urgent necessity for developing individualized national energy transition strategies that account for each country's unique energy profile and economic circumstances. The methodology established provides a framework for evidence-based policy formulation toward sustainable energy futures.

1 INTRODUCTION

Energy systems today are at a critical point of transformation that will determine the future fate of humanity and the planet. Global challenges related to climate change and geopolitical upheavals put forward new requirements for understanding and managing energy processes [1]. According to the World Meteorological Organization, 2023 was unprecedented in climate history - the warmest over the entire period of instrumental observations, with a global temperature increase of almost 1.5°C [2]. Such cardinal changes create a complex mosaic of challenges for the world community: from environmental consequences to the transformation of economic models.

The modern energy landscape is characterized by paradoxical trends: simultaneous growth in fossil fuel consumption and unprecedented development of renewable energy. The International Renewable Energy Agency report confirms that the share of green electricity continues to grow rapidly, creating a real alternative to traditional energy models [3].

According to the International Energy Agency, there were radical changes in industry indicators. In 2023, global energy demonstrated impressive dynamics of recovery and growth, returning to pre-coronavirus development trends [4]. A key feature of the past year was a significant revival in energy markets, primarily due to the weakening of logistics constraints and the restoration of economic activity. China played a decisive role in this process, canceling strict anti-epidemic restrictions and significantly increasing energy consumption. Statistics demonstrate unprecedented indicators: global oil consumption exceeded 100 million barrels/day for the first time; coal demand reached a historical maximum; renewable energy consumption grew six times faster compared to overall primary energy consumption; electricity demand increased 25% more intensively than total primary consumption [5]. These trends confirm the gradual transformation of the global energy landscape with the strengthening of renewable sources and diversification of energy portfolios of various countries.

At the same time, according to global monitoring studies, approximately 750 million people, every tenth inhabitant of the planet, are still deprived of basic access to electricity. This makes elementary needs impossible: lighting homes, storing food, and protection from rising temperature loads. Moreover, about 2.6 billion people continue to use extremely toxic biofuel sources for heating and cooking, such as charcoal, coal, and organic animal waste. In 2023, researchers recorded significant geographical differences in the relationships between regional population size and energy consumption volumes, which underscores the uneven nature of global energy development [4].

Today, studies aimed at a comprehensive understanding of structural transformations in global energy, and identification of regional characteristics and patterns of national energy systems development are gaining particular relevance. The purpose of our study is to analyze trends and patterns of modern energy profile formation in countries worldwide, identifying systemic interconnections between fuel balance structure, consumption volumes, and carbon emission levels. The obtained knowledge can provide a theoretical basis for countries to develop effective energy policies aimed at achieving a balance between economic development and environmental sustainability.

2 RELATED WORKS

Research on risks and opportunities created by the energy transition has been the focus of scholarly debates in recent years. Many scholars have dedicated work to identifying the main factors shaping the future of the global energy system [6–7]. O. A. Osobajo et al. investigated the impact of energy usage and economic growth on CO_2 emissions. The findings showed that the study variables (population, capital stock, and economic growth) are mutually linked to CO_2 emissions, whereas energy consumption has a one-way causal effect [8]. M. Ahmed et al. examined the relationship between energy consumption and CO_2 emissions (CO_2e) across China, India, and the USA. The study analyzed key factors including population growth, energy consumption per capita, and income levels. Their findings revealed a significant correlation between energy consumption per capita (p.c.) and CO_2 emissions in these countries [9]. M. Kameni Nematchoua and J. A. Orosa investigated strategies for reducing CO_2 concentrations and minimizing energy demands at the district level [10]. The researchers J. Li et al. examined the relationship between CO_2 emissions, energy consumption, mortality, life expectancy, and GDP in the top five carbon-emitting countries, revealing a strong positive correlation between CO_2 emissions and energy consumption [11]. However, most scientific research in this field relates to regional studies [12–15]. Further research on the efficiency and environmental performance indicators of countries' energy systems remains relevant for developing effective energy strategies.

3 METHODOLOGY

Analysis of global carbon dioxide emission trends demonstrates a clear correlation between primary energy consumption volumes and the level of anthropogenic CO_2 emissions. China has the maximum primary energy consumption level (Fig. 1).

China also significantly leads other countries in carbon dioxide emissions from energy [4]. The USA, India, and Russia are next in the ranking of countries with high primary energy consumption. This trend is associated with intensive industrial development and high electricity consumption in these countries, where traditional fossil energy sources such as coal, oil, and natural gas, the main sources of greenhouse gas emissions, predominate.

However, no correlation has been established between carbon dioxide emissions and energy dependence on primary energy consumption p.c. of countries worldwide (Fig. 2). This confirms the complexity and multi-factor nature of energy transformation processes in the modern global economic space and requires further research.

To identify groups of countries with similar energy consumption and environmental impact models, we conducted a cluster analysis of 66 countries worldwide based on selected indicators of efficiency and environmental friendliness of countries' energy systems [4].

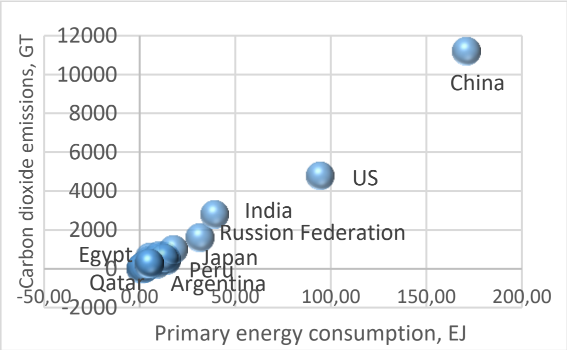


Figure 1: Chart of carbon dioxide emissions from energy dependence on primary energy consumption.

3.1 Data and Variables

The data set comprised efficiency and climate sustainability metrics of energy infrastructures from 66 countries worldwidey, presented in Table 1.

The countries include: Canada, Mexico, US, Argentina, Brazil, Chile, Colombia, Ecuador, Peru, Trinidad and Tobago, Venezuela, Austria, Belgium, Czech Republic, Finland, France, Germany, Greece, Hungary, Italy, Netherlands, Norway, Poland, Portugal, Romania, Spain, Sweden, Switzerland, Turkiye, Ukraine, United Kingdom, Azerbaijan, Belarus, Kazakhstan, Russian Federation, Turkmenistan, Uzbekistan, Iran, Iraq, Israel, Kuwait, Oman, Qatar, Saudi Arabia, UAE, Algeria, Egypt, Morocco, South Africa, Australia, Bangladesh, China, China Hong Kong SAR, India, Indonesia, Japan, Malaysia, New Zealand, Pakistan, Philippines, Singapore, South Korea, Sri Lanka, Taiwan, Thailand, Vietnam.

Empirical research was conducted using Rapid Miner software [16] based on the values of the aforementioned indicators, published in the official report “2023 Statistical Review of World Energy” [4]. *K*-means clustering [17] was used to identify homogeneous models of energy behavior and environmental impact of countries worldwide. This enabled grouping countries with similar energy consumption characteristics and carbon dioxide emissions into homogeneous groups.

Table 1: Study variables.

Variable	Variable
CDEE, Gt	Carbon dioxide emissions from energy
CDEEPMF, $GtCO_2e$	Carbon dioxide equivalent emissions from energy, process emissions, methane and flaring
ENGF, $GtCO_2e$	Emissions from natural gas flaring
PEC, EJ	Primary energy consumption
PECPC, EJ	Primary energy consumption p.c.
Consumption by fuel:	
CC, EJ	Coal
CH, EJ	Hydroelectricity
CNG, EJ	Natural Gas
CNE, EJ	Nuclear Energy
CO, EJ	Oil
CR, EJ	Renewables

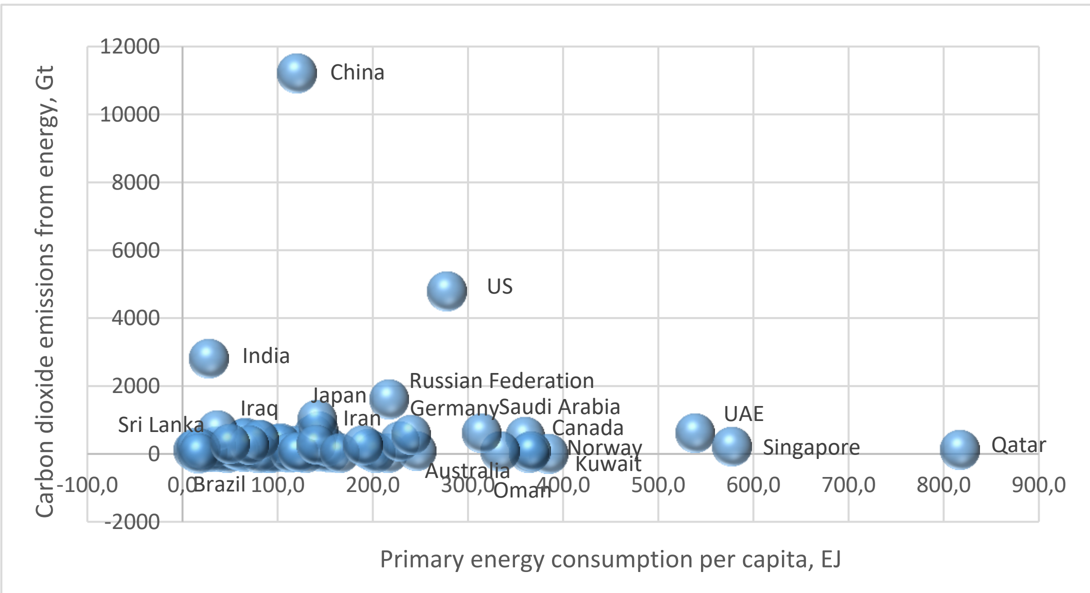


Figure 2: Chart of carbon dioxide emissions from energy dependence on primary energy consumption per capita.

3.2 Method

K-means clustering represents a fundamental unsupervised machine-learning technique for partitioning datasets into predetermined clusters [17]. The method minimizes within-cluster variability by iteratively assigning data points to the nearest centroid and recalculating cluster centers. At its core, the algorithm seeks to reduce the total within-cluster sum of squared distances, which can be expressed mathematically as the objective function:

$$J = \sum_{k=1}^{K} \sum_{x \in C_k} \left\| x - \mu_k \right\|^2, \quad (1)$$

where K denotes the number of clusters, C_k represents a specific cluster, x indicates individual data points, and μ_k represents the cluster centroid. The Euclidean distance between a point and its cluster centroid is calculated using the standard distance formula:

$$d(x, \mu_k) = \sqrt{\sum_{i=1}^{n} \left(x_i - \mu_{k_i}\right)^2}. \quad (2)$$

The clustering process involves several iterative steps beginning with centroid initialization. Initially, K centroids are randomly selected from the dataset, with each centroid representing the initial cluster center. The algorithm then alternates between two primary operations: assignment and update.

In the assignment phase, each data point is assigned to the closest centroid using the Euclidean distance. Mathematically, this assignment can be represented as:

$$C_i = arg \min_j \left\| x_i - \mu_j \right\|^2. \quad (3)$$

The update phase recalculates cluster centroids by computing the arithmetic mean of all points assigned to each cluster:

$$\mu_j = \frac{1}{C_j} \sum_{x_i \in C_j} x_i. \quad (4)$$

Convergence is achieved when centroids stabilize, typically determined by minimal changes in cluster assignments or reaching a predefined maximum number of iterations. The overall optimization is expressed through the convergence criterion:

$$|J(t+1) - J(t)| < \varepsilon, \quad (5)$$

where ε represents a small threshold value indicating minimal change between successive iterations.

4 RESULTS AND DISCUSSION

As a result of applying the k-means clustering method, 3 groups of countries were identified with distinct models of energy behavior and environmental impact (Fig. 3). The optimal number of clusters (k=3) was determined using the elbow method. The largest differences in mean cluster values were found in the variables CDEEPMF (carbon dioxide equivalent emissions from energy, process emissions, methane and flaring) and CDE (carbon dioxide emissions from energy). However, the variables PEC (primary energy consumption) and CNG (natural gas consumption) have the greatest influence on the distribution of countries into groups (Fig. 4). This means that the structure and scale of the energy system are the primary determinants of environmental impact. Emissions are a secondary consequence of overall energy consumption and energy balance diversification.

Cluster No. 0 included all 62 analyzed countries worldwide, except for China, the Russian Federation, and India. This group is characterized by significantly lower mean values for all analyzed indicators compared to countries in other clusters, except primary energy consumption p.c. (Table 2).

Table 2: Cluster Means.

Variable	Cluster 0	Cluster 1	Cluster 2
PECC	152.827	119.800	246.950
PEC	4.711	170.740	62.785
CDEE	263.141	11218.400	3206.450
ENGF	2.900	4.900	40.200
CDEEPMF	282.011	12603.500	1367.900
CO	1.720	32.730	21.535
CNG	1.212	14.570	24.120
CC	0.907	91.940	6.015
CNE	0.275	3.900	4.635
CH	0.315	11.460	2.045
CR	0.383	16.130	4.440

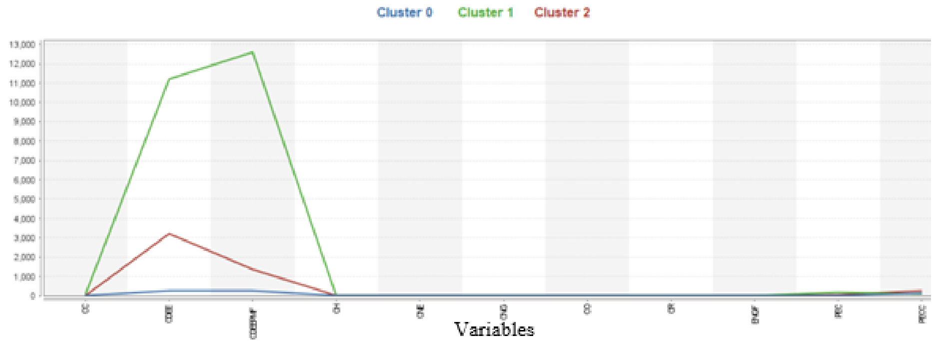


Figure 3: Centroid chart.

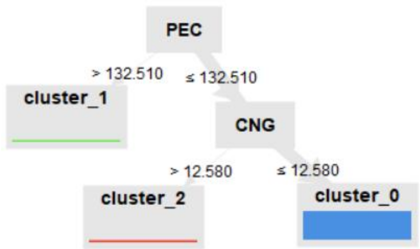


Figure 4: Cluster tree.

Cluster No. 0 effectively represents the “baseline” group of countries with moderate energy and environmental parameters.

Cluster No. 1 included only China. Its energy policy is characterized by significantly higher values, compared to the average of other identified country clusters, in the following indicators: primary energy consumption, coal consumption, oil consumption, hydroelectricity consumption, carbon dioxide emissions from energy, and carbon dioxide equivalent emissions from energy, process emissions, methane, and flaring. At the same time, China has the lowest primary energy consumption p.c. compared to the average values of this indicator in the other two clusters. This indicates the specificity of its energy model – high total consumption with relatively low p.c. consumption.

Cluster No. 2 included the USA, Russian Federation, and India. The average values of nuclear energy consumption, natural gas consumption, emissions from natural gas flaring, and primary energy consumption p.c. for countries in this group are significantly higher than the corresponding averages in the other two identified clusters. This cluster represents countries with a diversified and intensive energy policy.

To assess the model's resilience, we conducted a sensitivity analysis by introducing controlled variations in the input data (±5%) and re-executing the clustering. Results showed stability of the main clusters with an average object reassignment error of 4.2%, confirming the reliability of the classification. This indicates sufficient model robustness to minor changes in input data and the validity of the obtained results.

The obtained estimates confirm the hypothesis of the heterogeneity of global energy systems and the need for an individual approach to assessing their efficiency and environmental impact. The results of the empirical analysis demonstrate the complex interaction between energy consumption and emissions indicators during country clustering. Although the largest statistical differences between clusters are observed in carbon dioxide emissions indicators, which reflect the ecological aspect of energy systems, the determining factors for grouping countries were the volumes of primary energy consumption and natural gas. This means that the structure and scale of the energy system are the primary determinants of environmental impact. This result emphasizes the systemic nature of the relationship between economic scale, energy policy, and environmental impact, where total energy consumption serves as a fundamental basis for further differentiation.

Table 3 presents the correlation assessments between each pair of variables included in the cluster model.

A strong direct relationship was identified between primary energy consumption and the following indicators: carbon dioxide emissions from energy (0.992), renewables consumption (0.970), oil consumption (0.939), and carbon dioxide emissions from energy (0.904); carbon dioxide emissions from energy and renewables consumption (0.961); carbon dioxide emissions from energy and carbon dioxide equivalent emissions from energy, process emissions, methane, and flaring (0.934); coal consumption and carbon dioxide equivalent emissions from energy (0.986); coal consumption and carbon dioxide emissions from energy (0.946); oil consumption and renewables consumption (0.908). A strong positive

Table 3: Correlations.

	CDEE	CDEEPEF	CH	CNE	CNG	CO	CR	ENGF	PEC	PECC
CC	0.946	0.986	0.889	0.468	0.406	0.714	0.888	0.042	0.904	-0.057
CDEE		0.934	0.884	0.693	0.671	0.898	0.961	0.185	0.992	0.006
CDEEPEF			0.885	0.442	0.427	0.697	0.856	0.154	0.893	-0.056
CH				0.527	0.486	0.727	0.897	0.120	0.878	-0.009
CNE					0.870	0.875	0.731	0.270	0.764	0.100
CNG						0.872	0.661	0.555	0.743	0.131
CO							0.908	0.263	0.938	0.071
CR								0.070	0.970	-0.017
ENGF									0.221	0.006
PEC										0.019

correlation is observed between hydroelectricity consumption and the following indicators: coal consumption (0.889), carbon dioxide equivalent emissions from energy, process emissions, methane, and flaring (0.885), and carbon dioxide emissions from energy (0.884). Primary energy consumption correlates with carbon dioxide equivalent emissions from energy, process emissions, methane, and flaring (0.893) and hydroelectricity consumption (0.878); hydroelectricity consumption correlates with oil consumption (0.908)

The results of the correlation analysis reveal systemic interconnections between various parameters of the energy system, demonstrating that a change in one indicator almost determines changes in others. The exceptionally strong relationship between primary energy consumption and carbon dioxide emissions, as well as renewable energy sources, confirms a direct dependence between the scale of energy consumption and environmental impact.

In particular, the strong correlation between coal consumption and carbon emissions indicates that coal-based energy remains a major source of anthropogenic environmental impact. The high correlation of hydroelectricity consumption with emission indicators and the consumption of fossil energy sources is noteworthy, suggesting systemic complementarity between different energy sectors in a global context.

These findings confirm the complex nature of energy systems, where changes in one parameter trigger cascading changes in other components.

5 CONCLUSIONS

The study presents a comprehensive analysis of global energy systems, uncovering fundamentally new aspects of the interrelationship between the scale of energy consumption, the structure of the fuel balance, and environmental impacts. The application of cluster analysis made it possible, for the first time, to systematically differentiate countries into three unique groups representing fundamentally different models of energy behavior.

Particularly notable is the unique model of China, which has developed a distinctly specific energy consumption strategy: extremely high total volumes combined with relatively low p.c. consumption. It was found that China represents a unique cluster characterized by the highest indicators of primary energy consumption, coal-based generation, hydroelectricity, and carbon dioxide emissions.

The key scientific outcome of the study is the identification of systemic patterns within the global energy landscape. It was found that primary energy consumption and the structure of the fuel balance are the determining factors of environmental impact, rather than emission volumes directly. The clustering of countries revealed the heterogeneity of global energy systems: from the moderate model of developed countries to China’s unique strategy of extremely high total consumption, and the diversified energy policies of the United States, Russia, and India.

Russia’s full-scale invasion of Ukraine has further highlighted global energy challenges, demonstrating the critical dependence of European countries on Russian fossil fuels. The study shows that the Russian Federation belongs to the group of countries with diversified and intensive energy policies, characterized by high levels of nuclear and gas generation. Geopolitical upheavals have underscored the urgent need to accelerate the transition to renewable energy sources as a strategy for ensuring national and economic security.

Europe’s energy dependence on Russia has caused widespread consequences for the energy balance of other regions. Following Russia's invasion of Ukraine, a global redistribution of energy flows occurred, with Asian countries, especially China and India, increasing purchases of Russian energy resources at reduced prices. This led to increased competition in liquefied natural gas markets, where European countries began seeking alternatives to Russian gas, creating additional pressure on suppliers from the Middle East, Africa, and the United States. At the same time, this accelerated the development of renewable energy in many regions of the world and stimulated the diversification of energy sources in developing countries seeking to reduce their vulnerability to geopolitical upheavals.

The study empirically confirmed exceptionally strong correlations between key parameters of energy systems. In particular, it was established that primary energy consumption almost determines the levels of carbon emissions, renewable energy consumption, and oil consumption, with correlations exceeding 0.9. A particularly important scientific conclusion is that coal-based energy remains the primary source of anthropogenic impact, emphasizing the urgent need for the transformation of energy models.

The results obtained hold significant theoretical and practical value for shaping effective national and global energy transition strategies. The study convincingly demonstrates that balancing economic development and environmental sustainability requires an individualized approach that accounts for each country's unique characteristics. The presented cluster analysis methodology can be successfully adapted for further studies on the dynamics of global energy systems.

REFERENCES

[1] O. Kovalchuk, M. Karpinski, L. Babala, M. Kasianchuk, and R. Shevchuk, "The canonical discriminant model of the environmental security threats," Complexity, vol. 2023, p. 5584750, 2023.

[2] State of the Global Climate 2023, World Meteorological Organization, 2024, [Online]. Available: https://library.wmo.int/viewer/68835/download?file=1347_Global-statement-2023_en.pdf&type=pdf&navigator=1.

[3] Renewable Energy Statistics 2024, International Renewable Energy Agency, 2024, [Online]. Available: https://www.irena.org/Publications/2024/Jul/Renewable-energy-statistics-2024.

[4] Statistical Review of World Energy 2024, Energy Institute, 2024, [Online]. Available: https://www.energyinst.org/__data/assets/pdf_file/0006/1542714/684_EI_Stat_Review_V16_DIGITAL.pdf.

[5] Global Electricity Review 2024, EMBER, 2024, [Online]. Available: https://ember-energy.org/app/uploads/2024/05/Report-Global-Electricity-Review-2024.pdf.

[6] Z. Bi, R. Guo, and R. Khan, "Renewable adoption, energy reliance, and CO2 emissions: A comparison of developed and developing economies," Energies, vol. 17, no. 13, p. 3111, 2024.

[7] S. G. Nnabuife, C. K. Darko, P. C. Obiako, B. Kuang, X. Sun, and K. Jenkins, "A comparative analysis of different hydrogen production methods and their environmental impact," Clean Technol., vol. 5, no. 4, pp. 1344-1380, 2023.

[8] O. A. Osobajo, A. Otitoju, M. A. Otitoju, and A. Oke, "The impact of energy consumption and economic growth on carbon dioxide emissions," Sustainability, vol. 12, no. 19, p. 7965, 2020.

[9] M. Ahmed et al., "The effect of energy consumption, income, and population growth on CO2 emissions: Evidence from NARDL and machine learning models," Sustainability, vol. 15, no. 15, p. 11956, 2023.

[10] M. K. Nematchoua and J. A. Orosa, "Low carbon emissions and energy consumption: A targeted approach based on the life cycle assessment of a district," Waste, vol. 1, no. 3, pp. 588-611, 2023.

[11] J. Li et al., "The relationship between energy consumption, CO2 emissions, economic growth, and health indicators," Int. J. Environ. Res. Public Health, vol. 20, no. 3, p. 2325, 2023.

[12] R. G. Alajmi, "Energy consumption and carbon emissions: An empirical study of Saudi Arabia," Sustainability, vol. 16, no. 13, p. 5496, 2024.

[13] L. I. Patiño, E. Padilla, V. Alcántara, and J. L. Raymond, "The relationship of energy and CO2 emissions with GDP per capita in Colombia," Atmosphere, vol. 11, no. 8, p. 778, 2020.

[14] C. D. Sabău-Popa et al., "Do energy consumption and CO2 emissions significantly influence green tax levels in European countries?," Energies, vol. 17, no. 9, p. 2186, 2024.

[15] K. B. Keerthana, S.-W. Wu, M.-E. Wu, and T. Kokulnathan, "The United States energy consumption and carbon dioxide emissions: A comprehensive forecast using a regression model," Sustainability, vol. 15, no. 10, p. 7932, 2023.

[16] RapidMiner Documentation, 2023, [Online]. Available: https://docs.rapidminer.com.

[17] A. Migliore and C. Rossi-Lamastra, "Cluster analysis," in Methodological Approaches for Workplace Research and Management, 2023, pp. 95-107.

Automation of Plagiarism and AI Detection in IT Students Papers with a Software Tool Using API

Oleksandr Sharyhin[1,2,3], Oksana Klochko[1,6] and Ihor Tverdokhlib[2,4,5]
[1] *Vinnytsia Mykhailo Kotsiubynskyi State Pedagogical University, Ostrozhskogo Str. 32, 21100 Vinnytsia, Ukraine*
[2] *Dragomanov Ukrainian State University, Pyrogova Str. 9, 01601 Kyiv, Ukraine*
[3] *Miratech, Vatslav Havel Avenue 6z, 03124 Kyiv, Ukraine*
[4] *Institute of Pedagogy of the NAES of Ukraine, Sichovyh Streltsiv Str. 52-D, 04053 Kyiv, Ukraine*
[5] *Taras Shevchenko National University of Kyiv, Volodymyrska Str. 60, 01033 Kyiv, Ukraine*
[6] *Academy of cognitive and Natural Sciences, 54 Universytetskyi Ave., Kryvyi Rih, 50086, Ukraine*
{*exhaustic, klochkoob, i.a.tverdokhlib*}*@gmail.com*

Keywords: Academic Integrity, Plagiarism, Automation, Student Work Verification, Information Technology Students, Professional Training, Artificial Intelligence, Information Law, Natural Language Processing, Intelligent Information System.

Abstract: The article considers various approaches to automating the checking of student works by lecturers for plagiarism and the presence of fragments of text/code generated by Artificial Intelligence. As a result of testing various text fragments related to IT, it was concluded that some of existing services can be considered as good candidates for plagiarism detection, but there's no obvious choice for AI detection. Services (or combination of them) may act as a support mechanism, but final decisions should be made by lecturers. A software application has been developed to automate the routine work of lecturers of higher education institutions in checking student works for plagiarism and AI detection. Developed software application uses various APIs to search for plagiarism and fragments of text/code generated by AI. The practical value of the study lies in the possibility of using the proposed method to check student works for originality. Further research on this topic may include multi–user support, integration with existing educational platforms, adding UI components for visualization of AI/plagiarism detection results.

1 INTRODUCTION

The development of information technologies and their mass introduction into most spheres of human activity, in particular into the educational process, has not only positive aspects. Currently, the Internet contains a large amount of information, the amount of which is constantly growing. Systems based on Artificial Intelligence are actively developed and used, which are capable of generating almost any information content. All this creates the prerequisites for unscrupulous researchers who present materials found on the Internet or generated by generative Artificial Intelligence (AI) as their own scientific results.

In the conditions of the Information Society, the training of future specialists in information technologies plays an important role. They are the driver of the further development of the sphere of information technologies, playing a key role in most areas of human activity: economy, production, medicine, service sector, education, etc. However, due to the conditions defined above, the training of future specialists in information technologies is accompanied by the conscious or accidental use of borrowings in writing scientific papers, program code, borrowing other people's ideas. Therefore, the problem of detecting plagiarism in students works is a very important and urgent task of today, and the solution to this task usually falls on the shoulders of lecturers of higher education institutions.

A review of scientific sources on solving the problem of anti-plagiarism verification of scientific and current works of students was conducted. According to the results of the analysis, the main directions of scientific research of scientists were identified. For more than ten years, the problem of detecting plagiarism and developing programs for anti-plagiarism verification has been put to the fore. In scientific works devoted to the study of these issues, the authors raise the issues of detecting plagiarism in scientific and pedagogical activities [1, 2], the use of AI systems by students for writing papers [3, 4], and a review and comparative analysis of existing programs for anti-plagiarism verification.

An important role in observing the principles of academic integrity in the training of future IT specialists is played by the detection of plagiarism in the program code during the lecturer's verification of student works or educational projects. In the article [5] the authors present a privacy-preserving protocol for plagiarism detection that eliminates the need for code disclosure during similarity computation. In the works [6, 7] scientific investigations on ways to automate the verification of the source code in student scientific works were published. A group of developers in [8] developed a web-based application that contains a set of tools for detecting and preventing plagiarism in educational program code.

The active use of AI systems in all spheres of human activity [9] in recent years contributes to an increase in the generation of borrowings by students in their scientific papers and program code. On the other hand, the use of AI methods and systems contributes to the automation of plagiarism detection. Developments have also been underway in this direction in recent years. In particular, it is worth highlighting a number of works [10, 11] that consider methods for detecting plagiarism based on semantic text analysis and deep learning algorithms. In work [12], the authors proposed a fundamentally new approach for the estimation of the time complexity of an algorithm. However, the issues of automating the verification of student scientific and practical papers using several anti-plagiarism verification services and combining different types of services (detecting matches with information sources on the Internet, using artificial intelligence systems to analyze texts for borrowings or fabrication) to check the work for plagiarism remain insufficiently developed.

The authors provide a list of processes that can be automated [13, 14] and a comparative analysis of ready-made software products for automated testing.

2 LITERATURE REVIEW

According to the analysis of scientific sources, popular science publications and normative documents, we have identified the main concepts that form the methodological basis of the research.

On the basis of a review of normative documents and scientific sources, the authors define the definition of plagiarism as the use of a part of other people's public works or scientific papers, texts, images, ideas without attribution in order to present them as one's own research results. Detection of plagiarism plays an important role in the educational sector, especially when it comes to the presentation of research results by students.

There are various classifications of plagiarism types. Summarizing the results presented in [15, 16], the following main types of plagiarism can be distinguished:

- Complete Plagiarism is a type of plagiarism when you present a work written by another person as your own.
- Direct Plagiarism – copying individual parts of someone else's work (words, sentences, paragraphs or chapters) and adding them to your work without citing the author.
- Paraphrasing Plagiarism is one of the most common types of plagiarism, which involves borrowing someone else's idea, paraphrasing it in your own words and placing it in your work without attribution.
- Plagiarizing Yourself – reusing your own ideas and parts of texts that have already been published.
- Accidental Plagiarism is a type of plagiarism according to which parts of other works are unintentionally used in writing a part of a work without giving a reference to them. This is usually done unconsciously or accidentally.
- Mosaic Plagiarism is the use by an author of very small parts of the works of other authors in combination with his or her own thoughts.
- Source-based Plagiarism is the use of borrowings together with references to their source. In this case, the source may be incorrect, not be the original source, or not exist.

In recent years, revolutionary processes have been observed related to the active distribution and use of artificial intelligence-based technologies in various fields of human activity. Artificial Intelligence is also actively penetrating into the educational sphere. Students of higher education institutions are increasingly using and abusing AI-based tools. In [17], it is noted that "...students put the morality of the act in whether they manipulate the tool or not, and not so much in the intellectual property of what is delivered". Since AI is currently a tool that has no moral principles and cannot independently ensure the principles of academic integrity, all responsibility for the use of AI tools lies with its users [17].

The World Economic Forum notes that our society is navigating a second era of digital technologies, which include publicly available generative artificial intelligence [18]. Such systems allow students to generate texts for written work quite easily. Thus, developments in the field of AI have benefited many industries, but for the educational and scientific field, they have created new challenges for students to use AI tools to generate scientific and educational texts, they have reinforced the problem of academic integrity in the modern information society.

Today there are international organizations to combat cheating, plagiarism, and academic dishonesty in higher education. Such an organization is The International Center for Academic Integrity [19]. It offers assessment services, resources, and consultations to its member institutions. The European Network of Academic Integrity (ENAI) holds an annual conference on ethics and integrity in academia. At the last conference significant debate was generated by the session on artificial intelligence (AI) detection tools based on recent research by a team of ENAI members concluding that they don't work reliably [20].

The European Commission issued the first act on the regulation of AI [21]. It is the first comprehensive legal framework for artificial intelligence worldwide. It regulates the rules for the use of AI in education and professional training. Content that is either generated or modified with the help of AI — images, audio or video files (for example, deepfakes) — needs to be clearly labeled as AI-generated so that users are aware when they come across such content. Thus, it is planned to regulate the issue of academic integrity regarding the use of AI systems.

Plagiarism detection software was developed in the 1990s. These programs were focused on detecting plagiarism in the form of copying and pasting data from the Internet [2]. However, this problem has now become much more complicated due to the ability of artificial intelligence systems to generate texts and program code. The search strategies used to detect fabrications made by artificial intelligence differ from those used to detect plagiarism [3]. Plagiarism detection is a part of natural language processing (NLP). Currently, there are many solutions for detecting lexical or syntactic plagiarism based on NLP methods, in particular the concept of using language dictionaries such as WordNet [22].

Nowadays, scientists have made a significant contribution to solving the problem of automated verification of students' academic texts, but the sources and types of plagiarism are constantly changing and new ones are emerging. Therefore, the problem of automating the detection of plagiarism in students' academic papers is relevant and will remain so in the future.

The purpose of the study is to automate the routine work of higher education lecturers in checking IT student papers for plagiarism and AI detection by developing a software application using the API to search for plagiarism and fragments of text/code generated by AI.

3 METHODOLOGY

The research was conducted through an empirical analysis of various API-based services to detect plagiarism and AI-generated content in academic texts. All text fragments that we used in this research are related to IT. The methodology involved the creation and classification of text datasets and the systematic testing of external services using these text datasets.

For the purposes of this study, two distinct datasets were compiled: one is for evaluating AI-generated content detection, and another designed to assess the accuracy of plagiarism detection. Three categories of text fragments were prepared for AI detection:

1) Data subset DS11. AI-generated texts – generated using previous versions of ChatGPT (GPT-3.5 and earlier). Our assumption was that existing services would handle AI detection for these cases, showing results close to
   a) 80 text fragments were generated.
2) Data subset DS12. AI-generated texts – generated using latest version of ChatGPT (GPT-4o). This service is newer and is available in paid plans. Our assumption was that existing services would handle AI detection for these cases, showing results close to 100. 60 text fragments were generated.
3) Data subset DS13. Original academic texts
   a) from dissertations written before the emergence of generative AI tools i.e. there is not even a theoretical possibility that text fragments were generated with the help of AI. 112 text fragments were prepared.

For plagiarism detection, a separate dataset was constructed, it contains three categories of text fragments:

1) Data subset DS21. Existing scientific articles, dissertation works. We expect to find plagiarism in these documents because this information already exists and it's public. 134 text fragments were prepared.

2) Data subset DS22. Fragments of new texts which were not published yet. We expect not to find plagiarism in these fragments. 56 text fragments were prepared.
3) Data subset DS23. Text fragments generated by different versions of AI text generators. We expect not to find plagiarism in these fragments. 31 text fragments were generated.

The analysis of API service outputs on various categories of text fragments reveals the occurrence of both Type I errors (false positives) and Type II errors (false negatives).

To prove the statistical significance of the results and justify that the selected number of fragments is sufficient for reliable conclusions, we will conduct a power analysis using the Cohen method [23, 24]. For its calculation we will use statsmodels.stats.power of the statsmodels Python package:

```
from statsmodels.stats.power \
import FTestAnovaPower
analysis = FTestAnovaPower()
k_groups = 3
n_per_group_large_75 = \
analysis.solve_power(effect_size=0.4, \
alpha=0.05, power=0.75, k_groups=3)
print(f"(f=0.4): {n_per_group_large_75:.1f}")
```

The calculated result is 57. When expecting a large effect, the sample is sufficient for the study because the power of subsets DS11, DS12, DS13 exceeds the obtained value.

4 RESULTS AND DISCUSSION

In this paper, we reviewed several API services that provide plagiarism checking and detection of text fragments built with the help of AI.

Turnitin is a well–known plagiarism detection service but is not directly accessible to individual users [25]. Access to Turnitin requires an institutional li-cense. If a university or educational institution does not hold such a license, its lecturers and students cannot use the platform for originality checking or academic integrity purposes. Thus, there is a need to consider other tools for detecting plagiarism and AI–generated text.

An important requirement for such services is the availability of an API that allows them to be integrated into the educational environment and a set of tools used by the teaching staff who evaluate students' work.

The result of our work is an API which combines existing services, its main processes are shown in Figures 1 and 2.

During the research process we conducted an AI detection using each of the services for data subsets DS11 (results are shown in Table 1), DS12 (Table 2) and DS13 (Table 3).

Table 1: Statistic results of AI detection services execution data subset DS11.

AI detection services	Average	Min	Max
GptZero	76	62	100
OpenAI	76.2	75	80
Writer	3.3	2.0	5/0
CopyLeaks	91.6	81.2	100
Sapling	0.1	0.0	0.2
Grammarly	0.0	0.0	0.0
ZeroGPT	51.4	39.9	62.8
Undetectable.ai	87.5	87.5	87.5
EdenAI	52.7	41.1	70.0

Table 2: Statistic results of AI detection services execution for data subset DS12.

AI detection services	Average	Min	Max
GptZero	76.6	58.0	93.0
OpenAI	77.0	65.0	85.0
Writer	1.2	0.0	3.0
CopyLeaks	80.0	0.0	100.0
Sapling	0.0	0.0	0.0
Grammarly	0.0	0.0	0.0
ZeroGPT	46.0	42.0	48.0
Undetectable.ai	83.8	68.8	85.0
EdenAI	53.8	38.7	68.1

Table 3: Statistic results of AI detection services execution for data subset DS13.

AI detection services	Average	Min	Max
GptZero	74.6	62.0	98.0
OpenAI	39.5	15.0	57.5
Writer	2.3	0.0	4.0
CopyLeaks	0.0	0.0	0.0
Sapling	0.0	0.0	0.0
Grammarly	0.0	0.0	0.0
ZeroGPT	40.5	29.0	48.5
Undetectable.ai	10.5	0.0	31.3
EdenAI	48.8	43.0	62.4

Our testing shows that Grammarly, Sapling and Writer do not allow detecting AI in text fragments. Possible reasons could be unsupported languages, specific subject areas of text fragments.

Results shown in Table 1 and 2 demonstrate that

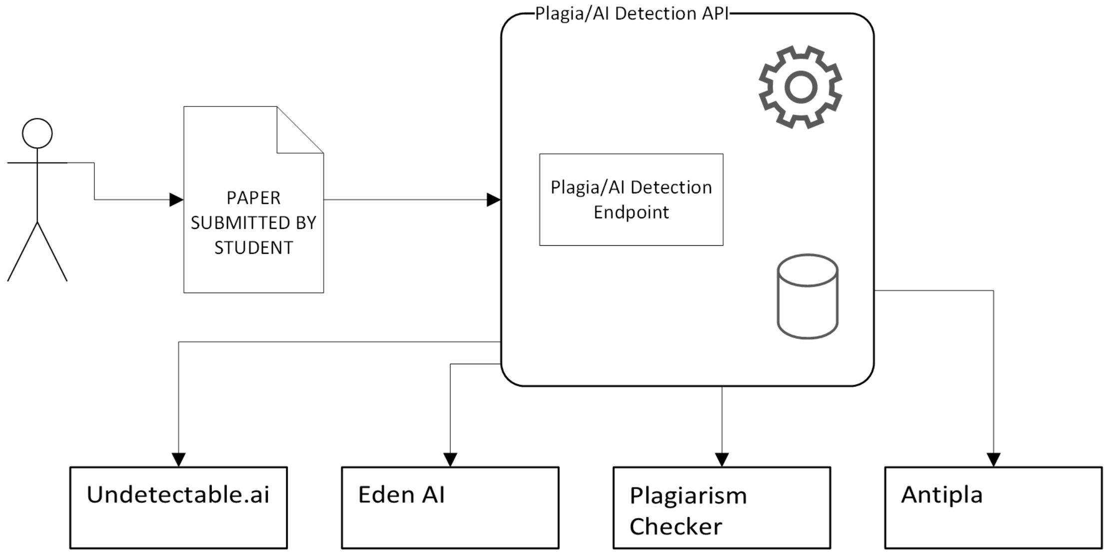


Figure 1: The process of initiating processing of a document uploaded by student.

differences between texts generated by GPT-4o and older versions of ChatGPT are not relevant.

GptZero, EdenAI results have a large number of false positives for papers which do not contain AI generated text. The best results are obtained using Undetectable.ai, there's only 10.5% false positives here. Results given by OpenAI and ZeroGPT are worse, but also acceptable.

The main conclusion is that there's no services which can be used as an automated tool for AI detection. The decision must be made by humans (lecturer which checks student work), AI detectors (or a combination of them) can only act as an advising mechanism.

In this paper we reviewed the following services for plagiarism detection: Eden AI [26], GPTZero [27], OpenAI [28], Grammarly [16, 29], ZeroGPT [30].

During the research processwe conducted a plagiarism detection using each of the services for data subsets DS21 (results are shown in Table 4), DS22 (Table 5) and DS23 (Table 6).

Results show that almost all the services do not generate false positives, there are issues with OpenAI only which tries to find sources even for original texts (Table 7). The same results were obtained for text fragments generated by AI – all services except-ing OpenAI found no plagiarism in them.

GptZero, Winston.ai and EdenAI results can be considered as good candidates for plagiarism detection because the percentages of errors are less than 11% (Table 7). It saves time for lecturers who check student papers.

In addition, these technologies are developing at a rapid pace and in a short time the service that is most optimal may change. At the same time, the task that needs to be solved by the lecturer who evaluates the student's work is quite stable. In most cases, information about the percentage of plagiarism and AI–generated content is sufficient.

That is why we decided to create our API and integrate it into services which are used for communication between students and lecturers in the educational process. The service we are developing is an aggregator and wrapper over a number of external services described above. Development stack is C#, ASP .NET.

An example of typical API that we use is EdenAI which provides a list of services, including Ai Detection and Plagia Detection [26].

AI Detection service is an aggregator of other services, so it supports Originality AI [31], WinstonAI [32], Sapling [33]. It is possible to specify which 3rd party services to use, it is important because the cost of operation depends on it. Response for each provider contains resulting ai_score and a list of text fragments with ai_score and prediction for each of them. A prediction is one of two values: "AI-generated" and "original". For each provider response may also contain original response from each provider.

Plagia Detection service supports 3rd party services such as Originality AI and Sapling. Response for each provider consists of the resulting plagia_score and a list of objects. Every object contains text and a list of candidates each consisting of url (where this fragment is placed), plagia_score, prediction ("plagia-

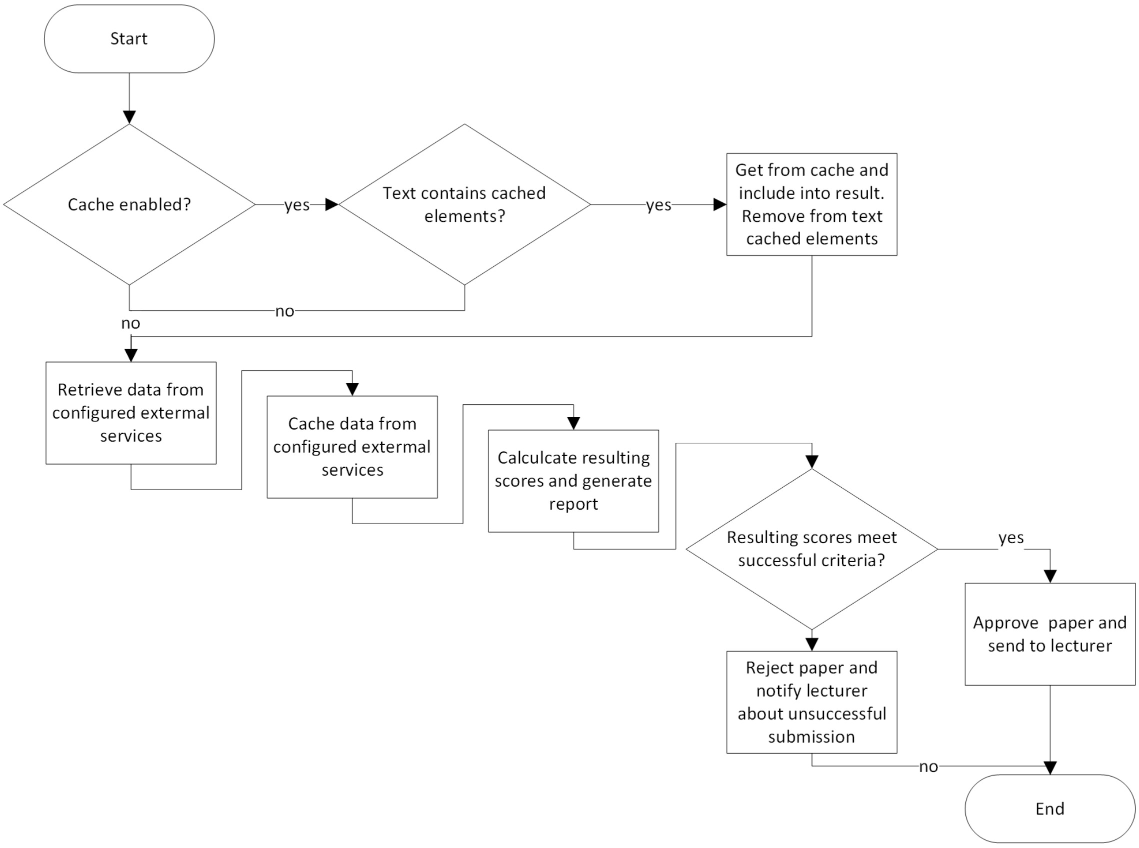


Figure 2: The flowchart of AI/plagiarism detection in a document uploaded by a student.

Table 4: Statistic results of plagiarism detection services execution for data subset DS21.

Plagiarism detection services	Incorrect detections/errors	Incorrect detections/errors, %
Grammarly	113.0	84.3
GptZero	16.0	11.9
OpenAI	57.0	42.5
Plagiarism detector.net	53.0	39.6
Winston.ai	24.0	17.9
Originality.ai	127.0	94.8
EdenAI	24.0	17.9

Table 5: Statistic results of plagiarism detection services execution for data subset DS22.

Plagiarism detection services	Incorrect detections/errors	Incorrect detections/errors, %
Grammarly	0.0	0.0
GptZero	0.0	0.0
OpenAI	11.0	19.6
Plagiarism detector.net	0.0	0.0
Winston.ai	0.0	0.0
Originality.ai	0.0	0.0
EdenAI	0.0	0.0

Table 6: Statistic results of plagiarism detection services execution for data subset DS23.

Plagiarism detection services	Incorrect detections/errors	Incorrect detections/errors, %
Grammarly	0.0	0.0
GptZero	0.0	0.0
OpenAI	2.0	6.5
Plagiarism detector.net	0.0	0.0
Winston.ai	0.0	0.0
Originality.ai	0.0	0.0
EdenAI	0.0	0.0

Table 7: Statistic results of plagiarism detection services execution for all test cases.

Plagiarism detection services	Type 1 errors (false positives)	Type 2 errors (false negatives)	Type 1 errors, %	Type 2 errors, %	Errors, %
Grammarly	0.0	113.0	0.0	51.1	51.1
GptZero	0.0	16.0	0.0	7.2	7.2
OpenAI	13.0	57.0	5.9	25.8	31.7
Plagiarism detector.net	0.0	53.0	0.0	24.0	24.0
Winston.ai	0.0	24.0	0.0	10.9	10.9
Originality.ai	0.0	127.0	0.0	57.5	57.5
EdenAI	0.0	24.0	0.0	10.9	10.9

rized” or “original”) and plagiarized_text.

Back to our API, the main endpoint is POST /v1/document. It allows the uploading of a document and starts a process of AI and plagiarism detection. The figure 1 shows the process of initiating document processing, it is triggered by a document submitted by student.

As it was mentioned above it should be possible to manage usage of external APIs – enable or disable them, set API key etc. For the current version of API we implemented this functionality based on application settings. Below is an example of appsettings.json fragment:

```
 "AIDetection": {
"SuccesfulScore": 0.95,
"Undetectable.ai": {
  "Enabled": true,
  "ApiKey": "***"
},
"EdenAI": {
  "Enabled": true,
  "ApiKey": "***",
  "Subservices": {
    "OriginalityAI": {
      "Enabled": true
   },
   "WinstonAI": {
      "Enabled": false
    },
    "Sapling": {
      "Enabled": true
    }
  }
},
...
   },
  "PlagiaDetection": {
 "SuccesfulScore": 0.8,
 // settings are similar to AIDetection section
  },
  "cacheSettings": {
 "Enabled": true,
 "ExpirationMinutes": 1440
  }
```

It consists of 3 main sections: settings for AI detection, settings for plagiarism detection and cache settings. In an example above Undetectable.ai, and EdenAI (OriginalityAI and Sapling) APIs are used for AI detection. Antipla and EdenAI/WinstonAI are disabled in this configuration. SuccesfulScore = 0.95 means that 5% of paper is allowed to be detected as AI generated content.

Caching of responses allows to save locally results retrieved from external APIs. It allows to minimize costs needed for paper processing. Let’s consider the situation when a student submits a paper, but it contains a high percentage of plagiarism. Some changes are applied to the next version of a paper, but the majority of it remains the same. Caching allows us to reuse results of processing from previous attempts. It means that external API will be called only for new or edited text fragments.

The flowchart of this process is shown on figure 2. This flowchart is simplified and just shows the concept. The real process is asynchronous, the initial endpoint finishes its work after asynchronous calls of external services. Responses from external services trigger the second part of flow.

5 CONCLUSIONS

In this paper, we reviewed the problem of plagiarism and AI detection in the educational process. Automation of this process allows lecturers to don't do routine work, but spend time on truly creative aspects.

We analyzed existing services for plagiarism and AI detection and tested them on a number of documents grouped by certain criterias. Based on our testing we consider GptZero, Winston.ai and EdenAI as good candidates for plagiarism detection. There are no services which demonstrate ideal results for AI detection, but from our point of view it can be achieved by recommendation based on a combination of Undetectable, OpenAI and ZeroGPT with a final decision made by a human (lecturer).

We developed an API that acts as an aggregator and wrapper for external APIs. A developed API can be integrated into tools and pipelines that are used for communication between students and lecturers. An API supports caching, it allows not to do additional calls retrieving the same data as on previous attempts. This API is configurable so it allows defining which services can be used. We provide a default config file based on the results of our testing, but it can be changed by a person who uses the API.

The proposed system helps reduce the time spent by the lecturer on checking student work and simplifies the process as a whole. This system is a convenient tool for maintaining academic integrity by checking the originality of student work.

Further research in this direction may focus on:
- multi-user support (currently software supports 1 API key per platform);
- integrate with existing educational platforms;
- add UI components for visualization of AI/plagiarism detection results.

REFERENCES

[1] D. Fanelli, "How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data," PLoS ONE, vol. 4, no. 5, 2009.

[2] J. Canzonetta, "Repurposing plagiarism detection services for responsible pedagogical application and (in)formative assessment of source attribution practices," Assessing Writing, vol. 50, p. 100563, 2021.

[3] F. R. Elali and L. N. Rachid, "AI-generated research paper fabrication and plagiarism in the scientific community," Patterns, vol. 4, no. 3, p. 100706, 2023.

[4] V. M. Fedorets, O. V. Klochko, I. A. Tverdokhlib, and O. A. Sharyhin, "Cognitive aspects of interaction in the 'Human - Artificial Intelligence' system," J. Phys.: Conf. Ser., vol. 2871, no. 1, p. 012023, Oct. 2024.

[5] N. Bhimrajka et al., "Privacy-preserving plagiarism checking," in Proc. 24th Int. Conf. Cryptol. India (INDOCRYPT 2023), Goa, India, 2023, pp. 105-125.

[6] N. Tahaei and D. C. Noelle, "Automated plagiarism detection for computer programming exercises based on patterns of resubmission," in Proc. 2018 ACM Conf. Int. Comput. Educ. Res., Espoo, Finland, 2018, pp. 178-186, [Online]. Available: https://doi.org/10.1145/3230977.3231006.

[7] S. Myronenko and Y. Myronenko, "Detecting plagiarism in student assignments using source code analysis," WSEAS Trans. Comput. Res., vol. 12, pp. 367-376, 2024.

[8] R. Maertens et al., "Discovering and exploring cases of educational source code plagiarism with dolos," SoftwareX, vol. 26, p. 101755, 2024.

[9] A. V. Slobodianiuk and S. O. Semerikov, "Advances in neural text generation: A systematic review (2022-2024)," CEUR Workshop Proc., vol. 3917, pp. 332-361, 2025.

[10] E. M. Hambi and F. Benabbou, "A new online plagiarism detection system based on deep learning," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 9, 2020, [Online]. Available: https://doi.org/10.14569/IJACSA.2020.0110956.

[11] K. Vani and D. Gupta, "Detection of idea plagiarism using syntax-semantic concept extractions with genetic algorithm," Expert Syst. Appl., vol. 73, pp. 11-26, 2017, [Online]. Available: https://doi.org/10.1016/j.eswa.2016.12.022.

[12] O. A. Sharyhin and O. V. Klochko, "Automation of checking student assignments in IT-related subjects based on AI systems," CEUR Workshop Proc., vol. 3781, pp. 88-97, 2024, [Online]. Available: https://ceur-ws.org/Vol-3781/paper07.pdf.

[13] V. I. Klochko, O. V. Klochko, and V. M. Fedorets, "Visual modeling technologies in the mathematical training of engineers in modern conditions of digital transformation of approaches to creating visual content," in Proc. 3rd Workshop Digit. Transform. Educ., Lviv, Ukraine, 2024, pp. 57-73, [Online]. Available: https://ceur-ws.org/Vol-3771/paper20.pdf.

[14] O. V. Klochko et al., "Anthropologically oriented strategies of interaction in the Human-Computer system," J. Phys.: Conf. Ser., vol. 2611, no. 1, p. 012018, Oct. 2023.

[15] Originality.ai, "7 Common Types Of Plagiarism - Examples And How To Avoid," 2025, [Online]. Available: https://originality.ai/blog/common-typesof-plagiarism.

[16] Grammarly Inc., "7 Common Types of Plagiarism, With Examples," 2024, [Online]. Available: https://www.grammarly.com/blog/plagiarism/typesof-plagiarism/.

[17] A. Benito, "Academic Integrity and the use of Artificial Intelligence (AI)," 2023, [Online]. Available: https://academicintegrity.org/resources/blog/119-2023/november-2023/471-academic-integrity-andthe-use-of-artificial-intelligence-ai.

[18] L. Volante, C. DeLuca, and D. A. Klinger, "ChatGPT and cheating: 5 ways to change how students are graded," World Econ. Forum, 2023, [Online]. Available: https://www.weforum.org/stories/2023/03/chatgptand-cheating-5-ways-to-change-how-students-aregraded/.

[19] ADG Creative, "International Center for Academic Integrity," 2023, [Online]. Available: https://academicintegrity.org/.

[20] S. Dann, M. Draper, and I. Glendinning, "Reflections on the European Network for Academic Integrity Conference," Qual. Assur. Agency, 2023, [Online]. Available: https://www.qaa.ac.uk/en/newsevents/blog/reflections-on-the-european-network-foracademic-integrity-conference.

[21] European Union, "AI Act," 2025, [Online]. Available: https://digitalstrategy.ec.europa.eu/en/policies/regulatoryframework-ai.

[22] Princeton University, "WordNet. A Lexical Database for English," 2025, [Online]. Available: https://wordnet.princeton.edu/.

[23] J. Cohen, Statistical Power Analysis for the Behavioral Sciences. Routledge, 2013.

[24] J. Buckley, "Conducting power analyses to determine sample sizes in quantitative research: A primer for technology education researchers using common statistical tests," J. Technol. Educ., vol. 35, no. 2, pp. 81-109, 2024.

[25] Turnitin LLC, "Higher education: Everything you need to uphold integrity at your institution," 2025, [Online]. Available: https://www.turnitin.com/solutions/higher-education.

[26] Eden AI, "The Full Stack AI Platform for Developers," 2025, [Online]. Available: https://docs.edenai.co/.

[27] GPTZero, "More than an AI detector," 2025, [Online]. Available: https://gptzero.me/.

[28] OpenAI, "Ask ChatGPT anything," 2025, [Online]. Available: https://openai.com/.

[29] Grammarly, "Grammarly: Responsible AI that ensures your writing and reputation shine," 2025, [Online]. Available: https://www.grammarly.com/.

[30] ZeroGPT, "AI Detector for ChatGPT, GPT-4o, Gemini, Claude AI Generated Text," 2025, [Online]. Available: https://zerogpt.org/.

[31] Originality.ai, "Originality AI Plagiarism and Fact Checker," 2025, [Online]. Available: https://originality.ai/.

[32] Winston AI, "The Most Trusted AI Detector," 2025, [Online]. Available: https://gowinston.ai/.

[33] Sapling, "AI Detector," 2025, [Online]. Available: https://sapling.ai/ai-content-detector.

Global DevOps Market Dynamics and Workforce Analysis

Oksana Vasylenko[1], Kateryna Shulakova[1,2], Liliia Bodnar[3], Eduard Siemens[1], Roman Tsarov[2] and Oleksandr Vudvud[4]

[1]*Anhalt University of Applied Sciences, Bernburger Str. 57, Köthen, Germany*
[2]*State University of Intelligent Technologies and Telecommunications, Kuznechna Str. 1, Odesa, Ukraine*
[3]*South Ukrainian National Pedagogical University, Staroportofrankyvska Str. 26, Odesa, Ukraine*
[4]*Odesa Polytechnic National University., Shevchenko Avenue 1, 65044 Odesa, Ukraine*
bodnarl79@pdpu.edu.ua, katejojo29@gmail.com, oksana.vasylenko@hs-anhalt.de, eduard.siemens@hs-anhalt.de, rcarev@gmail.com, o.m.vudvud@op.edu.ua

Keywords: DevOps Market Growth, Regional Adoption, Salary Analysis, IT Workforce Demographics, DevOps Implementation.

Abstract: This study delivers a comprehensive analysis of the global DevOps market, dissecting its growth trajectory, regional adoption variations, and salary dynamics as observed in 2024. Synthesizing data from recent industry surveys and market reports, the research evaluates the market's current valuation, forecasts projected growth rates, and identifies the key factors driving regional disparities in DevOps implementation. Beyond mere market metrics, this study delves into the intricacies of salary ranges for DevOps engineers worldwide, illuminating the significant impact of both market demand and regional economic conditions on compensation structures. Further enriching the analysis, the paper explores the demographic and firmographic characteristics of IT professionals, offering critical insights into the composition of the workforce, the distribution of roles within organizations, and the strategic implications for talent management. By integrating quantitative market data with qualitative insights into workforce demographics, this research aims to equip IT professionals, business leaders, and policy-makers with actionable intelligence, enabling them to effectively navigate the evolving DevOps environment and make informed strategic decisions that align with current and future market demands. This study not only highlights the current state of the DevOps market but also provides a forward-looking perspective, emphasizing the importance of adapting to the rapid changes in technology and workforce dynamics.

1 INTRODUCTION

The DevOps market has seen significant growth in recent years, driven by the increasing demand for faster software delivery, enhanced collaboration between development and operations teams [1], and the proliferation of cloud-native applications. As organizations undergo digital transformation, DevOps has become a key methodology for achieving agility, automation, and continuous integration and delivery [2]. Analysts predict [3] that this market will continue to expand, influenced by technological advancements in automation, artificial intelligence, and evolving enterprise needs.

Despite the widespread global adoption of DevOps, there is a lack of consolidated data that thoroughly examines market dynamics, regional adoption patterns, and workforce compensation trends. This study aims to fill this gap by integrating diverse data sources to present a comprehensive overview of the DevOps domain as of 2024. The analysis combines market growth projections with regional adoption statistics and compensation insights, providing valuable perspectives for organizations and professionals engaged in DevOps.

2 MARKET TRENDS AND SALARY INSIGHTS

DevOps holds significant value as it streamlines the integration of development and operations, enhancing the speed, security, and efficiency of software development and delivery. It leverages automation and collaboration, crucial in managing increasingly complex IT infrastructures that blend physical, virtual, and cloud environments [4]. By optimizing software deployment operations, DevOps enables organizations to achieve superior outcomes and continuous improvement. The demand for DevOps specialists steadily grows, and this trend is expected to continue. Alongside demand, the responsibilities of DevOps engineers expand [5]. To truly grasp its impact, one should examine global statistics and apply relevant mathematical and statistical methods. Reports on the DevOps market provide in-depth insights into market dynamics, technological advancements, and regional analyses, guiding businesses in aligning their strategies with current and future trends.

2.1 Market Overview and Growth Projections

In 2023, the global DevOps market was valued at \$10.5 billion according to data [6, 7], and it is expected to grow at an estimated rate of 19.50-21.20% during the forecast period, depending on the region (Fig. 1). The market is expected to grow from \$10.45 billion in 2023 to \$25.53 billion in 2028 and to \$52.16 billion in 2032. North America is the largest DevOps market, accounting for 38.5% of the global market in 2023-2024.

DevOps is becoming an increasingly popular approach in the IT industry. Companies that implement DevOps usually achieve better results, such as higher product quality, faster time to market, and lower costs. For example, it is the most popular process framework in IT organizations, with 49% of respondents using it. But when asked to name the biggest technical skills gap in their teams, 37% of IT leaders cited DevOps and DevSecOps [6]. 61% of organizations report that DevOps has improved the quality of their products, and they can invest 33% more time in infrastructure improvements.

2.2 Regional Adoption Patterns

The distribution of DevOps technology services reveals notable regional differences. In the United States, 61.21% of companies have adopted DevOps practices, compared to 8.77% in the UK and 6.8% in India [6]. Several factors contribute to these variations:

Leadership in Technology. The United States, with its significant investment in IT infrastructure and advanced technology, leads in DevOps adoption. The country's market maturity and the presence of numerous tech giants and startups contribute to its high adoption rate. The widespread use of cloud services also drives the rapid implementation of DevOps in the US.

The UK's Tech Ecosystem. Despite being smaller than the US, the UK boasts a strong technology sector, particularly in financial services, telecommunications, and e-commerce. The need to stay competitive in these industries fuels DevOps adoption, though on a smaller scale compared to the US.

India's Growing Industry. India is experiencing rapid growth in its technology sector, with increasing adoption of DevOps driven by the expansion of software development and IT services companies. However, the lower adoption rate compared to the US reflects a more gradual transition from traditional methods to agile DevOps practices.

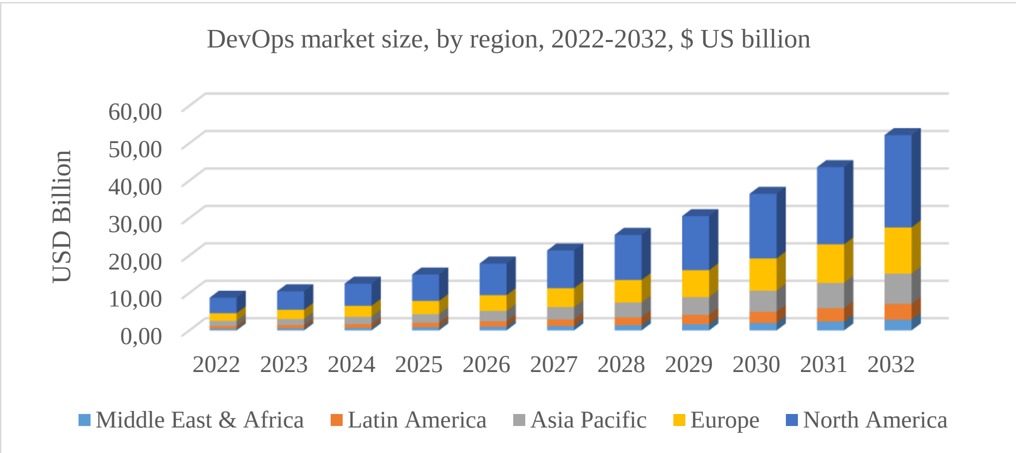


Figure 1: DevOps market size by region, 2022-2032 (USD billion).

2.3 Devops Engineer Salaries Worldwide

To provide a comprehensive understanding of DevOps engineer salaries worldwide, the following data outlines compensation levels across various countries. This information is based on recent surveys and industry reports, reflecting the diverse salary ranges for DevOps professionals in different regions. Detailed salary information in US dollars [8, 9, 10] is presented in Figure 2.

Global Overview. According to a Stack Overflow survey, the average salary for DevOps engineers is approximately $80,158.50. For instance, in the UAE, the average annual salary for a DevOps engineer is $82,006, which translates to an hourly wage of $39.43. The salary range for a DevOps engineer in the UAE is around $56,547 [10].

In the United States, there are 17,000 advertised DevOps engineer roles. The average salary for these positions is $109,300, with a typical range between $96,600 and $122,000. In Germany, the average annual salary for a DevOps engineer is $96,461, with a maximum salary potential of up to $115,555. In Poland, the average salary stands at $50,226. In the UK, the average salary for a DevOps engineer is $84,545, with a maximum salary of $101,274. Meanwhile, DevOps engineers in Australia earn an average salary of $98,795, surpassing the salary levels in the UK, Germany, and Canada [8, 10].

2.3.1 Factors Influencing Salary Disparities

The rising demand for DevOps engineers has significantly impacted salaries in the tech industry. Due to the scarcity of professionals with the specialized skill set required for automating processes, optimizing CI/CD pipelines, and maintaining seamless IT operations, companies are offering higher compensation to attract and retain top talent. This trend is particularly evident in regions like the United States, where competition among employers has pushed salaries to higher brackets ($96,600–$122,000).

Salaries for DevOps engineers vary widely across different regions, reflecting local market conditions, economic factors, and industry demand [9]:

- Europe: The salary range spans from $50,000 in Poland to $96,461 in Germany, illustrating how regional economic conditions and the presence of multinational tech companies influence compensation levels.
- Technology Hubs: Countries with advanced technology sectors such as Germany, Australia, and the United States tend to offer higher salaries due to the concentration of tech firms and startup ecosystems requiring DevOps expertise.
- Cost of Living Adjustments: Salaries are also adjusted to reflect the cost of living in a particular region. For instance, while DevOps salaries in Australia are higher than in Germany, they align with Australia's higher cost of living.

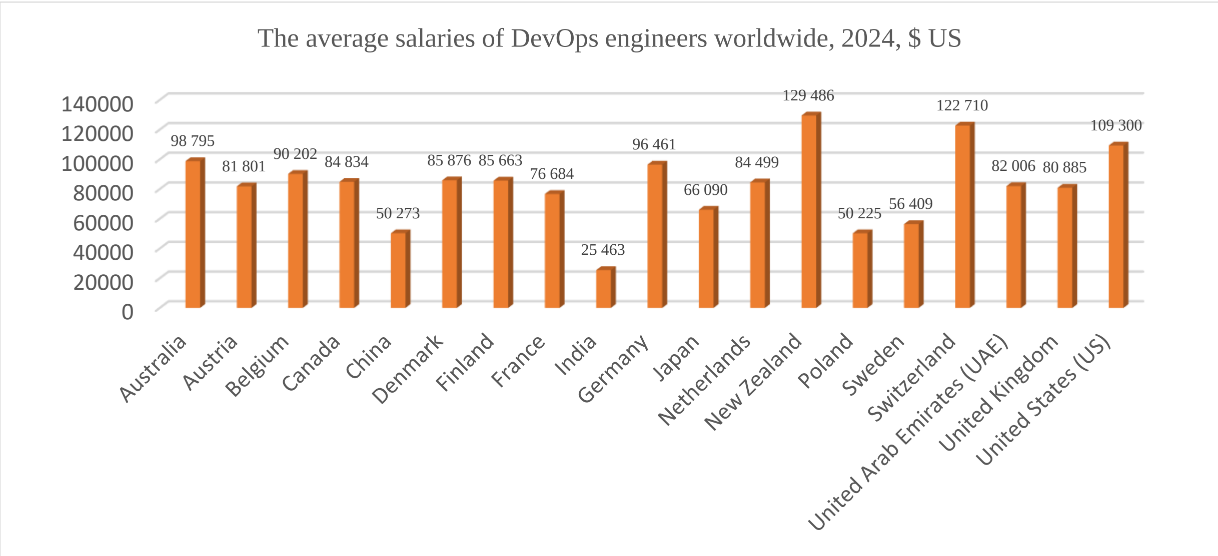


Figure 2: Average salaries for DevOps engineers worldwide.

2.3.2 Strategic Importance of DevOps Roles

The role of DevOps engineers is becoming increasingly strategic as companies strive to enhance efficiency, reduce time-to-market, and improve IT system reliability [5]. As responsibilities expand, particularly in areas like cloud computing, security automation, and infrastructure as code (IaC), the value of DevOps professionals rises, further driving up salaries.

Additionally, broader economic trends and advancements in technology - including the rise of AI-driven automation and containerized environments (e.g., Kubernetes, Docker) [11] - will sustain the strong demand for DevOps engineers, ensuring continued salary growth.

2.3.3 Analysis of Employment Trends and Salaries

Table 1 outlines projected employment and compensation trends in the DevOps industry for 2024 [9]. It categorizes roles, team sizes, and revenue correlations, offering a granular view of salary expectations and organizational structures.

Key insights from Table 1 reveal significant trends in job role distribution, team size, and salary correlations within the DevOps industry [9]. DevOps Engineers constitute only 2% of roles, indicating that while specialized DevOps professionals remain a niche group, they are in high demand due to their critical role in automation and IT operations. Leadership positions, such as Chief Information Officers (14%) and Chief Technology Officers (16%), dominate the job market, underscoring the strong influence and high salaries commanded by executive roles. Additionally, IT Managers (23%) and Directors of IT (21%) make up a substantial portion of the workforce, highlighting the increasing importance of managerial roles in tech-driven organizations.

The correlation between team size and salary structures further emphasizes organizational dynamics. Smaller teams, typically with fewer than 10 employees, generate less than $1 million in revenue, whereas companies with over 10,000 employees significantly contribute to multi-billion-dollar revenues. Engineering team size also plays a crucial role in salary trends, as larger teams tend to offer higher pay scales due to bigger budgets and strategic priorities.

Revenue brackets provide additional insights into salary distribution. Companies generating between $500 million and $1 billion in revenue house the largest proportion of DevOps professionals, suggesting that mid-to-large enterprises prioritize DevOps investments for scalability and efficiency. Interestingly, organizations earning over $10 billion employ fewer DevOps engineers, which may indicate a shift toward automation and cloud-based solutions, reducing the need for additional hires. These insights demonstrate how company size, revenue, and leadership structures collectively shape the demand and compensation for DevOps professionals.

Table 1: Roles and pay scale of DevOps developers in 2024.

Category	Indicator	Value/ Percentage
Roles	DevOps Engineer	2%
	Senior Developer	1%
	Team Leader of Application Engineering	1%
	Build/Automation Manager	1%
	Director of Developer Tools	2%
	CIO	14%
	CTO	16%
	IT Manager	23%
	IT Ops Manager	2%
	Director of IT	21%
	Architect	1%
	Other	14%
Number of Employees	101–249	8%
	250–499	12%
	500–1,000	24%
	1,001–4,999	30%
	5,000–9,999	11%
	10,000+	14%
Revenue ($)	$0–$5 million	3%
	$5.1–$10 million	7%
	$10.1–$25 million	8%
	$25.1–$50 million	8%
	$50.1–$100 million	11%
	$100.1–$250 million	9%
	$250.1–$500 million	9%
	$500.1–$1 million	19%
	$1.1–$5 billion	13%
	$5.1–$10 billion	5%
	$10.1 billion +	6%
	Don't know	2%
Engineering Team Size	Less than 10	6%
	11–20 employees	16%
	21–50 employees	24%
	51–100 employees	24%
	101+ employees	29%

The increasing demand for DevOps engineers, combined with evolving industry needs, is driving salaries upward. Organizations with larger revenue streams and larger team sizes tend to offer more competitive salaries, while the adoption of automation and cloud computing continues to shape job roles and compensation trends. Future projections indicate sustained demand for DevOps expertise, ensuring that these professionals remain highly valued in IT [5].

2.4 Demographic and Firmographic Analysis of it Professionals

The study [9] presents a detailed analysis of the demographic and firmographic characteristics of IT professionals based on survey data (Table 2). The demographic overview reveals that a substantial majority of respondents are male, accounting for 72% of the sample, while females represent 28%. Age distribution indicates that the predominant age group is 35-44 years, comprising 46% of the respondents, followed by those aged 25-34 years, who make up 27%. Geographically, the majority of participants are from the Northeast region of the country. These demographic insights are crucial for understanding the workforce composition and can inform organizational strategies and workforce planning.

In terms of firmographics, the analysis highlights the distribution of job titles within the surveyed organizations. The data, presented through a bar graph, shows that Chief Technology Officers (CTOs) represent 13% of the respondents, while Directors constitute 30% and Managers/Team Leads make up 39%. Notably, 77% of respondents are engaged in both building and operating applications. Further examination through a pie chart reveals that 15% focus exclusively on building applications, 8% on operating applications, and the remaining 77% are involved in both activities. This distribution of responsibilities provides a nuanced view of the roles within IT departments.

Finally, the departmental roles analysis indicates that an overwhelming 90% of participants are employed within IT departments, with the remaining 10% engaged in Software Development & Programming roles. This distribution underscores the significant concentration of IT professionals within IT departments and highlights the specific responsibilities related to technology management and development. The insights derived from this data are invaluable for researchers and practitioners seeking to understand organizational structures, departmental functions, and the distribution of technology-related roles in contemporary settings.

Table 2: DevOps statistics of demographics and monitoring.

Category	Subcategory	Percentage
Gender	Male	72%
	Female	28%
Age	18-24	2%
	25-34	24%
	35-44	47%
	45-54	19%
	55-64	7%
	65 or older	1%
Region	Northeast	24%
	South	18%
	Midwest	33%
	West	25%
Level	Chief Executive Officer	4%
	Chief Information Officer	8%
	Chief Technology Officer	13%
	President	1%
	Vice President	3%
	Director	31%
	Manager/Team Lead	39%
Title	Architect	1%
	Director of DevOps	1%
	Senior Software Developer	4%
	DevOps Engineer	1%
	Team Leader of Application	1%
	Director of IT	30%
	IT Manager	40%
	IT Ops Manager	3%
	Manager, Operations and Release	1%
	CTO	6%
Devops responsibility	Building apps	15%
	Operating apps	17%
	Both	66%
	Other	1%
Department	IT	90%
	Software Development	10%

Analysis of Table 2 indicates several key trends in IT workforce demographics and roles. Gender representation reveals a 72% male dominance, consistent with industry norms, while the 28% female representation suggests a positive shift toward diversity, likely driven by corporate inclusion initiatives. The age group distribution highlights that 47% of professionals are aged 35-44, underscoring the industry's reliance on mid-career experts with substantial experience. Regional trends show that

33% of IT professionals are based in the Midwest, signaling the rise of emerging tech hubs beyond traditional coastal centers, possibly due to lower living costs and growing tech investments. Leadership representation is also notable, with 39% of respondents in Manager/Team Lead roles and 31% in Director positions, indicating strong career progression opportunities in the sector. DevOps responsibilities are broadly distributed, with 66% of professionals involved in both building and operating applications, reflecting the widespread adoption of continuous integration and deployment practices. Finally, the overwhelming presence of IT professionals (90%) within IT departments, rather than software development, emphasizes the crucial role of IT operations in infrastructure management, cloud services, and cybersecurity across organizations.

3 CONCLUSIONS

The DevOps market continues to expand, driven by the growing need for automation, agility, and seamless collaboration in software development and IT operations. The analysis of market trends, salary insights, and regional adoption patterns highlights the increasing strategic importance of DevOps in modern enterprises. Companies leveraging DevOps methodologies report significant benefits, including faster time-to-market, enhanced software quality, and cost efficiency. However, regional disparities in adoption and skill shortages remain key challenges that organizations must address to fully capitalize on DevOps capabilities.

As AI-driven automation reshapes DevOps, new methodologies such as GitOps and AI-based DevOps optimization will redefine industry practices. Companies that proactively adopt these innovations will gain a competitive edge, streamlining development cycles and improving software reliability. The continued integration of AI and cloud-native solutions will solidify DevOps as an essential component of modern IT infrastructures, ensuring sustained demand for skilled professionals in the field.

This research provides valuable insights for a diverse audience, including DevOps engineers, IT managers, HR professionals, and business leaders seeking to optimize their IT strategies. Understanding current trends and future projections will enable organizations to make informed decisions regarding talent acquisition, technological investments, and operational improvements. The continued integration of artificial intelligence and cloud-based solutions will further solidify DevOps as a crucial component of modern IT infrastructures [11], ensuring a steady demand for skilled professionals in this field.

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. Sánchez-Gordón and R. Colomo-Palacios, "Characterizing DevOps Culture: A Systematic Literature Review," in I. Stamelos et al. (eds), Software Process Improvement and Capability Determination. SPICE 2018. Communications in Computer and Information Science, vol. 918, pp. 3-15, Springer, Cham, 2018, [Online]. Available: https://doi.org/10.1007/978-3-030-00623-5_1.

[2] D. Belcher, "Three Trends That Will Transform DevOps in 2023," 2023, [Online]. Available: https://devops.com/three-trends-that-will-transform-devops-in-2023, [Accessed: 21 Dec 2024].

[3] M. Modi, "The Future of DevOps: 2024 and Beyond," 2023, [Online]. Available: https://www.knowledgehut.com/blog/devops/future-of-devops, [Accessed: 23 Dec 2024].

[4] T. Leppanen, A. Honkaranta, and A. Costin, "Trends for the DevOps Security. A Systematic Literature Review," in B. Shishkov (eds), Business Modeling and Software Design. BMSD 2022. Lecture Notes in Business Information Processing, vol. 453, pp. 200–217, Springer, Cham, 2022, [Online]. Available: https://doi.org/10.1007/978-3-031-11510-3_12.

[5] L. Bodnar, M. Bodnar, K. Shulakova, O. Vasylenko, R. Tsarov, and E. Siemens, "Practical Experience in DevOps Implementation," Proceedings of International Conference on Applied Innovation in IT, vol. 12, no. 1, pp. 33-39, 2024, doi: 10.25673/115639.

[6] M. Michalowski, “Top 47 DevOps Statistics 2024: Growth, Benefits, and Trends,” Spacelift, Aug. 05, 2024. [Online]. Available: https://spacelift.io/blog/devops-statistics.

[7] “DevOps Market Share, Size, Trends, Industry Analysis Report, By Organizational Size (Small and Medium Enterprises, Large Enterprises); By Industry; By Deployment; By Offering; By Region; Segment Forecast, 2024 - 2032,” Polaris Market Research, Jan. 2024. [Online]. Available: https://www.polarismarketresearch.com/industry-analysis/devops-market.

[8] S. Spiro, “120 DevOps statistics: Principles, success, & future endeavors,” Spacelift, Mar. 06, 2024. [Online]. Available: https://hutte.io/trails/devops-statistics/.

[9] D. Bhavsar and T. Greek, “DevOps Statistics: Key Figures and Trends You Must Know in 2024,”

Bacancy, Jul. 03, 2024. [Online]. Available: https://www.bacancytechnology.com/blog/devops-statistics#devops-adoption-statistics.

[10] "DevOps Engineer Salaries by Country," Salary Expert, 2024. [Online]. Available: https://www.salaryexpert.com/salary/browse/countries/devops-engineer.

[11] 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," in Proceedings of International Conference on Applied Innovation in IT, vol. 12, no. 2, pp. 19–25, 2024. doi: 10.25673/118105.2.

Digital Marketing: Tools, Trends and Perspectives

Iryna Polishchuk[1], Yuriy Dovhan[1], Oksana Yakushevska[1], Larysa Dovhan[1], Oksana Kravchenko[2], and Viktoriia Petrova[1]
[1] *Vinnytsia Institute of Trade and Economics of the State University of Trade and Economics, Soborna Str. 87, 21050 Vinnytsia, Ukraine*
[2] *Manychat Inc., 8605 Santa Monica Blvd, PMB 64372 West Hollywood, CA 90069, USA*
i.polishchuk@vtei.edu.ua, y.dovhan@vtei.edu.ua, o.yakushevska@vtei.edu.ua, l.dovhan@vtei.edu.ua, ok.oksana16@gmail.com, v.petrova@vtei.edu.ua

Keywords: Marketing, Consumer, Internet, Digitalization, Digital Marketing Tools.

Abstract: The article highlights the issues of modern marketing-oriented business response to new challenges associated with the rapid development of information and digital technologies and digitalization of the population. The purpose of the article is to highlight the results of researchers' approaches to understanding and content of digital marketing, reviewing its modern tools, trends, and prospects. Comparative, institutional, systemic, and structural-functional methods are used to conduct the research. The theoretical aspects of the application of marketing tools and the history of the emergence of digital marketing are briefly highlighted, the differences between traditional and digital marketing are identified. The researchers' approaches to understanding the content of digital marketing, its components, and its tools are reviewed. The key characteristics of digital marketing are identified. The main digital marketing tools that are currently used are described. Modern trends in their application are given. The possibilities of further development and prospects for the application of digital marketing tools in the near and more distant future are identified. It is emphasized that traditional marketing management support tools and marketing implementation tools are naturally transforming and becoming digital. It was concluded that modern businesses adapt to the digital environment and actively use digital marketing tools to meet the demands, maintain competitiveness, and achieve success.

1 INTRODUCTION

According to the Global Overview Report "Digital 2024", the world's population has grown to 8.08 billion, with 5.35 billion individuals (66.2% of the population) using the internet, 5.61 billion (69.4% of the population) unique mobile phone subscribers, and 5.04 billion (62.3% of the population) social media user identities. The data indicates that internet penetration continues to rise, now exceeding 25% in most of the world's regions [1]. Rapid development of the internet, information technologies, and digitalization of the population leads to new challenges in conducting business and marketing.

Nowadays, the economy is experiencing a rapid digitalization, which creates favorable conditions for increasing company efficiency and better identifying and satisfying consumer needs. Today, business and marketing must operate in a digital environment and apply digital tools to be competitive and successful.

2 TRADITIONAL AND DIGITAL MARKETING

Since the mid-20th century, a marketing concept has been the leading concept in business management and organization. Marketing encompasses the processes of research, creation, promotion, and distribution of goods and services aimed at meeting consumer needs and achieving business goals. It involves market analysis, target audience identification, competitive strategy development, demand generation, and customer relationship management.

Marketing utilizes tools to gather external market information, analyze it, and apply it to justify business decisions and actively influence consumer needs and market conditions through marketing instruments.

The scientific systematization of marketing tools began in the first half of the 20th century with the introduction of the so-called “prescriptive approach”, where the seller was seen as a marketing program manager capable of integrating all its components into a marketing mix. The marketing mix, in general terms, is a set of marketing tools used by a company to influence its target market and achieve a desired response. Traditionally, the marketing mix consists of four key elements: product, price, place (distribution), and promotion. Alternative models of the marketing mix, such as 7P, 4C, SIVA, and SAVE, also exist, but the tools used within them remain largely consistent with the classical model.

Marketing tools can be categorized into two groups: marketing management support tools, including market research, and marketing information systems, and implementation tools, including covering product, pricing, distribution, and communication strategies. These tools are selected and applied within the framework of the company’s marketing mix strategy [2].

Traditional marketing is a set of classical methods for promoting goods and services without the involvement of digital technologies. The emergence of digital marketing was a natural consequence of the development of the internet and digital technologies, which fundamentally changed the way brands interact with consumers. In the 1990s, with the advent of the first websites and search engines, businesses began utilizing the online space to promote their products and services. The development of information and digital technologies further expanded possibilities for analytics, automation, and the digitalization of marketing processes. Today, businesses strive to balance conventional marketing efforts with data-driven strategies in the digital era to maintain personalized, relevant, and customer-centric experiences.

Today, companies aim to balance conventional marketing efforts and the power of data in the digital era to maintain a customized, relevant, and customer-centric customer experience [3]. A changing consumer behavior has become a significant driver of the transition from traditional to digital marketing. People actively use the internet for information searches, shopping, and brand interactions, pushing companies to adapt their marketing strategies.

The development of digital marketing aims to strengthen the communication channels of traditional marketing [4]. While both traditional and digital marketing serve the same fundamental purpose, they have significant differences (Table 1).

Digital marketing overcomes certain limitations inherent in traditional marketing. Traditional marketing is typically more expensive, more complex in measuring effectiveness, less interactive and flexible, has limited customer interaction capabilities, and offers a less personalized approach. It operates in a one-way mode, restricting opportunities for feedback and rapid adaptation to changes. Advertising campaigns usually have longer implementation cycles, are planned and executed over extended periods, and are difficult and costly to modify. In contrast, digital marketing is not merely an updated version of traditional marketing but represents a new stage in its evolution, based on the use of digital data, technologies, and two-way communication. This leads to its increasing popularity and application.

Table 1: Traditional Marketing versus Digital Marketing.

Traditional Marketing	Digital Marketing
Expensive, time-consuming, and labor-intensive	Generally cheaper and faster
Results are difficult to measure	Results can be tracked easily and quickly
Limited audience reach	Broad audience reach
Weak interaction with consumers	High engagement with consumers
One-way communication	Two-way communication
Limited personalization	High-level personalization
Difficult to maintain 24/7/365 interaction	Enables continuous 24/7/365 engagement
Limited customer involvement due to the use of traditional technologies	High customer engagement through diverse digital tools
Marketing campaigns require long-term planning and are costly to adjust	Campaigns can be modified quickly and cost-effectively
Weak ability to go viral	Strong ability to go viral
Declining usage	Growing adoption and popularity

3 DIGITAL MARKETING TOOLS: LITERATURE REVIEW

In specialized literature on digital marketing, there is no unified perspective regarding its tools.

Makrides, Vrontis, and Christofi (2019) equate digital marketing with advertising in digital channels [5], however, this interpretation is rather narrow, since digital marketing and its tools extend beyond advertising.

Gupta (2020) describes digital marketing as a general term for various advertising methods used to reach customers in the digital environment, highlighting tactics such as website marketing, search engine optimization, pay-per-click advertising, e-mail marketing, social media marketing, affiliate marketing, mobile marketing, video marketing, and content marketing [6]. Although Gupta (2020) focuses on advertising methods in digital marketing, his list includes tools that are not strictly advertising-based.

David, Adepoju, and Akinyomi (2022) argue that digital marketing involves the marketing and promotion of products or services through technological tools available on the internet, as well as brand, service, and product promotion via various online media [7]. Saleh (2020) notes that digital marketing, also known as online marketing, encompasses the use of the internet to advertise and promote goods and services [8]. Similarly, Salhab (2024) fairly emphasizes that digital marketing includes a broad range of marketing activities that go beyond online marketing [9].

Pașcalău, Popescu, Bîrlădeanu, and Gigauri (2024) consider that digital marketing is not just advertising; today, the term "digital marketing" has changed and expanded from referring only to the promotion of goods and services through digital channels to referring to the use of digital technologies to increase sales by attracting new customers and satisfying their preferences [10].

Antczak (2024), Bermeo-Giraldo, Valencia-Arias, Ramos de Rosas, Benjumea-Arias, and Villanueva Calderón (2022), consider two different ways of marketing in modern business: digital marketing and social media [11, 12]. Digital marketing uses tools such as e-mail, social media, website design, and online advertising to reach customers and interact with them. In turn, social media marketing uses social media platforms [6]. The authors distinguish digital marketing tools and social media marketing. In our opinion, this distinction is rather conditional, since social media marketing uses digital marketing tools in social media channels [6].

Schaefer and Hetman (2019) describe digital marketing tools as "channels to reach consumers" and define internet marketing, mobile phone marketing, social media marketing, display advertising, search engine marketing, and other forms of digital media as key digital marketing tools [13].

Chattopadhyay (2020) defines digital marketing as a set of various strategies and communication channels with customers, encompassing all marketing activities that use the internet or electronic devices (such as search engine optimization, content marketing, social media marketing, and mobile marketing) [14].

Sundaram, Sharma, and Shakya (2020) note that digital marketing encompasses various channels and techniques, such as search engine optimization, pay-per-click advertising, e-mail marketing, social media marketing and mobile marketing [15].

Polydoros (2022) emphasizes that digital marketing uses a variety of strategies to achieve the best results for a business in the field of marketing and advertising. The techniques of digital marketing are as follows: marketing using electronic mail (e-mail marketing), affiliate marketing, viral marketing, search engine marketing, social media marketing, mobile marketing (m-marketing) etc. [4].

Munna, and Shaikh (2023) evaluate different tools and techniques of digital marketing, categorizing them as search engine optimization, pay-per-click, social media marketing, content marketing, e-mail marketing, influencer marketing, website optimization, mobile marketing, video marketing, marketing automation, analytics and data-driven insights, customer relationship management systems [16].

Bermeo-Giraldo, Valencia-Arias, Ramos de Rosas, Benjumea-Arias, and Villanueva Calderón (2022), along with Coman, Popica, and Rezeanu (2020), Amjad (2022) rightly point out that digital marketing relies on the use of data and analytics to understand consumer behavior and preferences. Digital marketing offers a wide range of tools and channels for reaching and interacting with consumers, which allows companies to build stronger relationships with their customers while driving growth and profitability [12, 17, 18]. Therefore, it is reasonable to expand the list of digital marketing tools to include these additional instruments.

Umadevi and Sundar (2023) examined digital marketing tools in e-commerce and determined that they encompass a variety of technologies and strategies that companies use to improve their online presence, engage with their target audience, and drive sales. The authors identified digital marketing tools as search engine optimization, social media marketing, e-mail marketing, content marketing, pay-per-click advertising, influencer marketing, analytics and data insights, chatbots and AI-powered customer service, mobile marketing, retargeting, and remarketing [19].

Dykha, Ustik, Krasovska, Pilevych, Shatska, and Iankovets (2022) identified social media marketing, search engine optimization, contextual and banner advertising, communication marketing, blogging, video marketing, remarketing, and event marketing as the most proven and successful digital marketing tools in e-commerce [20].

Joshi, Com, and Phill (2022) define digital marketing as a general term for marketing products or services using digital technologies, primarily on the internet, but also including mobile phones, display advertising and another digital medium. The authors also include social media marketing, pay-per-click and search engine marketing, e-mail marketing, marketing automation, inbound marketing, affiliate marketing, website marketing, content marketing [21].

The review allows us to highlight key characteristics that, if present, let us consider marketing as “digital marketing”: two-way communication via the internet and any other electronic digital tools; implementation in the digital environment through digital channels and using tools to engage consumers to the online environment from offline one; utilization of marketing digital data, methods, technologies, techniques, and tools; application of marketing digital strategies, tactics, and activities.

The classification of digital marketing methods, technologies, techniques, tools, strategies, tactics, and activities requires further study and improvement.

4 OVERVIEW OF DIGITAL MARKETING TOOLS

The digital marketing tools presented in this article were selected based on their prevalence in modern scientific and specialized literature, their effectiveness for business, which is confirmed by academic studies considered in the literature review, as well as industry reports (Statista, HubSpot) that reflect key trends in their application. The main ones are represented below.

Search Engine Optimization (SEO). It aims to improve the visibility of a website in search engine results. The main goal is to increase organic traffic by optimizing content, technical aspects of the site, and external factors, which allows the site to occupy higher positions in search, increase traffic, and increase brand awareness without spending on paid advertising.

In 2024, the average website has a bounce rate of 37% and an SEO click-through rate of 13% [22].

Social Media Marketing (SMM). It focuses on using social networks to promote the brand, attract an audience, and increase sales. The main goal is to create interesting, useful and visually attractive content that promotes interaction with subscribers, increases brand awareness and stimulates consumer loyalty. Through targeted advertising, organic promotion, and analytics, companies can effectively reach their target audience and manage their reputation in real time.

Digital Advertising. It is online advertising that uses digital platforms and technologies to promote products, services, or brands. It includes banner advertising, search advertising, contextual advertising, video advertising, social media advertising, and e-mail advertising.

In 2024, 31% of the audience paid attention to advertising on social networks. Social media advertising spending increased by 15% to $247 billion. 17% of users made purchases through social networks [1].

Website Marketing. It focuses on promoting a company, brand, or product through its own website in order to attract visitors, increase conversions, and improve the user experience.

A successful example is the strategy of Airbnb (airbnb.com). Their website is an example of an effective combination of user experience, personalization, and marketing tools.

Mobile Marketing. It focuses on interacting with consumers via mobile devices (smartphones, tablets) and includes SMS and MMS mailings, push notifications, mobile applications, advertising on social networks, responsive websites, wi-fi marketing, geotargeting, use of NFC, QR codes, and voice assistants.

Mobile advertising spending is projected to reach nearly $400 billion by 2024 [23].

E-mail Marketing. It includes sending commercial offers, newsletters, personalized promotions, trigger letters, and automated campaigns. E-mail marketing allows brands to build long-term relationships with consumers by providing them with useful and targeted content that can drive repeat purchases, increase loyalty and engagement.

Email marketing revenue is expected to exceed $9.5 billion by 2024. More than a half of marketers surveyed reported a 100% improvement in the return on investments (ROI) from their email marketing campaigns [24].

Content Marketing. It focuses on creating, distributing, and promoting valuable, relevant, and useful content to attract and retain a target audience.

In 2024, 29% of marketers actively used content marketing [22].

Retargeting and Remarketing. They aim to return users who have previously interacted with the brand but did not take a targeted action (such as a purchase, registration or inquiry). Used to remind about products or services through advertising or e-mail campaigns. Retargeting is usually implemented through banner or contextual advertising, remarketing is more often used in the form of e-mail newsletters with personalized offers.

Inbound Marketing. It aims to attract potential customers through the creation of valuable and relevant content, rather than through aggressive advertising or pushy sales. The main idea is to help consumers discover the brand themselves by offering them useful information, educational content, and solutions to their problems. This is achieved through content marketing, SEO, social media, blogging, webinars, e-mail marketing, and marketing automation.

Inbound Marketing is a strategy developed by HubSpot, which pioneered this approach. It engages potential customers with a wealth of useful, free content like blogs, eBooks, webinars, online courses, and tools (like free CRMs or template generators) that address the needs of your target audience.

Video Marketing. The main types of video marketing include commercials, educational videos, product reviews, live streams, webinars, video stories (stories), content from influencers and viral videos.

In 2024, 89% of companies used video as a marketing tool. 95% of video marketers consider video as an important part of their overall strategy [25].

Influencer Marketing. It involves brands working with thought leaders (influencers) to promote products or services to their audience. Influencers can be bloggers, celebrities, experts in a certain niche, or popular personalities on social networks.

In 2024, the market for influencer marketing grew to $24 billion. 68% of marketers collaborated with niche influencers (less than 100.000 followers) due to the higher loyalty of their audience [26].

Affiliate Marketing. It involves cooperation between a company (advertiser) and partners (affiliates) who promote its products or services and receive a commission for each successful conversion (purchase, subscription, click, etc.). Affiliates use various marketing channels, such as blogs, social networks, e-mail newsletters, channels, advertising, and specialized websites to attract potential customers. The models related to affiliate marketing are the following: cost per action, cost per sale, revenue sharing, cost per thousand or cost per view, cost per install.

Booking.com is an example of successful digital affiliate marketing. It has implemented a large-scale affiliate program that allows website owners, travel bloggers, travel agencies, and other partners to earn commissions on bookings made through their affiliate links or widgets.

Communication Marketing. Communication in digital marketing involves using social media messaging tools, chatbots, e-mail, SMS, MMS, and even business voice messaging [20]. A successful example of communication marketing is Nike’s campaign in the Nike Training Club App and across social media. During the COVID-19 pandemic, Nike opened up free access to premium content in the App, and also actively communicated with the audience via Instagram, YouTube, and used push notifications, supporting users’ motivation to exercise at home.

Digital Event Marketing. This approach includes planning, promoting, and hosting online events to attract audiences, enhance brand awareness, and generate leads. Digital events include webinars, online conferences, virtual exhibitions, masterclasses, panel discussions, and live streams conducted on online platforms. These events allow businesses to interact with a global audience without geographical restrictions, increasing participant engagement through interactive elements such as chats, polls, and Q&A sessions.

16% of marketers plan to try experiential marketing this year, while many others are increasing their event budgets or trying to get more out of their event marketing strategies [22].

Marketing Analytics and Data Insights Instruments. These tools facilitate the collection, analysis, and interpretation of user behavior data in digital environments to optimize marketing strategies. Using web analytics platforms, social media analytics tools, CRM systems, and other data platforms, companies gain insights into traffic, conversions, content engagement, lead sources, and the effectiveness of advertising campaigns. This enables marketers to track trends, understand audience needs, and make data-driven decisions to improve campaign performance.

ASUS has successfully used marketing analytics tools such as Improvado to analyze data from various channels, saving 90 hours per week on data-related tasks and achieving up to 30% optimization of marketing spend [27].

The use of Big Data allows businesses to automate the analysis of vast datasets, identify hidden patterns, and predict the effectiveness of future marketing actions. Big Data technologies have laid the foundation for personalization systems, which are now widely applied across nearly all industries [28].

Artificial Intelligence and Machine Learning (AI, ML). These technologies are used in digital marketing to analyze large volumes of data, predict user behavior, and automate processes. With artificial intelligence, marketers can personalize content, optimize advertising, automate customer interactions, and enhance the user experience. Some of the most common applications of AI and ML in marketing include chatbots, voice assistants, dynamic content, automated audience segmentation, predictive analytics, and image recognition. The use of AI and ML in marketing enables companies to work more efficiently, respond quickly to market changes, and increase consumer engagement. Thus, Amazon uses Amazon Personalize and Amazon Bedrock to analyze customer behavior and create personalized product recommendations, which significantly increases engagement and sales.

Marketing Automation. This involves the use of specialized software and technologies to automate marketing processes, helping companies interact more effectively with their audience, reduce manual labor, and improve productivity. Automation covers processes such as e-mail marketing, lead generation, audience segmentation, social media management, personalized offers, and user behavior analysis. Popular digital platforms allow businesses to create complex automated sales funnels and trigger-based campaigns that launch based on user actions.

One of successful examples is the use of the Salesforce Marketing Cloud platform by Coca-Cola, which implements personalized marketing campaigns in real time, focusing on consumer behavior, their geolocation, and previous interactions with the brand.

The list of digital marketing tools provided here is not exhaustive and continues to expand as technology advances.

5 DIGITAL MARKETING TRENDS AND PERSPECTIVES

Today, more than 69.4% of the world's population uses mobile devices, and since the beginning of 2023, their number has increased by 138 million (+2.5%). Over 66% of people use the internet, with 97 million new internet users (+1.8%) added in 2023 alone. Meanwhile, the total number of social media users increased by 266 million (+5.6%). More users are turning to the internet and social media for product and brand research, purchasing, and gathering information about goods and services [1]. Businesses and marketing professionals cannot afford to ignore these digitalization trends. Most recognize the need to enhance their analytical and communication efforts in the digital space.

Maintaining competitiveness and business success in digital marketing today is driven by trends that are currently associated with the growing capabilities of artificial intelligence in data analytics and marketing execution, the continuing competition among social media platforms for consumer attention and success metrics, the advancement of search engine optimization techniques and paid digital advertising models, and the development of new digital marketing strategies and tactics [29].

Key modern instrumental trends in digital marketing include AI-driven personalization, voice search optimization, video marketing evolution, influencer marketing maturity, privacy-focused advertising, interactive content, sustainability and ethical marketing, omnichannel marketing integration, social commerce expansion, and content experience.

The future growth and application of digital marketing tools should be assessed in the context of technological, economic, and socio-cultural developments.

From a technical and technological perspective, the advancement of digital marketing depends on three key components, particularly, the state of information technology development, the evolution of digital technologies and hardware.

The development of digital marketing is impossible without technological progress. Information technology is the foundation of digital marketing, as it enables the processing, storage, analysis, and transmission of data necessary for effective marketing management. Digital technologies define the formats, channels, and tools of digital marketing, while hardware advancements determine its speed, quality, and scale. More powerful processors, graphic cards, and mobile devices allow for the processing of large data volumes, support high-quality video content, AR/VR, 3D animation, and interactive web applications. With the advancement of smartphones, IoT (Internet of Things), and mobile devices (smartwatches, VR glasses), marketers gain new touchpoints with their audience, enabling more precise targeting and personalized advertising. Information technologies facilitate data collection and analysis, digital technologies open new communication channels, and hardware innovations make marketing campaigns more interactive, engaging, and accessible to consumers.

From an economic perspective, digital marketing continues to lower entry barriers for businesses, giving small and medium enterprises access to global markets while increasing competition and the demand for effective marketing strategies. In the future, digital technologies will continue transforming economic models.

The rise of digital marketing has profoundly changed society, making communication between companies and consumers more interactive, personalized, and accessible. However, the digital divide between different social groups and issues of data privacy remain critical challenges that need to be addressed in the future.

The globalization of digital marketing has opened up new opportunities for companies to enter international markets, but at the same time it requires a deeper understanding of the cultural characteristics of different regions. Adaptation of content, consideration of cultural differences, local traditions, and ethical standards are becoming critical factors in a successful marketing strategy.

Therefore, the overall development and application of digital marketing tools will continue to depend on the directions of development of society, science, and economy.

6 CONCLUSIONS

Our consumer society is becoming much more digitalized. The rapid development of digital technologies enhances the efficiency of companies, allowing them to analyze and meet consumer needs more effectively. Modern businesses adapt to the digital environment and actively use digital marketing tools to satisfy demands, maintain competitiveness, and achieve success. Traditional marketing management support tools and marketing implementation tools are naturally transforming and becoming digital.

Digital marketing is not merely a modernized version of traditional marketing. It represents a qualitatively new stage in its evolution, based on the use of digital technologies, data analysis, and interactive communication with consumers. The classification of methods, technologies, techniques, tools, approaches, strategies, and tactics in digital marketing requires further research and refinement, considering its continuous expansion and adaptation to digital technology advancements and changes in consumer behavior.

This study conducted a comparative analysis of traditional and digital marketing, identified key characteristics of digital marketing, systematized digital marketing tools, and analyzed trends and prospects for its development. All this as well as the identification of gaps in the classification of digital marketing tools will allow the academic community to develop a clearer systematization, which will contribute to the standardization and comparability of research in this field.

The further development and future prospects of digital marketing tools should be examined in the context of improving digital analytics, interactivity, and two-way communication. Their evolution can only be predicted by considering the specifics and dynamics of technical, economic, and socio-cultural progress.

REFERENCES

[1] S. Kemp, “Digital 2024: Global Overview Report,” [Online]. Available: https://datareportal.com/reports/digital-2024-global-overview-report, [Accessed: 28 Feb 2025].

[2] I. Polishchuk, Yu. Dovhan, O. Hromova, L. Dovhan, and I. Hnydiuk, “Tools for Marketing Management of the Innovation and Investment Activity,” Financial and Credit Activity: Problems of Theory and Practice, vol. 2, issue 43, pp. 228-237, 2022, doi: 10.55643/fcaptp.2.43.2022.3654.

[3] A.S. Bist, V. Agarwal, Q. Aini, N. Khofifah, “Managing Digital Transformation in Marketing: “Fusion of Traditional Marketing and Digital Marketing,” International Transactions on Artificial Intelligence, vol. 1, issue 1, pp. 18-27, 2022, doi: 10.34306.

[4] G. Polydoros, “Digital Marketing Techniques and Tools in Covid-19,” International Journal of Research and Innovation in Social Science, vol. 6, isssue 11, pp. 86-91, 2022.

[5] A. Makrides, D. Vrontis, and M Christofi, “The Gold Rush of Digital Marketing: Assessing Prospects of Building Brand Awareness Overseas,” Business Perspectives and Research, vol. 8, issue 1, pp. 4-20, 2019, doi: 10.1177/2278533719860016.

[6] N. Gupta, “Digital Marketing: Trends, Opportunities, and Challenges,” Asian Journal of Management, vol. 11, issue 4, pp. 434-440, 2020, doi: 10.5958/2321-5763.2020.00066.9.

[7] O.L. David, O.O. Adepoju, and E.R. Akinyomi, “Digital Marketing Tools and Sales Performance of Shoprite, Shopping Mall Akure, Ondo State Nigeria,” Open Journals of Management Science, vol. 3, issue 2, pp. 21-34, 2022, doi: 10.52417/ojms.v3i2.409.

[8] H. Saleh, “Enhance Small Medium Enterprise (Smes) Family Business in Malaysia Through E-Marketing Strategies,” International Journal of Scientific &

Technology Research, vol. 9, issue 2, pp. 3374-3377, 2020.

[9] H.A. Salhab, "Analytical Study on the Impact of Digital Marketing Strategies and Performance of Small and Medium-sized Companies," International Journal of Religion, vol. 5, issue 7, pp. 248-255, 2024, doi: 10.61707/aars0p68.

[10] S.-V. Pașcalău, F.-A. Popescu, G.-L. Bîrlădeanu, and I. Gigauri, "The Effects of a Digital Marketing Orientation on Business Performance," Sustainability, vol. 16, issue 15, 6685, 2024, doi: 10.3390/su16156685.

[11] B. Antczak, "The Influence of Digital Marketing and Social Media Marketing on Consumer Buying Behavior," Journal of Modern Science, vol. 56, issue 2, pp. 310-335, 2024, doi: 10.13166/jms/189429.

[12] M.C. Bermeo-Giraldo, A. Valencia-Arias, J.D. Ramos de Rosas, M. Benjumea-Arias, and J.A. Villanueva Calderón, "Factors Influencing the Use of Digital Marketing by Small and Medium-Sized Enterprises during COVID-19," Informatics, vol. 9, issue 86, 2022, doi: 10.3390/informatics9040086.

[13] M. Schaefer and O. Hetman, "Effective Tools of Digital Marketing Implementation," University Economic Bulletin, vol. 41, pp. 67-74, 2019, doi:10.31470/2306-546X-2019-41-67-74.

[14] P. Chattopadhyay, "A Study on Various Digital Marketing Tools and its Applications: Descriptive Approach," International Journal of Research Publication and Reviews," vol. 1, issue 7, pp. 46-50, 2020.

[15] R. Sundaram, R. Sharma, and A. Shakya, "Power of digital marketing in building brands: A review of social media advertisement," Journal of International Management, vol. 11, pp. 244-254, 2020.

[16] A.S. Munna, and S.I. Shaikh, "Impact of Digital Marketing Tools and Methods on SMEs and Proposed Recommendations," pp. 173-191, 2023, doi: 10.4018/978-1-6684-7735-9.ch008.

[17] C. Coman, M.M. Popica, and C. Rezeanu, "The adoption of digital marketing by SMEs entrepreneurs," in T. Antipova, and A. Rocha (Eds.), Digital science 2019, vol. 1114, pp. 431-441, 2020, doi: 10.1007/978-3-030-37737-3_37.

[18] T. Amjad, "Digital entrepreneurial marketing: A bibliometric analysis reveals an inescapable need of business schools," The International Journal of Management Education, vol. 20, issue 2, 100655, 2022, doi: 10.1016/j.ijme.2022.100655.

[19] B. Umadevi, and D. Sundar, "Revolutionizing Retail: The Impact of Digital Marketing Tools on E-Commerce Promotion," International Research Journal of Modernization in Engineering Technology and Science, vol. 5, issue 12, pp. 3895-3903, 2023, doi: 10.56726/IRJMETS47929.

[20] M. Dykha, T. Ustik, O. Krasovska, D. Pilevych, Z. Shatska, and T. Iankovets, "Marketing Tools for the Development and Enhance the Efficiency of E-Commerce in the Context of Digitalization," Estudios de Economía Aplicada, vol. 39, issue 5, 2021, doi: 10.25115/eea.v39i5.5234.

[21] G.R. Joshi, M. Com, and M. Phill, "Study on Comparison Between Traditional Marketing v/s Digital Marketing," Journal of Emerging Technologies and Innovative Research, vol. 9, issue 1, pp. 544-555, 2022.

[22] HubSpot, "State of Marketing Report 2024," 2024. [Online]. Available: https://www.hubspot.com/marketing-statistics. [Accessed: 30 April 2025].

[23] Statista, "Mobile advertising spending worldwide from 2007 to 2024," 2024. [Online]. Available: https://www.statista.com/statistics/303817/mobile-internet-advertising-revenue-worldwide/ [Accessed: 30 April 2025].

[24] Statista, "E-mail Marketing Worldwide - Statistics & Sacts," 2024. [Online]. Available: https://www.statista.com/topics/1446/e-mail-marketing/ [Accessed: 30 April 2025].

[25] Wyzowl, "Video Marketing Statistics 2025," 2025. [Online]. Available: https://www.wyzowl.com/video-marketing-statistics/ [Accessed: 30 April 2025].

[26] Influencer Marketing Hub, "Influencer Marketing Benchmark Report 2025," 2025. [Online]. Available: https://influencermarketinghub.com/influencer-marketing-benchmark-report/ [Accessed: 30 April 2025].

[27] Improvado, "ASUS Unified Marketing Data Across All Regions of Presence," [Online]. Available: https://improvado.io/customer/asus. [Accessed: 30 April 2025].

[28] I. Polishchuk, Yu. Dovhan, I. Kramar, L. Dovhan, and O. Yakushevska, "Marketing Logistics and its Digitalization," Proceedings of the 12th International Conference on Applied Innovations in IT, vol. 12, issue 2, pp. 199-205, 2024, doi: 10.25673/118134.

[29] E. Prunty, "Key Digital Marketing Trends for 2024," [Online]. Available: https://digitalmarketinginstitute.com/blog/key-digital-marketing-trends-for-2024. [Accessed: 28 Feb 2025].

Impact of Artificial Intelligence on Customer Relationship Management: a Bibliometric Analysis

Aleksy Kwilinski[1], Nataliia Trushkina[2], Yuliia Remyha[3] and Tamila Patlachuk[2]
[1]*The London Academy of Science and Business, Mary At Hill Str. 6-7, EC3R 8EE London, England*
[2]*Research Center for Industrial Problems of Development of the National Academy of Sciences of Ukraine, Inzhenernyi Lane 1a, 61165 Kharkiv, Ukraine*
[3]*State Non-commercial Company State University "Kyiv Aviation Institute", Liubomyra Huzara Avenue 1, 03058 Kyiv, Ukraine*
a.kwilinski@london-asb.co.uk, nata_tru@ukr.net, remyga_julia@ukr.net, tpatlachuk@ukr.net

Keywords: Artificial Intelligence, Customer Relationship Management, Logistics Service, Customer Experience, Loyalty, Customer Orientation, Digital Transformation, Information Environment, Digital Marketing, Digital Technology, Digital Innovation, Bibliometric Analysis.

Abstract: In modern conditions, the strategic task of logistics management of companies is to form an appropriate customer relationship management system that is able to quickly respond to crisis situations, risks, threats, transformation of the business environment, as well as flexibly adapt to unstable demand and constant changes in customer needs and preferences. In this regard, it is advisable to introduce qualitatively new management and customer-oriented approaches, smart technologies, customer experience management methods, digital marketing tools, and digital innovations. At the same time, in recent years, the use of artificial intelligence as the most influential digital innovation in customer relationship management has played a significant role. Therefore, the purpose of this article is to identify the relationship between artificial intelligence and customer relationship management by characterizing the evolution of key patterns of scientific publications on this problem. To achieve the goal, a relevant sample of scientific articles was formed based on identifying periods of publication activity and bibliometric analysis of keyword coincidences to identify promising areas of research in this area. The formed sample of publications for the study includes 797 documents indexed by the database Scopus for the period 2001-2024. Bibliometric analysis and visualization of its results were carried out using the VOSviewer software product. Based on visualization maps, seven clusters were identified and characterized by the content coincidence of keywords in publications and five stages of evolutionary development of the customer relationship management system using artificial intelligence. Based on the analysis of empirical data, an exponential growth in the number of publications on the selected issues was confirmed (the annual increase in the number of scientific papers on this topic is 15.3%). The results of the analysis can be used in further research to substantiate and develop a strategy for the digital transformation of the customer relationship management system.

1 INTRODUCTION

In the era of digitalization and automation of business models and processes, the issues of information support for the customer relationship management system are becoming increasingly relevant. This is due to the digital transition in the logistics management of companies [1] due to changes in the needs, preferences, demand and behaviour of consumers. In this regard, there is an urgent need to find and apply innovative approaches, smart and intelligent systems [2], customer-oriented customer relationship management technologies [3], methods and tools of digital marketing.

As stated in the Zendesk Report [4], customer loyalty can contribute to business success due to increased fierce competition and economic uncertainty and instability. Therefore, improving the quality of customer service stimulates loyalty. Thus, 74% of customers are loyal to a particular brand or company.

A 2024 study of various aspects of digital customer interactions by CMSWire (the world's leading community of digital customer experience professionals) [5] found that 81% of respondents consider digital customer service to be a business priority and the most important tool for companies. According to Gartner [6], 91% of companies plan to develop artificial intelligence to manage relationships with consumers in the future.

According to surveys by IBM Corporation [7], 35% of companies are already using artificial intelligence technologies in business. Another 42% of enterprises are exploring the possibilities of their implementation. Today, 28% of companies are implementing a holistic artificial intelligence development strategy, while 25% have a strategy that focuses only on limited or specific areas of activity. According to Statista [8], the global market for artificial intelligence in marketing was 15.8 billion dollars in 2021. It is predicted that by 2028, this figure will increase by 6.8 times to 107.5 billion dollars [8].

According to research, in recent years, artificial intelligence tools have been embedded in the digital marketing landscape. More than 80% of digital marketing experts say they integrate some form of artificial intelligence technology into their online marketing activities [8].

According to Statista [9], artificial intelligence was used in 2024 to manage customer experience in the following areas: open feedback analysis (50% of respondents); content creation (50%); machine learning and predictive analytics (26.6%); consumer behaviour forecasting (21.4% of respondents).

In a study conducted by AIPRM Corporation [10], artificial intelligence is most often used in business for customer service (56% of companies surveyed).

A survey conducted in 2024 identified the biggest obstacles for companies around the world in using artificial intelligence to serve customers. For example, over 40% of respondents emphasize the lack of experience to implement this tool and the need to obtain specialized knowledge. However, some organizations (23%) show resistance to adopting this technology and changing their processes [11].

Therefore, the issues of digital transformation of the customer relationship management system are extremely relevant and timely and require in-depth scientific research and development.

Given the above, it is relevant and necessary to analyze the impact of artificial intelligence on customer relationship management using a bibliometric approach.

2 LITERATURE REVIEW

In the scientific and educational community, the business environment, there is a steady attention to the issues of customer relationship management, which is due to the constant development of the process of providing logistics services in the direction of ensuring customer requirements in receiving orders, on the one hand, and profit from managing logistics activities, on the other.

Analysis of the scientific literature indicates a variety of approaches of scientists to defining the essence and content of the concept of “customer relationship management”. Scientists use different concepts, namely: “customer orientation”, “customer centricity”, “customer-oriented approach”, “customer focus”, “logistics service”, “loyalty”, “consumer service”, “customer interaction”, “customer involvement”, “customer engagement”, “customer experience”, “customer experience management”, etc. That is, there are many interpretations of these terms, which are based on various scientific concepts, theories and provisions.

In recent years, many foreign scientists (R. Abduljabbar et al. [12]; P. B. Acharjee et al. [13]; S. Chatterjee et al. [14]; S. Chatterjee & R. Chaudhuri [15]; Y. Fu, G. Guo & T. S. Huang [16]; K. Krishnareddy et al. [17]; V. Kumar et al. [18]; D. Ozay et al. [19] and others) have been investigating the issue of integrating artificial intelligence into the customer relationship management system. This, according to researchers, will provide transformative opportunities for improving customer interaction and increasing work efficiency.

Thus, many scientific publications indicate the interest of scientists in studying various aspects of customer relationship management as an important component of logistics management of companies. However, despite a wide range of research on the given topic, theoretical issues regarding customer relationship management in accordance with the challenges associated with the digital transformation of logistics management and the rapid use of artificial intelligence remain insufficiently studied.

Therefore, the purpose of the article is to identify the relationship between artificial intelligence and customer relationship management by characterizing the evolution of key patterns of scientific publications on this problem.

3 METHODOLOGY

The theoretical and methodological basis of the study is the provisions of the theories of systems, information society, network economy, digital economy; concepts of logistics and marketing management, customer relationship management.

The information base of the study is analytical materials from CMSWire, Gartner, IBM Corporation, Statista, Zendesk, which highlight the results of surveys and statistical analysis on the problems of the impact of artificial intelligence on customer relationship management.

It is worth noting that today many methods have been developed for assessing the digital transformation of business processes of companies of various industries, one of the components of which is customer-centricity (customer experience, digital logistics service, omnichannel, digital marketing and communications). Such methods include: the Deloitte digital maturity model; the digital transformation index (Arthur D. Little agency); the digital maturity index of enterprises; Digital Piano model (Global Centre for Digital Business Transformation initiated by IMD and Cisco); Digital Transformation Change Index proposed by Ionology.

The following general scientific methods were used in the research process: analysis and synthesis, expert survey, bibliometric analysis, comparison, classification, system approach, structural-logical generalization.

The study selected bibliometric analysis as a method that reveals the connection between artificial intelligence and customer relationship management. This type of analysis is based on mathematical graph theory, clustering methods, and scientific visualization, which makes it widely applicable in various fields of science [20].

Based on the structuring of a large volume of metadata of scientific publications, bibliometric analysis allows us to identify the essence of the subject area and its conceptual foundations and to substantiate the evolution of the research area [20]. The research methodology included the following main stages: data collection and analysis, selection of a visualization tool, graphical representation of the identified relationships, and interpretation of the results.

This was implemented using the VOSviewer v.1.6.19 software product. The functionality of this program involves creating keyword visualization maps based on compatibility data, maps of authors or countries based on the number of citations, etc. [21]. In addition, network visualization maps in VOSviewer are displayed in several ways (for example, by content criterion, by publication period) [20].

An important stage of bibliometric analysis, which ensures its quality, is the selection of a data source and the formation of a relevant sample of publications. The Scopus database was chosen as a data source due to the breadth of its coverage of such subject areas as computer science; business, management and accounting; decision sciences; economics, econometrics and finance; social sciences, etc. To form the data sample, the chronology of the release of 839 publications for the search query “Artificial Intelligence” (or AI) and “Customer Relationship Management” (or CRM) for the years 1989-2025 was first determined, and then the period of publication activity of the research topic, namely from 2001 to 2024. The key categories for selecting scientific publications were the title, abstract and keywords for them. Thus, the studied sample of publications included 797 works that met the above criteria, published in the period 2001-2024 and indexed by the Scopus database.

4 RESULTS

A quantitative analysis of the formed sample of research papers showed an exponential growth of research in the context of the relationship between artificial intelligence and customer relationship management during the period 2001-2024 (Fig. 1). At the same time, it can be assumed that by the end of 2025 there will also be a trend of significant growth in the number of publications on this topic compared to the previous year. On average, the annual growth of publications is 15.3%.

As the analysis shows, the following keywords are mostly used in publications: Artificial Intelligence (487 documents), Customer Relationship Management (208), Public Relations (205), Sales (197), Information Management (85), Customer Satisfaction (83), Machine Learning (78), Decision Making (78), Decision Support Systems (70), CRM (60), Learning Systems (55), Customer Experience (37), Big Data (37), Supply Chain Management (36), Marketing (34), Electronic Commerce (32), AI (31), Customer Relationship Management Systems (30 documents) etc.

In addition, it is appropriate to quantitatively analyse publications from the perspective of countries of origin. The size of the circle on the visualization map (Fig. 2) is proportional to the volume of publications indexed in the Scopus database, and the colour of the circle corresponds to the average number of citations of scientific works. For example, the largest number of publications belongs to authors affiliated with Indian institutions and organizations (176 documents of the studied sample), the second place in this parameter is occupied by scientists from the USA (111), the third – from China (101), the fourth – from the UK (47), and the fifth place belongs to representatives from France (35 documents). However, the overwhelming number of publications of representatives of India and China is inferior to the USA, the UK and France in terms of the level of citations. Thus, each article of scientists from the USA, the UK and France is cited on average 25 times, which is more than twice the value of this parameter for researchers from India (on average 15 times) and China (12 times). The highest citation rates are among scientists from Finland, Denmark, and Cyprus, with each publication cited in other authors' works an average of 42 times.

Key publications that publish papers on the topic of customer relationship management using artificial intelligence technologies include: Lecture Notes In Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics (31 documents); Lecture Notes In Networks And Systems (17); Communications In Computer And Information Science (15); Advances In Intelligent Systems And Computing (9); Ceur Workshop Proceedings (9 documents).

The main organizations involved in solving the selected problem are: Indian Institute of Technology Kharagpur (19 documents); Amity University (14); Symbiosis International Deemed University (10); The Hong Kong Polytechnic University (8); Indian Institute of Management Mumbai (8); University of Nicosia (8); Chinese Academy of Sciences (6); K L Deemed to be University (6); Indian Institute of Management Ranchi (6); CNRS Centre National de la Recherche Scientifique (5 documents).

By type of documents, the works can be ranked as follows: 1st place is taken by Conference Paper (350 documents or 41.7% of the total number of publications on the selected research topic); 2nd place – Article (311 or 37.1%); 3rd place – Book Chapter (92 or 11%); 4th place – Conference Review (37 or 4.4%); 5th place – Review (21 or 2.5%); 6th place – Book (18 documents or 2.1% of the total number of publications).

Mostly scientific works that consider the impact of artificial intelligence and customer relationship management are published in the following fields of knowledge: Computer Science (521 documents or 29.8% of the total number of publications on this topic); Business, Management and Accounting (260 or 14.9%); Engineering (246 or 14.1%); Decision Sciences (143 or 8.2%); Mathematics (137 or 7.8%); Economics, Econometrics and Finance (106 or 6.1%); Social Sciences (79 or 4.5%). This indicates that the research topic is interdisciplinary and multifaceted.

The main sponsors that finance scientific publications on the selected topics include the following: National Natural Science Foundation of China (21 documents); European Commission (8); Ministry of Science and Technology of the People's Republic of China (8); Fundamental Research Funds for the Central Universities (6); Horizon 2020 Framework Programme (5); Ministry of Education (5); National Key Research and Development Program of China (5); National Research Foundation of Korea (5); UK Research and Innovation (5 documents) etc.

Based on the results of the analysis of the coincidences and closeness of the relationship between the keywords of the selected sample of publications, network visualization maps were constructed (Fig. 3 and Fig. 4), and 7 clusters on the research topic were identified and characterized (Fig. 3). The first cluster (red, Fig. 3) contains the largest number of terms (namely 17 items), among which the following can be mentioned: "artificial intelligence", "customer relations", "digital transformation", "digitalization", "cloud computing", "Industry 4.0", "management", "supply chain management". It is worth noting the keyword "artificial intelligence", the frequency of its co-use in the studied sample is 487, and the strength of the connection is 2943.

The second cluster (12 items, green, Fig. 3) combines such terms as "AI", "blockchain", "customer relationship management", "e-commerce", "personalization", "predictive analytics", "customer experience", "customer engagement". In this cluster, the keyword "customer relationship management" has the highest frequency – the ratio is 208, while the strength of the association is 1631.

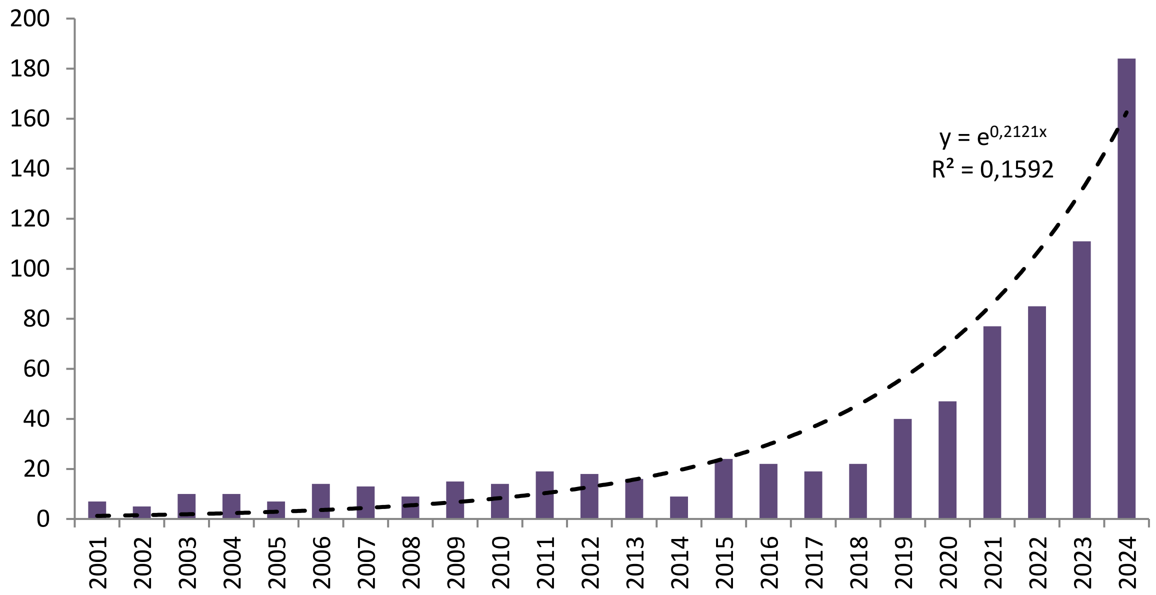


Figure 1: Dynamics of publications on artificial intelligence and customer relationship management for 2001-2024.

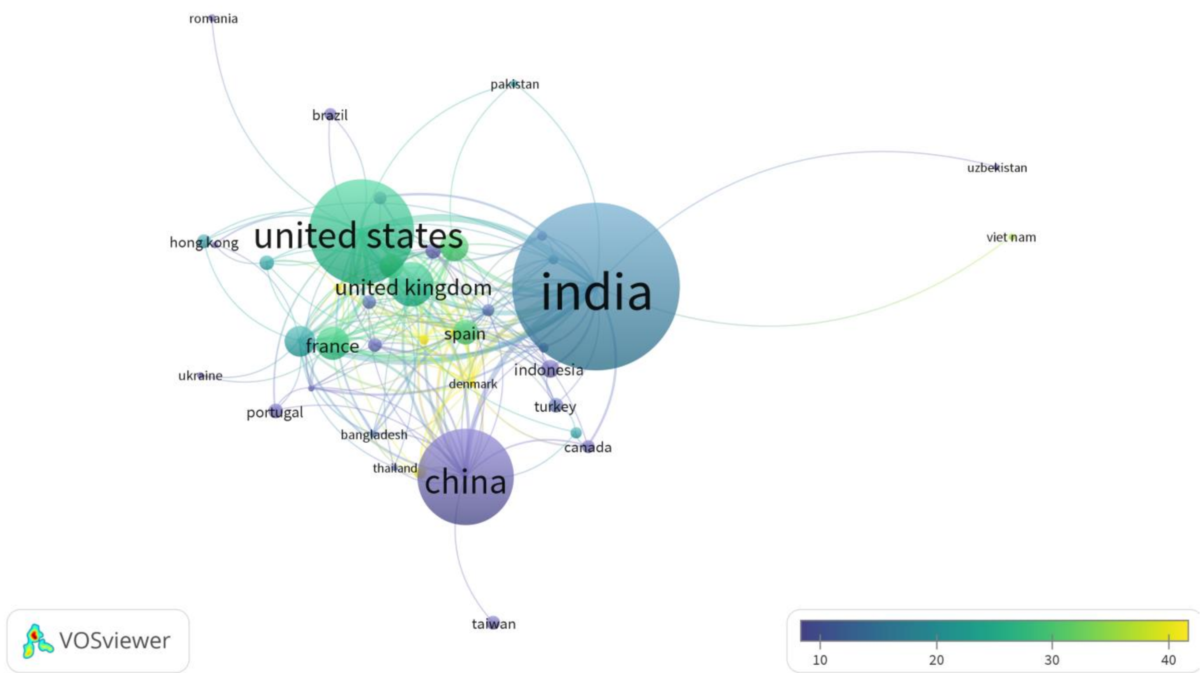


Figure 2: Countries of origin of publications on artificial intelligence and customer relationship management.

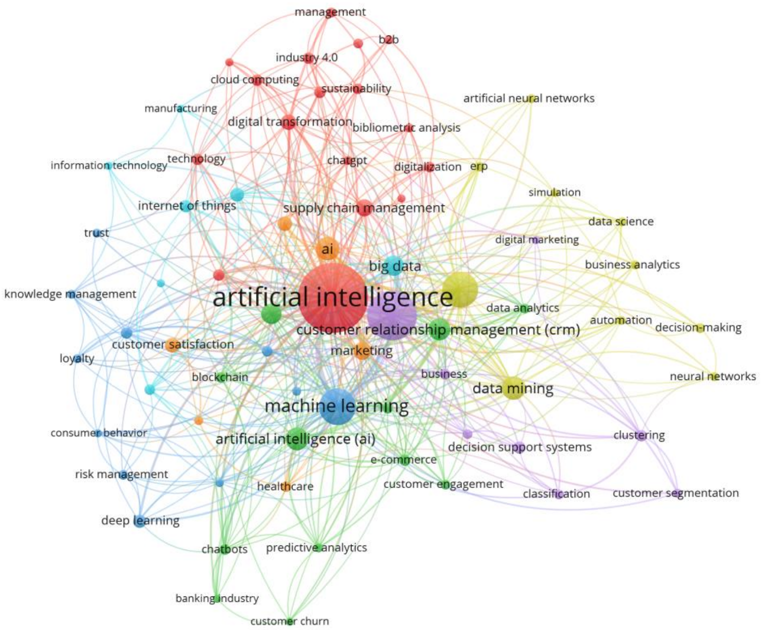


Figure 3: Visualization map of the bibliometric analysis of publications devoted to the application of artificial intelligence for customer relationship management (content aspect).

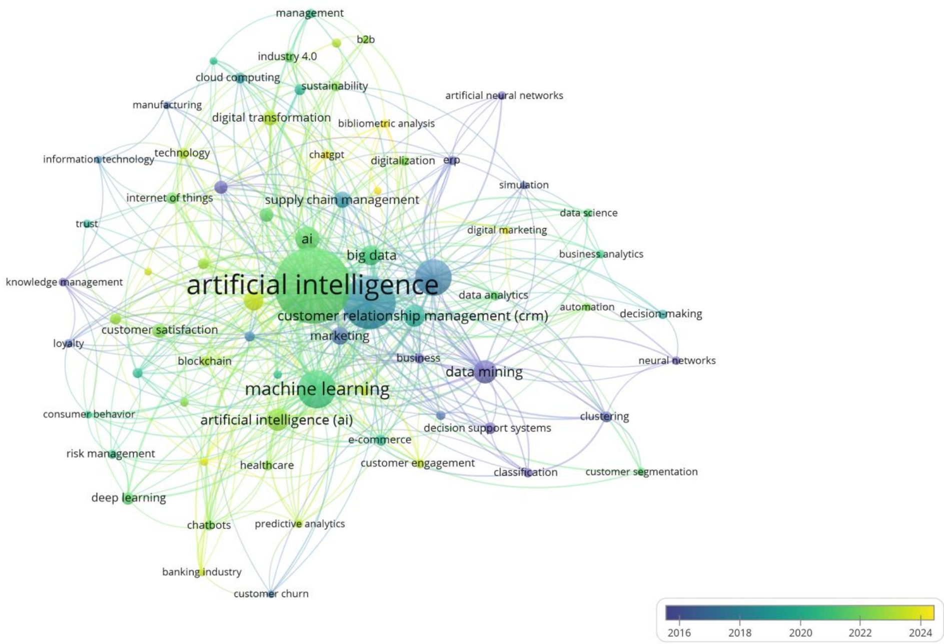


Figure 4: Visualization map of the bibliometric analysis of publications devoted to the application of artificial intelligence for customer relationship management (evolutionary and temporal aspects).

The third cluster (11 items, blue, Fig. 3) describes the connection between artificial intelligence and customer relationship management with the following terms: “consumer behaviour”, “loyalty”, “trust”, “machine learning”, “risk management”. The main key term in this cluster is “machine learning”, the frequency of its co-use in the studied sample of scientific publications is 79, and the strength of the association is 569.

The fourth cluster (10 elements, yellow, Fig. 3) is associated with the following categories: “CRM”, “business analytics”, “artificial network”, “decision-making”. The keyword “CRM” has the highest frequency of co-use – 61, while the strength of the association is 372.

The fifth cluster (8 elements, purple, Fig. 3) covers the following categories: “clustering”, “classification”, “customer segmentation”, “digital marketing” and others. The frequency of co-use of the main keyword “clustering” in this cluster is 9, and the strength of the association is 65.

The sixth cluster (7 elements, turquoise, Fig. 3) is associated with the concepts of customer service, information technology, Big Data, Internet of Things. The seventh cluster (6 categories, orange colour, Fig. 3) includes such concepts as marketing, customer satisfaction, service quality, chatbots, etc.

Thus, based on the constructed terminological map of categories and the highlighted most significant keywords related to issues of artificial intelligence in customer relationship management, it can be argued about the multidimensionality and cross-dependence of the studied areas, since there are numerous connections between the terms, as well as their high prevalence in research.

According to the results of bibliometric analysis in the evolutionary and temporal dimensions, it can be stated that in the development of scientific research related to customer relationship management using artificial intelligence tools, five most significant stages can be distinguished (Fig. 4).

The first stage of development was observed until 2016, when most publications considered general issues of forming networks and decision support systems using data analysis technologies.

In the second stage, which lasted from 2016 to 2018, the focus of scientists' research shifted to the terms “customer relationship management”, “supply chain management”, “information technologies”.

From 2018 to 2020, that is, at the third of the selected stages, the dominant key terms were “machine learning”, “customer segmentation”, “consumer behaviour”, “risk management”, “business analytics”, “e-commerce”.

The fourth stage (from 2020 to 2022) is characterized by the predominance of the terms “artificial intelligence”, “Big Data”, “cloud computing”, “customer satisfaction”, “chatbots”, “Internet of Things”.

The last, fifth stage began from 2022 to 2024. According to the results of its analysis, it can be stated that the main terms in the study were: “digital transformation”, “blockchain”, “technology”, “customer experience”, “customer engagement”, “predictive analytics”, “digital marketing”, “ChatGPT” and others.

Thus, summing up the above, we can trace the change in emphasis in scientific publications, caused by the development and improvement of information technologies from the already traditional ones: “Internet”, “software”, “information systems”, widespread at the first – third stages (Fig. 4) to the increasing importance of artificial intelligence tools at the fourth – fifth stages. The widespread use of modern technologies has led to the emergence of new phenomena and areas of scientific research, such as “digital marketing”, “e-commerce”, “customer experience”, etc., which determines the importance of simultaneously studying the subject area in the context of digital transformations.

5 CONCLUSIONS

Based on the purpose and results of the study, we can conclude that there is a high level of closeness between the concepts of “artificial intelligence” and “customer relationship management”. Thus, empirical data showed an exponential growth of research in the context of the relationship between artificial intelligence and customer relationship management (the annual increase in publications on this topic is 15.3%). It can be assumed that, in accordance with the dialectical law, the accumulated quantitative changes will turn into qualitative ones, which will lead to the complementarity and mutual stimulation of these factors.

This article performs a bibliometric analysis of key aspects of scientific publications on the application of artificial intelligence technologies to improve the efficiency of customer relationship management. Using the VOSviewer software, network visualization maps of keyword matches of publications indexed by the international scientometric database Scopus from 2001 to 2024 were created.

According to the semantic correspondence of the keywords of the studied sample, seven clusters were identified and described, and the presence of five most significant stages of the development of scientific research dedicated to the transformation of customer relationship management systems using artificial intelligence was established.

A comprehensive analysis of the identified research areas showed that the introduction of artificial intelligence, tools, methods and technologies of digital marketing are strategic at the international level for improving the customer relationship management system. The widespread use of modern digital technologies, one of which is artificial intelligence, leads to the emergence of fundamentally new areas of scientific research, which determines the importance of simultaneously studying the subject area in the context of global digital changes.

Thus, the results of the study will allow us to obtain a holistic view of the current state and prospects for the transformation of the customer relationship management system, which can contribute to the formation of a more effective marketing strategy and sales policy of enterprises of various industries.

The practical significance of the study results lies in the fact that they can be used in substantiating and developing a strategy for the digital transformation of the customer relationship management system. This will become the direction of further scientific research.

REFERENCES

[1] A. Kwilinski et al., “Managing the Logistic Activities of Agricultural Enterprises under Conditions of Digital Economy”, Virtual Econ., vol. 5(2), pp. 43-70, 2022. https://doi.org/10.34021/ ve.2022.05.02(3).

[2] Y. Remyha et al., “Energy-saving technologies for sustainable development of the maritime transport logistics market”, IOP Conf. Series: Earth and Environmental Science, vol. 1126, 012037, 2023. https://doi.org/10.1088/1755-1315/1126/1/012037.

[3] A. Kwilinski et al., “Organizational and Economic Mechanism of the Customer Relationship Management under the Era of Digital Transformations”, E3S Web of Conf., vol. 456, 05002, 2023. https://doi.org/10.1051/e3sconf/ 202345605002.

[4] The Zendesk Customer Experience Trends Report 2020. Zendesk, 2020. https://www.zendesk.com/ blog/zendesk-customer-experience-trends-report-2020/.

[5] C. Scott, “Top Customer Experience Trends You Should Watch in 2025”, CMSWire, Dec. 4, 2024. https://www.cmswire.com/customer-experience/top-customer-experience-trends-you-should-watch.

[6] K. Costello, “Gartner Says the Future of Self-Service Is Customer-Led Automation”, Gartner, May 28, 2019. https://www.gartner.com/en/newsroom/press-releases/2019-05-28-gartner-says-the-future-of-self-service-is-customer-l#targetText.

[7] Global Data from IBM Shows Steady AI Adoption as Organizations Look to Address Skills Shortages, Automate Processes and Encourage Sustainable Operations. IBM Corporation, May 19, 2022. https://newsroom.ibm.com/2022-05-19-Global-Data-from-IBM-Shows-Steady-AI-Adoption-as-Organizations-Look-to-Address-Skills-Shortages,-Automate-Processes-and-Encourage-Sustainable-Operations.

[8] Market value of artificial intelligence (AI) in marketing worldwide from 2020 to 2028 (in billion U.S. dollars). Statista, Dec. 10, 2024. https://www.statista.com/statistics/1293758/ai-marketing-revenue-worldwide/.

[9] AI adoption for customer experience uses in 2024. Statista, Dec. 10, 2024. https://www.statista.com/ statistics/1490150/ai-customer-experience-adoption/.

[10] AI Statistics 2024. AIPRM. https://www.aiprm.com/ ai-statistics/.

[11] Main challenges encountered in implementing AI for customer experience uses in 2024. Statista, Sep. 3, 2024. https://www.statista.com/statistics/1490167/ai-implementation-customer-experience-challenges/.

[12] R. Abduljabbar et al., "Applications of artificial intelligence in transport: An overview”, Sustainability (Switzerland), vol. 11, iss. 12, 189, 2019. https://doi.org/10.3390/su11010189.

[13] P. B. Acharjee et al., “Artificial Intelligence (AI) in CRM (Customer Relationship Management): A Sentiment Analysis Approach”, Proc. of the 2nd IEEE International Conf. on Trends in Quantum Computing and Emerging Business Technologies (TQCEBT 2024), 10545168, 2024. https://doi.org/ 10.1109/TQCEBT59414.2024.10545168.

[14] S. Chatterjee et al., “Assessing Organizational Users’ Intentions and Behavior to AI Integrated CRM Systems: a Meta-UTAUT Approach”, Information Systems Frontiers, vol. 25, iss. 4, pp. 1299-1313, 2023. https://doi.org/10.1007/s10796-021-10181-1.

[15] 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. https://doi.org/10.1007/978-3-031-19711-6_8.

[16] Y. Fu, G. Guo, and T. S. Huang, “Age synthesis and estimation via faces: A survey”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, iss. 11, pp. 1955-1976, 2010. https://doi.org/10.1109/TPAMI.2010.36.

[17] K. Krishnareddy et al., “AI-based Fuzzy Clustering System for Improving Customer Relationship Management”, Proc. of the 6th International Conference on I-SMAC (IoT in Social, Mobile,

Analytics and Cloud, I-SMAC 2022), pp. 673-677, 2022. https://doi.org/10.1109/I-SMAC55078.2022.9987262.

[18] 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. https://doi.org/10.1177/0008125619859317.

[19] 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, iss. 7, 2351869, 2024. https://doi.org/10.1080/17517575.2024.2351869.

[20] A. Kwilinski, "The Relationship between Sustainable Development and Digital Transformation: Bibliometric Analysis", Virtual Econ., vol. 6(3), pp. 56-69, 2023. https://doi.org/10.34021/ve.2023.06.03(4).

[21] V. Khaustova, M. Kyzym et al., "Digital transformation of energy infrastructure in the conditions of global changes: bibliometric analysis", Proc. of the 12th Int. Conf. on Applied Innovations in IT. Koethen: Anhalt University of Applied Sciences, vol. 12, iss. 1, pp. 135-142, 2024. http://dx.doi.org/10.25673/115664.

Priority Directions for Life Insurance Development in Uzbekistan

Mirzaeva Azizakhon[1,2] and Mirzamakhmudova Madina[1]
[1]*Department of Macroeconomics, National University of Uzbekistan, Universitet Str. 4, 100174 Tashkent, Uzbekistan*
[2]*University of Tashkent for Applied Sciences, Gavhar Str. 1, 100149 Tashkent, Uzbekistan*
aziza.mirzaeva73@mail.ru, madinabonu1311@mail.ru

Keywords: Insurance, Insurance Market, Life Insurance, Insurance Companies, Life Insurance Programs, Social Security, GDP Per Capita, Investment, Inflation, Insurance Demand.

Abstract: The article was prepared within the framework of the practical project of AL-7823051642 “econometric modeling and forecasting of demand for insurance services in Uzbekistan on the basis of socio-economic and demographic factors” performed at the National University of Uzbekistan named after Mirzo Ulugbek.In this paper, the authors analyze the formation of life insurance in Uzbekistan, the ongoing reform of the insurance market of the republic, and ways of accelerating its development. It also looks at the development of higher-demand life insurance services. The paper examines the main drivers of insurance industry modernization, innovative product lines, regulatory modernization, and the importance of new digital technologies for increasing availability and quality of insurance services. The main risks in the activities of life insurance companies in Uzbekistan are insufficient literacy rate of society, frequent changes in laws and regulatory documents regulating the insurance industry in Uzbekistan, increased prices for raw materials and services of insurance companies, the presence of high financial, natural disasters and environmental risks in Uzbekistan,. it was found to be technology-related risks. In order to study the development of the insurance market in Uzbekistan, the relationship between GDP, inflation and demand for insurance services in the country was analyzed. According to him, an increase in GDP develops the insurance market, increases demand and volume. Inflation, on the other hand, increases risks in the insurance market, assuming a balance between premiums and payments. The results intended for calling attention to strategic ways how to increase the strength of the insurance sector, to support the economic stability and to increase insurance awareness of life insurance in Uzbekistan.

1 INTRODUCTION

The progress of the insurance industry is a key element of the financial system, which in turn affects macroeconomic stability as well as the long-term growth of the economy [1]. As nations pursue economic resilience in the face of global uncertainties, the role of insurance in risk management, savings mobilization, and social protection has never been greater. New research underscores the relevance of novel insurance products and the role of trust in consumers in sustaining mature markets, as well as in promoting financial inclusion [2]. In this regard, emerging economies such as Uzbekistan must address the issue of developing their insurance markets in line with global requirements and the ever-changing conditions of today’s technology.

In Uzbekistan, aligning the insurance market with international best practices and fostering long-term sectorial growth requires reforming insurance and providing new services that meet the needs of the population [3]. The continued restructuring of the economy, together with the policy of liberalization and development of financial institutions, provides an enabling environment for the insurance industry to contribute to economic stability. The sector’s growth is further supported by improved socio-economic education levels and the need to enhance risk absorption capacity.

Studies have shown that adopting modern technologies, such as digital platforms and fintech solutions, improves service delivery, access, transparency, and operational efficiency [4]. Digital transformation not only lowers operational costs but also extends insurance products to underserved and rural areas, promoting financial inclusion. Additionally, big data analytics and AI enhance risk assessment accuracy, personalize products, and expedite claim payments, all critical for market development.

Further, promoting public vigilance and expanding regulatory frameworks are key to maintaining market stability and competitiveness [5]. Positive regulation fosters investment, innovation, and consumer protection. In Uzbekistan, applying internationally recognized regulatory standards will build consumer trust and stimulate foreign investment in the sector. Regulatory bodies and insurers must also enhance their capacities to address emerging challenges and technological advancements.

Combining global best practices with local economic realities will help create a dynamic and resilient insurance market in Uzbekistan, ultimately enhancing financial inclusion and economic resilience [6]. As the industry evolves, it will play a central role in risk-sharing, social safety nets, and wealth generation, supporting broader goals like poverty alleviation and sustainable development.

2 MATERIALS AND METHODS AND RELEVANCE OF STUDY

Emergent or unexpected incidents when extra financial resources are needed to meet a need; like medical treatment, funeral expenses, buying a house, or replacement of revenue when breadwinner has gone, any of these situations is an inseparable part of the day-today life of an individual in the present-day world. [1] states that listed above situations becomes a trouble for families with low living standards and lack of financial preparedness typical for the countries with developing economics. Notwithstanding, the significance level of the GDP per capita of a country in development of the insurance market is one the highest among other indicators. As a country's GDP per capita rises, people tend to have more disposable income and a greater need for financial protection against various risks. This can translate into higher demand for life, health, property, and other insurance products [7].

Moreover, developed countries' economies demonstrate that the most superior form of protection for families' financial stability is a life insurance policy. Life insurance contributes to long-term financial protection, minimizes risks associated with unforeseen events and grantees an opportunity for wealth accumulation, with a system of savings and investment part of insurance products [8]. Market share of insurance sector in many developed countries accounts for 4-6% of GDP, which indicates the critical role that industry plays in social protection and economic safety systems.

2.1 Methodology

The pervious and the presented study is of complex nature involving the technique of quantitative and qualitative analysis of the development potential of life insurance in Uzbekistan [9]. The resulting quantitative methods are in: statistical study of official statistics provided by the National Bank and the Insurance Association of Uzbekistan; regression modeling; correlation analysis identification of factors affecting insurance penetration and quality of service. Intermediate Qualitative techniques, Intermediate techniques include interviews with experts, benchmarking against international practices, all of which help form a holistic picture of the sector's status and potential for growth. The main regression analysis is based on OLS regression model in Stata, where we are considering insurance percentage out of GDP amount as a dependent variable and main independent variable is GDP and control variables are inflation rate and investment amount.

2.2 Relevance and Rationale

The importance of the study depends on continuous increase of the role of life insurance both in social and economic security. Amidst changing demographics and evolving socio-economic dynamics, there is increasing need for pioneering, cost-effective and dependable insurance solutions. In addition, the adoption of international standards and best practices, and technological innovations are needed to enhance efficiency in the sector and to build greater confidence in the public.

Distinguishing the clear causal relationship between all our variables we need to extract the context of global economic difficulties (especially the consequences of the COVID-19 pandemic, inflationary pressures, and demographic changes), considering the present situation and determining the strategic ways of development of the life insurance in Uzbekistan is of particular importance. Thus, the regression analysis includes the database for the time period from 2020 to 2024. This study is expected to provide policy enabling recommendations to policymakers, industries, consumers for enhancing sector growth, financial inclusion and social protection.

3 DETAILED ANALYSIS OF THE PECULIARITIES OF THE MODERN TENDENCIES OF THE DEVELOPMENT AND PROSPECTS OF THE UZBEK INSURANCE MARKET

3.1 The Structure of Markets, Capitalization and Segmentation

As shown in Table 1, the insurance market in Uzbekistan has been growing steadily in recent years, and the life insurance segment experienced dramatic ups and downs due mainly to legislative changes, macroeconomic factors and market forces. At the end of 2023, 38 insurance companies were registered, 7 of which were exclusively focused on life insurance, estimated at up to 22% of total industry capital [10]. That’s a market that is still in the development phase, with a small and possibly high-growth portion.

The data in Table 1 clearly demonstrates that despite the insurance sector's total capital growing almost 22.6% between 2023 and 2024, the life insurance sub-segment has seen a significant reduction in its contribution, one that came predominantly with legislative reform that ended the tax advantage of life policies funded voluntarily. Accordingly, at the end of 2024 the ratio of life insurance over aggregate capital had declined from 24% in 2022 to 3% in 2024, denoting a move in the wrong direction.

3.2 Volume and Sectoral Dynamics

The trends presented in Table 2 show that the aggregate sum of insurance premia in 2024 amounted to 9.8 trillion sums or 21.2% more than the previous year. The decrease in the life insurance sphere is especially dramatic: premiums on voluntary life insurance sank by 11.2% to 0.29 trillion sums in 2024 compared to 0.3 trillion sums in 2023 [10]. This decrease is directly tied to the legal changes and law reforms in 2023-2024 [11].

The precipitous decrease in life insurance segment values reveals weaknesses in the current policy regime and market trust. Once tax subsidies ended in April 2023, consumer interest plummeted, aggravated by macroeconomic volatility, inflation orders of magnitude higher than in Japan, and currency devaluation [1, 2, 3].

3.3 Macroeconomic Constraints

The industry has a number of macroeconomic challenges that it must face to preserve its growth potential. High inflation, which rose by more than 15% in 2022, undercuts the purchasing power of consumers, and weakens demand for voluntary coverage [10]. Currency devaluation – and specifically the devaluation of the Uzbek sum – causes [insurers with obligations in that currency to have] increased liabilities, therefore increasing the risk of insolvency, and lowers investor confidence".

The analysis in Table 3 reveals that at the same time, sluggishness of domestic investment and price volatility in the global commodities market (particularly the energy commodities) present additional uncertainties [6]. In conjunction with political instability and policy, these macroeconomic forces impact the stability and long-term health of the insurance sector.

At the microeconomic level, insurers experience a set of internal barriers that make them unable to successfully invest, including mis-estimation of the risks, absence of technical infrastructure, and low awareness among the public [4]. Advanced analytics continues to be absent (and much needed) in many insurance companies, which makes it difficult to accurately assess and price risk. Also, consumer awareness in this area is low as much of the population are unaware of the advantages of life insurance, causing limited geographical reach [12].

3.4 Results of Regression Analysis

According to the three models it is preferable to consider that GDP is the one of the most important variables with the perfect significance level where the p-value is equal to 0.009. Moreover, the first model demonstrates the reverse relationship between the variables insurance and inflation as the coefficient and t-value have the minus sign. This result towards relationship between inflation rate and insurance level is explained by the tendency of citizens to keep their money for the basic needs while the value of the national currency drops in international market [3]. Furthermore, this phenomenon decreases the level of trust in national currency and the demand for insurance expenses fades into the secondary needs, which is obvious for the developing countries economy.

As shown in Table 4, the R-squared of all three models shows perfect reliability with the reality remaining all other factors are constant and stable. However, the models two and three show less significance level with p-values of gdp and investment equal to 0.756; 0.125 and 0.040 respectively.

Tests for autocorrelation, heteroscedasticity and multicollinearity showed that none of these problems are present in dataset used for regression analysis [9]. To assess and address autocorrelation, heteroscedasticity, and multicollinearity in regression analysis, several tests are used. Autocorrelation is tested using the Durbin-Watson test or Breusch-

Table 1: Market composition and capital dynamics (2020–2024).

Year	Number of Insurance Companies	Total Capital (trillion sums)	Life Insurance Capital (trillion sums)	Share of Life Insurance in Total Capital (%)
2020	40	1.4	0.3	10.5
2021	42	1.6	0.7	18.5
2022	41	1.9	1.5	24.0
2023	38	2.3	0.3	4
2024	33	2,9	0,2	3

Table 2: Sectoral premiums and growth rates (2020–2024).

Indicator	2020	2021	2022	2023	2024	Change in 2024 (%)
Total premiums (trillion sums)	1.9	3.0	4.7	7.7	9.5	+22.6
Premiums from life insurance (trillion sums)	0.3	0.7	1.5	0.3	0.29	-11.2
Share of life insurance premiums in total (%)	15.8	23.3	31.9	3.9	3	—

Table 3: Key macroeconomic risks affecting the sector (2020–2024).

Risk Factor	2020	2021	2022	2023	2024	Impact on Insurance Sector
Inflation Rate	13 %	10.9 %	11.4 %	17.3%	9.6%	Reduces consumer purchasing power, limits premium affordability
Currency Exchange Rate	1 USD = 10000-10500 sums	1 USD = 10500-11000 sums	1 USD = 11000-11500 sums	1 USD = 11500-12000 sums	1 USD = 12000-12900 sums	Increases liabilities for foreign currency obligations
Investment Levels	2.87 % of GDP	3.28 % of GDP	3.27 % of GDP	2.36 % of GDP	10.3 % of GDP	Low investment hampers product innovation and sector expansion
Raw Material Prices	Volatile	Volatile	Volatile	Volatile	Volatile	Affects macroeconomic stability and sector confidence

Table 4. Regression analysis results of insurance market development factors.

v1	v2	v3	v4
	(1)	(2)	(3)
VARIABLES	Model 1	Model 2	Model 3
gdp (GDP in mln UZS)	0.001***	0.000	0.003
	(0.0000)	(0.0008)	(0.0013)
invest (investment percentage)	0.597***	1.217**	
	(0.0063)	(0.2524)	
inf (inflation level in percent)	-0.270***		
	(0.0024)		
Constant	4.319***	-2.413	5.047***
	(0.0627)	(1.5736)	(0.8314)
Observations	5	5	5
R-squared	1.000	0.968	0.599

Godfrey test, while the Breusch-Pagan test and visual inspection of residuals are common methods for detecting heteroscedasticity. Multicollinearity is identified using Variance Inflation Factor (VIF) tests or by examining correlation matrices.

3.5 Strategic Focus and Policy Recommendations

Uzbekistan needs a multifaceted approach to revitalize and speed up development in this sector. According to the Decree of the President of the Republic of Uzbekistan PP No. 5265 dated October 23, 2021 "On additional measures for the digitalization of the insurance market and the development of the life insurance sector" [10], provides 50% decrease of tax rates for life insurers within 2022–2025 which is a move forward. But more is needed, including:

Strengthening the existing and new legislative and regulatory framework to ensure transparency and trust towards the consumer [13].

- Launching digital platforms and new distribution channels to widen reach [4].
- Promoting the positive aspects of life insurance by creating financial literacy programs for the public [2].
- Rebalancing macroeconomic policies to stabilize inflation and exchange rate movements [3].
- Strengthening partnership with FALIA and OLICD to share best practices and develop professionals [8].

4 DETAILED RESEARCH OF STRATEGIC DIRECTIONS THE LIFE INSURANCE ON UZBEKISTAN DEVELOPMENT: THE BASIS OF THE STABLE SECTOR

4.1 Aims and Evidence for Priorities

Taking into account the comprehensive study of the industry, expert opinions, and the best international experience, the following strategic areas are considered important for the sustainable development and stability of the life insurance market in Uzbekistan. The priorities serve to address current difficulties, tackle opportunities, and build robustness against international economic variances.

4.2 Strengthening the Financial Cognition of the Population

Boosting the financial literacy of the population and insurance awareness are also essential to the increase in the use of services. Higher financial literacy is positively related to making decisions in favor of the use of insurance services and making more informed financial decisions or create better risk management [14]. Propagating both at the school-level in curricula and courses in higher education across the nation, and in the public through initiatives such as mobile applications and online courses, opens awareness on a massive scale, particularly for younger generations. The theory of financial resilience [15] suggests that, by increasing financial literacy, information asymmetry is reduced, the efficiency of the market is improved, and consumer confidence is strengthened, eventually leading to faster sector growth.

4.3 Theoretical and Practical Aspects to Innovative Insurance Products Designing

The creation of special insurance products aimed at various age groups – from the young to the family to pensioners, is crucial to appeal to a wider range of customers. They must be flexible, digital-based, and personalized by integrating modern solutions like AI and big data analytics – as [16] have highlighted. For instance, hybrid policies that integrate protection and investment features, and which offer automatic premiums and tailored coverage choices, are known to better engage and satisfy consumers. As per the portfolio diversification theory [17], diversified products in turn, reduce the overall sectoral risk and add to the financial security, particularly, when it is in tune with the demographic shifts.

4.4 Harnessing Digital Technology: Trends and Strategies for Integration

Use of digital platforms for policy sales, customer service, and claims is essential for modernization of the industry. As per the definition by [18], digitalization cuts costs and enhances customer experiences whilst making operations more efficient. Introduction of online portals, mobile apps, chatbots and automated risk assessment tools would ensure open, transparent and efficient services, appealing to tech friendly generations and spread coverage geographically. To realize the benefit and trust, we also need to consider data security, compliance with international cyber-security standards, and user-friendly design.

4.5 State Support and Regulatory Frameworks: Policy Creation and Institution Building

Establishing transparent and predictable legal and regulatory regimes is essential to developing this sector. The recent legislative steps, i.e. the tax rates reduction on activities of insurance for 50% for the period of 2022-2025 [13], give stimuli for insurance companies. There is also the potential for increased transparency, enhanced risk management, and greater investor confidence by conforming to international standards, like IFRS 17 on insurance contracts and principles from Solvency II. According to [19] successful regulation will lead to decreased information asymmetry and moral hazard, and will 29 promote market stability which in turn attracts domestic and foreign investment.

4.6 International Cooperation and Transfer of Knowledge

Working with international organizations like FALIA (The Foundation for the Advancement of Life & Insurance Around the World) and Japan's OLICD promotes information exchange and staff development and encourages a best practice approach. Such international co-operations help Uzbekistan to implement advanced risk management methodologies, product innovations and digital solutions and promote sector modernization. The globalization of markets [20] insists on the fact that international cooperation can expedite the diffusion of innovation, help cut standards reference for companies while it can also enhance competitiveness.

4.7 Adjusting to Demographic Shift: Product and Market Innovation.

Insurance products need to evolve, given changes in demographics including aging societies and burgeoning youth populations. For example, the creation of pension products offering flexible contribution requirements, micro-insurance for low-income segments and insurance products for migrants could lead not only to increased product diversity but also to a broader market participation. Population cycle theory [21] suggests that if insurance options are carefully linked to demographic circumstances, the sector will remain viable and the guarantee of social protection will have substantive credibility.

4.8 Mission to Enhance Service Quality and Professional Capability

Quality of Customer Service Trust is derived from a high-quality service and can determine how loyal customers will be. Training programs, progress of standards, and feedback systems need to be institutionalized. Implementing CRM systems, automation of business processes and use of data analytics will improve the efficiency of operations and customer satisfaction. As [22], quality significantly impacts customer loyalty and company image that play an essential role in industry growth.

4.9 Measures to be Taken for Implementation and Technical Innovation

To achieve these strategic directions, the detailed action plan would comprise:

- Stimulating economic growth and providing basic infrastructure and social security, which will ensure the basic needs of the population and rise to the next level of living standards;
- Designing educational programs and financial literacy campaigns aimed at schools, universities and community centers;
- Market research and customer surveys to discover unmet wants and to shape (PDP);
- Development of IT infrastructure, user-friendly mobile apps, online insurance platforms, and automated process;
- Enhance legal and regulatory frameworks consistent with international standards, also with consumer protection laws;
- Developing strategic alliances with offshore insurers and financial institutions;
- Conducting insurance professionals training workshops & certification programs;
- Launching marketing strategies to inform and draw in new customers;
- Developing monitoring and evaluation systems to monitor the use of these measures and adjust approaches.

A comprehensive, multi-vector solution represented by legal changes, development of technology, educational initiatives, and international collaboration shall ajust the necessary frameworks for the sustainable development of life insurance in Uzbekistan. By instilling in consumers a greater level of confidence in the market, encouraging product innovation, and improving service quality, life insurance will move from a new-niche market to an indispensable tool for social stability and protect consumers against financial risk. Although today some of those opportunities have been shadowed by amendments to the law and macroeconomic instability, we believe that the sector's future is still bright if its strategic priorities are successfully implemented. The government and sector must collaborate, while the sector evolves in accordance with world best practice and the needs of the country's socio-economic context.

5 CONCLUSIONS

This article focuses on the formation of the life insurance market in Uzbekistan, the ongoing reform of the sector and the measures needed for its rapid development. The main message is that the insurance industry is crucial for ensuring macroeconomic stability, social well-being and financial development. Creativity, digitalization and standardization will be key to further development of the industry in the future. However, there is still much room for growth, especially through active stimulation of economic growth and strengthening of public-private partnerships. Especially considering the results of the regression analysis, where economic growth in the form of GDP growth and investment growth are the basic factors for improving the living standards of the state's residents.

High standards and sufficiently high earnings of citizens instill confidence and trust in the government system. In turn, improving the standard of living requires more decent ways to ensure economic and social security, namely life and property insurance. In general, further progress in the insurance market of Uzbekistan. Continuous improvement of regulatory regimes, the use of new and digital technologies and the development of competition have a complementary effect against the background of sustainable economic development and ensuring economic security for success. These measures will not only stabilize the sector but will also increase consumer confidence, which will further strengthen the country's economy. Whether there is a dual causal effect among these factors is the next question that requires more sophisticated data analysis and research.

REFERENCES

[1] World Bank, Financial Sector Development and Macroeconomic Stability. Washington, D.C.: World Bank Publications, 2022.

[2] OECD, Insurance Market Trends and Consumer Confidence. Paris: OECD Publishing, 2023.

[3] IMF, Reforming Insurance Markets in Emerging Economies. Washington, D.C.: International Monetary Fund, 2021.

[4] KPMG, Digital Transformation in Insurance. London: KPMG International, 2022.

[5] PricewaterhouseCoopers, Global Insurance Industry Outlook. London: PwC, 2023.

[6] Asian Development Bank, Financial Inclusion and Sector Development. Financial Inclusion in Central Asia. Manila: ADB Publications, 2022.

[7] Oxford Business Group, "PNG's niche markets require cultural insight to insure," 2015, [Online]. Available: https://oxfordbusinessgroup.com/reports/papua-new-guinea/2015-report/economy/pngs-niche-markets-require-cultural-insight-to-insure, [Accessed: 28 Jan. 2025].

[8] International Association of Insurance Supervisors (IAIS), Global Insurance Market Report. Geneva: IAIS, 2021.

[9] M. Mirzamakhmudova, "Oʻzbekistonda hayot sugʻurtasining joriy etilishi va unga ta'sir etuvchi demografik omillar ta'siri," Yashil iqtisodiyot va taraqqiyot jurnali, vol. 4, pp. 1422–1427, Apr. 2025.

[10] Uzbekistan Insurance Market Report, Annual Review of the Insurance Sector in Uzbekistan. Tashkent: State Committee of the Republic of Uzbekistan on Statistics, 2023.

[11] Law of Uzbekistan "On Insurance Activities". Legal Framework for Insurance Sector. Tashkent, 2020.

[12] A. Kuznetsov and R. Ismailov, "Modern Approaches to Life Insurance Development in Central Asia," Journal of Financial Regulation, vol. 15, no. 3, pp. 45–62, 2021.

[13] Presidential Decree № 5265, "On Additional Measures for Digitization and Development of the Insurance Sector," Tashkent, 2021.

[14] A. Lusardi and O. S. Mitchell, "The Economic Importance of Financial Literacy: Theory and Evidence," Journal of Economic Literature, vol. 52, pp. 5–44, 2014, doi: 10.1257/jel.52.1.5.

[15] R. C. Merton, "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates," Journal of Finance, vol. 29, pp. 449–470, 1974, doi: 10.1111/j.1540-6261.1974.tb03058.x.

[16] Deloitte, "2022 insurance industry outlook," [Online]. Available: https://www.insurtechinsights.com/deloitte-2022-insurance-industry-outlook, [Accessed: 28 Dec. 2023].

[17] H. Markowitz, "Portfolio Selection," The Journal of Finance, vol. 7, pp. 77–91, 1952, doi: 10.1111/j.1540-6261.1952.tb01525.x.

[18] E. Brynjolfsson and A. McAfee, The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. New York: W. W. Norton & Company, 2014.

[19] J. E. Stiglitz, "Markets, Market Failures, and Development," American Economic Review, vol. 79, pp. 197–203, 1989, doi: 10.3386/w2961.

[20] T. L. Friedman, The World Is Flat: A Brief History of the Twenty-First Century. New York: Farrar, Straus and Giroux, 2005.

[21] J. C. Caldwell, "Toward a Restatement of Demographic Transition Theory," Population and Development Review, vol. 2, pp. 321–366, 1976, doi: 10.2307/1971615.

[22] A. Parasuraman, V. A. Zeithaml, and L. Berry, "A Conceptual Model of Service Quality and Its Implications for Future Research," Journal of Marketing, vol. 49, pp. 41–50, 1985, doi: 10.2307/1251430.

Digital Marketing Tools for Preserving Cultural Heritage in the Context of Sustainable Urban Development

Tatyana Oklander[1], Anastasiia Pandas[1], Mykhailo Oklander[2], Oksana Yashkina[2] and Hlib Chaban[1]
[1]Department of Economic and Entrepreneurship, Odesa State Academy of Civil Engineering and Architecture, Didrihsona Str. 4, 65052 Odesa, Ukraine
[2]Department of Marketing, Odesa Polytechnic National University, Shevchenko Avenue 1, 65044 Odesa, Ukraine
oklander@odaba.edu.ua, anastasypandas@gmail.com, oklander@op.edu.ua, yashkina@op.edu.ua, glebchaban@gmail.com

Keywords: Municipal Marketing. Digital Marketing. Digital Technologies. Investment Resources. Branding. Brand Marketing. Marketing Strategy. Cultural Heritage. Sustainable Urban Development. HYPERVSN.

Abstract: The article is concerned with researching the possibilities of using digital marketing tools in the area of cultural heritage preservation in the context of sustainable urban development. The main focus is on innovative technologies, in particular, on the use of the holographic system “HYPERVSN”, which provides broadcasting of a three-dimensional image with an expressive emotional effect. The methodological basis of the study was general scientific approaches to analysis and systematisation, as well as quantitative methods, including a sociological survey, which made it possible to determine the level of public perception and interest in the use of such technologies in public space. The results indicate the growing openness of consumers to the latest formats of visual communication, as well as the potential of digital media as an effective tool for promoting cultural objects. The research outlines both the advantages, such as innovation, attractiveness, and the ability to create sustainable interest in a city brand, and the risks associated with possible visual overload or inappropriate content of advertising. The originality of the work lies in the integration of a marketing approach to the tasks of cultural heritage protection and in the substantiation of its role in shaping an attractive city image. The practical value lies in the possibility of using the presented findings to develop municipal strategies focused on combining digital technologies, cultural policy and sustainable development.

1 INTRODUCTION

Analysing global challenges and a focus on sustainable development, the preservation of cultural heritage is increasingly seen not only as a matter of historical value, but also as an important factor in social interaction, tourist attractiveness and economic sustainability of the urban environment. Today, there is a change in communication approaches, an increase in the role and relevance of visual technologies in public space, which requires new tools for interaction with society, especially in the context of promoting cultural heritage sites and the need to preserve them.

Recently, digital marketing has become particularly relevant as an effective and efficient tool for sustainable city image management. It creates a vivid, emotionally rich visual context that not only informs but also engages the audience in interaction. It is the younger generations: millennials and representatives of Generation Z – who determine trends in consumer behaviour, particularly in the perception of advertising and urban space. For them, classical forms of promotion are gradually losing their relevance, while innovative solutions based on interactivity can create lasting impressions and evoke emotional attachment to a brand or place [1,2,3].

Holographic technologies, which until recently were perceived as experimental or decorative, are gradually becoming quite applied [4]. They are successfully used by leading brands, including Coca-Cola, Adidas, Vodafone, Walmart and Puma. Examples of successful campaigns using three-dimensional images in public space demonstrate the ability of these solutions not only to attract attention but also to change consumer practices. “HYPERVSN” technology plays a special role in this process, allowing dynamic holograms to be broadcast in 3D, creating the effect of presence and engagement.

Unlike traditional advertising media, “HYPERVSN” allows not only to demonstrate visual content, but also to build a holistic communication experience that can be easily integrated into the city environment - on the streets, in tourist areas, near cultural sites. It can be concluded that holographic advertising is not only a commercial tool, but also a means of strengthening the city brand, promoting urban space, visualising historical monuments and stimulating interest in cultural heritage.

The relevance of the study lies in the need to analyse and highlight the significant role of digital marketing tools in preserving and promoting cultural heritage and, as a result, promoting cities and encouraging sustainable urban development. The purpose of this article is to analyse the potential of innovative technologies, in particular the “HYPERVSN” system, for integration into municipal sustainable development strategies, as well as to study the perception of such tools by residents and tourists in the urban environment.

2 RESEARCH METHODOLOGY

The article uses general scientific methods of analysis and systematisation to study the regulations on cultural heritage to prove that the municipal authorities are systematically engaged in the preservation of the city's cultural heritage and the development of cultural tourism. Quantitative methods in the form of a sociological survey among residents and tourists to determine the level of awareness, interest, perception of cultural heritage and the introduction of innovative digital communication tools using “HYPERVSN” technology in the urban context.

3 ANALYSIS OF RESEARCH RESULTS OR SCIENTIFIC PROBLEM

3.1 Cultural Heritage Management and Urban Branding Strategies

Odesa is the tourist capital of Ukraine and is called the "pearl by the sea". It is a major European city with an architectural, cultural, balneological and recreational industry. The concept of city construction in the late 18th century was developed by representatives of many European countries and embodied the best European achievements in urban planning. The Spaniard O. Deribas, the Frenchmen E. Richelieu and A. Langeron supervised the construction, and the Dutchman F. de Volan developed the city's project. The architecture embodies many classical European architectural styles. The historic center of the city is a cultural heritage. On January 25, 2023, the World Heritage Committee decided to inscribe the historic center of Odesa on the World Heritage List.

Audrey Azoulay, Director-General of UNESCO, said: "Odesa, a free city, a world city, a legendary port that has left its mark on cinema, literature and art, is now under increased protection by the international community. As the war continues, the inscription of this site on the World Heritage List embodies our collective determination to protect this city, which has always withstood global upheaval, from further destruction" [5].

The Odesa City Council systematically preserves cultural heritage. The Department of Culture, International Cooperation and European Integration is in charge of this within the Executive Committee. Within the structure of this department, there are the Department of Culture, the Department of Tourism, the Department of International Relations, the Department for Work with UNESCO and the Protection of Cultural Heritage.

Currently, the Department of Culture, International Cooperation and European Integration of the Odesa City Council is developing a project "Support for the Implementation of the Odesa Cultural Development Strategy" to preserve cultural heritage and develop tourism in Odesa. The project is funded by the EU under the EU4Culture project, implemented by the Goethe Institute in partnership with the Danish Institute of Culture, Czech Centers, and the French Institute in Georgia. The project for the implementation of the city's Cultural Development Strategy is to be developed in the fall of 2024. “The issue of preserving historical heritage has become particularly relevant against the backdrop of Russian aggression against Ukraine and, as a result, damage to cultural and historical monuments, as well as the receipt of the status of a UNESCO World Heritage Site by the historic center of Odesa,” - notes Ivan Liptuga, Director of the Department of Culture, International Cooperation and European Integration of the Odesa City Council [6].

Odesa City Council adopted the investment strategy of the city “Odesa 5T”.

“Odesa 5T” is an investment brand of the city that unites five main vectors of the city's development:

- Transportation.
- Trade.
- Technology (construction, IT, medicine).
- Tourism.
- Trust (transparent economy, trust, financial services market).

Investment Agency “Odesa 5T” was created for effective meetings of domestic and foreign investors and entrepreneurs. Tourism, transport, technology, trade, and trust - the five main areas of promising and effective investment in one of the main economic centers of Ukraine. Investment Agency “Odesa 5T” was established as a one-stop shop to facilitate investment attraction, promotion and implementation of priority projects for cities. The team analyzes the economic and legal conditions of projects, taking into account the interests of cities and investors, monitors and controls the implementation of initiatives that have already been launched, introduces systemic measures by improving the investment climate, cooperation with executive authorities, and entrepreneurial responsibility.

In addition, the municipal institution “Odesa 5T Grant Office” operates. It became the legal successor of the municipal enterprise “Agency for Development Programs of the City of Odesa” already operating under the mayor's office. The municipal institution makes extensive use of digital marketing tools.

To create conditions for the effective functioning of the city's economic complex and for a decent life for people, it is necessary to skillfully combine the efforts of various municipalities. A system of marketing activities is needed to attract new economic investments to the city through the distribution of information and materials about the city, targeted visits by city leaders, and meetings with potential investors. The goals of municipal marketing are to improve (preserve) the competitiveness of industrial and service enterprises in the region; to attract new businesses to the region and, as a result, to obtain funds for the preservation of cultural heritage. An extremely crucial component is the city brand, which is an integral set of features that contain unique original characteristics of the city and imagery and perception that allow identifying this city among others in the eyes of target groups. City brand gives a first impression of the value promised in terms of the experience a potential user can receive [7]. The brand should perform the following functions [8]:

1) identify the city among others;

2) to form a sense of attachment to the city and identity with it among target audiences;

3) to be a guarantor of the city's quality.

Municipal marketing is actively developing in the city, which has become the basis for shaping the city's brand and image. At the initiative of the local authorities, competitions were held to create a tourist slogan, symbols and logo for the city, which were later used as elements of the city's identity.

It is crucial to emphasise that the core of any city is its inhabitants who live on or near the heritage site, while other facilitators such as authorities, experts and economic actors are considered to be the wider community [9,10]. Local communities can share responsibility for integrating heritage conservation into sustainable urban development through cooperation and community empowerment [11,12,13].

3.2 Introduction to the Research Objects

Digital marketing is a model of 21st century marketing, mass individualization marketing, which allows targeted interaction with target market segments in virtual and real environments through digital channels using digital methods. Digital marketing could offer a crucial platform that is community - based for sustainable and holistic heritage conservation [14]. Municipal authorities use the following tools:

1) Local networks of enterprises or districts, which are self-sufficient information systems. Now, local networks are gradually being integrated with the Internet.
2) Mobile devices with branded applications or organization of WOW-calls to the phone (WOW-call is a platform that combines the Internet and telephony).
3) Interactive screens, POS-terminals located in stores and on the streets. They are gradually replacing standard outdoor advertising, as they allow for closer interaction with the consumer, delivering messages or helping to make purchases through a POS-terminal.
4) Digital art is an art form in which a computer is used to create or reproduce an artistic work: drawing, sound, animation, video, games, website, performance, installation.

Trigger of preservation of cultural heritage is the effective functioning and development of the recreational complex of the city [15]. This complex is part of ensuring stable and dynamic economic growth [16]. It is worth noting that the recreational complex is an integral part of the national economy, the role of which is to ensure the production and sale of socially useful products, the object of which is tourist and recreational services [17]. World experience shows that the recreational complex is becoming increasingly important in stimulating urban development. Consumers of recreational services overcome distances in order to meet the demand. After that, the country experiences what is called "primary" economic revitalization in the form of utilizing the available potential of the recreational complex. Next, a multiplier effect occurs. There is a "secondary" revitalization, but already in other sectors of the economy. The total revitalization due to the multiplicative effect, as it can be seen from Fig. 1, can give a strong revitalizing result.

3.3 Analysis of Tools for the Development of the City's Recreational Complex

To advertise tourist attractions, it is proposed to use "HYPERVSN" to visualize three-dimensional images. The technology based on the use of holographic images is revolutionary for outdoor advertising. A three-dimensional diorama uses diodes to reproduce a full-fledged and realistic 3D image. The hologram displays all sides of the observed object, conveys its volume, relief, and depth. The advanced solution allows you to first create and display unique three-dimensional photo or video content, and then manage it.

The WOW-effect of a hologram promotes deep memorization of information about a product, including on an emotional level, and from the technical side, it gives a potential buyer much more information about the advertised product than "flat" advertising. The consumer can see everything from all sides. This inspires trust.

"HYPERVSN" technology is a cost-effective alternative that will make it possible to use small and medium-sized holograms. The effect of a 3D hologram is created by blades with LEDs placed on them, which rotate at high speed. This is how the effect of image visualization is achieved.

Marketing research is another tool for the development of the city's recreational complex. It is worth noting that the organised system of marketing tools forms an effective comprehensive tool for implementing the strategy for the development of the city's recreational complex [18]. A survey was conducted to assess the effectiveness of the proposed innovative advertising technology. This method of marketing research was chosen to identify the perception of the city residents of the use of

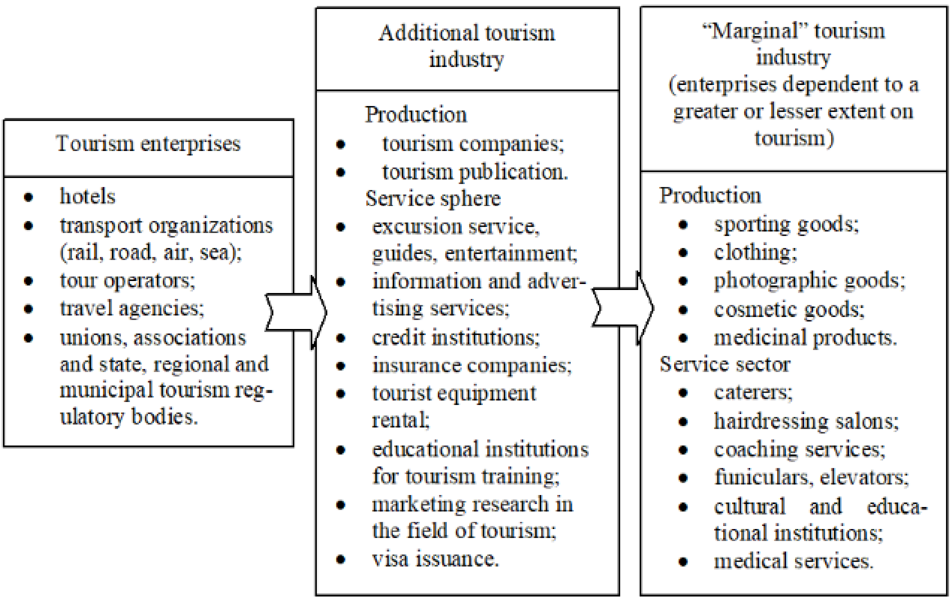


Figure 1: Scheme of multiplier effect as a result of the development of recreational complex of the city.

advertising in the historic part of the city. The respondents who took part in the survey were also offered a questionnaire to survey the use of a new type of advertising in the urban space. It contained the following questions:

1) Will you be interested in this technology and why?
2) To what extent is the technology noticeable among other types of advertising media in the historic part of the city (banners, billboards)?
3) What negative aspects of this technology can you name?
4) What positive aspects of this technology can you name?
5) Would you make a video of the "HYPERVSN" and share it with your friends?
6) What would you like to see from the goods/services/brands on a 3D hologram of the "HYPERVSN"?
7) Do you think that your attitude towards the product/service/brand that will be advertised on this technology has changed? Why?
8) Where would you like to see this technology in action? (In supermarkets, on the street, near the advertised product, etc.).

The answers of the respondents are shown in Figure 2, Figure 3, Figure 4.

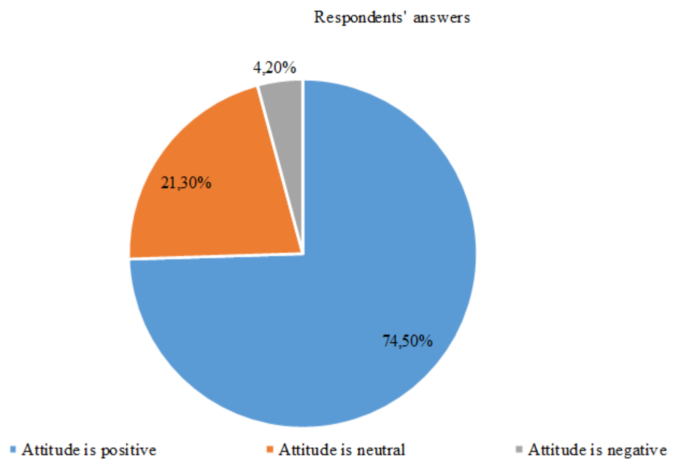


Figure 2: Respondents' interest in "HYPERVSN" technology.

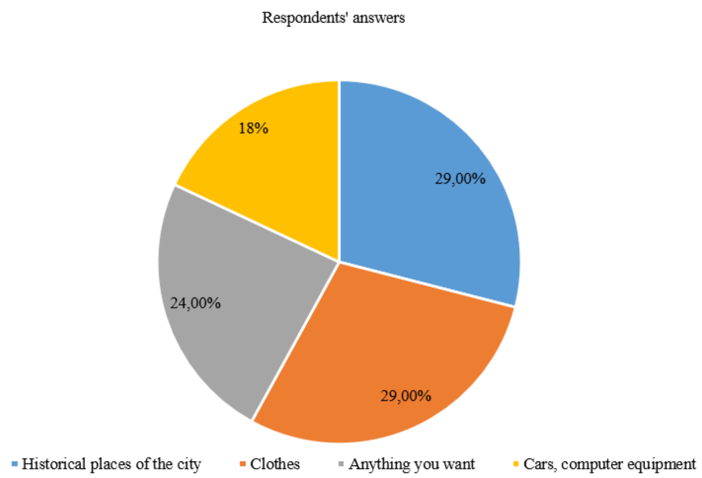


Figure 3: Respondents' attitudes towards the use of "HYPERVSN" technology.

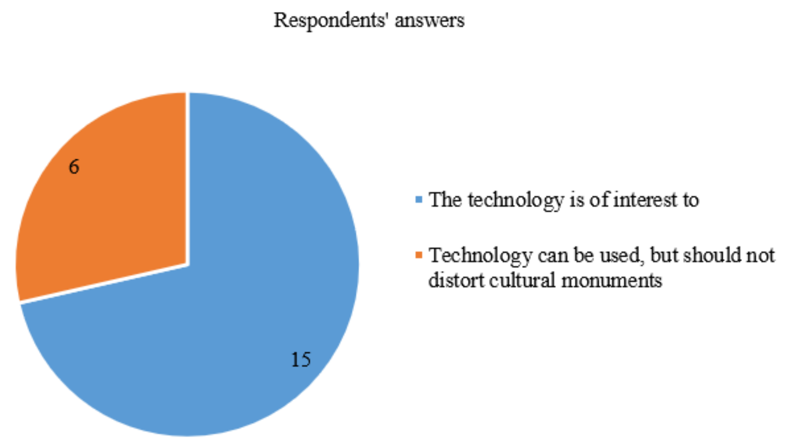


Figure 4: What goods and services would respondents like to see advertised with 3D holograms of the “HYPERVSN” technology?

A total of 47 respondents took part in the survey, including 23.4% of men and 76.6% of women. Age range of respondents: 59.6% of respondents were under 20 years old, 38.3% were aged 21-35, and the rest were over 35. 74.5% of respondents have a positive attitude towards new technologies, 21.3% are neutral and 4.3% are negative. All respondents said they would definitely share a video with such technology with their friends.

“HYPERVSN” continues to evolve. In 2023. Michael Blackman, Managing Director of Integrated Systems Europe, joined the Talent Congress in Barcelona in the form of a 3D hologram from 4,000 miles away in Oman, thanks to “HYPERVSN”. The demo version of this innovative solution at ISE 2023b demonstrated the difference between the required "human contact" and visualisation from traditional flat-screen calls and live broadcasts. This real-time streaming solution ensured maximum realism and impact on the audience. Interactivity is an important new factor in increasing the impact on consumers.

4 CONCLUSIONS

The research shows the feasibility of using digital marketing tools, in particular HYPERVSN holographic technology, in the process of preserving and promoting cultural heritage in the context of sustainable urban development. The obtained results confirm the high level of audience interest in innovative forms of visual communication, which opens up new opportunities for increasing public and tourist awareness of cultural sites.

The scientific novelty of the study lies in the integration of the concepts of municipal marketing, city branding and digital communications in the context of cultural heritage preservation. For the first time, holographic technology is considered not only as a commercial advertising tool, but also as a potential mechanism for forming an emotional connection between citizens and the historical and cultural environment. The use of innovative visual solutions allows us to rethink traditional approaches to the promotion of urban spaces and heritage values.

The practical significance of the results obtained lies in the possibility of their use by local governments, city departments of culture, tourism and marketing to develop modern strategies for communication with the public. The proposed technologies can be effectively integrated into the information and cultural infrastructure of the city, contributing to the growth of tourist attractiveness, strengthening the brand of the territory and more active participation of citizens in the processes of preserving cultural heritage.

Taking this into account, the advantage of the study is its novelty and interdisciplinary approach: the topic of using marketing tools to preserve cultural heritage in the context of sustainable urban development is still poorly understood. The paper reveals the potential of the latest digital technologies for promoting intangible and cultural factors of the urban environment in the tourism market, which creates the basis for a new model of communication between the city and target audiences.

The limitations of the study include the insufficient amount of statistical material due to the staged nature of the article. In view of this, the use of mathematical and statistical methods for an in-depth analysis of the factual base is planned in further empirical research.

Further research will be related to the study of the possibility of using such marketing tools as virtual reality, augmented reality, mixed reality in the preservation of cultural heritage.

REFERENCES

[1] R. Baltezarevic, I. Baltezarevic, Benefits of using holograms in marketing communication, Conference: 7. International Izmir economics congress, Izmir, Turkey, 2023, pp. 784-791

[2] C. R. Cunha, L. Lopes, V. Mendonça, Immersive digital marketing for smart cities focusing tourism. Smart Innovation, Systems and Technologies, 2022, pp. 605 - 614.

[3] N. Vangelov, Digital Marketing and Outdoor Advertising in Smart Cities, Smart Cities and Regional Development, Vol. 6 No. 3, 2022, pp. 81-91

[4] M. Nusenkis, Holographic marketing, or how to solve the problems of brand advertising. Marketer.ua, [Internet]. Available from: https://marketer.ua/

[5] UNESCO, [Internet]. Available from: https://www.unesco.org/

[6] Official website of Odesa, The European Union finances the elaboration of the Odesa Cultural Development Strategy, [Internet]. Available from: https://omr.gov.ua/

[7] R. Sutriadi et al, Place branding as the development of thematic city digital era, IOP Conf. Ser.: Earth Environ. Sci. 592 012032, 2020

[8] O. Bilovodska, N. Gaidabrus, The theoretic-methodological basics of brand creation. Marketing and Management of Innovations 1, 2012, pp. 35-43

[9] Li, J. Krishnamurthy, S. Pereira Roders, A. van Wesemael, P. State-of-the-practice, Assessing community participation within Chinese cultural World heritage properties. Habitat Int. 2020, 96, pp. 102-107

[10] X. Liang, Lu, Y. Martin, A Review of the Role of Social Media for the Cultural Heritage Sustainability. Sustainability, 13, 1055, 2021

[11] T.C. Ferreira, Bridging planned conservation and community empowerment: Portuguese case studies. J. Cult. Herit. Manag. Sustain. Dev. 2018.

[12] G. MacRae, Universal heritage meets local livelihoods: ‘awkward engagements’ at the world cultural heritage listing in Bali. Int. J. Herit. Stud. 2017, 23, pp. 846-859.

[13] G. Verdini, F. Frassoldati, C. Nolf, Reframing China's heritage conservation discourse. Learning by testing civic engagement tools in a historic rural village. Int. J. Herit. Stud. 2017, 23, pp. 317-334.

[14] A. Ginzarly, P. Roders, J. Teller, Mapping historic urban landscape values through social media. J. Cult. Herit. 2019, 36, pp. 1-11.

[15] H. Keyvan, A. Stefaniec, and S. Parisa Hosseini, World Heritage Sites in developing countries: Assessing impacts and handling complexities toward sustainable tourism. Journal of Destination Marketing & Management 20, 2021: 100616.

[16] O. Pryhara, S. Bryk, I. Holubets, Development of tourist and recreational complex in conditions of sustainable growth of economy. Economy and Society 33, 2021.

[17] A. Alijeva, Orghanizacijno-ekonomichni zasady stalogho rozvytku turystychno-rekreacijnoji sfery v Ukrajini (in Ukrainian), Lutsk, 2021.

[18] V. Kazlouski, U. Ganski, The marketing of historical and cultural heritage and its specifics in conditions of culture commercialization. Market Infrastructure 19, 2018, pp. 39-44

Multivariate Statistical Analysis of the Drivers of Women's Empowerment in Labor Market

Drishtanti Saha, Pratik Sarkar, Markus Holz, Cornelia Scott and Halina Falfushynska
Anhalt University of Applied Sciences, Strenzfelder Allee 28, 06406, Bernburg, Germany
{Drishtanti.Saha, Pratik.Sarka}r@student.hs-anhalt.de, {markus.holz, cornelia.scott, galyna.falfushynska}@hs-anhalt.de

Keywords: Female Labor Force Participation, Education and Employment, Maternal Health, Gross National Income per Capita, Gross Domestic Product.

Abstract: The current study seeks to explore the time series perspective of Indian women's involvement in available work seasoned with other factors determined for their employment from 1991 to 2023. Using some sophisticated statistical techniques, such as Regression models, Principal Component Analysis and Partial Least Squares-Discriminant Analysis, the research showed that the reduction of the number of deaths of children under the age of one year per 1000 live births has been a major contributing factor in the increase in the percentage of women participating in the labor force. In addition, economic advancement, maternal death rates, and life expectancy were also relevant in determining these trends. From 1991 to 1999, factors such as the lifetime risk of maternal mortality, the increasing number of women with lower levels of education participating, and changing economic conditions were critical in determining the extent of women's labor force participation. From 2000 to 2009, more women graduates began to join the labor force as every society begun to change gradually. From 2010 to 2019, the growth of GNI per capita and increasing expansion in women participation in the labour force paid dividends. This research highlights the importance of continued socio-economic progress in promoting gender equality in employment.

1 INTRODUCTION

Rise in female labour force participation is a significant economic transformation of the previous century. Historically, men have participated in labour markets more frequently than women worldwide, but this participation landscape is rapidly evolving. The global labour force participation rate for women is currently less than 47%, whereas for men it is 72% [1].

Considerably in particular regions, this gap reaches 50%, underscoring the persistent challenge of attaining gender equality in the labour market. It is believed that female labour force participation rates are higher in many of the world's poorest and highest-income countries based on several factors [2].

Although India ranks as the fifth-largest economy globally by nominal GDP and the third-largest by purchasing power parity (PPP) [2] its female labour force participation rate was very low at approximately 37% in 2022 [3] which is significantly below the global rate.

The dynamics of female labour force participation in India represent a complex interaction of socio-economic factors, illustrating the multifaceted nature of the quest for equal opportunity. In contrast to the generally observed linear models of structural transformation, India's economic narrative reveals a distinct trajectory [4], wherein the anticipated growth in the manufacturing sector did not occur following the collapse in agriculture. The services sector predominantly accommodated the economic shift, witnessing significant expansion over the past twenty years. However, this economic progress has not been directly reflected in employment possibilities, notably impacting women.

Evidence demonstrates that empowering women not only offers individual advantages such as decreased fertility rates and postponed marriage but also serves as an indicator of wider developmental results. Following 2019, a significant rise in the female labour force participation corresponds with the consistent increase in India's Gross National Income (GNI) per capita [5] indicating a mutually reinforcing link that may support progress in women's education and entrepreneurship. This synergy underscores the transformative potential of targeted policies in fostering both economic growth and gender equality.

The study aims to explore the dynamics of female labor force participation in India, highlighting the complex socio-economic, cultural, and policy factors that influence women's employment. The research investigates how these factors interact to impact gender equality in the labor market and assesses the broader economic implications.

2 MATERIALS AND METHODS

For this statistical analysis we have collected data from the central repository of World Bank Database [6].

The main variable of interest includes Labor force female, Expected years of schooling for female students, Expected years of schooling for all gender students, Female labor force with different levels of education, and Gross national income in terms of PPP. The dependent variable was considered to be the count of labor force female and we have analysed how other crucial factors in a socio-economic environment effects its variations from year to year and in some cases among decades. Expected year of schooling for all genders and especially for female have been analysed as two separate variables to observe their independent effects on the dependent variable. The independent variable ‘Infant Mortality Rate’ is defined as the number of live births per thousand people per year. Independent variable named ‘GNI per Capita in terms of PPP’ is measured on current US Dollar. ‘Lifetime Risk of Maternal Death’ this independent variable is measured in a scale of hundred.

Data cleaning and pre-processing are routinely performed on the collected information. To prevent inaccurate interpretations, all indicators with missing values were excluded from subsequent analyses. Normality and homogeneity of variance were assessed using the Kolmogorov–Smirnov and Levene’s tests, respectively. Data were adjusted using the Box–Cox transformation method whenever possible. To analyse possible relation between parameters correlation heatmap based on Pearson correlation test was created. To determine which parameters significantly affect Labor Force Female multiple regression was performed. Normalized, auto-scaled data underwent principal component analysis (PCA) to distinguish individuals based on their personal traits. PCA is the conventional method for component extraction and for visualizing the similarities among samples. It aims to identify a low-dimensional representation of the data that combine the majority of the variance. Only variables exhibiting a correlation greater than 0.60 with the first two dimensions of the PCA (Factor Loadings) have been included. Statistical calculations were performed using MetaboAnalyst v. 6.0 (https://www.metaboanalyst.ca/) and MS Excel 365 with Data Analysis ToolPak. Google Sheet machine learning algorithm was used to predict the missing values for unavailable data points based on other available datasets.

3 RESULTS AND DISCUSSION

The findings from the correlation and regression analysis pointed the complex interactions among socio-economic and demographic factors, highlighting their combined impact on female labour force participation. The correlation matrix (Fig. 1) demonstrates interrelated associations among principal indicators. The positive correlation between female life expectancy at birth and expected years of schooling for females underscores the interdependent relationship between educational achievement and health consequences.

Furthermore, GDP Per Capita (PPP) exhibits a positive correlation with both Life Expectancy (r= 0.366, $p<0.01$) and Female Labor Force Participation (r=0.405, $p<0.01$), emphasizing the role of economic expansion in enhancing living standards and fostering women's involvement in the labor market.

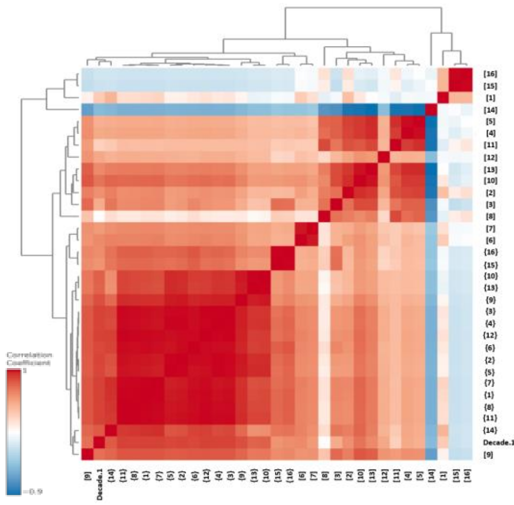


Figure 1: Correlation Heatmap.

The number and their corresponding variable names are as follows: [1]-Labor force female percentage of total labor, [2]-Labor force female, [3]-Employers female, [4]-Expected years of schooling for all genders, [5]-Expected years of schooling for female, [6]-Labor force female with basic education, [7]-Labor force female with advanced education, [8]-Life expectancy at 60 for female, [9]-Paid maternity leave length calendar days, [10]-GNI per capita in PPP($), [11]-Life expectancy at birth years, [12]-Government expenditure in education percentage of GDP, [13]-GDP growth per capita, [14]-Mortality rate of infant per 1000 live births, [15]-Lifetime risk of maternal death percentage, [16]-Number of maternal deaths, {1}-Average of LFF percentage of total labor, {2}-Average of Labor force female, {3}-Average of employers female, {4}-Average of expected years of schooling, {5}-Average of expected years of schooling for female, {6}-Average of LFF with basic education, {7}-Average of LFF with advanced education, {8}-Average of life expectancy at 60 for female, {9}-Average of paid maternity leave length in calendar days, {10}-Average of GNI per capita PPP($), {11}-Average of life expectancy at birth years, {12}-Average of Govt expenditure in education as percentage of GDP, {13}-Average of GDP growth per capita, {14}-Average of mortality rate of infant per 1000 live births, {15}-Average of lifetime risk of maternal death percentage, {16}-Average of number of maternal deaths.

Conversely, the negative correlations exhibited by variable such as the Mortality Rate (Infant Per 1000 Live Births) ($r=-0.499$, $p<0.001$) and positive correlations with expected years of schooling for a female ($r=+0.765$, $p<0.01$) with Life Expectancy and Educational Outcomes suggest the adverse effects of poor health metrics on socio-economic development can be improved by increasing the average study period for a woman. Further, Government Expenditure on Education is positively correlated with educational and health variables (Fig. 1), emphasizing the importance of public investment in nurturing human capital. Similar findings were pointed for South Asia countries in general [7]. In particular, a one percent increase in trade openness results with a 0.0152% rise in female employment, and empirical data supports long-term cointegration among education, trade, urbanization, male labor-force participation, per capita income, and employment. Also, 10% increases in the work parity index score were significantly associated with a decrease of 14.6 maternal deaths per 100,000 live births and an increase of 0.9 years in female life expectancy at birth, independent of county income level [8]. Therefore, public investment in education and work parity is essential, as government expenditure is positively correlated with improved educational and health outcomes. This tendency is consistent across different regions, regardless of a country's income level, reinforcing the concept of integrated growth strategies.

The Multi-Linear Regression analysis enhances these data by elucidating the degree, direction, and statistical significance of each variable's impact on socio-economic outcomes. The preliminary regression study, covering the period from 1991 to 2023 and omitting two maternal health variables, reveals a strong model with an R-squared value of 0.952, with Gross National Income (GNI) per Capita identified as a substantial positive factor influencing socio-economic outcomes:

Labor Force Female = 191959083 - 1621217.78×Expected Years of Schooling + 447867.24×Expected Years of Schooling Female + 231179.21×Life Expectancy at 60 Female - 90388.28×Paid Maternity Leave Length + 14558.93×GNI per Capita* - 194442.39×Life Expectancy at Birth - 1781077.94×Govt. Expenditure in Education - 40717.61×GDP Growth Per Capita* -1064582.39×Mortality Rate, $F = 50.816$, $R^2 = 0.952$ (* indicates a parameter that has a significant effect on the Labor Force Female as the dependent variable, with $p<0.05$).

At the same time GDP Growth per Capita shows a significant, albeit negative, relationship, suggests that while economic growth generally enhances overall living conditions, it can also contribute to disparities in specific contexts. In developing countries, for instance, high female labor force participation rates are often associated with economic hardship rather than empowerment [9], as many women enter the workforce out of necessity rather than choice. Moreover, while the gender gap in labor market participation tends to be smaller in developing economies compared to high-income countries [10], women in these regions are more likely to be employed in lower-paid jobs. This is largely due to limited access to education and skill development opportunities, which constrains their upward mobility within the labor market.

These findings highlight the crucial role of education in optimizing female labor market participation. By improving educational attainment among women, not only can their employment prospects and earnings potential be enhanced, but they may also gain greater societal recognition and agency. In the long run, such improvements could significantly improve economic growth. Notably, if all other factors remain constant, a 1% increase in female labor force participation is estimated to contribute to a 3.06% rise in overall economic growth [11].

Enhancing the analysis from 2000 to 2023, by incorporating the Percentage of Lifetime Risk in Maternal Death and the Count of Maternal Death, elucidates the intricate influences of maternal health indicators, offering a revised perspective that mirrors current trends and healthcare progress. The revised model demonstrates a robust fit, with a R^2 value of 0.986, highlighting the significance of GNI per capita and infant mortality rate as key predictors.

Labor force female = - 188965438 - 1904267×Expected Years of Schooling + 4207734×Expected Years of Schooling Female - 2237822×Life expectancy at 60 Female + 42365×Paid Maternity Leave Length + 16892×GNI Per Capita* + 2277591×Life Expectancy at birth - 47943×Govt Expenditure in education - 13243×GDP growth per capita + 3015358×Mortality rate infant* + 121705×Lifetime risk of maternal death - 529×Number of maternal deaths. $R^2 = 0.986$, $F = 78.475$, $p = 0.05$ (* indicates a parameter that has a significant effect on the Labor Force Female as the dependent variable, with $p<0.05$)

The analysis of Indian households reveals parallels with observations from other areas, such as Kenya, where financial strains associated to healthcare and the loss of essential family members significantly affect households [12]. In India, economic pressures compel households to reallocate resources and time, frequently resulting in women managing significant home responsibilities alongside income generation. Based on these ideas, policies like improving access to financial safety nets, promoting female labor participation, and strengthening healthcare facilities are crucial for alleviating economic shocks and creating sustained socio-economic growth.

Recognizing that PCA provides a more comprehensive synthesis by reducing the dataset's dimensionality and highlighting the most influential variables, we applied this technique and identified two principal components with the highest eigenvalue that represent the total amount of variance that can be explained by a given principal component. Two PCs account for 62.69% and 14.26% of the total variance in the dataset. The biplot analysis illustrates socio-economic and temporal trends in the dataset (Fig. 2).

These clusters reflect temporal changes, with the data for earlier years (1990 and 2000) distinctly separated from later years (2010 and 2020). This suggests that significant shifts occurred in the underlying variables over time. For example, indicators such as life expectancy, education levels, and mortality rates may have improved over the decades, as suggested by the clustering pattern and the positioning of observations along PC1. PC2, in turn, highlights gender-specific labor dynamics, influenced by variables such as female employers and female labor force. PC1 captures disparities in socio-economic development, with strong positive correlations observed for variables like life expectancy at birth for a new born and GDP growth per capita while average schooling years for a woman shows a negative correlation, reflecting its association with lower socio-economic progress.

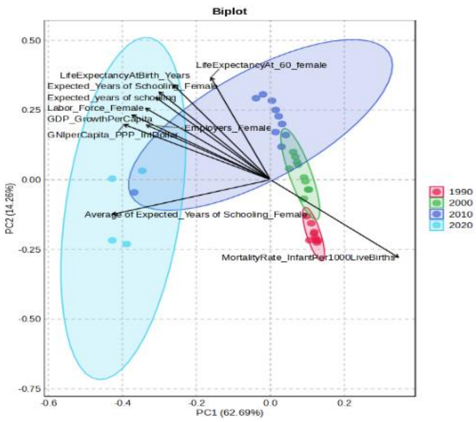


Figure 2: PCA Analysis of socio-economic and temporal parameters that affect Labor force female

Temporal analysis reveals that 2020 observations align with improved indicators, including higher GDP per capita and life expectancy, whereas 1990 data is associated with higher infant mortality rates. This indicates significant socio-economic and health progress over time.

In India, family is the major social unit, and all the important decisions in the family are made by the head of the household. Therefore, it is important to know how the education of women affects family-related issues [13]. All these factors contribute directly or indirectly to the working habit of a woman in a society like India. Increased family planning to reduce the unmet need (for spacing and limiting births) by amounts ranging from 25% to 100%, reduced maternal deaths by amounts ranging from 7.0% to 28.1% in rural India and 5.8% to 23.5% in urban India. In rural India, eliminating the unmet need for family planning decreased the TFR (Total Fertility Rate) from 2.97 to 2.14, the proportion of deaths that are pregnancy related from

16.4% to 12.3%, and the lifetime risk of maternal death from 1 in 65 to 1 in 90 [14].

Over the past two decades, government expenditure on education as a percentage of GDP has been a key driver of female labor force participation in India, enabling access to quality education, skill development, and economic empowerment. Despite the *Prevention of Child Marriage Act 2016*, early marriages persist, but women marrying after 21 years fare better in employment and earnings. Increasing education levels have also fuelled a rise in female entrepreneurship, with India ranking as the most active country for women entrepreneurs, particularly in manufacturing [15]. Life expectancy further impacts women's workforce participation by encouraging long-term investment in skills and improving health outcomes. Together, these factors foster socio-economic empowerment, transforming women's roles in economic and social development [16].

4 CONCLUSIONS

The progression of women's labor force participation in India from 1991 to 2023 underscores a complex interaction of socio-economic, cultural, and policy influences. Health and education have become essential factors influencing female labor force participation. Decreases in infant mortality rates and enhancements in maternal health have been associated with greater workforce inclusion. Access to higher education has equipped women with skills and confidence, facilitating their transfer from agriculture to industry and services, where formal work opportunities are expanding.

Economic factors such as GNI per capita and GDP growth have impacted women's labor force participation, although rural regions have persistent structural obstacles. Economic progress has broadened options, especially in entrepreneurship; nonetheless, cultural norms and gender biases persist in constraining women's potential. Policy interventions in education, healthcare, and maternity leave have yielded beneficial outcomes; nevertheless, their effectiveness is compromised by implementation deficiencies, particularly in marginalized areas. Cultural influences, such as familial arrangements, substantially influence workforce involvement, with Hindu women experiencing comparatively greater autonomy.

ACKNOWLEDGMENTS

This research was funded by the BMBF Project “Professor*innengewinnung und Nachwuchsentwicklung zur Etablierung eines Centers of Advanced Scientific Education (CASE).

Also, 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] "Keys to understanding the world of work," [Online]. Available: https://webapps.ilo.org/infostories/en-GB/Stories/Employment/barriers-women#global-gap, [Accessed: 15 May 2025].

[2] "Our World in Data – Women's Employment," [Online]. Available: https://ourworldindata.org/female-labor-supply, [Accessed: 15 May 2025].

[3] "Economy of India," [Online]. Available: https://en.wikipedia.org/wiki/Economy_of_India, [Accessed: 15 May 2025].

[4] R. Lahoti et al., "Economic development and female labor force participation in India," IIM Bangalore Research Paper, no. 414, 2013.

[5] A. Costagliola, "Labor participation and gender inequalities in India: Traditional gendernorms in India and the decline in the Labor Force Participation Rate (LFPR)," The Indian Journal of Labour Economics, vol. 64, no. 3, pp. 531-542, 2021.

[6] "World Bank Group | Gender Data Portal – India," [Online]. Available: https://genderdata.worldbank.org/en/economies/india#:~:text=In%20India%2C%20the%20labo, [Accessed: 15 May 2025].

[7] L.C. Voumik et al., "The impact of female education, trade openness, per capita GDP, and urbanization on women's employment in South Asia: application of CS-ARDL model," Systems, vol. 11, no. 2, p. 97, 2023.

[8] "Gender parity at scale: Examining correlations of country-level female participation in education and work with measures of men's and women's survival," [Online]. Available: https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(20)30043-2/fulltext, [Accessed: 15 May 2025].

[9] "Female Labor Force Participation in Developing Countries," IZA World of Labor, no. 87, 2014, [Online]. Available: https://wol.iza.org/articles/female-labor-force-participation-in-developing-countries/long, doi:10.15185/izawol.87.

[10] International Labour Organization, "Economic Impacts of Reducing the Gender Gap," What Works: Research Brief No. 10, 2017, [Online]. Available:

https://www.ilo.org/media/422531/download, [Accessed: 15 May 2025].

[11] R. Akhtar et al., "Economic growth, gender inequality, openness of trade, and female labour force participation: A nonlinear ARDL approach," Economic Change and Restructuring, vol. 56, no. 3, pp. 1725-1752, 2023.

[12] A. Kes et al., "The economic burden of maternal mortality on households: evidence from three sub-counties in rural western Kenya," Reproductive Health, vol. 12, p. 1-10, 2015.

[13] K. Borkotoky et al., "Female education and its association with changes in socio-demographic behaviour: evidence from India," Journal of Biosocial Science, vol. 47, no. 5, pp. 687-706, 2015.

[14] S.J. Goldie et al., "Alternative strategies to reduce maternal mortality in India: a cost-effectiveness analysis," PLoS Medicine, vol. 7, no. 4, p. e1000264, 2010.

[15] R. Singh et al., "Exploring Reasons for Low Female Labour Force Participation in Regular Salaried Jobs: Evidence from Young Lives Longitudinal Study in India," Indian Journal of Human Development, vol. 16, no. 2, pp. 267-285, 2022.

[16] S. Manshani et al., "Startup women in startup India: A study of women entrepreneurs in India," International Journal of Business Management (IJBGM), vol. 6, no. 4, pp. 91-100, 2017.