Enhancing Machine Learning Model Accuracy with Effective Data Scaling Strategies

Wasan Khairallah Ali
Department of Chemistry, College of Science, University of Mosul, 41002 Mosul, Iraq
wasankhairallahali@uomosul.edu.iq

Keywords: Data Scaling, Machine Learning Algorithms, Prediction of Heart Disease, Model Performance Metrics, Preprocessing Strategies.

Abstract: The study on the impact of data scaling techniques on machine learning algorithms for predicting heart disease highlights the importance of preprocessing in enhancing model performance. Data scaling is essential when dealing with datasets that have diverse attribute ranges, as it can significantly influence the effectiveness of various machine learning models. In this investigation, eleven widely used algorithms, including K-Nearest Neighbors (KNN) and Logistic Regression, were evaluated using three scaling methods: Min-Max scaling, Z-score standardization, and MaxAbs scaling. The performance was assessed through precision, recall, and F1 score metrics across multiple experiments.The findings indicate that several algorithms performed better with MaxAbs scaling, particularly those sensitive to data distribution, such as KNN and Logistic Regression. This suggests that the choice of scaling technique is crucial for achieving accurate and consistent predictions in machine learning applications related to heart disease. The results emphasize the need for careful selection of scaling methods to optimize the performance of machine learning models in medical diagnostics.

1 INTRODUCTION

As a result of their ability to analyze large datasets and make predictions, machine learning models have gained widespread adoption across many industries, including healthcare and finance. Creating accurate and efficient machine learning models, however, requires data handling. A raw, unprocessed dataset's noise and inconsistency can have an adverse effect on algorithm performance. Preprocessing techniques like data scaling are essential for improving model performance. Algorithms are prevented from being biased against variables with larger scales by scaling data, which includes normalization and standardization. When not scaled, distance-based machine learning algorithms, like KNN, or gradient-based algorithms, like gradient descent, can produce inaccurate results, take longer to converge, or require longer training times [1]. Effective data scaling can improve model accuracy and performance, particularly when dealing with high-dimensional datasets. Scaling is an essential step for many machine learning algorithms, but the optimum scaling strategy will depend on the data's characteristics and the algorithm itself. Several data scaling techniques are examined, including Min-Max scaling, standardizing Z-scores, and robust scaling on machine learning models. Models can achieve greater accuracy and reliability if practitioners understand and apply the appropriate data scaling methods.

Electrocardiograms and blood tests are often required to evaluate heart disease symptoms properly. There are almost 12 million deaths caused by heart disease every year [2]. As a result, it is crucial to diagnose this disease at the earliest opportunity. Artificial intelligence (AI) has provided quick and alternative methods for diagnosing diseases, which may be beneficial in rural areas where doctors are few and diagnostic equipment is costly. An automated system that can be operated by non-medical personnel would therefore be helpful. Research shows that early diagnosis of heart disease can save patients time, money, and healthcare costs by using additional patient information and medical histories [3]. Machine learning approaches help develop decision support systems.

Learning online [4], scheduling [5], multiobjective optimization [6], and vehicle routing [7] are examples of AI-based algorithms (e.g., heuristics, metaheuristics). Medical diagnosis can be greatly improved using deep-learning-based approaches, according to several recent

studies [4],[5]. Deep learning technologies, such as image segmentation, now improve the efficiency and effectiveness of diagnosing diseases like diabetes, cancer, and SARS-CoV-2. Although SARS-CoV-2 caused a global pandemic, chest X-rays (X-rays) and computed tomography (CT) scans were used in several studies to identify COVID-19 symptoms [1]. A laptop of office grade and a small amount of data were used for this experiment.

By combining data classification techniques with nature-inspired algorithms, genetic programming [10] and swarm algorithms can be used to identify bacteria from viral meningitis [11]. In recent years, artificial intelligence has gained popularity among decision-support and optimization tools due to these advantages. The computational cost of deep learning and neural network approaches increases with larger datasets. Since traditional machine learning approaches consume less memory and computational power than deep learning approaches, they are frequently preferred unless necessary [6].

The development of a decision support system based on data analysis requires standard data, which is often the result of extensive preprocessing. Preprocessing involves the cleaning of data, pruning, slicing, and scaling of features before data can be analyzed. The effects of data scaling on overall model performance have been examined in many studies. Still, few have investigated the effects of different machine-learning algorithms and feature selection on overall model performance [7]. It examines ML algorithms for predicting symptoms of heart disease in patients using various data scaling methods. A robust, data-driven decision support system can be developed based on researchers' and practitioners' experimental results.

2 LITERATURE REVIEW

According to the referenced literature, machine learning algorithms (ML) are expressed in terms of accuracy. ML algorithms, however, perform differently in each study because of different approaches to ML. According to the Author [14], the Bagging algorithm achieved 81.14% accuracy, and the Decision Tree (DT) algorithm achieved 78.90% accuracy. According to the Author [15], 84.14% of patients with heart disease were correctly identified using a naïve approach. The Author [8] also reported an accuracy rate of 84.10% using a decision tree [23].

The computational accuracy of several references shows promising results, yet the UCI heart disease dataset shows a variation of almost 7–8% despite using the same dataset [9]. Due to the lack of mention of data scaling methods in any of the studies, it is impossible to identify the reasons for variations in DT accuracy between them. Perhaps the training/test sets are divided differently or the number of features is different. Moreover, accuracy is not always indicative of overall performance. Therefore, categorization matrices based on F1 scores and accuracy, precision, recall, and recall coefficients are more reliable [10].

As the preprocessing of data does not seem to affect prediction models, most studies examine feature selection rather than data scaling due to the lack of information about its effect on prediction models. However, data analysis cannot ignore the importance of feature selection. Using ML algorithms and features as examples, the Author [7] demonstrated different levels of accuracy based on different combinations of algorithms. Furthermore, the study showed that accuracy often drops by 14% due to the limited number of features in medical diagnosis, which is significant [11].

Various machine-learning approaches rely heavily on normalization, according to the Author [12]. A variety of machine learning algorithms, including 12 different ones, were used in the study in order to predict heart disease. Based on a study that used different normalization methods, there is a correlation between the performance of ML algorithms and the selection of normalization methods. Its accuracy is 78%, which is the highest of all eleven supervised algorithms. However, NaVe Bayes also has a high accuracy and low fitting time, according to this study.

In addition, Author [13] notes that data normalization and standardization techniques like MinMax normalization also play an important role in data analysis. Combining ML algorithms like KNN, Nave Bayesian, ANN, and SVM with RBF was used in this study. The performance of NB was more stable without data scaling techniques than that of SVMs, while that of KNN was more stable without data scaling techniques. According to their computation, MinMax scaling outperformed other algorithms with SVM, which contradicts the findings of the study. Although their studies do not synchronize, one could still conclude that data scaling affects overall performance even though they are not synchronized.

According to the Author [14], feature scaling and normalization techniques can improve classification accuracy and model convergence rates. Scaling methods include min-max normalization, Z-score standardization, and decimal scaling; the choice depends on the distribution of data and algorithm [15].

According to the Author, preprocessing is an essential step in knowledge discovery [16]. Unscaled data is a significant problem for several algorithms, including support vector machines, K-nearest neighbours, and neural networks. The researchers noted that standardization often improves the performance of models based on Gaussian-distributed features, whereas robust scaling is more suitable for datasets with outliers. The Author [17] conducted comparative experiments using various scaling techniques across different datasets and machine learning algorithms. Using imbalanced datasets as a case study, another study examined the effects of data preprocessing strategies. The authors primarily focused on class imbalance, but they also demonstrated how improper scaling can exacerbate minority class detection issues, particularly in distance-based models [18].

3 METHODOLOGY

A total of five benchmark datasets were used to examine the impact of min-max data normalization on the regression performance of three machine learning algorithms. Based on the methodology, the following conclusions can be drawn:

3.1 The UCI Datasets

UCI's machine learning repository includes five benchmark datasets [19]. All attributes have a wide range of records, which makes these datasets appealing. The Airfoil SelfNoise Dataset has big differences in its ranges, while power plant datasets have very similar ranges. Considering this variation, the min-max scaling method should have a greater impact on regression performance, which is the primary objective. This dataset contains data from physics, biology, engineering, and business applications.

3.2 Data Scaling in Classification Modelling

When it comes to machine learning and data mining, scaling and normalization refer to the same preprocessing procedure. Data consolidation or transfer is necessary to make mining and modelling possible. [20]. A scaled dataset has been shown to perform better than an unscaled dataset in models trained with scaled data. A distance-based method such as nearest neighbour classification and clustering relies heavily on scaling data. A neural network's stability and speed of learning are improved when the input data is normalized [20].

Confusions of gene expression normalization. Microarray technology makes it possible to diagnose a wide range of diseases and cancers based on gene expression data. It is usually necessary to perform a normalization step in order to identify and remove systematic variations in fluorescence measurements [21] prior to analyzing them. In our study, the normalization of gene expression is not the same as data scaling. In order to learn a classification model, normalized gene expression datasets are usually scaled and processed. In most cases, gene expression data-driven models outperform gene expression data-driven models without scaling by a wide margin.

3.3 Commonly Used Data Scaling Algorithms

3.3.1 Min-Max Algorithm

Algorithms such as Min-max use linear transformations. Our samples contain the minimum and maximum values of variables. x_{min} and x_{max}. Using the formula below, Min-Max converts a value, v, into a value, v';

It is commonly used to scale data using two algorithms

Algorithms based on min-max and z-score.

$$v' = \frac{v - x_{min}}{x_{max} + x_{min}} + x_{min}. \quad (1)$$

Training samples are scaled from $[x_{min}, x_{max}]$ to [-1, 1] (or 0, 1) by a lined plotting. In some applications, it may be problematic to have scaled values outside the interval [-1, 1] (or [0, 1]) if the unseen/testing samples are outside the training data range. Further, it is highly sensitive to outliers, as explained in the subsequent sections.

3.3.2 Z-Score Algorithm

The Z-score algorithm divides a variable's original value by its new value, v'.

$$v' = \frac{v - \ddot{x}}{\sigma_x}. \quad (2)$$

There are two variables in the training samples: $\bar{x}$ and σ_x, which represent their mean and standard deviation. With a mean of 0 and a standard deviation of 1, you will get new values with 0 and 1 as the mean and standard deviation, respectively. In addition to not using interval mapping, the algorithm is also

sensitive to outliers. The mean and standard deviation calculated from the data may not accurately reflect the true mean and standard deviation in a few cases, especially in biomedical research scenarios.

3.3.3 Data Saling Formula

Variables in the examples are modelled as random variables $(r.v.)\ X$:

$$v' = P_x(v) \quad , \tag{3}$$

X r.v. is represented by a cumulative density function (CDF) called $P_X(\cdot)$

The Histogram Equalization [21], which is used in Digital Image Processing to enhance image contrast, also uses a CDF as a mapping tool. We use the GL algorithm instead of Histogram Equalization by learning/approximating the CDF's functional expression, which can be used to scale unknown values according to the CDF.

3.4 Overview of Algorithms

3.4.1 Logistic Regression (LR)

Additionally, Logistic Regression can be used to classify data. LR is an effective machine-learning method for classifying objects. According to Verhulst's paper in proceedings of the Belgian Royal Academy, the logistic function has three parameters and a curve. In spite of the simplicity of this machine learning method, it has numerous applications. A logistic regression technique predicts binary classes based on statistical data. A Bernoulli distribution is applied to the dependent variable. Sigmoid functions, or logistic functions, are S-shaped curves that take values between 0 and 1. Observe that the curve will predict 1 if it reaches positive infinity and 0 if it reaches negative infinity [22].

3.4.2 Linear Discriminant Analysis (LDA)

Linear discriminant analysis was used in this study to extract features, while principal component analysis was also evaluated. Several techniques have been developed to reduce dimensionality in preprocessing for machine learning classification applications, including linear discriminant analysis (LDA) [23]. LDA optimizes class separation by minimizing variance between and within classes by transforming features into a lower-dimensional space. It is possible to project an N-dimensional space onto a subspace $K\ (K \leq n - 1)$ by using LDA without compromising class discrimination.

3.4.3 K-Nearest Neighbors (KNN)

There was an initial report on the use of k-Nearest Neighbors (KNN) to categorize text. As part of this method, a query's category is determined not only from its closest document in the document space but also from the k closest documents' categories. In this context, the Vector method is equivalent to the KNN method, where k = 1. Using a vector-based, distance-weighted matching function, this work calculates a document's similarity using the Vector method, as did Yang.

3.4.4 Classification and Regression Trees (CART)

Decision trees can be built using CART both for classifications and regressions. Using classification trees, we can separate datasets into two classes, while using regression trees, we can predict numeric or continuous values. The goal of a classification tree is to divide the dataset at hand into two parts based on data homogeneity. When splitting attributes in CART, impurity measures such as entropy and Gini index are used to determine where they should be split. Since regression trees have no classes for output attributes, predicting the value of output variables is more important than predicting their classes. The split point that gives the lowest sum of squared errors between the actual and predicted values is chosen as the root node, and then it is divided into two more nodes.

3.4.5 Naive Bayes (NB)

Based on the assumption that predictors are independent, this classification technique utilizes the Bayes Theorem. Using the Naïve Bayes classifier, one feature is assumed to be independent of another. Naïve Bayes mainly targets the text classification industry. The main purpose of this method is to cluster and classify in accordance with conditional probabilities.

3.4.6 Support Vector Machine (SVM)

Support vector machines (SVMs) are among the most common techniques for machine learning. It is an algorithm that uses support vector machines to classify and regress data. SVMs can also be classified nonlinearly using kernel tricks, which implicitly map inputs into high-dimensional feature spaces. Class margins are drawn in this way. Classification error is minimized by drawing margins that are as far away from the classes as possible.

3.4.7 Extreme Gradient Boosting (XGB)

A new gradient-boosting method called XGBoost has been introduced. As a result of XGBoost, gradient-boosting decision trees can be implemented. Enhancing existing models by adding new ones is an ensemble-based method of correcting errors. In boosting techniques, errors or misclassified observations are given weights to make the selection of subsamples more intelligent. Gradient boosting creates new models by predicting the errors or residuals of earlier models and then combining them to arrive at a final prediction. [12] Gradually boosting is used to reduce loss as new models are added.

3.4.8 Decision Tree (DT)

Graphs are used to represent choices and their results. There are nodes in the graph that represent events and edges that represent the rules or conditions that govern decisions. A node connects an individual branch or node of a tree. An attribute can be represented as a node, and a value can be represented as a branch in a classified group.

3.4.9 Random Forest (RF)

Classification and regression are both carried out using tree-based random forests. As a result of constructing multiple trees, the mean prediction would be the classification output.

3.4.10 Gradient Tree Boosting (GB)

Euclidean optimization methods cannot be used to optimize the tree ensemble model since functions are included as parameters. Models are instead trained in an additive manner.

3.4.11 Adaboost (AB)

Weights are maintained by the AdaBoost algorithm, and they are adjusted after each weak learning cycle. A weak learner's misclassifications will be increased in weight, and correctly classified samples will have their weight decreased.

3.4.12 Extra Trees (ET)

A classifier that uses ensembles of trees to classify data is the extra Trees Classifier. Multiple decision trees are combined into a single prediction by the random forest algorithm. The Extra Trees Classifier differs from traditional Random Forests in that instead of selecting the best split for each feature, and it uses random thresholds. As a result of this randomization process, we are left with a greater variety of trees so that we can make stronger final predictions.

3.5 Evaluating Models

Model performance was evaluated using established classification metrics derived from the confusion matrix framework presented in [24]. Four fundamental classification outcomes were computed: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Based on these values, accuracy, precision, recall, and F1-score were calculated according to the formulas presented in the aforementioned reference.

Accuracy quantifies the overall proportion of correctly classified samples relative to all predictions. Precision measures the relationship between correctly predicted positive instances and the total number of predicted positives. Recall, also denoted as the true positive rate (TPR), represents the proportion of actual positive cases that were correctly identified by the model. The F1-score combines precision and recall into a single performance metric, providing a balanced evaluation of model effectiveness. These metrics collectively provide a comprehensive assessment of the model's classification performance as outlined in [24].

4 RESULTS AND DISCUSSION

On the basis of eleven different data scaling techniques, Figure 1 illustrates the overall accuracy of eleven machine learning algorithms. The highest accuracy was obtained by SVM and CART (99%), even without scaling. Compared to KNN, it had the poorest accuracy, with just 75%. A significant improvement in overall accuracy was observed when MaxAbs scalers were applied. Most algorithms also exhibited largely consistent accuracy, regardless of scaling methods used, with the exception of Logistic Regression (LR), KNN, and SVM, which exhibited more pronounced differences.

Based on various data scaling techniques, Figure 2 shows the overall precision scores of eleven machine learning algorithms. CART's precision (100%) without scaling was the highest, while KNN's was the lowest (78%). A number of algorithms suffer from reduced performance when normalized, including the LR algorithm, the LDA algorithm, the

CART algorithm, the SVM algorithm, and the AB algorithm. The figure clearly illustrates that CART consistently achieves the highest precision, no matter how it is scaled. Conversely, KNN, SVM, and LDA were found to be the least precise. In addition to CART, Random Forest (RF) and Extra Trees (ET) demonstrated greater precision across different scaling methods.

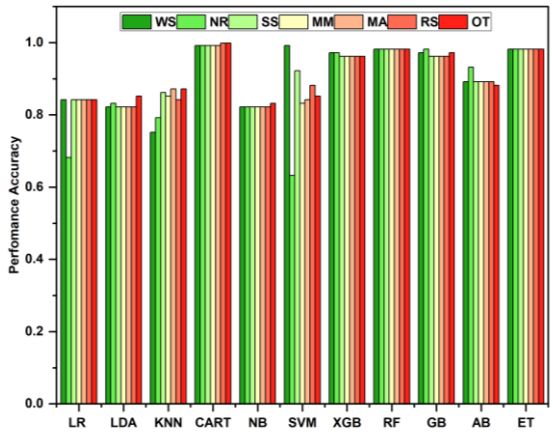

Figure 1: Various algorithms' accuracy performance.

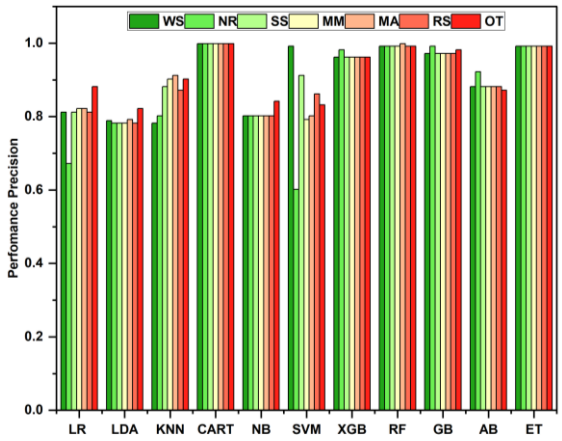

Figure 2: A comparison of algorithms based on their precision.

According to Figure 3, eleven machine learning algorithms were evaluated with various data scaling methods to determine their overall recall performance. With a recall close to one, CART achieved the highest performance. On the other hand, KNN (ranging from 0.72 to 0.83) and Logistic Regression (ranging from 0.74 to 0.89) recorded the lowest recall values.

A comparison of eleven machine learning algorithms is shown in Figure 4 in terms of the overall F1 score. Among the F1 teams, CART achieved the highest score of 100%. The lowest F1 scores were achieved by KNN, SVM, Logistic Regression (LR), and Naïve Bayes (NB) among the algorithms evaluated.

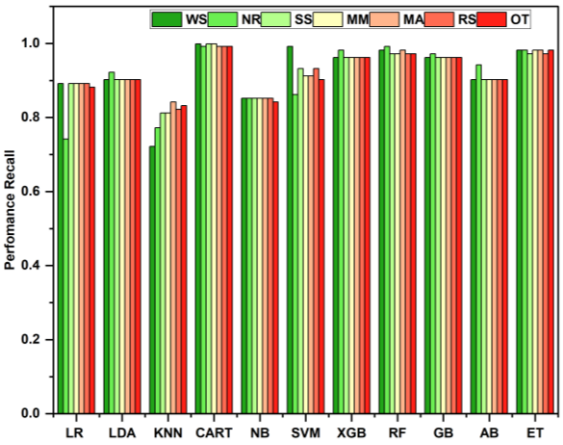

Figure 3: Algorithm performance based on recall.

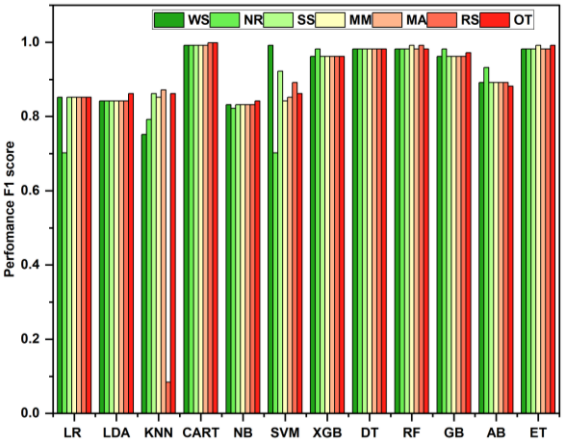

Figure 4: An analysis of the performance of different algorithms using F1 scores.

5 CONCLUSIONS

Machine learning models perform significantly better when data is scaled, particularly in medical diagnosis scenarios such as predicting heart disease, where precision is critical. Even without applying scaling techniques, the Classification and Regression Tree (CART) algorithm consistently outperformed eleven other machine learning algorithms tested across all evaluation metrics, demonstrating its robustness to data variations. Conversely, algorithms like K-Nearest Neighbors (KNN) and Logistic Regression exhibited substantial performance gains when scaling

methods such as MaxAbs scaling were employed, indicating their sensitivity to feature scales. Interestingly, certain algorithms showed minimal or negligible changes when scaling was applied, whereas normalization methods even resulted in degraded performance for specific models. This underscores the critical insight that no single scaling technique universally enhances model performance; rather, its effectiveness is inherently tied to the characteristics of the dataset and the specific algorithm employed. In healthcare applications, selecting an appropriate scaling strategy can significantly enhance model reliability, accuracy, and efficiency, thereby improving the quality and trustworthiness of data-driven clinical decision support systems, and ultimately, patient outcomes.

REFERENCES

[1] P. Rani, S. Verma, S. P. Yadav, B. K. Rai, M. S. Naruka, and D. Kumar, "Simulation of the lightweight blockchain technique based on privacy and security for healthcare data for the cloud system," International Journal of E-Health and Medical Communications, vol. 13, no. 4, pp. 1-15, 2022, [Online]. Available: https://doi.org/10.4018/IJEHMC.20221001.oa8.

[2] J. Soni, U. Ansari, D. Sharma, and S. Soni, "Predictive data mining for medical diagnosis: An overview of heart disease prediction," International Journal of Computer Applications, vol. 17, no. 8, pp. 43-48, 2011.

[3] W. P. Lord and D. C. Wiggins, "Medical Decision Support Systems," in Advances in Health Care Technology, G. Spekowius and T. Wendler, Eds., vol. 6. Dordrecht: Kluwer Academic Publishers, 2006, pp. 403-419, doi: 10.1007/1-4020-4384-8_25.

[4] M. M. Ahsan, T. E. Alam, T. Trafalis, and P. Huebner, "Deep MLP-CNN Model Using Mixed-Data to Distinguish between COVID-19 and Non-COVID-19 Patients," Symmetry, vol. 12, no. 9, p. 1526, Sep. 2020, [Online]. Available: https://doi.org/10.3390/sym12091526.

[5] M. M. Ahsan et al., "Detecting SARS-CoV-2 From Chest X-Ray Using Artificial Intelligence," IEEE Access, vol. 9, pp. 35501-35513, 2021, [Online]. Available: https://doi.org/10.1109/ACCESS.2021.3061621.

[6] P. Rani, P. N. Singh, S. Verma, N. Ali, P. K. Shukla, and M. Alhassan, "An implementation of modified blowfish technique with honey bee behavior optimization for load balancing in cloud system environment," Wireless Communications and Mobile Computing, vol. 2022, pp. 1-14, 2022.

[7] M. S. Amin, Y. K. Chiam, and K. D. Varathan, "Identification of significant features and data mining techniques in predicting heart disease," Telematics and Informatics, vol. 36, pp. 82-93, Mar. 2019, [Online]. Available: https://doi.org/10.1016/j.tele.2018.11.007.

[8] M. Shouman, T. Turner, and R. Stocker, "Integrating decision tree and k-means clustering with different initial centroid selection methods in the diagnosis of heart disease patients," in Proceedings of the International Conference on Data Science, 2012, p. 1, [Online]. Available: http://world-comp.org/p2012/DMI9007.pdf.

[9] G. Ansari, P. Rani, and V. Kumar, "A novel technique of mixed gas identification based on the group method of data handling (GMDH) on time-dependent MOX gas sensor data," in Proceedings of International Conference on Recent Trends in Computing, Springer, 2023, pp. 641-654.

[10] A. Singh et al., "Blockchain-Based Lightweight Authentication Protocol for Next-Generation Trustworthy Internet of Vehicles Communication," IEEE Transactions on Consumer Electronics, vol. 70, no. 2, pp. 4898-4907, May 2024, [Online]. Available: https://doi.org/10.1109/TCE.2024.3351221.

[11] S. Verma et al., "An automated face mask detection system using transfer learning based neural network to preventing viral infection," Expert Systems, p. e13507, 2024.

[12] L. Shahriyari, "Effect of normalization methods on the performance of supervised learning algorithms applied to HTSeq-FPKM-UQ data sets: 7SK RNA expression as a predictor of survival in patients with colon adenocarcinoma," Briefings in Bioinformatics, vol. 20, no. 3, pp. 985-994, May 2019, [Online]. Available: https://doi.org/10.1093/bib/bbx153.

[13] A. Ambarwari, Q. Jafar Adrian, and Y. Herdiyeni, "Analysis of the Effect of Data Scaling on the Performance of the Machine Learning Algorithm for Plant Identification," Journal of RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 4, no. 1, pp. 117-122, Feb. 2020, [Online]. Available: https://doi.org/10.29207/resti.v4i1.1517.

[14] S. G. K. Patro and K. K. Sahu, "Normalization: A Preprocessing Stage," International Advanced Research Journal in Science, Engineering and Technology, pp. 20-22, Mar. 2015, doi: 10.17148/IARJSET.2015.2305.

[15] P. Rani, U. C. Garjola, and H. Abbas, "A Predictive IoT and Cloud Framework for Smart Healthcare Monitoring Using Integrated Deep Learning Model," NJF Intelligent Engineering Journal, vol. 1, no. 1, pp. 53-65, 2024.

[16] J. Han, M. Kamber, and J. Pei, "Data mining: Concepts and Techniques," Morgan Kaufmann Publishers, 2012, [Online]. Available: http://homes.di.unimi.it/ceselli/IM/2012-13/slides/02-KnowYourData.pdf.

[17] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," Neural Networks, vol. 106, pp. 249-259, Oct. 2018, [Online]. Available: https://doi.org/10.1016/j.neunet.2018.07.011.

[18] S. P. Yadav et al., "An improved deep learning-based optimal object detection system from images," Multimedia Tools and Applications, vol. 83, no. 10, pp. 30045-30072, 2024.

[19] D. Dua and C. Graff, "UCI machine learning repository," 2017, [Online]. Available: https://archive.ics.uci.edu/ml.

[20] J. Han, J. Pei, and H. Tong, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2022, [Online]. Available: https://books.google.com/books?hl=en&lr=&id=NR1oEAAAQBAJ.

[21] S. Dudoit, Y. H. Yang, M. J. Callow, and T. P. Speed, "Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments," Statistica Sinica, pp. 111-139, 2002.

[22] A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien, "Linear discriminant analysis: A detailed tutorial," AI Communications, vol. 30, no. 2, pp. 169-190, 2017.

[23] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," Machine Learning, vol. 63, no. 1, pp. 3-42, Apr. 2006, [Online]. Available: https://doi.org/10.1007/s10994-006-6226-1.

[24] D. M. W. Powers, “Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation,” J. Mach. Learn. Technol., vol. 2, pp. 37–63, 2011, doi: 10.9735/2229-3981.