Available online at website: https://jurnal.iaii.or.id/index.php/RESTI

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 9 No. 3 (2025) 519 - 524 e-ISSN: 2580-0760

Obesity Status Prediction Through Artificial Intelligence and Balanced Label Distribution Using SMOTE

Arif Riyandi[1*], Mahazam Afrad[2], M Yoka Fathoni[3], Yogo Dwi Prasetyo[4]
[1,2,3,4] Department of Information System, Information System, Telkom University, Purwokerto, Indonesia
[1]arifriyandi@telkomuniversity.ac.id, [2]mahazamafrad@telkomuniversity.ac.id, [3]myokafathoni@telkomuniversity.ac.id, [4]yogoop@telkomuniversity.ac.id

Abstract

Obesity, a global health challenge influenced by genetic and environmental factors, is characterized by excessive body fat that increases the risk of various diseases. With over two billion individuals affected worldwide, addressing this issue is crucial. This study investigated the application of Artificial Intelligence (AI) to predict obesity status using a dataset of 1,610 individuals, including demographic and anthropometric data. Four AI algorithms were analyzed: Artificial Neural Network (ANN), K-Nearest Neighbors (KNN), Random Forest, and Support Vector Machine (SVM). The Synthetic Minority Over-Sampling Technique (SMOTE) was applied to address dataset imbalance. The results demonstrate that SMOTE significantly enhanced the models' performance, especially in recall and F1-score for minority classes, such as obesity. Random Forest achieved the highest accuracy (92%) and recall (92%) post-SMOTE. The ANN showed substantial improvement in recall, increasing from 77% to 89%, whereas the SVM achieved the highest precision (89%), minimizing false positives. Despite these improvements, KNN remained the least effective. The findings underscore the critical role of SMOTE in improving AI model accuracy for obesity prediction and highlight Random Forest as the most reliable algorithm for clinical decision-making. Limitations, such as dataset representativeness, suggest future research directions, including expanding data diversity and advanced feature selection techniques. This study provides valuable insights into leveraging AI and preprocessing methods for obesity management.

Keywords: obesity prediction; SMOTE; random forest; artificial neural network; AI in healthcare

How to Cite: A. Riyandi, Mahazam Afrad, M Yoka Fathoni, and Yogo Dwi Prasetyo, "Obesity Status Prediction Through Artificial Intelligence and Balanced Label Distribution Using SMOTE", *J. RESTI (Rekayasa Sist. Teknol. Inf.)* , vol. 9, no. 3, pp. 519 - 524, Jun. 2025.
*Permalink/*DOI: https://doi.org/10.29207/resti.v9i3.6204

Received: December 2, 2024
Accepted: May 26, 2025
Available Online: June 12, 2025

1. Introduction

Obesity is an escalating global health challenge with significant public health implications. The World Health Organization (WHO) reports that the prevalence of obesity has steadily increased over the past few decades, affecting various age groups [1]. Globally, obesity has more than doubled since 1990 and has become one of the leading causes of morbidity and mortality associated with non-communicable diseases, such as coronary heart disease, hypertension, type 2 diabetes, and certain types of cancer [2]. This increasing trend is particularly concerning in Southeast Asia, including Indonesia, where obesity is linked to a higher risk of chronic diseases and contributes to the growing burden on national healthcare systems [3].

In recent years, various efforts have been made to address obesity, including monitoring dietary patterns, physical activity, and developing technology-based predictive methods. Artificial intelligence (AI) has been widely used in the medical field to help predict an individual's health status by analyzing healthcare data [4]. One of the primary approaches involves utilizing machine learning models to analyze demographic and anthropometric data to identify individuals at risk of obesity [5].

However, one of the main challenges in applying AI for obesity prediction is data imbalance in healthcare datasets. This imbalance causes predictive models to be biased toward classifying individuals into majority categories, making minority groups, such as individuals with obesity, less likely to be accurately detected [6]. As a result, the model's performance is often suboptimal in accurately identifying obesity cases. To address this issue, the Synthetic Minority Over-Sampling Technique (SMOTE) has been introduced as a solution to enhance the representation of minority classes by generating synthetic samples to balance data distribution [7].

Several studies have explored the use of SMOTE in various healthcare applications. However, research specifically evaluating its impact on the performance of different AI algorithms in obesity prediction remains limited [8]. Previous studies have typically assessed only one or two algorithms without conducting broader comparative analyses, making it difficult to draw comprehensive conclusions about the effectiveness of this method across different machine learning approaches. Additionally, many studies have only tested SMOTE on small or specific datasets, limiting the generalizability of their findings to larger populations [9].

This study aims to evaluate and compare the performance of four machine learning algorithms Artificial Neural Network (ANN), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest in predicting obesity status before and after applying SMOTE. These algorithms were selected due to their distinct characteristics in handling healthcare data. ANN is highly capable of recognizing complex patterns in nonlinear datasets, SVM is highly effective in classifying high-dimensional data, KNN operates based on similarity between data points, and Random Forest is known for its robustness against noise and its ability to handle datasets with numerous variables [10].

Theoretically, this study contributes by confirming the effectiveness of SMOTE in improving the performance of various machine learning algorithms in the healthcare domain. Practically, the findings of this research are expected to provide recommendations for healthcare practitioners in selecting more accurate and reliable obesity prediction methods to support data-driven clinical decision-making. By leveraging advanced AI and data-balancing techniques such as SMOTE, early obesity detection can be enhanced, allowing for earlier medical interventions to prevent further complications [11].

With the growing health challenges posed by obesity, the development of accurate and reliable AI-based predictive models has become increasingly urgent. Therefore, this study offers new insights that contribute to the advancement of early obesity detection systems and opens opportunities for further exploration of AI integration in data-driven healthcare solutions [12].

2. Methods

This study employs a quantitative approach with an experimental design to test and evaluate the performance of AI algorithms in predicting obesity status. This approach facilitates an objective analysis of the effects of data imbalance and the application of SMOTE [13], [14].

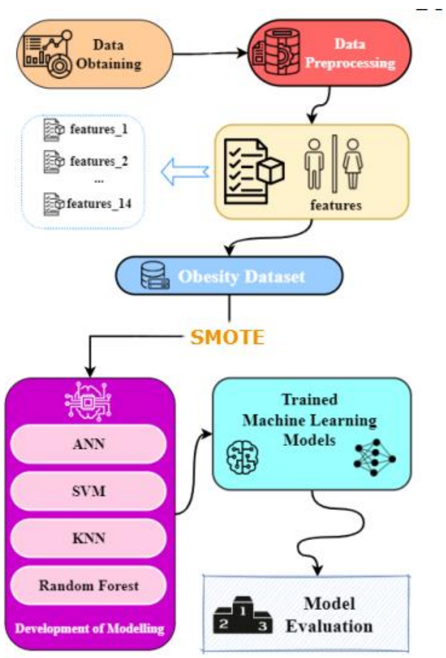

Figure 1. Flow chart of the study

2.1 Dataset

The dataset used in this study was obtained through an online survey involving 1,610 individuals from various regions in Turkey. The survey captured key variables relevant to obesity analysis, including gender, age, height, dietary habits, physical activity levels, smoking behavior, and transportation patterns. The study population comprised males and females aged 18 to 54 years. The sample included 712 males and 898 females, with the obesity class distribution as follows: underweight (73 individuals), normal weight (658 individuals), overweight (592 individuals), and obese (287 individuals). Sampling was designed to encompass diverse demographic, social, and lifestyle groups to ensure broad representation in predicting obesity status. [11], [15], [16].

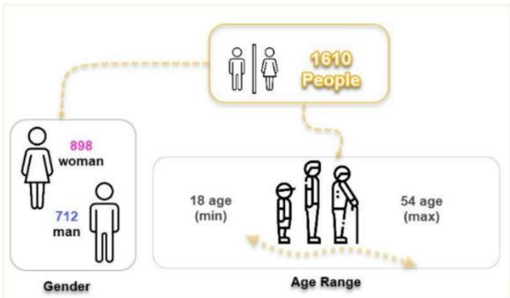

Figure 2. Information about the study group that created the obesity dataset

2.2 Preprocessing

The SMOTE technique was applied to address data imbalance by generating synthetic samples of minority classes through linear interpolation. This method has proven effective in enhancing the performance of machine learning algorithms on imbalanced datasets [9], [12].

The data collection process involved the following steps: (1). Downloading the dataset from a reliable public repository. (2). Cleaning the data to address anomalies and missing values. (3). Splitting the dataset into training and testing sets, with 70% allocated for training and 30% for testing. (4). Applying data preprocessing, such as normalization and standardization, to ensure consistency in the analysis [17].

2.3 Classification

Artificial Neural Networks (ANN) is a machine learning algorithm inspired by the workings of the human brain. It processes data and generates predictions using layers of neurons (input, hidden, and output). ANN is highly effective in recognizing non-linear and complex patterns in health datasets [18], [19]. It excels in analyzing high-dimensional data but may be prone to overfitting without the use of regularization techniques. Research has demonstrated that combining ANN with SMOTE improves the accuracy of obesity predictions for minority classes [20].

Support Vector Machines (SVM) employees supervised learning to address classification and regression tasks. The algorithm identifies the optimal hyperplane that separates classes within the data. SVM is particularly effective for high-dimensional datasets and those with complex distributions [21], [22]. However, its performance is sensitive to kernel parameter selection, making it more challenging to optimize compared to other algorithms. Applying SMOTE has been shown to enhance SVM performance on imbalanced datasets [23], [24].

K-Nearest Neighbors (KNN) is an instance-based algorithm that classifies new data points by evaluating their proximity to the K nearest neighbors in the feature space. It is simple and effective for small and uniformly distributed datasets [25], [26]. However, KNN is less stable when applied to large or noisy datasets. Recent studies have shown that preprocessing techniques, such as normalization and SMOTE, can significantly improve KNN's performance in obesity prediction [27], [28].

Random Forest employs an ensemble approach using multiple decision trees. It is renowned for its robustness in handling large datasets with numerous features [29], [30], [31]. The algorithm is naturally resistant to overfitting due to its aggregation of results from individual trees. Studies indicate that integrating Random Forest with SMOTE enhances its sensitivity in detecting minority obesity classes [32].

The data analysis was conducted in four stages: (1). Data Preprocessing: Preparing the dataset by applying normalization, standardization, and SMOTE. (2). Artificial Intelligent Algorithm Implementation: Developing predictive models using ANN, SVM, KNN, and Random Forest. (3). Model Evaluation: Assessing algorithm performance using metrics such as accuracy, precision, recall, and F1-score, both before and after applying SMOTE. (4). Performance Comparison: Comparing the results of the algorithms to identify the best model for handling imbalanced datasets [10].

3. Results and Discussions

This study aims to evaluate the effectiveness of machine learning algorithms Artificial Neural Network (ANN), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest in predicting obesity status. The analysis was conducted on imbalanced datasets, both before and after the application of the Synthetic Minority Over-sampling Technique (SMOTE) as a data balancing method. The study presents the performance of each algorithm based on several evaluation metrics, including accuracy, precision, recall, and F1-score.

The discussion focuses on interpreting the findings and aligning them with the objectives of the study. It also examines the relationships between the results, explores generalizability, and identifies the potential limitations. This approach provides in-depth insights into the impact of SMOTE on the performance of machine learning algorithms and its relevance in detecting obesity cases.

The following sections provide a detailed presentation of the study's results and discussion, based on the data analysis performed.

3.1 Results

The results of this study systematically evaluate and compare the performance of Artificial Neural Network (ANN), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest algorithms in predicting obesity status, both before and after the application of SMOTE for data balancing. The evaluation was conducted using metrics such as accuracy, recall, precision, and F1-score.

The findings, presented in Table 1, demonstrate a significant improvement in algorithm performance following the implementation of SMOTE. This highlights the effectiveness of data balancing in enhancing predictive accuracy, particularly for minority classes.

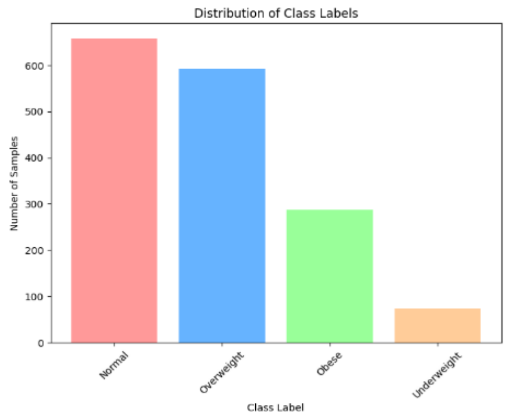

Figure 2. Distribution of Class Label Before Using SMOTE

Table 1. Classification Results Before Applying SMOTE

Classifier	Accuracy	Precision	Recall	F1-Score
K-Nearest Neighbors	0.6925	0.6902	0.6925	0.6909
Support Vector Machine	0.7732	0.7789	0.7732	0.7712
Random Forest	0.8509	0.8528	0.8509	0.8509
Artificial Neural Network	0.7763	0.7821	0.7763	0.7752

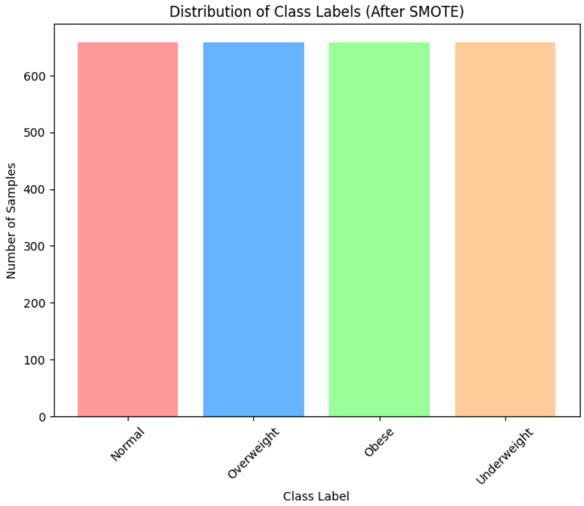

Figure 3. Distribution of Class Label After Using SMOTE

Table 2. Classification Results After Applying SMOTE

Classifier	Accuracy	Precision	Recall	F1-Score
K-Nearest Neighbors	0.8197	0.8145	0.8197	0.8145
Support Vector Machine	0.8444	0.8432	0.8444	0.8435
Random Forest	0.9222	0.9235	0.9222	0.9216
Artificial Neural Network	0.8975	0.8968	0.8975	0.8965

Based on Table 1 and Table 2, the accuracy of all algorithms improved after applying SMOTE. Random Forest achieved the highest accuracy at 92%, up from 85% before SMOTE. ANN showed a significant increase from 77% to 89%. Although KNN demonstrated improvement, it maintained the lowest accuracy among the algorithms, reaching 81% after SMOTE.

Recall exhibited the most substantial improvement, as SMOTE directly targeted the minority class (obesity cases). ANN recorded the highest recall increase, from 77% to 89%. Random Forest had the highest overall recall post-SMOTE, achieving 92%.

Precision also improved across all algorithms, reflecting a reduction in false positive predictions after SMOTE. Random Forest maintained the highest precision at 92%, followed by ANN at 89%. SVM and KNN also showed significant precision enhancements.

F1-score, which balances precision and recall, increased for all algorithms. Random Forest led with an F1-score of 92% after SMOTE, followed by ANN at 89% and SVM at 84%. KNN remained the lowest with an F1-score of 81%.

These results demonstrate that SMOTE effectively enhanced the predictive capabilities of all algorithms, particularly in detecting the minority obesity class. Random Forest emerged as the best performing algorithm overall, while ANN also achieved notable improvements, underscoring the value of SMOTE in addressing imbalanced datasets.

3.2 Discussions

This study highlights the importance of addressing data imbalance in processing datasets using the Synthetic Minority Over-sampling Technique (SMOTE) to enhance the performance of machine-learning algorithms in predicting obesity. SMOTE proved effective in improving recall, a key metric for detecting minority cases such as obesity. By generating synthetic data for the minority class, SMOTE creates a more balanced data distribution, reduces bias toward the majority class and enables algorithms to learn more effectively. Notably, SMOTE significantly improved the performance of the K-Nearest Neighbors (KNN) algorithm, with recall increasing by 13%, from 69% to 81%.

Among the tested algorithms, Random Forest emerged as the top performer, excelling across all metrics both before and after applying SMOTE. Their ability to handle complex datasets with diverse features makes them a reliable choice for obesity prediction. Meanwhile, KNN showed significant improvements in recall and F1-score following the SMOTE application. However, its overall accuracy remained lower than that of the Random Forest, likely owing to its sensitivity to overfitting and the need for precise parameter tuning. Despite this, the KNN demonstrated strong capabilities in capturing nonlinear patterns in health data.

Support Vector Machine (SVM) maintained stable performance, with significant improvements in recall and precision after SMOTE. The hyperplane approach used by SVM in high-dimensional spaces made it effective for complex datasets. The increase in precision following SMOTE indicates that SVM is suitable for scenarios where reducing false positives is critical. In contrast, KNN exhibited more moderate improvements compared to the other algorithms. KNN faced limitations in handling noise and high-dimensional datasets, though SMOTE still contributed to enhanced precision and recall.

The findings also emphasize SMOTE as an essential preprocessing step to improve algorithm capability in detecting obesity cases. Random Forest stands out as the best choice for applications requiring high accuracy and recall, such as clinical decision support tools. However, its computational complexity could pose challenges in resource-limited settings. SVM and ANN offer viable alternatives, balancing performance and computational efficiency, making them suitable for similar applications.

Despite the demonstrated effectiveness of SMOTE, several limitations of this study should be noted. The demographic composition of the dataset may limit the generalizability of the findings. Future research should consider expanding the dataset to include a more diverse population. Additionally, integrating SMOTE with advanced feature selection techniques could further enhance model performance.

With these findings, this study provides valuable insights into the impact of SMOTE on machine learning algorithm performance in predicting obesity while offering practical guidance for applications in healthcare.

4. Conclusions

This study evaluates the effectiveness of machine learning algorithms in predicting obesity status using an imbalanced dataset, both before and after the application of SMOTE. The results indicate that SMOTE significantly improves the performance of Artificial Neural Network (ANN), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest, particularly in terms of recall and F1-score, demonstrating its success in detecting the minority obesity class. These findings are consistent with previous studies that also reported improved classification model performance following the application of SMOTE in healthcare datasets. This study extends prior research by conducting a comparative analysis of different algorithms and affirming that the choice of algorithm influences the extent to which SMOTE enhances model performance.

Among the tested algorithms, Random Forest exhibited the best performance, achieving 92% accuracy and 92% recall after SMOTE application. This superior performance can be attributed to Random Forest's ability to handle complex datasets with diverse features and its high resistance to overfitting. In contrast, although KNN showed performance improvement, it remained the least effective algorithm. This could be due to KNN's heavy reliance on data distribution, which can lead to performance degradation in high-dimensional datasets.

Despite the promising results, several limitations of this study should be acknowledged. The dataset used has demographic composition constraints, limiting the generalizability of the findings to broader populations. Additionally, the limited sample size may affect the stability of the experimental results. Therefore, future research should incorporate larger and more geographically and demographically diverse datasets to ensure the validity of the findings.

Furthermore, although a substantial increase in accuracy and recall was observed after SMOTE application, this study did not assess the statistical significance of the performance differences. Statistical tests, such as t-tests or confidence intervals, should be conducted in future studies to confirm that performance improvements are not due to random variation.

As a direction for future research, it is recommended to explore alternative balancing techniques, such as cost-sensitive learning or deep learning with synthetic data augmentation, to enhance obesity prediction. Additionally, integrating SMOTE with ensemble-based or hybrid learning approaches could be a potential solution for improving model generalization in more complex datasets.

References

[1] A. N. M. S. Islam, H. Sultana, M. Nazmul Hassan Refat, Z. Farhana, A. Abdulbasah Kamil, and M. Meshbahur Rahman, "The global burden of overweight-obesity and its association with economic status, benefiting from STEPs survey of WHO member states: A meta-analysis," Oct. 01, 2024, *Elsevier Inc.* doi: 10.1016/j.pmedr.2024.102882.

[2] M. M. Ali, S. Parveen, V. Williams, R. Dons, and G. I. Uwaifo, "Cardiometabolic comorbidities and complications of obesity and chronic kidney disease (CKD)," Jun. 01, 2024, *Elsevier B.V.* doi: 10.1016/j.jcte.2024.100341.

[3] E. S. Tee and S. H. Voon, "Combating obesity in Southeast Asia countries: Current status and way forward," Sep. 01, 2024, *KeAi Communications Co.* doi: 10.1016/j.glohj.2024.08.006.

[4] T. Khater, H. Tawfik, and B. Singh, "Explainable artificial intelligence for investigating the effect of lifestyle factors on obesity," *Intelligent Systems with Applications*, vol. 23, Sep. 2024, doi: 10.1016/j.iswa.2024.200427.

[5] A. Agarwal, G. Singh, S. Jain, and P. Mittal, "Beyond boundaries: Charting the frontier of healthcare with big data and ai advancements in pharmacovigilance," *Health Sciences Review*, vol. 14, p. 100214, Mar. 2025, doi: 10.1016/j.hsr.2025.100214.

[6] C. Bunkhumpornpat, E. Boonchieng, V. Chouvatut, and D. Lipsky, "FLEX-SMOTE: Synthetic over-sampling technique that flexibly adjusts to different minority class distributions," *Patterns*, Nov. 2024, doi: 10.1016/j.patter.2024.101073.

[7] A. X. Wang, V. T. Le, H. N. Trung, and B. P. Nguyen, "Addressing imbalance in health data: Synthetic minority oversampling using deep learning," *Comput Biol Med*, vol. 188, Apr. 2025, doi: 10.1016/j.compbiomed.2025.109830.

[8] S. M. Ganie, B. B. Reddy, H. K, and M. Rege, "An investigation of ensemble learning techniques for obesity risk prediction using lifestyle data," *Decision Analytics Journal*, vol. 14, Mar. 2025, doi: 10.1016/j.dajour.2024.100539.

[9] D. Cahn *et al.*, "Novel application of deep learning to evaluate conversations from a mental health text support service," *Natural Language Processing Journal*, vol. 9, p. 100119, Dec. 2024, doi: 10.1016/j.nlp.2024.100119.

[10] T. B. Zannah, S. I. Tonni, M. A. Sheakh, M. S. Tahosin, A. H. Sarower, and M. Begum, "Comparative performance analysis of ensemble learning methods for fetal health classification," *Inform Med Unlocked*, vol. 56, Jan. 2025, doi: 10.1016/j.imu.2025.101656.

[11] D. Panteli *et al.*, "Artificial intelligence in public health: promises, challenges, and an agenda for policy makers and public health institutions," May 01, 2025, *Elsevier Ltd*. doi: 10.1016/S2468-2667(25)00036-2.

[12] A. O. Widodo, B. Setiawan, and R. Indraswari, "Machine Learning-Based Intrusion Detection on Multi-Class Imbalanced Dataset Using SMOTE," in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 578–583. doi: 10.1016/j.procs.2024.03.042.

[13] M. F. Shahzad, S. Xu, W. M. Lim, X. Yang, and Q. R. Khan, "Artificial intelligence and social media on academic performance and mental well-being: Student perceptions of positive impact in the age of smart learning," *Heliyon*, vol. 10, no. 8, Apr. 2024, doi: 10.1016/j.heliyon.2024.e29523.

[14] U. Hasanah, A. M. Soleh, and K. Sadik, "Effect of Random Under sampling, Oversampling, and SMOTE on the Performance of Cardiovascular Disease Prediction Models," *Jurnal Matematika, Statistika dan Komputasi*, vol. 21, no. 1, pp. 88–102, Sep. 2024, doi: 10.20956/j.v21i1.35552.

[15] L. Ezquerro, R. Coimbra, B. Bauluz, C. Núñez-Lahuerta, T. Román-Berdiel, and M. Moreno-Azanza, "Large dinosaur egg accumulations and their significance for understanding nesting behaviour," *Geoscience Frontiers*, vol. 15, no. 5, Sep. 2024, doi: 10.1016/j.gsf.2024.101872.

[16] W. Zhang, M. Xu, Y. Feng, Z. Mao, and Z. Yan, "The Effect of Procrastination on Physical Exercise among College Students—The Chain Effect of Exercise Commitment and Action Control," *International Journal of Mental Health Promotion*, vol. 26, no. 8, pp. 611–622, 2024, doi: 10.32604/ijmhp.2024.052730.

[17] A. Ebadi, M. Kaur, and Q. Liu, "Hyperparameter optimization and neural architecture search algorithms for graph Neural Networks in cheminformatics," *Comput Mater Sci*, vol. 254, May 2025, doi: 10.1016/j.commatsci.2025.113904.

[18] H. Liu, H. Gu, J. Li, Y. Fang, S. Yang, and G. Liang, "Evaluating the relationship between environmental chemicals and obesity: Evidence from a machine learning perspective," *Ecotoxicol Environ Saf*, vol. 300, p. 118457, Jul. 2025, doi: 10.1016/j.ecoenv.2025.118457.

[19] Y. Zhang, F. Li, and S. Huang, "Data preprocessing and its impact on AI performance," *IEEE Access*, vol. 11, pp. 134567–134578, 2023.

[20] H. Matsukawa, Y. Miyagi, and K. Otake, "ANN-based estimation of pure-component parameters of PC-SAFT equation of state using quantum chemical data," *Fluid Phase Equilib*, vol. 596, Sep. 2025, doi: 10.1016/j.fluid.2025.114444.

[21] F. Y. P. Assalé, A. F. A. Kouao, and M. T. Kessé, "Machine learning and neural networks in predicting grain-size of sandy formations," *Results in Earth Sciences*, vol. 3, p. 100084, Dec. 2025, doi: 10.1016/j.rines.2025.100084.

[22] M. A. El-Omairi, M. El Garouani, and A. El Garouani, "Enhanced lithological mapping via remote sensing: Employing SVM, random trees, ANN, with MNF and PCA transformations," *Egyptian Journal of Remote Sensing and Space Science*, vol. 28, no. 1, pp. 34–52, Mar. 2025, doi: 10.1016/j.ejrs.2024.12.001.

[23] N. S. Bajaj, A. D. Patange, R. Jegadeeshwaran, S. S. Pardeshi, K. A. Kulkarni, and R. S. Ghatpande, "Application of metaheuristic optimization based support vector machine for milling cutter health monitoring," *Intelligent Systems with Applications*, vol. 18, May 2023, doi: 10.1016/j.iswa.2023.200196.

[24] D. J. Maulana, S. Saadah, and P. E. Yunanto, "54-61 Data in Classifying Financial Distress Companies using SVM and Naïve Bayes," *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 10, no. 1, pp. 54–61, 2024, doi: 10.29207/resti.v8i1.5150.

[25] F. J. Gomez-Gil, V. Martínez-Martínez, R. Ruiz-Gonzalez, L. Martínez-Martínez, and J. Gomez-Gil, "Vibration-based monitoring of agro-industrial machinery using a k-Nearest Neighbors (kNN) classifier with a Harmony Search (HS) frequency selector algorithm," *Comput Electron Agric*, vol. 217, Feb. 2024, doi: 10.1016/j.compag.2023.108556.

[26] S. A. More and A. V Kachavimath, "SDN Intrusion Detection using Meta-Heuristic Optimization and K-Nearest Neighbors Classifier," *Procedia Comput Sci*, vol. 260, pp. 1137–1144, 2025, doi: 10.1016/j.procs.2025.03.299.

[27] P. A. Jusia, A. Rahim, H. Yani, and J. Jasmir, "Improving Performance of KNN and C4.5 using Particle Swarm Optimization in Classification of Heart Diseases," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 3, pp. 333–339, Jun. 2024, doi: 10.29207/resti.v8i3.5710.

[28] Wikky Fawwaz Al Maki, Khaidir Mauladan, and Indra Bayu Muktyas, "Date Fruit Classification using K-Nearest Neighbor with Principal Component Analysis and Binary Particle Swarm Optimization," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 6, pp. 1456–1463, Dec. 2023, doi: 10.29207/resti.v7i6.4839.

[29] M. Kasahun and A. Legesse, "Machine learning for urban land use/ cover mapping: Comparison of artificial neural network, random forest and support vector machine, a case study of Dilla town," *Heliyon*, vol. 10, no. 20, Oct. 2024, doi: 10.1016/j.heliyon.2024.e39146.

[30] V. Distefano, M. Palma, and S. De Iaco, "Multi-class random forest model to classify wastewater treatment imbalanced data," *Socioecon Plann Sci*, vol. 95, Oct. 2024, doi: 10.1016/j.seps.2024.102021.

[31] B. R. Ramos Collin *et al.*, "Random forest regressor applied in prediction of percentages of calibers in mango production," *Information Processing in Agriculture*, 2024, doi: 10.1016/j.inpa.2024.12.002.

[32] Y. Zhao and C. Teng, "Classification of soil layers in Deep Cement Mixing using optimized random forest integrated with AB-SMOTE for imbalance data," *Comput Geotech*, vol. 179, Mar. 2025, doi: 10.1016/j.compgeo.2024.106976.