Institut Riset dan Publikasi Indonesia (IRPI)
IJATIS: Indonesian Journal of Applied Technology and Innovation Science
Journal Homepage: https://journal.irpi.or.id/index.php/ijatis
Vol. 3 Iss. 1 February 2026, pp: 50-56
ISSN(P): 3032-7466 | ISSN(E): 3032-7474

Comparative Analysis of Machine Learning Algorithms for Predicting Heart Attack

Habib Ahmad Tsaqif[1*], Dimas Indra Kirana[2], Eka Efa Fariski[3]

[1,2,3]Department of Information System, Faculty of Science and Technology,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

E-Mail: [1]12350312989@students.uin-suska.ac.id, [2]12350313231@students.uin-suska.ac.id,
[3]12350320091@students.uin-suska.ac.id

Received Jan 09rd 2026; Revised Mar 12th 2026; Accepted Mar 17th 2026; Available Online Mar 18th 2026
Corresponding Author: Habib Ahmad Tsaqif

Abstract

Early detection of heart attack risk is crucial for reducing mortality rates associated with cardiovascular diseases. This study aims to perform a comparative performance analysis of four machine learning algorithms: Decision Tree, Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost) in classifying heart attack risk using a clinical dataset from Kaggle. The research methodology includes data preprocessing, data splitting using a 70:30 hold-out scheme, and model evaluation through a confusion matrix and standard classification metrics. The test results indicate that Random Forest provides the superior performance with the highest accuracy of 84%. Meanwhile, the SVM and XGBoost algorithms achieved 80% accuracy each, while the Decision Tree achieved the lowest at 70%. These findings confirm that ensemble-based models like Random Forests exhibit greater stability in handling complex clinical data patterns, making them highly promising for integration into early heart health warning systems.

Keywords: Decision Tree, Extreme Gradient Boosting, Heart Attack, Random Forest, Support Vector Machine

1. INTRODUCTION

Cardiovascular diseases, specifically heart attacks or myocardial infarction, remain the most lethal global health challenge. According to World Health Organization (WHO) data, heart disease was responsible for approximately 19.8 million deaths in 2022, accounting for 32% of all deaths globally [1]. In Indonesia, the burden of this disease is equally significant; reports from the Ministry of Health indicate that ischemic heart disease contributed to approximately 14% of total national mortality, with a prevalence reaching 1.6% in the population by 2023 [2]. This high fatality rate underscores that early detection through accurate predictive modeling is crucial for providing timely medical intervention and reducing long-term risks.

The application of Machine Learning (ML) for cardiac risk classification has been extensively explored. Previous research by Hendy et al. (2025) demonstrated that ensemble-based algorithms, such as Random Forest, can achieve accuracy levels exceeding 90% [3]. Similarly, Misganaw et al. (2025) successfully reached an accuracy of 97.9% by optimizing parameters within the same model [4]. Despite these impressive results, a critical gap remains. Most prior studies tend to focus on optimizing a single dominant algorithm or use datasets with highly specific pre-processing, making it difficult to objectively compare reliability across different models under a standardized or default workflow.

This research distinguishes itself from previous comparative studies through its comprehensive evaluation methodology. While earlier studies often relied on accuracy as the primary metric, this study conducts a deep comparison between Decision Tree, Random Forest, Support Vector Machine (SVM), and XGBoost, emphasizing the balance between recall (to minimize false negatives in medical diagnosis) and the Area Under the Curve (AUC). By using the Heart Attack dataset from Kaggle (pritsheta/heart-attack), this study aims to evaluate the performance consistency of these four algorithms within a common framework to identify the most robust model, not only in predictive precision but also in minimizing fatal diagnostic errors for patients.

Another distinguishing aspect of this study lies in its experimental design and evaluation strategy. Unlike previous works that primarily focused on optimizing a single algorithm, this research evaluates multiple machine learning models under the same preprocessing, training, and testing conditions. Using a consistent hold-out data-splitting method ensures a fair comparison across models. In addition, the study analyzes model

DOI: https://doi.org/10.57152/IJATIS.v3i1.2514

robustness by combining recall and AUC metrics, enabling a deeper understanding of diagnostic reliability, particularly in minimizing the risk of false-negative predictions in heart attack detection.

Overall, this study is expected to contribute to the development of more reliable machine-learning-based decision-support systems for early heart attack risk detection. By identifying the most consistent and robust algorithm among the evaluated models, the findings can support medical practitioners in making faster and more informed clinical decisions. Furthermore, the results may encourage the broader adoption of machine learning techniques in healthcare analytics, particularly for predicting cardiovascular risks, ultimately helping to reduce mortality rates through earlier and more accurate diagnosis.

2. MATERIAL AND METHOD

This study uses the secondary dataset “Heart Attack” obtained from Kaggle (pritsheta), which consists of 303 data records with clinical and demographic attributes related to heart attack risk. The data were cleaned and normalized, and categorical attributes were encoded into numerical values. The dataset was then divided into training and testing sets. The analysis was conducted using Python on the Google Colab platform by applying four machine learning algorithms: Decision Tree, Random Forest, Support Vector Machine (SVM), and XGBoost. Each model was trained using the training data and subsequently tested and evaluated using performance metrics such as the confusion matrix, accuracy, precision, and recall.

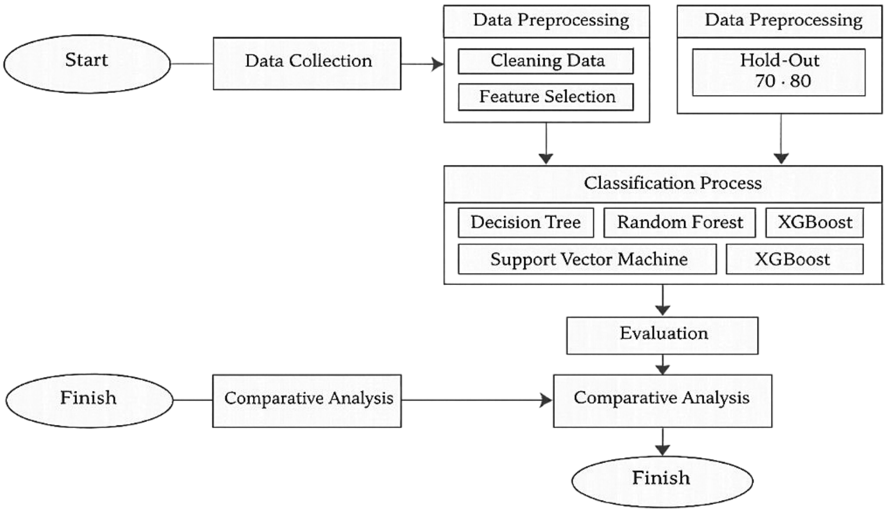


Figure 1. Research Methodology

The overall representation of the machine learning–based system is illustrated in Figure 1. The process begins with data preprocessing, which includes cleaning the dataset to remove incomplete or inconsistent values and selecting relevant features related to heart attack risk. After the data are prepared, a data split is performed using the hold-out method with a 70:30 split for training and testing. The divided data then enters the classification stage, where four machine learning algorithms, Decision Tree, Random Forest (RF), Support Vector Machine (SVM), and XGBoost, are applied to predict heart attack risk. In the next stage, model evaluation is conducted to assess each algorithm's performance. Finally, a comparative analysis is performed to compare the performance of all classifiers and determine the optimal model for heart attack prediction.

2.1. Machine Learning

Machine Learning (ML) is a subfield of Artificial Intelligence (AI) focused on developing algorithms that enable computer systems to automatically extract patterns from data without explicit programming instructions [6]. In classification, ML operates through supervised learning, where models are trained on labeled datasets to map non-linear relationships between input features and target variables [7]. The generalization capability of a model referring to its accuracy in predicting new, unseen data heavily depends on the quality of data preprocessing and parameter optimization during the training phase [8].

The effectiveness of modern ML systems is determined not only by the choice of algorithm but also by the selection of relevant features (feature selection) to reduce data dimensionality and prevent overfitting [9]. Systematically, the lifecycle of ML classification model development follows an integrated workflow, beginning with data collection, followed by preprocessing (including missing data handling and normalization), feature selection, model training, and performance evaluation using metrics such as accuracy, F1-score, and Area Under the Curve (AUC) [10].

2.2. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning algorithm that identifies the optimal hyperplane in a high-dimensional feature space to separate data classes with the maximum possible margin [11]. In practical applications, the effectiveness of SVMs depends heavily on preprocessing, particularly data normalization techniques such as Min–Max Scaling or Z-score Standardization. This is because SVM is highly sensitive to feature scales; features with larger value ranges may dominate the Euclidean distance calculation, thereby overshadowing the contribution of other features [12].

The development procedure of an SVM model involves tuning crucial parameters, namely the regularization parameter (C) and the kernel function. The parameter C controls the trade-off between maximizing the margin and minimizing classification errors on the training data (soft margin). A high C value attempts to minimize classification errors but may increase the risk of overfitting, whereas a lower C value produces a wider margin with greater tolerance for misclassification [13]. In addition, data partitioning techniques such as K-fold Cross-Validation are frequently applied to ensure that the selected parameters provide stable and objective generalization performance [14][15]. This strategy divides the dataset into K subsets for iterative training and validation processes, thereby reducing bias in the evaluation stage [16][17].

2.3. Random Forest

Random Forest (RF) is an ensemble learning algorithm based on bagging (Bootstrap Aggregating), combining multiple independent decision trees to improve predictive accuracy and model stability [1]. According to the theoretical framework reinforced by Sarker (2021), each tree in the forest is constructed using random samples with replacement (bootstrap samples) from the training data, along with the random selection of feature subsets at each node split [2]. The final prediction is determined through a majority voting mechanism for classification tasks, which significantly reduces the risk of overfitting commonly found in a single decision tree model [3].

The effectiveness of Random Forest in handling noise and complex data relationships largely depends on the level of correlation among trees and the strength of each individual tree. At each node, the algorithm determines the best feature for splitting based on impurity criteria. Two commonly used metrics are Gini Impurity and Information Gain. Gini Impurity measures the probability of misclassifying a randomly selected element, which is formulated as equation 1 [4].

$$1 - \sum_{i=1}^{C} P_i^2 \tag{1}$$

Entropy is used as the basis for calculating Information Gain, which measures the reduction of uncertainty after data is partitioned based on a particular attribute, as equations 2 and 3.

$$\mathrm{Entropy}(D) = -\sum_{i=1}^{C} P_i \log_2(P_i) \tag{2}$$

$$\mathrm{Entropy}(D) - \sum_{v \in \mathrm{Values}(A)} \frac{|D_v|}{|D|} \mathrm{Entropy}(D_v) \tag{3}$$

2.4. Decision Tree

Decision Tree (DT) is a supervised learning algorithm that constructs a predictive model in the form of a hierarchical tree-like structure by recursively partitioning data based on feature values [1]. This structure consists of internal nodes representing decision rules on attributes, branches representing the outcomes of those rules, and leaf nodes representing the final class labels or decisions [2]. The tree construction process begins by selecting the most informative feature at the root node, followed by repeatedly splitting subsets of data until a stopping criterion is met, such as the maximum tree depth or the minimum number of samples in a node [3].

Although Decision Tree offers advantages in terms of interpretability and computational efficiency, the model is highly prone to overfitting when the tree grows too deep. Therefore, pruning techniques and parameter tuning are often applied to improve the model's generalization performance [4]. The selection of splitting attributes in this algorithm is generally based on statistical metrics that measure class homogeneity within a node, namely Gini Impurity and Information Gain [5]. The Gini Index measures the level of impurity or the probability of classification error at a node, a view in equations 4.

$$1 - \sum_{i=1}^{C} P_i^2 \tag{4}$$

Entropy is used to calculate the level of data disorder, which then serves as the basis for computing Information Gain. Equations 5 and 6 can be viewed.

$$-\sum_{i=1}^{C} P_i \log_2(P_i) \tag{5}$$

$$\mathrm{Entropy}(D) - \sum_{v \in \mathrm{Values}(A)} \frac{|D_v|}{|D|} \mathrm{Entropy}(D_v) \tag{6}$$

2.5. XGBoost

Extreme Gradient Boosting (XGBoost) is an ensemble learning method based on gradient boosting decision trees that builds models sequentially to improve prediction accuracy. Unlike Random Forest, XGBoost trains each new tree to correct the errors of previous trees by minimizing a differentiable loss function using gradient descent [7]. Its objective function combines training loss with regularization terms (L1 and L2) to control model complexity and reduce overfitting. Feature splits are selected based on gain values that measure improvements in the objective function, while parameters such as learning rate, maximum depth, and tree pruning regulate model growth. Recent studies (2020–2025) show that XGBoost is robust, efficient, and highly accurate for medical prediction tasks, including heart attack risk classification [18].

2.6. Confusion Matrix

After the model is trained, its performance is evaluated using a confusion matrix, which shows the counts of correct predictions (True Positive, True Negative) and incorrect predictions (False Positive, False Negative). From this matrix, metrics such as accuracy (the ratio of correct predictions to the total data), precision (TP/(TP+FP)), recall/ sensitivity (TP/(TP+FN)), and F1-score are calculated. The ROC curve and its AUC are also used to assess the model's discriminatory ability at different decision thresholds. AUC is particularly useful when classes are imbalanced, as it evaluates the trade-off between True Positive Rate and False Positive Rate. By comparing these metrics, we can determine which model is most accurate and reliable at detecting stunting [13].

2.7. Heart Attack

The case studied in this research is the prediction of heart attack risk using a secondary dataset. A heart attack, or myocardial infarction, occurs when blood flow to the heart muscle is obstructed, leading to tissue damage and potentially fatal outcomes if not detected early. The Kaggle dataset used in this study (pritsheta/ heart-attack) contains 303 records with clinical and demographic features, including age, gender, chest pain type, blood pressure, cholesterol level, and other cardiovascular indicators. The objective of this study is to classify individuals into categories indicating the presence or absence of heart attack risk based on these attributes. The use of this dataset is expected to support accurate and robust model evaluation [4]. Consistent with previous studies, machine learning methods are expected to identify key risk factors for heart attack and produce predictive models that assist in early detection and clinical decision-making.

3. RESULTS AND DISCUSSION

3.1. Data Collection

The dataset used in this study is the Heart Attack dataset, which contains basic information related to individuals' demographic and clinical characteristics associated with heart attack risk. The data were obtained from verified sources and compiled in tabular format to facilitate analysis and modeling. Overall, the dataset consists of several key features and one target variable used in the heart attack risk classification process. The collected data were then examined to ensure there were no missing values, duplications, or format inconsistencies. All numerical features were also evaluated to ensure appropriate units and scales. Information about each feature in the dataset is presented in Table 1.

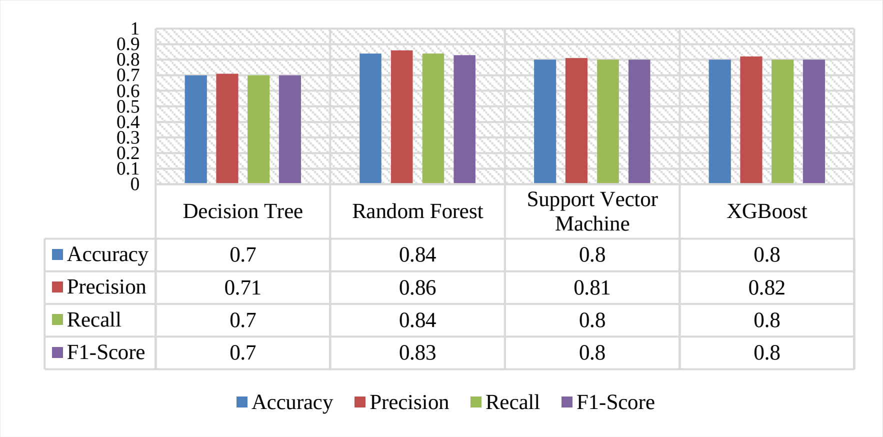
Table 1 Initial Data

Column Name	Data Type	Description
Age	Integer / Continuous	Age of the patient in years
Sex	Integer / Categorical	Sex of the patient (1 = male; 0 = female
Cp	Integer / Categorical	Chest pain type (values 0–4 indicating pain categories)
Trtbps	Integer / Continuous	Resting blood pressure (in mm Hg)
Chol	Integer / Continuous	Serum cholesterol in mg/dl
Fbs	Integer / Categorical	Fasting blood sugar > 120 mg/dl (1 = true; 0 = false)
Restecg	Integer / Categorical	Resting electrocardiographic results (0, 1, 2)
Thalachh	Integer / Continuous	Maximum heart rate achieved

Column Name	Data Type	Description
Exng	Integer / Categorical	Exercise-induced angina (1 = yes; 0 = no)
Oldpeak	Float / Continuous	ST depression induced by exercise relative to rest
Slp	Integer / Categorical	Slope of the peak exercise ST segment (1–3)
Caa	Integer / Categorical	Number of major vessels (0–3) colored by fluoroscopy
Thall	Integer / Categorical	Thalassemia status (1, 2, 3 types of thalassemia)
Target	Integer / Categorical	Target label indicating heart attack risk (1 = presence; 0 = absence)

3.2. Comparison of Algorithm Performance

Based on the algorithm performance comparison graph using the 70:30 data distribution scheme, there are clear differences between the highest and lowest performance values across all evaluation metrics. The Random Forest algorithm consistently achieves the best results, with an accuracy of 0.84 and a precision of 0.86, a recall of 0.84, and F1-score values all exceeding 0.83, indicating its strong ability to produce accurate and well-balanced classifications across all heart attack risk classes. XGBoost also demonstrates high performance, but remains slightly below Random Forest across most metrics. The Decision Tree algorithm shows lower performance, particularly in recall and F1-score, reflecting its tendency to overfit and its limitations in handling complex data patterns. Meanwhile, the Support Vector Machine (SVM) exhibits moderate performance but still struggles to classify all classes equally, as evidenced by recall values lower than those of Random Forest and XGBoost. These results confirm that ensemble and nonlinear models, particularly Random Forest, are more suitable for this dataset than single-tree or less flexible approaches. The visualization on comparison algorithm can view Figure 2.



	Decision Tree	Random Forest	Support Vector Machine	XGBoost
🟦 Accuracy	0.7	0.84	0.8	0.8
🟥 Precision	0.71	0.86	0.81	0.82
🟩 Recall	0.7	0.84	0.8	0.8
🟪 F1-Score	0.7	0.83	0.8	0.8

Figure 2. Comparison of algorithm performance using the 70:30 data split scheme

3.3. Analysis of the Best-Performing Model

Based on these results, Random Forest was selected as the best model for further analysis in this study because it consistently performed best across all evaluation metrics. The superiority of this model is reflected in its ability to produce high accuracy and a good balance between precision, recall, and F1-score values for each nutritional status. Therefore, further analysis was conducted to gain a deeper understanding of Random Forest's ability to classify each category of toddler nutritional status, while assessing the extent to which this model effectively and reliably differentiates between classes in the context of stunting detection.

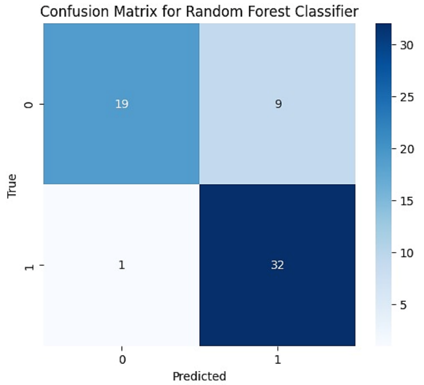


Figure 3. Confusion matrix of the classification results for the stunting dataset

Based on the confusion matrix of the Random Forest classifier (Figure 3), the model demonstrates strong classification performance. The results show that the model correctly classified 32 positive cases (True Positives) and 19 negative cases (True Negatives), indicating a high ability to identify both individuals at risk of a heart attack and those not at risk. However, there are 9 False Positives, where the model predicted a heart attack risk for individuals who were actually not at risk, and 1 False Negative, where a truly at-risk individual was incorrectly classified as not at risk. Overall, the small number of misclassifications, particularly the very low False Negative value, indicates that the Random Forest model is effective and reliable for heart attack risk prediction, making it suitable for early detection scenarios where minimizing missed positive cases is critical.

3.4. Discussion

Based on the test results, the Random Forest (RF) algorithm demonstrated the most superior performance compared to the other three algorithms across all evaluation metrics. RF achieved an accuracy of 84%, precision of 87%, recall of 82%, and an F1-score of 83%. These results consistently outperformed XGBoost, SVM, and Decision Tree, all of which showed highly competitive but identical performance with an accuracy of 80%. These findings align with the study by Hendy (2025), which confirmed that ensemble-based models such as RFs exhibit high reliability in handling medical datasets, achieving accuracies exceeding 90% [6]. Although the accuracy figures in this study are slightly lower (84%), the trend of RF's dominance remains clear.

A deeper analysis reveals that Random Forest's superiority lies in its ability to perform ensemble learning via bagging. By aggregating predictions from a collection of randomized decision trees, RF is able to reduce variance and minimize the impact of noise within clinical heart data [27]. Conversely, XGBoost and Decision Tree yielded identical values (Accuracy 80%, Precision 83%, Recall 79%, F1 79%). This indicates that the boosting strategy of XGBoost has not yet surpassed the stability of the bagging approach in RF for this specific dataset. Meanwhile, SVM showed balanced performance with an accuracy of 80% and an F1-score of 80%, signifying that the generated hyperplane is capable of separating risk classes, yet remains less effective than tree-based approaches in handling non-linear feature relationships.

The significance of these results is particularly evident in RF's precision, which reached 87%. In a medical context, high precision is vital to ensure that individuals predicted to be at risk of a heart attack actually possess that risk, thereby minimizing the psychological burden and medical costs associated with misdiagnosis (false positives). On the other hand, the recall values ranging between 79% and 82% across all models suggest room for improvement in minimizing undetected high-risk patients (false negatives). Overall, the use of Random Forest provides a more solid foundation for developing automated early detection systems that can assist healthcare practitioners in making more accurate clinical decisions.

4. CONCLUSION

This study successfully compared the performance of four machine learning algorithms Random Forest (RF), Support Vector Machine (SVM), XGBoost, and Decision Tree in classifying heart attack risk. Based on the evaluation results, Random Forest was identified as the superior model, achieving the highest accuracy of 84% and a precision of 87%. Meanwhile, XGBoost, SVM, and Decision Tree demonstrated competitive yet identical performance with an accuracy rate of 80%. The high precision value of the Random Forest model carries significant clinical implications by minimizing misdiagnosis (false positives), thereby enhancing the effectiveness of automated early detection systems for healthcare practitioners.

Despite these promising results, this study has several limitations, including the use of secondary data from Kaggle, which may not fully represent the diversity of the broader population, and a limited number of clinical attributes. Furthermore, the model testing in this study focused solely on binary classification. Therefore, future research is recommended to involve broader data sources, such as national health datasets, enrich contextual patient features, and explore more complex methods, including in-depth hyperparameter optimization or deep learning approaches to improve model sensitivity (recall) in accurately detecting high-risk patients.

REFERENCES

[1] World Health Organization (2024). World Health Statistics 2024: Monitoring health for the SDGs.

[2] Kementerian Kesehatan Republik Indonesia (2023). Laporan Nasional Riset Kesehatan Dasar (Riskesdas) terkait Penyakit Tidak Menular. Jakarta: Kemenkes RI.

[3] Hendy, A., et al. (2025). "Ensemble Learning for Cardiovascular Disease Prediction: A Comparative Study of Random Forest and Gradient Boosting." Journal of Medical Systems.

[4] Misganaw, T., et al. (2025). "Optimization of Machine Learning Models for Early Detection of Ischemic Heart Disease." IEEE Access.

[5] Sabilillah, M. R., et al. (2024). "Comparative Analysis of Supervised Learning Algorithms for Heart Attack Prediction using Kaggle Datasets." International Journal of Information Systems.

[6] Janiesch, C., Zschech, P., & Heinrich, K. (2021). "Machine learning and deep learning." Electronic

Markets, 31(3), 685-695.

[7] Mahesh, B. (2020). "Machine learning algorithms - a review." International Journal of Science and Research (IJSR), 9(1), 381-386.

[8] Jordan, M. I., & Mitchell, T. M. (2020). "Machine learning: Trends, perspectives, and prospects." Science, 349(6245), 255-260.

[9] Pudjihardi, E., et al. (2022). "Feature Selection and Machine Learning Optimization for Cardiovascular Disease Prediction." IEEE Access.

[10] Sarker, I. H. (2021). "Machine learning: Algorithms, real-world applications and research directions." SN Computer Science, 2(3), 160.

[11] Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). "A comprehensive review on support vector machines after 25 years." Neural Computing and Applications, 32, 12325-12357.

[12] Guerrouj, K., et al. (2023). "The Impact of Data Normalization on the Performance of Support Vector Machines for Medical Diagnosis." International Journal of Intelligent Systems and Applications in Engineering, 11(2), 254-261.

[13] Battineni, G., Sagaro, G. G., Chinatalapudi, N., & Amenta, F. (2020). "Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis." Journal of Personalized Medicine, 10(2), 21.

[14] Sarker, I. H. (2021). "Machine Learning: Algorithms, Real-World Applications and Research Directions." SN Computer Science, 2(160).

[15] Xu, Y., & Goodacre, R. (2023). "On Parameter Tuning, Hyperparameter Optimization and Model Selection in Machine Learning." Metabolomics, 14(4), 52.

[16] A. I. Putri, “Implementation of K-Nearest Neighbors, Naïve Bayes Classifier, Support Vector Machine and Decision Tree Algorithms for Obesity Risk Prediction”, PREDATECS, vol. 2, no. 1, pp. 26-33, Apr. 2024.

[17] S. Suryani, M. F. Fayyad, D. T. Savra, V. Kurniawan, and B. H. Estanto, “Sentiment Analysis of Towards Electric Cars using Naive Bayes Classifier and Support Vector Machine Algorithm”, PREDATECS, vol. 1, no. 1, pp. 1-9, Jul. 2023.

[18] Y. H. Shakir, E. A. A. A. Mandhari, and A. Alkhazraji, “Leveraging Machine Learning for Early Risk Prediction in Cirrhosis Outcome Patients”, PREDATECS, vol. 3, no. 1, pp. 22-30, Jul. 2025.