IT Journal Research and Development (ITJRD)
Vol.10, No.2, March 2026, E-ISSN : 2528-4053 | P-ISSN : 2528-4061
DOI : 10.25299/itjrd.2025.24703

Enhancing Early Heart Disease Detection Through Comparative Analysis of Random Forest, Decision Tree, and K-NN Models

Kelvin Leonardi Kohsasih[1], Daniel Smith Sunario[2], Alvin[3], Fedro Laurendio[4]
Departement of Informatics Engineering, STMIK TIME
Kelvinleonardi@stmik-time.ac.id, danielsunario5@gmail.com[2], alvin181849@gmail.com[3], fedrolaurendioo@gmail.com[4]

Article Info

Article history:

Received Aug 25, 2025
Revised Sep 19, 2025
Accepted Nov 25, 2025

Keyword:

Cardiovascular Disease
Machine Learning
Decision Tree
K-Nearest Neighbors
Random Forest

ABSTRACT

Heart disease is a leading cause of mortality worldwide and its rising prevalence challenges health systems. This study evaluates Decision Tree, k Nearest Neighbors, and Random Forest using the Heart Failure Prediction Dataset from Kaggle with 918 records and 12 demographic, clinical, and lifestyle features. The target variable indicates the presence of heart disease. Data preprocessing included cleaning, transformation, and scaling. Hyperparameters were tuned with stratified five fold cross validation to prevent data leakage. Performance was assessed using accuracy, precision, recall, F1 score, ROC AUC, PR AUC, Matthews Correlation Coefficient, and Brier score each estimated with 95 percent confidence intervals via bootstrap. k Nearest Neighbors achieved the highest accuracy at 90.2 percent, followed by Random Forest at 87.5 percent and Decision Tree at 85.3 percent. Calibration and decision curve analyses indicated that k Nearest Neighbors and Random Forest provided better calibrated probabilities and higher clinical utility across plausible thresholds. The study offers a reproducible evaluation pipeline and supports the use of machine learning for early detection of heart disease while encouraging future work on larger datasets and more advanced models.

Corresponding Author:

Kelvin Leonardi Kohsasih
Departement of Informatics Engineering
STMIK TIME
Jalan Merbabu No.32 aa-bb, Medan, Indonesia
Email : kelvinleonardi@stmik-time.ac.id

1. INTRODUCTION

Heart disease remains one of the leading causes of mortality worldwide, and epidemiological evidence consistently shows that its prevalence continues to rise, establishing it as a critical global public health concern [1]. This upward trend is reflected in estimates from the Global Burden of Disease reported by the World Heart Federation, which indicate that the number of deaths from CVD worldwide increased from 12.1 million in 1990 to approximately 20.5 million in 2021 [2]. While further evidence from the Atlas of Epidemiology by the Heart Failure Association (2019) underscores the substantial burden of heart disease across all continents, noting that both the number of patients and the associated management workload continue to increase [3].

Conventional diagnostic procedures for heart disease are often requires considerable time, cost, and medical expertise to evaluate complex parameters, which makes early detection challenging

and highlights the need for more efficient alternatives. Against this background, advances in medicine and the availability of large-scale clinical data have opened new opportunities for developing decision support systems [4], where automated software can facilitate parameter evaluation, training, and testing to assist physicians in producing more accurate and efficient clinical decisions [5], [6].

In particular, the integration of Artificial Intelligence (AI) and Machine Learning (ML) technologies into healthcare has introduced innovative pathways that enhance both the speed and accuracy of disease detection and diagnosis [7], [8]. enabling ML-based approaches to uncover complex patterns in clinical data, reduce the risk of misdiagnosis, and support earlier identification of heart disease conditions [9]. In turn, these methods provide data-driven recommendations that may reduce mortality associated with delayed detection.

Building on this potential, numerous prior studies have investigated the use of machine learning approaches for the detection and diagnosis of heart disease. Yang et al. (2020) developed a cardiovascular disease prediction model for a high-risk population in eastern China using the Random Forest algorithm. The proposed model achieved an AUC of 0.787, outperforming other machine learning methods such as CART, Naïve Bayes, Bagged Trees, and AdaBoost, and demonstrated significant improvement over the multivariate regression benchmark [10]. More recently, Barus et al. (2023) applied machine learning to classify heart failure patients using the Naive Bayes algorithm. The results indicated that the proposed approach achieved a classification accuracy of 74.58%, a precision of 97.67%, a sensitivity of 75%, and an AUC of 0.857, indicating excellent performance and potential application as an early warning system for individuals at risk of heart failure [11].

In a related study, Iwendi et al. (2023) proposed an advanced machine learning model for the early detection of cardiovascular disease using the XGBoost algorithm. The model achieved an accuracy of 91.67% and outperformed other evaluated algorithms, including Random Forest, Logistic Regression, and K-Nearest Neighbors. These results further highlight the effectiveness of machine learning in enhancing predictive accuracy and supporting timely clinical decision-making for cardiovascular disease management [12]. Similarly, another study developed a Decision Tree model for heart disease prediction using a dataset of 1,025 records, split into 70% for training and 30% for testing. The model identified chest pain as the most significant indicator of heart disease and achieved an accuracy of 90.5%, demonstrating its potential for early detection and clinical decision support [13].

Although prior studies have demonstrated the potential of machine learning in detecting heart disease, several gaps remain evident. Most existing research has primarily relied on traditional algorithms such as Naïve Bayes, Random Forest, or Decision Trees, which, while effective, often lack robustness across diverse datasets and offer limited interpretability for clinical decision-making. Building upon these limitations, the present study aims to develop a more comprehensive predictive framework that not only achieves high accuracy but also ensures interpretability and adaptability for clinical practice. The novelty of this research lies in its integration of advanced machine learning techniques with rigorous validation strategies, offering a balanced approach that emphasizes both predictive performance and practical usability. Addressing this gap is urgent, given the global rise in heart disease prevalence and mortality, highlighting the pressing need for reliable, accurate, and clinically applicable early detection systems to support timely interventions and improve patient outcomes.

2. RESEARCH METHOD

This study focuses on the detection of heart disease using various Machine Learning (ML) methods with the Heart Failure Prediction Dataset from Kaggle, consisting of 918 records. The research process includes data collection, preprocessing, splitting data into training and testing sets, model selection, training and testing, evaluation with performance metrics such as accuracy, precision, recall, F1-score, and AUC, as well as the implementation and analysis of the best-performing model to assess its predictive capability for early detection and clinical decision support.

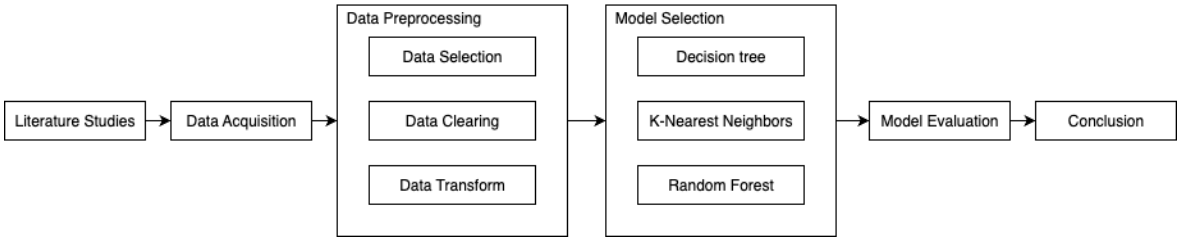

Figure 1. Research Methodology Flow Diagram

2.1. Data Acquisition

The data used in this study were obtained from the Heart Failure Prediction Dataset available on the Kaggle repository. This dataset consists of 918 observations with 12 attributes representing various clinical and demographic factors that may be associated with the occurrence of stroke. These characteristics include key variables such as Age, Sex, ChestPainType, RestingBP, Cholesterol, FastingBS, RestingECG, MaxHR, ExerciseAngina, Oldpeak, and ST_Slope, along with the HeartDisease indicator as the target label [14]. Out of the total records, 410 patients (44.7%) were labeled as not having heart disease, while 508 patients (55.3%) were labeled as having heart disease, indicating a relatively balanced class distribution. The distribution of patients with and without heart disease is illustrated in Figure 2 using a pie chart, which highlights the slightly higher proportion of positive cases. With its diverse range of variables and balanced classes, the dataset is considered highly suitable for predictive modeling in this research.

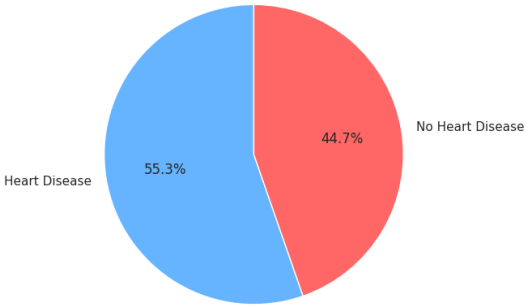

Figure 2. Distribution of Patients With and Without Heart Disease

2.2. Data Preprocessing

Data preprocessing is a critical step in transforming raw datasets into a format suitable for machine learning models. This stage not only involves data cleaning and transformation but may also include augmentation techniques to address common issues such as overfitting and class imbalance [15]. Recent evaluations of large-scale data cleaning tools have shown that their effectiveness largely depends on the application domain and the volume of data being analyzed [16]. In the semiconductor manufacturing domain, for example, the integration of methods such as missing value handling, dimensionality reduction, resampling, and feature scaling has proven effective in preventing data leakage and improving predictive accuracy [17]. Furthermore, recent studies emphasize that the impact of preprocessing on machine learning performance can be quantitatively measured through evaluation metrics, highlighting its crucial role in ensuring the reliability of predictive data-driven systems [18].

2.3. Model Development

In this study, three machine learning algorithms were selected: Decision Tree (DT), k-Nearest Neighbors (k-NN), and Random Forest (RF). The Decision Tree algorithm was chosen for this study due to its transparency and ease of interpretation, making it particularly effective for clinical decision-making in healthcare applications [19]. The k-Nearest Neighbors (k-NN) algorithm was

employed because of its ability to effectively model local neighborhood structures and handle non-linear decision boundaries in patient data [20] . Random Forest was incorporated as a robust ensemble method that reduces overfitting and noise through aggregation of multiple trees, resulting in improved generalization and accuracy [21] .

1. Decision Tree (DT)

The Decision Tree (DT) algorithm is a supervised learning technique that has gained significant attention due to its transparency, interpretability, and simplicity in both classification and regression tasks. Its tree-like structure partitions datasets into smaller subsets, with internal nodes representing feature tests and leaf nodes representing decision outcomes, thereby providing explainable decision rules that are easy to interpret, especially in sensitive fields such as healthcare[19], [22].

One of the strengths of DT lies in its interpretability, where each decision path corresponds to a combination of feature-based conditions. This makes DT particularly effective in healthcare decision support, allowing medical practitioners to validate automated predictions against clinical reasoning [23]. In addition, DTs have been widely adopted because of their adaptability and scalability in handling structured, high-dimensional, and heterogeneous data [24].

The construction of a DT commonly relies on information theory concepts, where the best attribute for splitting is chosen using measures such as information gain. The entropy $E(S)$ of a dataset S is defined as:

$$E(S) = -\sum_{i=1}^{c} p_i \log_2(p_i) \quad (1)$$

where p_i is the proportion of instances belonging to class i, and c is the total number of classes. The information gain of an attribute A is then calculated as:

$$Gain(S, A) = E(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} E(S_v) \quad (2)$$

where S_v is the subset of S for which attribute A has value v. The attribute with the highest information gain is selected as the splitting criterion, enabling the tree to progressively reduce uncertainty in classification.

2. K-Nearest Neighbors (KNN)

The k-Nearest Neighbors (KNN) algorithm is a simple yet widely used supervised learning method, primarily applied in classification and regression tasks, that operates on the principle of local similarity by classifying a query sample according to the majority label of its k nearest training samples, determined by a distance metric such as Euclidean, Manhattan, or Minkowski distance. [25]. The effectiveness of KNN is strongly influenced by the selection of the parameter k and the distance metric, where a very small k may cause overfitting and make the model sensitive to noise, while setting k too large can lead to underfitting with overly smoothed decision boundaries [26].

3. Random Forest (RF)

Random Forest (RF) algorithm is an ensemble learning method that builds upon the principle of bagging (bootstrap aggregating), where multiple decision trees are trained on different bootstrap samples of the dataset and combined to form a more robust predictive model. At each node of a tree, only a random subset of features is considered for splitting, which introduces additional diversity among the trees and reduces the correlation between them [27]. The final prediction is obtained by aggregating the outputs of all trees using majority voting in classification tasks or averaging in regression tasks. Formally, for a query instance q , the RF classifier prediction is given by:

$$\hat{y}(q) = \arg\max_{c} \sum_{b=1}^{B} 1\{T_b(q) = c\} \quad (3)$$

where B is the number of trees, $T_b(q)$ denotes the prediction of the b-th tree, and $1\{\cdot\}$ is the indicator function. For regression tasks, the prediction is expressed as:

$$\hat{y}(q) = \frac{1}{B}\sum_{b=1}^{B} T_b(q) \quad (4)$$

The main advantage of RF lies in its ability to reduce variance and overfitting compared to individual decision trees, while maintaining strong predictive performance across diverse datasets. It can effectively handle both numerical and categorical features, manage missing values, and remain robust against noisy or imbalanced data [28].

2.4. Model Evaluation

Model evaluation is a crucial stage in assessing the performance and reliability of machine learning models for heart disease prediction. In clinical contexts, this process becomes even more critical, as the costs of misclassification can lead to significant consequences for patient outcomes. To ensure a comprehensive assessment of predictive capability, this study employed several evaluation metrics, including the confusion matrix (CM), precision, recall, F1-score, and accuracy, in line with recommendations from recent medical AI literature [29], [30], [31].

A confusion matrix is a structured table that illustrates how predicted class labels correspond to the actual labels. It is composed of four elements: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Based on these values, various performance metrics can be calculated. Accuracy reflects the proportion of correctly classified samples out of the total dataset, while precision indicates the proportion of true positive cases among all predicted positives. Recall (or sensitivity) measures the proportion of actual positive cases that were correctly identified. The F1-Score, defined as the harmonic mean of precision and recall, provides a balanced assessment between the two. These measures are commonly used in medical AI research to ensure comprehensive evaluation of classification models. The mathematical formulations of accuracy, precision, recall, and F1-score are provided in Equations (5)–(8).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (8)$$

Additionally, AUC (Area Under the ROC Curve) evaluates the trade-off between true positive and false positive rates across thresholds. A higher AUC indicates better discrimination ability, which is crucial in clinical predictive settings

3. RESULTS AND ANALYSIS

The dataset used in this study is the Heart Failure Prediction Dataset obtained from Kaggle, consisting of 918 records and 12 attributes, including demographic, clinical, and lifestyle-related features that may influence cardiovascular disease risk. Several preprocessing steps were carried out prior to model training, including data cleaning to handle missing values, feature selection to retain relevant attributes, and data transformation to ensure all features were in numerical form. Exploratory Data Analysis (EDA) was conducted to examine the statistical properties and relationships among variables. The correlation heatmap, shown in Figure 3, illustrates the degree of association between clinical attributes and the target variable HeartDisease. For example, Age (r = 0.28) and Oldpeak (r = 0.40) exhibited positive correlations with heart disease, while MaxHR (r = –0.40) and Cholesterol (r = –0.23) showed negative correlations. These findings confirm that no single feature strongly dominates the prediction, emphasizing the importance of multivariate modeling in heart disease detection.

Figure 3. Correlation of The Features

After preprocessing, the dataset was partitioned into training and testing subsets with an 80:20 ratio. The training set was employed to build and optimize the machine learning models, while the testing set was utilized to assess the generalization capability of the trained models. Since KNN and ensemble methods are highly sensitive to differences in feature scales, feature scaling was performed using standardization to ensure that all variables contributed equally to distance calculations and splitting criteria during model training.

The first model evaluated was the Decision Tree (DT) classifier. To enhance performance and mitigate overfitting, hyperparameter tuning was performed on max_depth and min_samples_split. The validation curves in Figure 4 illustrate how these parameters influence model accuracy. Increasing the maximum depth improved training accuracy but widened the gap with validation accuracy, indicating overfitting in deeper trees. In contrast, larger values of min_samples_split stabilized validation accuracy but reduced training performance. An optimal balance was therefore selected to maintain generalization.

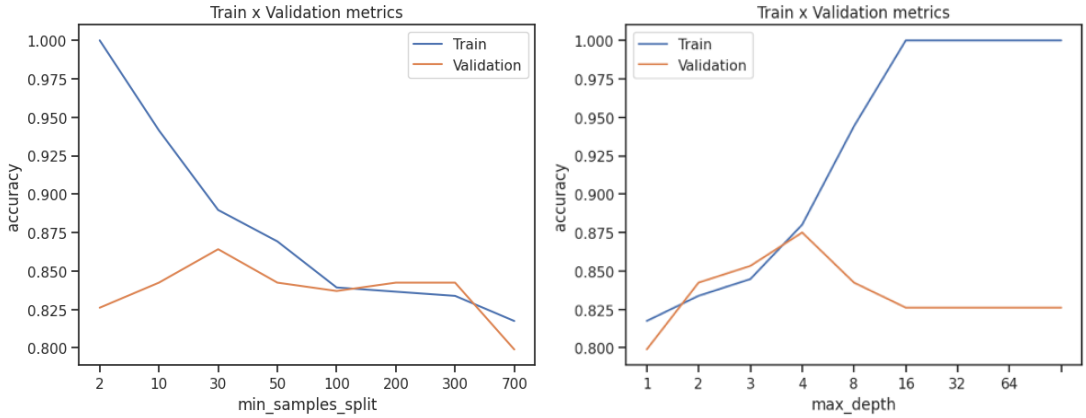

Figure 4. Validation Curves For Max Depth and Min Samples Split

The confusion matrix of the DT model (Figure 5) showed 69 true positives (TP), 88 true negatives (TN), 19 false positives (FP), and 8 false negatives (FN). The classification report indicated that DT achieved an accuracy of 85%, precision of 86%, recall of 85%, and an F1-score of 85%. While the DT offered interpretable decision rules and identified features such as chest pain and glucose level as significant predictors, it also exhibited signs of overfitting and limited robustness compared to other models.

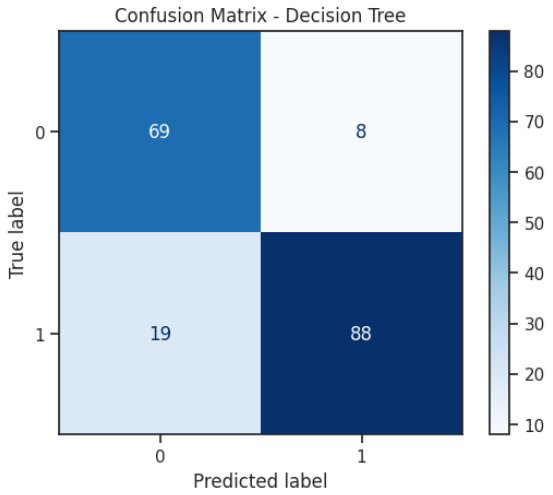

Figure 5. Confusion Matrix of Decision Tree

The second model evaluated was the k-Nearest Neighbors (KNN) algorithm, a distance-based method that classifies a new instance according to the majority label of its k nearest neighbors. Because KNN heavily depends on feature scale, its performance was assessed under three different preprocessing strategies: without scaling, with normalization, and with standard scaling. As shown in Figure 6, when trained without any scaling, the accuracy of KNN fluctuated substantially between 65–71%, indicating that features with larger numerical ranges dominated the distance calculation. This instability highlighted the inability of the model to effectively capture neighborhood structures in unscaled data.

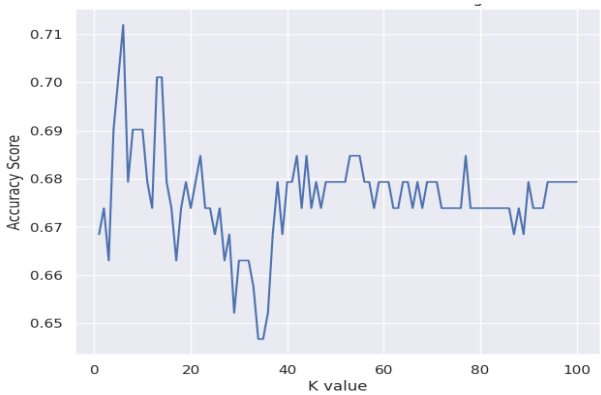

Figure 6. Performance of The Trained KNN Classifier Without Feature Scaling

When normalization was applied, the model's accuracy improved significantly, reaching 85–86%. Normalization rescaled features into a smaller and more uniform range, reducing the disproportionate influence of high-magnitude attributes such as cholesterol or glucose level. However, the accuracy curve remained less stable, and the performance plateaued beyond k ≈ 20, indicating limited generalization capability. The best performance was obtained using standard scaling, as shown in Figure 7, where the accuracy peaked at approximately 90% when k = 24. Standard scaling transformed the data into a distribution with mean 0 and unit variance, ensuring that all attributes contributed equally to the Euclidean distance metric. This not only enhanced the model's stability but also provided smoother decision boundaries, reducing sensitivity to noisy or redundant features. The hyperparameter tuning of KNN was carried out through a grid search over the number of neighbors k ranging from 1 to 100, with weights set to distance and Euclidean distance as the metric. The optimal configuration was obtained at k = 24 under standard scaling.

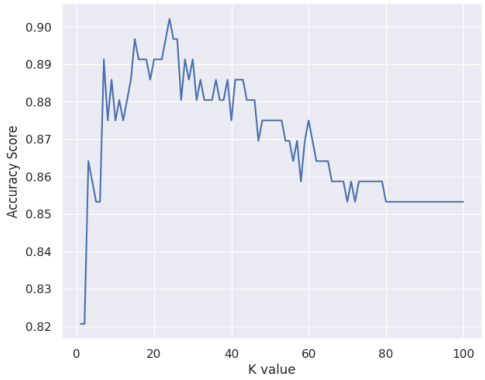

Figure 7. Performance of The Trained KNN Classifier With Feature Scaling

The confusion matrix for KNN, presented in Figure 8, confirmed these improvements, yielding 70 true negatives (TN), 96 true positives (TP), 7 false positives (FP), and 11 false negatives (FN). Based on these results, the model achieved an accuracy of 90.2%, precision of 90.4%, recall of 90.2%, and an F1-score of 90.2%.

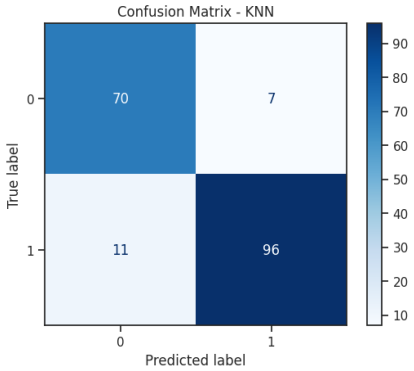

Figure 8. Confusion Matrix of KNN

The third model evaluated was the Random Forest classifier, which aggregates multiple decision trees using bootstrap sampling and random feature selection. Hyperparameter optimization was performed with a grid search on the training data using cross-validation. The search space covered the number of trees with values 50, 100, 150, and 200; the maximum tree depth with values 5, 10, 15, and no limit; and the minimum number of samples at a leaf with values 1, 5, and 10. The best configuration was then refit on the full training data and evaluated on the test set. The confusion matrix of RF, shown in Figure 9, reported 66 true negatives (TN), 95 true positives (TP), 11 false positives (FP), and 12 false negatives (FN). Based on these results, the RF model achieved an accuracy of 87.5%, precision of 87.5%, recall of 87.5%, and an F1-score of 87.5%.

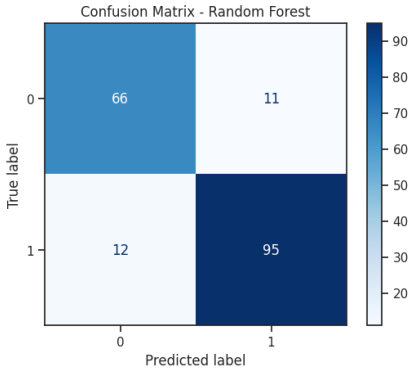

Figure 9. Confusion Matrix of Random Forest

Calibration and decision curve analysis were applied to evaluate the reliability and clinical usefulness of the models. Figure 10 presents the calibration plots on the left and the decision curves on the right for Decision Tree, Random Forest, and K Nearest Neighbors. The Decision Tree curve generally follows the 45 degree reference line, Random Forest shows slight miscalibration in the mid probability region but aligns well at higher probabilities, and K Nearest Neighbors is more variable at lower probabilities yet stable at higher ones. The decision curves indicate that within thresholds from 0.1 to 0.6 all three models achieve positive net benefit, with Random Forest and K Nearest Neighbors providing consistently higher benefit than Decision Tree.

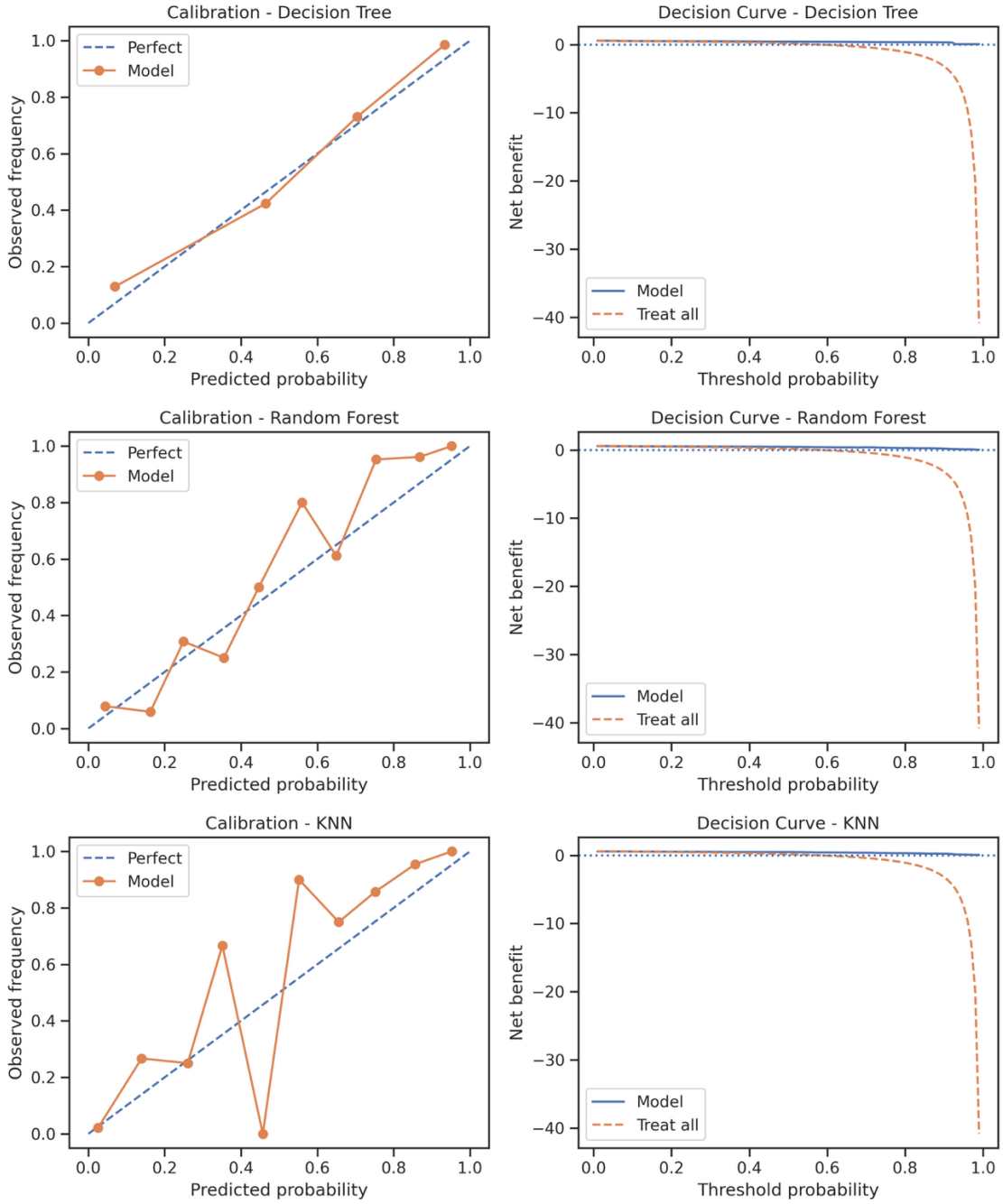

Figure 9. Calibration and Decision Curve Analysis for DT, RF and KNN

A detailed summary of the performance metrics obtained from all three models is presented in Table 1dalam , highlighting the comparative outcomes of the tested algorithms.

Table 1. Comparative Result of Model Evaluation

Params	Decision Tree	k-Nearest Neighbors	Random Forest
Accuracy	0.853	0.902	0.875
Precision	0.861	0.904	0.875
Recall	0.853	0.902	0.875
F1-Score	0.854	0.902	0.875
ROC-AUC	0.917	0.949	0.936
PR-AUC	0.930	0.964	0.955
MCC@0,5	0.710	0.801	0.744
Brier	0.108	0.091	0.102

Based on the performance comparison in Table 3, the three algorithms achieve competitive results with accuracy above 85 percent. Decision Tree attains 85.3 percent accuracy with 86.1 percent precision, 85.3 percent recall, and an 85.4 percent F1 score. Although Decision Tree offers good interpretability, its generalization capability is limited and it tends to overfit the training data. Random Forest, an ensemble of multiple decision trees, slightly outperforms Decision Tree with 87.5 percent accuracy and balanced precision, recall, and F1 score of 87.5 percent. This reflects reduced variance relative to a single tree, although the improvement is modest. In contrast, k-Nearest Neighbors delivers the strongest results with 90.2 percent accuracy, 90.4 percent precision, 90.2 percent recall, and a 90.2 percent F1 score. These findings show that with appropriate scaling k-Nearest Neighbors can capture complex patterns and produce more consistent predictions. Overall, k-Nearest Neighbors is the most effective model for this dataset, followed by Random Forest, while Decision Tree shows the lowest performance yet remains reasonably competitive.

4. CONCLUSION

This study aimed to enhance the early detection of cardiovascular disease using machine learning methods, as outlined in the Introduction. The results and discussion chapters demonstrated that the applied models Decision Tree (DT), k-Nearest Neighbors (KNN), and Random Forest (RF) successfully achieved the expected outcomes by classifying patients with a relatively high degree of accuracy. Among the three algorithms, KNN achieved the best performance with an accuracy of 90.2%, precision of 90.4%, recall of 90.2%, and F1-score of 90.2%, followed by Random Forest with 87.5% and Decision Tree with 85.3%. These findings validate the expectation that machine learning can serve as an effective tool for supporting clinical decision-making and providing early warnings in the detection of heart disease.

The study also revealed that preprocessing, particularly feature scaling, plays a critical role in improving the performance of distance-based models such as KNN. This highlights an important gap in existing research and demonstrates the necessity of robust preprocessing strategies for reliable predictive modeling in healthcare.

Looking ahead, the outcomes of this research open several prospects for further development. Future studies may integrate larger and more diverse datasets, explore additional algorithms such as deep learning or hybrid ensemble models, and incorporate real-time patient monitoring systems to improve predictive accuracy and clinical applicability. Furthermore, the application of explainable artificial intelligence (XAI) methods can strengthen the interpretability of machine learning predictions, increasing trust and adoption among medical practitioners. Ultimately, the findings of this study not only contribute to academic research but also hold practical potential to support healthcare systems in providing earlier, faster, and more precise detection of cardiovascular disease.

ACKNOWLEDGEMENTS

The authors would like to express their deepest gratitude to STMIK TIME for providing academic and research support throughout this study. Special thanks are extended to supervisors and colleagues for their constructive feedback and encouragement during the research and writing process. The authors also wish to acknowledge the valuable collaboration between faculty members and students, which made this work possible.

REFERENCES

[1] V. L. Roger, "Epidemiology of Heart Failure: A Contemporary Perspective," *Circ Res*, vol. 128, no. 10, pp. 1421–1434, May 2021, doi: 10.1161/CIRCRESAHA.121.318172.

[2] A. Singh, H. Mahapatra, A. K. Biswal, M. Mahapatra, D. Singh, and M. Samantaray, "Heart Disease Detection Using Machine Learning Models," in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 937–947. doi: 10.1016/j.procs.2024.04.089.

[3] P. M. Seferović *et al.*, "The Heart Failure Association Atlas: Heart Failure Epidemiology and Management Statistics 2019," *Eur J Heart Fail*, vol. 23, no. 6, pp. 906–914, Jun. 2021, doi: 10.1002/ejhf.2143.

[4] A. K. Gárate-Escamila, A. Hajjam El Hassani, and E. Andrès, "Classification models for heart disease prediction using feature selection and PCA," *Inform Med Unlocked*, vol. 19, Jan. 2020, doi: 10.1016/j.imu.2020.100330.

[5] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

[6] G. Alwakid, F. Ul Haq, N. Tariq, M. Humayun, M. Shaheen, and M. Alsadun, "Optimized machine learning framework for cardiovascular disease diagnosis: a novel ethical perspective," *BMC Cardiovasc Disord*, vol. 25, no. 1, Dec. 2025, doi: 10.1186/s12872-025-04550-w.

[7] T. Mahmood, A. Rehman, T. Saba, L. Nadeem, and S. A. O. Bahaj, "Recent Advancements and Future Prospects in Active Deep Learning for Medical Image Segmentation and Classification," *IEEE Access*, vol. 11, pp. 113623–113652, Sep. 2023, doi: 10.1109/ACCESS.2023.3313977.

[8] M. G. Hanna *et al.*, "Future of Artificial Intelligence—Machine Learning Trends in Pathology and Medicine," *Modern Pathology*, vol. 38, no. 4, p. 100705, Apr. 2025, doi: 10.1016/j.modpat.2025.100705.

[9] K. M. Shiwangi, J. K. Sandhu, and R. Sahu, "Effective Heart-Disease Prediction by Using Hybrid Machine Learning Technique," *Proceedings of the International Conference on Circuit Power and Computing Technologies, ICCPCT 2023*, pp. 1670–1675, 2023, doi: 10.1109/ICCPCT58313.2023.10245785.

[10] L. Yang *et al.*, "Study of cardiovascular disease prediction model based on random forest in eastern China," *Sci Rep*, vol. 10, no. 1, Dec. 2020, doi: 10.1038/s41598-020-62133-5.

[11] O. P. Barus, K. Lauwren, J. J. Pangaribuan, and Romindo, "Implementation of the Naive Bayes Algorithm to Predict the Safety of Heart Failure Patients," *Conference Series*, vol. 4, no. 1, pp. 172–177, Dec. 2023, doi: 10.34306/conferenceseries.v4i1.651.

[12] M. Ozcan and S. Peker, "A classification and regression tree algorithm for heart disease modeling and prediction," *Healthcare Analytics*, vol. 3, Nov. 2023, doi: 10.1016/j.health.2022.100130.

[13] P. C. C. Clarite and I. V. G. V. G. Palma, "Heart Disease Prediction Using Decision Tree Analysis," *Journal of Engineering and Science Application*, vol. 1, no. 2, pp. 26–32, Oct. 2024, doi: 10.69693/jesa.v1i2.9.

[14] "Heart Failure Prediction Dataset." Accessed: Jul. 22, 2025. [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction?resource=download

[15] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/j.gltp.2022.04.020.

[16] P. Martins, F. Cardoso, P. Váz, J. Silva, and M. Abbasi, "Performance and Scalability of Data Cleaning and Preprocessing Tools: A Benchmark on Large Real-World Datasets," *Data (Basel)*, vol. 10, no. 5, p. 68, May 2025, doi: 10.3390/data10050068.

[17] H. J. Park, Y. S. Koo, H. Y. Yang, Y. S. Han, and C. S. Nam, "Study on Data Preprocessing for Machine Learning Based on Semiconductor Manufacturing Processes," *Sensors*, vol. 24, no. 17, Sep. 2024, doi: 10.3390/s24175461.

[18] Everleen Nekesa Wanyonyi and Newton Wafula Masinde, "The Impact of Data Preprocessing on Machine Learning Model Performance: A Comprehensive Examination," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 11, no. 2, pp. 3814–3827, Apr. 2025, doi: 10.32628/CSEIT25112854.

[19] H. A. Abdulqader and A. M. Abdulazeez, "Review on Decision Tree Algorithm in Healthcare Applications," *Indonesian Journal of Computer Science*, vol. 13, no. 3, Jun. 2024, doi: 10.33022/ijcs.v13i3.4026.

[20] D. B. Olawade, A. A. Soladoye, B. A. Omodunbi, N. Aderinto, and I. A. Adeyanju, "Comparative analysis of machine learning models for coronary artery disease prediction with optimized feature selection," *Int J Cardiol*, vol. 436, Oct. 2025, doi: 10.1016/j.ijcard.2025.133443.

[21] A. A. Soladoye, N. Aderinto, B. A. Omodunbi, A. O. Esan, I. A. Adeyanju, and D. B. Olawade, "Enhancing Alzheimer's disease prediction using random forest: A novel framework combining backward feature elimination and ant colony optimization," *Curr Res Transl Med*, vol. 73, no. 4, Dec. 2025, doi: 10.1016/j.retram.2025.103526.

[22] I. D. Mienye and N. Jere, "A Survey of Decision Trees: Concepts, Algorithms, and Applications," *IEEE Access*, vol. 12, pp. 86716–86727, 2024, doi: 10.1109/ACCESS.2024.3416838.

[23] H. Blockeel, L. Devos, B. Frénay, G. Nanfack, and S. Nijssen, "Decision trees: from efficient prediction to responsible AI," 2023, *Frontiers Media SA*. doi: 10.3389/frai.2023.1124553.

[24] E. F. Siddiqui, T. Ahmed, and S. K. Nayak, "A decision tree approach for enhancing real-time response in exigent healthcare unit using edge computing," *Measurement: Sensors*, vol. 32, p. 100979, Apr. 2024, doi: 10.1016/j.measen.2023.100979.

[25] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-10358-x.

[26] K. L. Kohsasih, "COMPARISON OF SVM, KNN, AND NAÏVE BAYES ALGORITHMS IN MONKEYPOX DISEASE CLASSIFICATION," *TAMIKA: Jurnal Tugas Akhir Manajemen Informatika & Komputerisasi Akuntansi*, vol. 4, no. 2, pp. 168–174, Dec. 2024, doi: 10.46880/tamika.Vol4No2(SEMNASTIK).pp168-174.

[27] H. A. Salman, A. Kalakech, and A. Steiti, "Random Forest Algorithm Overview," *Babylonian Journal of Machine Learning*, vol. 2024, pp. 69–79, Jun. 2024, doi: 10.58496/BJML/2024/007.

[28] K. Jurczuk, M. Czajkowski, and M. Kretowski, "From Random Forest to an interpretable decision tree - An evolutionary approach," in *Proceedings of the Companion Conference on Genetic and Evolutionary Computation*, New York, NY, USA: ACM, Jul. 2023, pp. 291–294. doi: 10.1145/3583133.3590732.

[29] S. A. Hicks *et al.*, "On evaluation metrics for medical applications of artificial intelligence," *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-09954-8.

[30] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-56706-x.

[31] B. Kocak *et al.*, "Evaluation metrics in medical imaging AI: fundamentals, pitfalls, misapplications, and recommendations," *European Journal of Radiology Artificial Intelligence*, vol. 3, p. 100030, Sep. 2025, doi: 10.1016/j.ejrai.2025.100030.