Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 2, April 2026, Page. 1853-1870
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5684

Machine Learning Decision Support System for Heart Disease Prediction with Optuna and Threshold Optimization

William Ramdhan*[1], Jeperson Hutahaean[2], Deny Jollyta[3], Abdul Karim[4]

[1,2]Information Systems, Universitas Royal, Kisaran, Indonesia
[3]Informatics, Institut Bisnis dan Teknologi Pelita Indonesia, Pekanbaru, Indonesia
[4]Information Technology, Universitas Labuhanbatu, Labuhan Batu, Indonesia

Email: [1]william.ramdhan052@gmail.com

Received : Jan 31, 2026; Revised : Mar 8, 2026; Accepted : Mar 10, 2026; Published : Apr 18, 2026

Abstract

Cardiovascular disease remains a major global health challenge, necessitating accurate and reliable decision support systems for early detection. This study proposes a machine learning–based decision support system that integrates ensemble learning, automated hyperparameter optimization using Optuna, and decision threshold tuning. The system was evaluated using several baseline machine learning models, including Logistic Regression, SVM, KNN, Decision Tree, and Random Forest, with the Random Forest model selected for optimization. Hyperparameter tuning with Optuna and decision threshold optimization led to a significant improvement in accuracy (95.0%) and ROC–AUC (0.977), with the optimized model outperforming all baseline models. This approach demonstrates improved sensitivity, reduced false negatives, and enhanced predictive performance, offering a clinically reliable tool for early heart disease detection. The results emphasize the importance of model optimization and decision threshold calibration in clinical decision support systems.

Keywords : *Heart disease prediction, Decision support system, Machine learning, Optuna, Hyperparameter optimization, Threshold optimization*

1. INTRODUCTION

Heart disease remains one of the most critical global health challenges and continues to be a leading cause of morbidity and mortality worldwide [1]–[5]. The prevalence of cardiovascular diseases has increased significantly over recent decades, affecting both developed and developing countries due to factors such as aging populations, lifestyle changes, and increasing exposure to health risk factors [6]–[10]. The diagnosis of heart disease is inherently complex, as it involves the interpretation of multiple clinical, demographic, and physiological parameters that often exhibit non-linear and interdependent relationships [11]–[14]. This complexity places substantial pressure on healthcare systems and medical practitioners, particularly in environments with limited diagnostic resources and high patient loads [15]–[19]. In response to these challenges, the integration of data-driven technologies into healthcare has gained considerable attention, with machine learning emerging as a promising approach for supporting medical diagnosis [20]–[23]. Machine learning–based decision support systems are capable of analyzing large volumes of clinical data and identifying hidden patterns that may not be easily recognized through conventional diagnostic methods [24]–[29]. Consequently, such systems have been increasingly applied to heart disease prediction tasks to enhance diagnostic accuracy, reduce human error, and support clinical decision-making [30]–[32]. Nevertheless, despite their growing adoption, the effectiveness of these systems largely depends on how well the underlying models are configured and how reliably they translate predictive outputs into clinically meaningful decisions [33]–[35].

Vol. 7, No. 2, April 2026, Page. 1853-1870
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5684

Although numerous studies have applied machine learning techniques for heart disease prediction, several fundamental problems remain inadequately addressed. Many existing approaches rely on default or manually selected hyperparameters, which limits the ability of machine learning models to achieve optimal performance and reduces their robustness across different datasets [36]–[38]. In addition, most studies employ a fixed decision threshold, typically set at 0.5, to classify patients into disease and non-disease categories, without considering the asymmetric risks associated with misclassification in medical diagnosis [39]–[41]. This practice can result in suboptimal diagnostic performance, particularly in cases where false negatives may lead to delayed treatment and severe clinical consequences. Furthermore, comparative analyses of multiple machine learning algorithms are often conducted without a systematic optimization framework, making it difficult to determine whether performance differences arise from algorithmic superiority or from suboptimal parameter configurations. The absence of integrated hyperparameter optimization and decision threshold tuning also affects the reproducibility and generalizability of prior research findings. As a result, existing machine learning–based decision support systems for heart disease prediction may fail to provide consistent and clinically reliable outcomes. These limitations highlight the need for an enhanced approach that systematically addresses both model optimization and decision threshold selection to improve diagnostic accuracy and reliability [42].

Several recent studies have explored the application of machine learning techniques for heart disease prediction, demonstrating promising yet still limited diagnostic performance. Vu et al. (2025) reported moderate accuracy and AUC values around 0.73 using Random Forest and Support Vector Machine models [43]. Saha et al. (2024) achieved a maximum accuracy of approximately 91% and ROC–AUC of 0.91 through comparative evaluation of multiple supervised algorithms [44]. Gayathri (2024) proposed a hybrid learning framework and reported accuracy close to 94%, emphasizing the sensitivity of model performance to parameter configuration [45]. Similarly, Dipu et al. (2025) obtained accuracy values around 90–92% across multiple datasets but noted inconsistencies affecting generalizability [46].

Despite these advancements, most existing studies primarily focus on improving algorithmic accuracy while still relying on default hyperparameter settings and conventional fixed probability thresholds (commonly 0.5) for classification decisions. A recent comprehensive review by Wan et al. (2025) highlighted that explicit decision threshold optimization remains rarely addressed in cardiovascular disease prediction research, despite the asymmetric clinical risks associated with misclassification [47]. Consequently, although performance metrics continue to improve, the absence of integrated hyperparameter optimization and adaptive threshold calibration limits the clinical reliability and practical deployment of many existing models.

Overall, these studies demonstrate that while machine learning models are effective for heart disease prediction, there remains a clear research gap in integrating adaptive hyperparameter optimization and decision threshold tuning within a unified decision support system to achieve more robust and clinically reliable performance, which motivates the proposed approach in this study.

To address the identified research gaps, this study proposes an enhanced machine learning–based decision support system for heart disease prediction that integrates systematic model comparison, automated hyperparameter optimization, and decision threshold optimization within a unified framework. Initially, multiple machine learning algorithms, including Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors, and Support Vector Machine, are evaluated as baseline models using standardized preprocessing and evaluation protocols to identify the most suitable classifier for heart disease prediction. Subsequently, the selected model is enhanced through Optuna-based hyperparameter optimization, which employs an adaptive and efficient search strategy to identify optimal parameter configurations that maximize predictive performance, particularly in terms of ROC-

Vol. 7, No. 2, April 2026, Page. 1853-1870
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5684

AUC. In addition to model-level optimization, this study incorporates decision threshold optimization to adjust the classification boundary based on predicted probabilities, enabling a more balanced trade-off between sensitivity and specificity that is critical in medical diagnosis contexts. Unlike conventional approaches that rely on default thresholds, the proposed method systematically determines the optimal threshold to improve diagnostic reliability and reduce misclassification risk. The main contributions of this research are threefold: first, it provides a comprehensive comparative analysis of multiple machine learning algorithms for heart disease prediction under a consistent experimental setting; second, it demonstrates the effectiveness of integrating Optuna-based hyperparameter optimization to enhance model performance beyond baseline configurations; and third, it introduces decision threshold optimization as an additional mechanism to further improve accuracy and ROC-AUC in a machine learning–based decision support system. The experimental results confirm that the proposed approach significantly improves predictive performance and offers a robust and clinically relevant solution to support heart disease diagnosis.

Despite high reported accuracy values in many studies, the lack of alignment between model optimization strategies and practical clinical decision thresholds limits their translational applicability in real healthcare workflows. From a practical clinical perspective, the gap in threshold selection and systematic model optimization directly reflects real-world challenges faced by healthcare practitioners when implementing machine learning–based decision support systems. In many clinical environments, predictive models are deployed with default configurations and fixed probability thresholds that do not account for the asymmetric consequences of diagnostic errors. Such implementations may result in missed high-risk patients (false negatives) or unnecessary follow-up examinations (false positives), both of which impose clinical and economic burdens. By explicitly integrating automated hyperparameter optimization and adaptive decision threshold calibration, the proposed framework aligns predictive modeling with clinical risk management principles. This approach enables the system to prioritize sensitivity in early screening contexts or balance specificity in resource-constrained settings, thereby improving its practical usability. By addressing the gap in threshold selection and model optimization, this study aims to offer clinicians a more reliable tool that reduces diagnostic errors, enhances early detection of high-risk patients, and supports safer and more evidence-based cardiovascular risk stratification.

2. METHOD

This study proposes an enhanced machine learning–based decision support system for heart disease prediction by integrating multi-model evaluation, best model selection, hyperparameter optimization using Optuna, and decision threshold optimization. The research methodology is systematically structured as illustrated in Figure 1.

Figure 1 illustrates the overall research workflow used in this study. The process begins with dataset acquisition from IEEE Dataport, followed by preprocessing steps including missing value inspection, data encoding, and feature scaling. The processed dataset is then divided into training and testing subsets using stratified sampling to maintain class distribution. Several baseline machine learning models are subsequently trained and evaluated using consistent performance metrics. Based on the baseline results, the best-performing model is selected for automated hyperparameter optimization using Optuna. After the hyperparameters are optimized, decision threshold calibration is applied to refine the classification boundary. The final stage evaluates the optimized model using multiple performance metrics to assess its suitability for clinical decision support.

Vol. 7, No. 2, April 2026, Page. 1853-1870
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5684

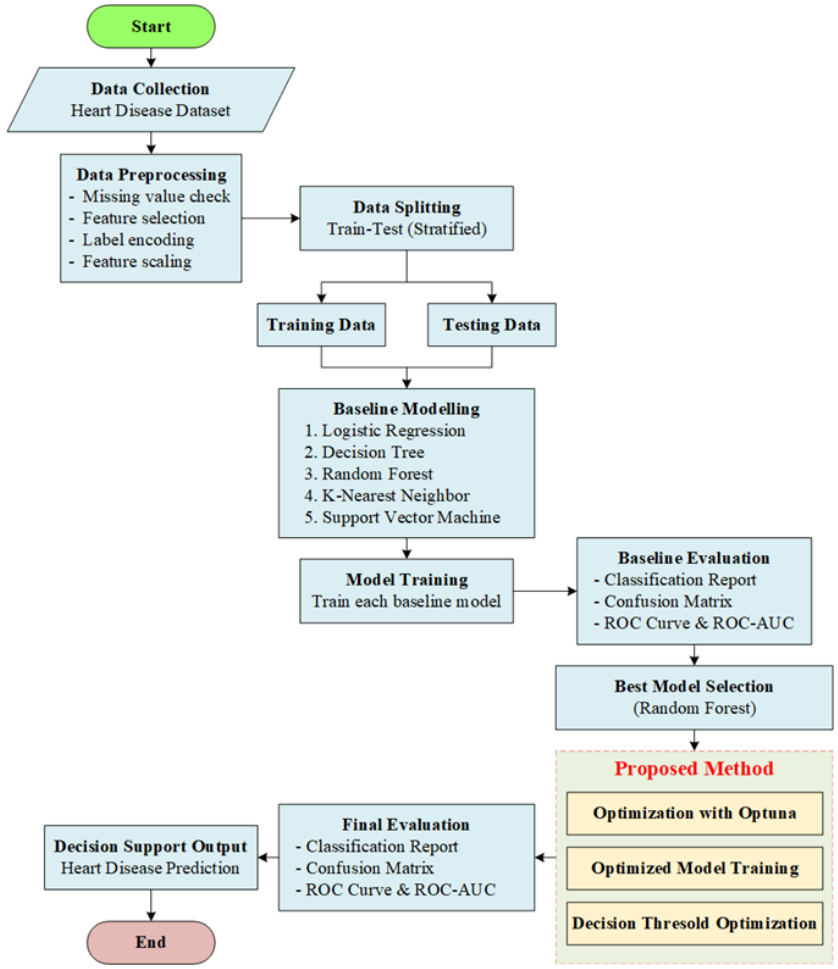


Figure 1. Research Flowchart

2.1. Data Collection

The dataset utilized in this study was obtained from an open-access repository provided by IEEE Dataport, entitled Heart Disease Dataset (Comprehensive), which is publicly accessible at: https://ieee-dataport.org/open-access/heart-disease-dataset-comprehensive [48].

This dataset was selected due to its completeness, reliability, and relevance for machine learning–based heart disease prediction. The dataset consists of approximately 1,190 patient records, each described by 11 predictive clinical features and one target variable. The features represent a combination of demographic and medical attributes that are widely used in cardiovascular risk assessment, including age, sex, resting blood pressure, serum cholesterol, fasting blood sugar, electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression, slope of the peak exercise ST segment, and the number of major vessels colored by fluoroscopy. Each record corresponds to an individual patient, making the dataset suitable for supervised learning approaches.

The target variable is defined as a binary class, where the label Absence indicates patients without heart disease, and Presence indicates patients diagnosed with heart disease. This binary classification structure supports the evaluation of various machine learning algorithms and facilitates the development of a decision support system for early heart disease diagnosis. The use of an openly accessible and well-documented dataset ensures experimental reproducibility and transparency. Moreover, the presence of clinically meaningful attributes allows the proposed system to reflect realistic diagnostic conditions, thereby strengthening its applicability in real-world decision support scenarios for heart disease prediction.

Vol. 7, No. 2, April 2026, Page. 1853-1870
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5684

2.2. Data Preprocessing

Data preprocessing is conducted to ensure data quality and model robustness. The preprocessing steps include missing value inspection, feature selection, label encoding, and feature scaling. The categorical target labels are transformed into numerical form as follows:

$$y = \begin{cases} 0, & \text{Absensi of heart disease} \\ 1, & \text{Presence of heart disease} \end{cases} \quad (1)$$

To eliminate scale discrepancies among features, standardization is applied using:

$$z = \frac{x-\mu}{\sigma} \quad (2)$$

where x represents the original feature value, μ is the mean, and σ is the standard deviation.

2.3. Data Splitting

The preprocessed dataset is divided into training and testing subsets using a stratified train–test split to preserve class distribution. A ratio of 80% for training and 20% for testing is applied. This strategy minimizes class imbalance effects during model evaluation.

2.4. Baseline Modelling

In this study, five supervised machine learning algorithms were implemented as baseline classifiers for heart disease prediction: Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM). These algorithms have been widely applied in cardiovascular disease prediction research due to their ability to model complex relationships in clinical data with varying degrees of interpretability and generalization ability [49].

Logistic Regression (LR) estimates the probability of heart disease occurrence using a logistic function defined as:

$$\mathrm{P}\,(y = 1|x) = \frac{1}{1+e^{-(\beta_0+\sum_{i=1}^{n}\beta_i x_i}} \quad (3)$$

where $\mathrm{P}\,(y = 1|x)$ represents the probability of heart disease presence, β_0 is the intercept, β_i are the learned coefficients, and x_i represents the input feature values. Logistic Regression has been demonstrated as a baseline method in heart disease and cardiovascular disease prediction studies [50].

Decision Tree (DT) classifies data by recursively splitting features based on impurity reduction, commonly measured using the Gini Index:

$$Gini = 1 - \sum_{i=1}^{c} p_{i^2} \quad (4)$$

where p_i denotes the proportion of class i in a node and c is the number of classes. Decision Trees are often included in comparative analyses for disease classification to assess interpretability and ease of clinical explanation [51].

Random Forest (RF) is an ensemble of decision trees that improves generalization by aggregating predictions through majority voting:

$$\hat{y} = mode\,(h_1(x), h_2(x), \ldots, h_m(x)) \quad (5)$$

where $h_j(x)$ is the prediction of the j-th tree and $\hat{y}$ is the ensemble prediction. Random Forest has demonstrated strong performance in heart disease prediction studies due to its robustness to overfitting and ability to handle nonlinear feature interactions [51].

Vol. 7, No. 2, April 2026, Page. 1853-1870
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5684

K-Nearest Neighbor (KNN) classifies a query point based on the majority class of its nearest neighbors using the Euclidean distance:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{n}(x_{ik} - x_{jk})^2} \quad (6)$$

where x_i and x_j are data samples and n is the number of features. The KNN method is non-parametric and relies on local neighborhood information for classification [50].

Support Vector Machine (SVM) identifies an optimal separating hyperplane by maximizing the margin between classes. A common formulation with hinge loss is:

$$min_{w,b}(\frac{1}{2}||w||^2 + C\sum_{i=1}^{N}\max(0,1 - y_i(w^T x_i + b)) \quad (7)$$

where w is the weight vector, b is the bias, C is a regularization hyperparameter, and $y_i \in \{-1,1\}$ are class labels. SVM has been shown to achieve competitive classification performance in heart disease prediction tasks [52].

The baseline algorithms were selected to represent diverse modeling paradigms commonly applied in cardiovascular disease prediction research. Logistic Regression was included due to its interpretability and long-standing use in clinical risk modeling. Support Vector Machine was chosen for its ability to construct optimal decision boundaries and handle complex patterns in structured datasets. K-Nearest Neighbor represents a distance-based approach that captures local similarity patterns among patients with similar clinical profiles. Decision Tree was included for its ability to model nonlinear relationships between clinical variables. Particular emphasis was placed on Random Forest because of its ensemble learning mechanism, which combines multiple decision trees to improve predictive stability. In medical datasets, where sample sizes are often moderate and feature interactions can be complex, single models may be prone to overfitting. Random Forest mitigates this issue by aggregating predictions from multiple bootstrapped trees, thereby reducing variance and improving generalization performance. These characteristics make Random Forest particularly suitable for heart disease prediction tasks.

2.5. Model Training

In this stage, each baseline machine learning model is trained using the preprocessed training dataset. The training process aims to enable each classifier to learn the underlying patterns and relationships between input features and the target variable. Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM) are trained independently to ensure a fair and unbiased comparison. All models are trained using the same training dataset and identical data partitioning strategy to maintain consistency in performance evaluation. For algorithms sensitive to feature scale, such as Logistic Regression, KNN, and SVM, standardized input features are utilized, while tree-based models are trained on the original feature space. This training strategy ensures that each algorithm operates under optimal conditions based on its inherent characteristics. The trained models serve as baseline classifiers whose performance is subsequently evaluated using multiple evaluation metrics. The comparative results obtained from this training phase provide the basis for selecting the most suitable model for further optimization in the proposed decision support system.

2.6. Baseline Evaluation

To comprehensively evaluate the performance of the proposed decision support system, several widely used evaluation metrics were employed, including Accuracy, Confusion Matrix, Classification

Vol. 7, No. 2, April 2026, Page. 1853-1870
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5684

Report, and Receiver Operating Characteristic–Area Under the Curve (ROC-AUC). These metrics were selected to provide a balanced assessment of classification performance, particularly for medical diagnosis tasks where both correct detection and error minimization are critical.

Accuracy measures the proportion of correctly classified instances among all predictions and is defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

where TP denotes true positives, TN represents true negatives, FP refers to false positives, and FN indicates false negatives.

The Confusion Matrix is used to present the distribution of prediction results across actual and predicted classes, enabling detailed analysis of misclassification patterns. This matrix provides insight into how well the model distinguishes between patients with and without heart disease. The Classification Report summarizes key performance indicators, including precision, recall, and F1-score. Precision is defined as:

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

where recall indicates the model's ability to correctly identify patients with heart disease. The F1-score, which balances precision and recall, is calculated as:

$$F1 = 2 \text{ x} \frac{Precission\ x\ Recall}{Precission+Recall} \quad (10)$$

Finally, the Receiver Operating Characteristic (ROC) curve and the corresponding Area Under the Curve (AUC) are used to evaluate the model's discriminative ability across various decision thresholds. The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR), defined as:

$$TPR = \frac{TP}{TP+FN} \qquad FPR = \frac{FP}{FP+TN} \quad (11)$$

where the ROC-AUC value represents the overall capability of the model to distinguish between positive and negative classes. A higher ROC-AUC indicates better classification performance, which is particularly important in medical decision support systems.

In heart disease prediction, ROC–AUC is an important metric because it measures how well the model can distinguish between patients with and without heart disease across different classification thresholds. Unlike accuracy, which depends on a single threshold, ROC–AUC evaluates the model's overall discrimination capability. A higher ROC–AUC indicates that the model is better at assigning higher prediction scores to patients who truly have heart disease compared to those who do not. This capability is particularly important in medical screening scenarios where early identification of high-risk patients is essential.

2.7. Best Model Selection

Based on the evaluation results obtained from the baseline model training and performance assessment, the Random Forest classifier demonstrates superior performance compared to other evaluated machine learning models. The selection of the best model is conducted by jointly considering classification accuracy, ROC-AUC value, and the stability of prediction results across the test dataset. Random Forest consistently achieves higher predictive performance due to its ensemble learning mechanism, which combines multiple decision trees to reduce variance and mitigate overfitting. By aggregating the outputs of individual trees through majority voting, Random Forest is capable of

Vol. 7, No. 2, April 2026, Page. 1853-1870
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5684

capturing complex and nonlinear relationships among clinical features, which are commonly observed in heart disease datasets. In addition to its strong classification performance, Random Forest exhibits robustness to noisy data and maintains reliable results without requiring extensive data scaling or feature transformation. These characteristics make it particularly suitable for medical decision support systems, where model reliability and generalization ability are critical. Therefore, Random Forest is selected as the base classifier for further enhancement in the proposed method. Subsequent optimization stages, including hyperparameter optimization using Optuna and decision threshold optimization, are applied exclusively to the Random Forest model in order to further improve diagnostic accuracy and discrimination capability.

2.8. Proposed Method

This study proposes an enhanced machine learning–based decision support system for heart disease prediction by integrating optimal model selection, hyperparameter optimization, and decision threshold tuning. Based on baseline evaluation, the Random Forest classifier is selected as the core predictive model due to its superior performance and robustness. Hyperparameter optimization is performed using Optuna to automatically identify the optimal configuration that maximizes the ROC-AUC score through cross-validation. This process improves the model's generalization ability and classification performance. Subsequently, decision threshold optimization is applied to adjust the classification boundary based on predicted probability distributions, enabling a better balance between sensitivity and specificity. The optimized Random Forest model with the tuned decision threshold is then employed as the final decision support system. This integrated optimization strategy enhances prediction accuracy and diagnostic reliability compared to baseline machine learning models.

2.9. Final Evaluation

The final evaluation is conducted to validate the performance of the optimized Random Forest model on the test dataset after hyperparameter optimization and decision threshold tuning. This stage focuses on assessing the generalization capability of the proposed model under unseen data conditions. Model performance is evaluated using accuracy and ROC-AUC, supported by confusion matrix and classification report analysis. These metrics provide an overall assessment of predictive reliability and class discrimination ability. The final results are then compared with baseline model performance to demonstrate the effectiveness of the proposed optimization strategy in improving heart disease prediction accuracy and robustness.

2.10. Decision Support Output

The decision support output presents the final classification results produced by the optimized Random Forest model to support clinical decision-making. Each patient record is assigned a binary prediction label, where 0 indicates the absence of heart disease (non-heart disease) and 1 indicates the presence of heart disease. The classification is determined based on the optimized decision threshold to balance sensitivity and specificity. This output provides clear and interpretable results, enabling healthcare practitioners to identify patients at potential risk of heart disease and to support early diagnosis and preventive decision-making.

3. RESULT

This section presents the experimental results of the proposed enhanced machine learning–based decision support system for heart disease prediction. The results are discussed in a structured manner, beginning with baseline model evaluation, followed by best model selection, optimization outcomes, and comparative analysis with prior studies. The evaluation focuses on accuracy, precision, recall, F1-score, confusion matrix, and ROC–AUC to ensure a comprehensive assessment of model performance.

Vol. 7, No. 2, April 2026, Page. 1853-1870
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5684

3.1. Experimental Evaluation Overview

All experiments were conducted using a stratified train–test split to ensure balanced class distribution. The evaluation focuses on both classification correctness and discriminative ability, which are essential in medical decision support systems. Performance is assessed using accuracy, precision, recall, F1-score, confusion matrix, and ROC–AUC. The dataset consists of two target classes:

Class 0: Absence of heart disease

Class 1: Presence of heart disease

To facilitate a structured evaluation, the experimental results are summarized using tables and supported by visual figures.

3.2. Baseline Model Performance Comparison

Table 1 presents the overall accuracy and ROC–AUC values for all baseline machine learning models evaluated in this study.

Table 1. Sample of Table

Model	Accuracy	ROC–AUC
Logistic Regression	0.84	0.9041
Decision Tree	0.88	0.8800
K-Nearest Neighbor	0.84	0.9179
Support Vector Machine	0.88	0.9352
Random Forest	0.92	0.9708

Table 1 indicates that ensemble-based Random Forest outperforms all other baseline models in both accuracy and ROC–AUC. Although SVM and KNN demonstrate competitive discriminative ability, their overall accuracy remains lower than that of Random Forest.

3.3. Detailed Classification Performance Analysis

To provide a deeper insight into class-wise performance, precision, recall, and F1-score values for each baseline model are reported in Table 2.

Table 2. Classification Performance Metrics of Baseline Models

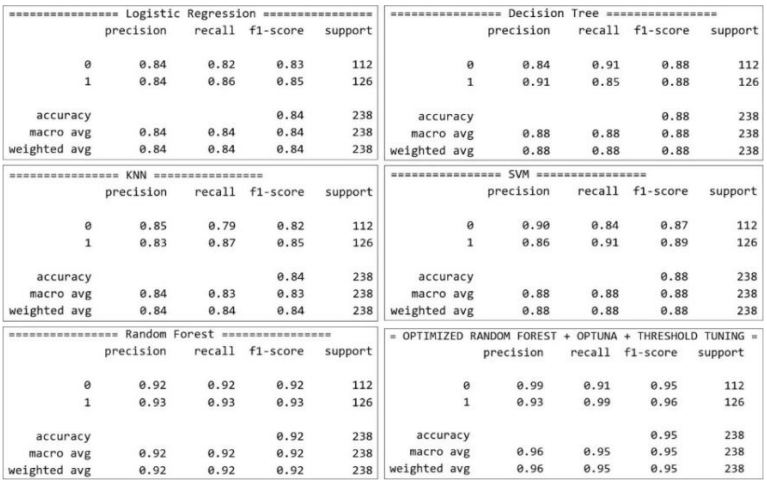
Model	Class	Precision	Recall	F1-score
Logistic Regression	0	0.84	0.82	0.83
	1	0.84	0.86	0.85
Decision Tree	0	0.84	0.91	0.88
	1	0.91	0.85	0.88
K-Nearest Neighbor	0	0.85	0.79	0.82
	1	0.83	0.87	0.85
Support Vector Machine	0	0.90	0.84	0.87
	1	0.86	0.91	0.89
Random Forest	0	0.92	0.92	0.92
	1	0.93	0.93	0.93

As shown in Table 2, Random Forest achieves consistently high precision, recall, and F1-score for both classes, indicating balanced performance. In contrast, Logistic Regression and KNN show relatively lower recall for class 0, while Decision Tree exhibits class imbalance tendencies.

In addition to the quantitative values reported in Table 2, the overall classification performance of all baseline models is visually summarized in Figure 6, which presents the complete classification reports for each classifier. Figure 2 provides a compact visualization of precision, recall, and F1-score

Vol. 7, No. 2, April 2026, Page. 1853-1870
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5684

distributions across both classes, enabling a direct visual comparison of class-wise performance consistency. The figure further confirms that Random Forest exhibits the most balanced and stable performance, with consistently higher F1-scores for both class 0 and class 1, whereas Logistic Regression and KNN show more variability in recall, particularly for the negative class.



Logistic Regression

	precision	recall	f1-score	support
0	0.84	0.82	0.83	112
1	0.84	0.86	0.85	126
accuracy			0.84	238
macro avg	0.84	0.84	0.84	238
weighted avg	0.84	0.84	0.84	238

Decision Tree

	precision	recall	f1-score	support
0	0.84	0.91	0.88	112
1	0.91	0.85	0.88	126
accuracy			0.88	238
macro avg	0.88	0.88	0.88	238
weighted avg	0.88	0.88	0.88	238

KNN

	precision	recall	f1-score	support
0	0.85	0.79	0.82	112
1	0.83	0.87	0.85	126
accuracy			0.84	238
macro avg	0.84	0.83	0.83	238
weighted avg	0.84	0.84	0.84	238

SVM

	precision	recall	f1-score	support
0	0.90	0.84	0.87	112
1	0.86	0.91	0.89	126
accuracy			0.88	238
macro avg	0.88	0.88	0.88	238
weighted avg	0.88	0.88	0.88	238

Random Forest

	precision	recall	f1-score	support
0	0.92	0.92	0.92	112
1	0.93	0.93	0.93	126
accuracy			0.92	238
macro avg	0.92	0.92	0.92	238
weighted avg	0.92	0.92	0.92	238

OPTIMIZED RANDOM FOREST + OPTUNA + THRESHOLD TUNING

	precision	recall	f1-score	support
0	0.99	0.91	0.95	112
1	0.93	0.99	0.96	126
accuracy			0.95	238
macro avg	0.96	0.95	0.95	238
weighted avg	0.96	0.95	0.95	238

Figure 2. Classification Reports of Baseline and Optimized Models

3.4. Confusion Matrix Analysis

The confusion matrices for all models are illustrated in Figure 3. To quantitatively support this visualization, Table 3 summarizes the number of correctly and incorrectly classified instances for each model.

Table 3. Confusion Matrix Summary of Baseline Models

Model	True Negative	False Positive	False Negative	True Positive
Logistic Regression	92	20	18	108
Decision Tree	102	10	19	107
K-Nearest Neighbor	89	23	16	110
Support Vector Machine	94	18	11	115
Random Forest	103	9	9	117

The Random Forest model produces the lowest number of misclassifications, particularly reducing false negatives, which is crucial in medical diagnosis where missing positive cases can have serious consequences.

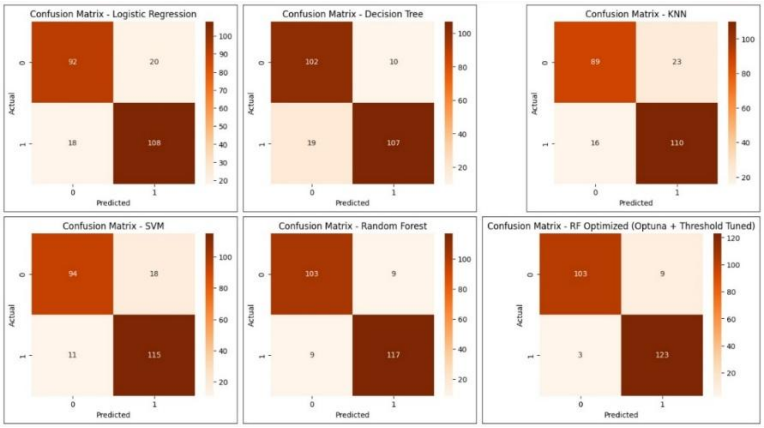


Figure 3. Confusion Matrix of Baseline and Optimized Models

Vol. 7, No. 2, April 2026, Page. 1853-1870
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5684

In Figure 3, the matrices present the distribution of true positive, true negative, false positive, and false negative predictions. The optimized Random Forest model exhibits a reduced number of misclassifications, particularly false negatives, highlighting its improved reliability for clinical decision support in heart disease prediction.

3.5. ROC Curve and Discriminative Ability

The ROC curves of all models are presented in Figure 3. The ROC–AUC values reported in Table 1 and visualized in Figure 4 demonstrate that Random Forest achieves the strongest discriminative capability among baseline models, with the ROC curve closely approaching the ideal upper-left corner.

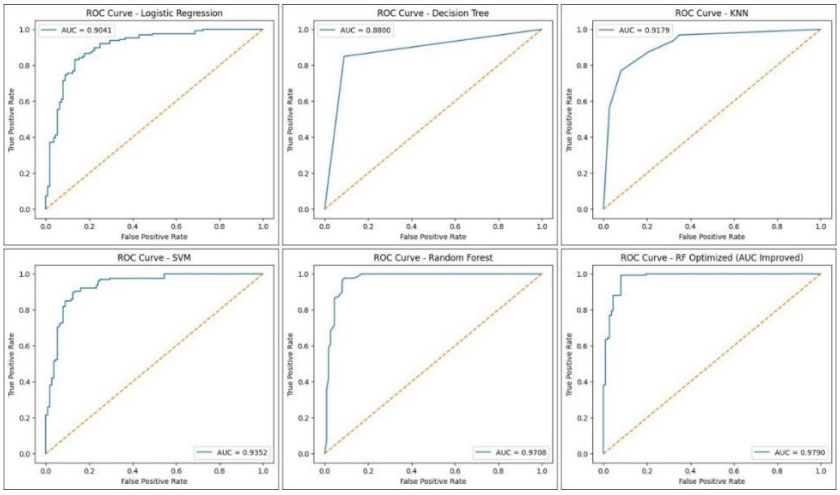


Figure 4. ROC Curves of Baseline and Optimized Models

Notably, while Logistic Regression and Decision Tree show acceptable discrimination, their ROC curves indicate weaker probability separation compared to ensemble-based and margin-based classifiers. The figure illustrates the trade-off between true positive rate and false positive rate across different decision thresholds. The optimized Random Forest model achieves the highest area under the ROC curve (ROC–AUC), indicating superior discriminative capability compared to other baseline classifiers.

3.6. Optimized Random Forest Performance

Following baseline evaluation, the Random Forest model was selected for further optimization based on its superior baseline performance. Hyperparameter optimization was performed using the Optuna framework to systematically explore the parameter space and identify the optimal model configuration. In addition, decision threshold optimization was applied to refine the classification boundary in order to balance sensitivity and specificity according to clinical risk priorities. The performance comparison between the baseline and optimized Random Forest models is presented in Table 4.

Table 4. Performance of Optimized Random Forest Model

Model	Accuracy	ROC–AUC
Baseline Random Forest	0.92	0.9708
Optimized Random Forest (Optuna + Threshold Tuned)	0.95	0.9790

Vol. 7, No. 2, April 2026, Page. 1853-1870
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5684

The optimized model demonstrates a clear improvement in both accuracy and ROC–AUC, confirming the effectiveness of the combined optimization strategy in enhancing predictive reliability and model generalization.

3.7. Optimized Model Classification and Error Reduction

The classification performance of the optimized Random Forest model shows a substantial reduction in misclassification errors, particularly false negatives in the heart disease class. The optimized decision threshold contributes to improved sensitivity toward positive cases, thereby minimizing the risk of undetected heart disease instances. This improvement is reflected in the confusion matrix and classification report visualizations (Figures 3 and 4), where more balanced class-wise performance and reduced error distribution can be observed. Table 5 presents the classification metrics of the optimized Random Forest model.

Table 5. Classification Performance of Optimized Random Forest

Class	Precision	Recall	F1-score
0 (No Heart Disease)	0.99	0.91	0.95
1 (Heart Disease)	0.93	0.99	0.96

Compared to the baseline model, the optimized Random Forest substantially reduces false negatives, as confirmed by the confusion matrix in Figure 3, where false negatives decrease from 9 to 3. This improvement enhances clinical safety by minimizing the risk of undetected heart disease cases. These results indicate that the integration of hyperparameter optimization and threshold tuning significantly enhances the clinical robustness of the model, making it more suitable for real-world diagnostic support applications. The improvement observed in accuracy and ROC–AUC indicates that the optimized model provides stronger discriminative capability in distinguishing between patients with and without heart disease. In the context of clinical decision support, this improvement is particularly important because higher discrimination performance enables more reliable identification of patients who are at potential cardiovascular risk.

In addition to overall accuracy, the increase in sensitivity and F1-score for the heart disease class has significant clinical implications. Higher sensitivity means that a larger proportion of patients with heart disease can be correctly identified by the system, thereby reducing the number of missed diagnoses (false negatives). In medical screening scenarios, minimizing false negatives is critical because undetected heart disease may delay treatment and increase the risk of severe complications.

Furthermore, the optimized decision threshold allows a better balance between sensitivity and specificity. While overall accuracy improves, the practical clinical value of the proposed approach lies in its ability to reduce false negative predictions while maintaining stable classification performance. This balance supports safer risk stratification and can assist clinicians in identifying high-risk patients who may require further diagnostic evaluation.

3.8. Summary of Experimental Results

Overall, the experimental results demonstrate that while baseline machine learning models provide reasonable predictive performance, the optimized Random Forest model achieves superior accuracy, discriminative capability, and classification reliability. The combined use of ensemble learning, Optuna-based hyperparameter optimization, and decision threshold tuning leads to a robust predictive framework that aligns with clinical decision support requirements. These results form the foundation of the proposed decision support output layer, enabling reliable risk classification for heart disease prediction.

Vol. 7, No. 2, April 2026, Page. 1853-1870
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5684

4. DISCUSSIONS

4.1. Clinical Interpretation and Practical Implications

The experimental results demonstrate that the proposed enhanced machine learning–based decision support system provides clinically meaningful improvements in heart disease prediction. The optimized Random Forest model achieves high accuracy and ROC–AUC, indicating strong discriminative capability in distinguishing between patients with and without heart disease. From a clinical perspective, this is particularly important because misclassification, especially false negatives, may lead to delayed diagnosis and increased cardiovascular risk.

The reduction in false negative cases observed in the optimized model directly contributes to improved patient safety. By increasing recall for the heart disease class, the proposed system enhances sensitivity toward patients at risk, which is essential for early screening and preventive intervention. In clinical workflows, such a system can be used as a secondary screening tool to support physicians in identifying high-risk individuals who may require further diagnostic testing, such as electrocardiography, echocardiography, or laboratory evaluation.

Furthermore, the integration of decision threshold optimization allows the system to be tailored to different clinical priorities. For example, in screening settings, a lower threshold may be adopted to prioritize sensitivity and minimize missed positive cases. In contrast, in resource-limited environments, a higher threshold may be selected to reduce false positives and optimize the allocation of diagnostic resources. This flexibility enhances the practical usability of the decision support system in real-world healthcare settings.

Another important consideration is the practical integration of the proposed framework into real-world clinical decision support systems. In clinical practice, data quality, interoperability with hospital information systems, and user acceptance by healthcare professionals may influence the successful deployment of machine learning models. The decision threshold optimization proposed in this study offers flexibility that can be adapted to different clinical priorities. For example, in screening scenarios, the threshold can be adjusted to prioritize sensitivity in order to minimize missed diagnoses, whereas in resource-limited environments it may be calibrated to balance sensitivity and specificity to avoid unnecessary follow-up examinations. This adaptability highlights the potential of the proposed approach to support diverse clinical decision-making contexts.

4.2. Implications for Decision Support System Design

The results highlight the importance of combining ensemble learning with systematic optimization strategies in clinical decision support systems. While baseline machine learning models already demonstrate reasonable predictive performance, the findings confirm that default hyperparameter configurations are insufficient to achieve optimal diagnostic reliability. The application of Optuna-based hyperparameter optimization enables automatic exploration of the parameter space, leading to a more robust and data-adaptive model configuration.

In addition, the inclusion of decision threshold tuning represents a critical extension beyond conventional machine learning evaluation. Many previous studies rely solely on fixed thresholds (typically 0.5), which may not align with the asymmetric costs of misclassification in medical diagnosis. By explicitly optimizing the decision threshold, the proposed system better reflects clinical risk management principles, where the cost of missing a true heart disease case is generally higher than that of generating a false alarm.

These design choices contribute to a more clinically aligned decision support system that not only maximizes statistical performance but also enhances interpretability and operational relevance. The confusion matrix and classification report visualizations further support transparency, enabling clinicians and system developers to understand error patterns and potential risks.

Vol. 7, No. 2, April 2026, Page. 1853-1870
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5684

4.3. Comparison with Previous Studies

Table 6. Comparison with Previous Studies on Heart Disease Prediction

Study	Method	Accuracy	ROC–AUC	Key Characteristics
Vu et al. (2025)	Random Forest, SVM	~0.73	~0.73	Baseline ML models with limited optimization
Saha et al. (2024	LR, KNN, SVM, RF	~0.91	~0.91	Comparative study using default configurations
Gayathri et al. (2024)	Hybrid ML + Reinforcement Learning	~0.94	N/A	Hybrid framework, sensitive to parameter settings
Dipu et al. (2025)	Ensemble & Gradient Boosting	0.90–0.92	High	Multi-dataset evaluation, performance variability
Wan et al. (2025)	Deep Learning Ensemble	~0.93–0.94	High	CNN-based ensemble, fixed decision threshold
Proposed Method (This Study) (2026)	**Optimized Random Forest + Optuna + Threshold Optimization**	**~0.95**	**~0.98**	**Optuna-based hyperparameter tuning + adaptive decision threshold**

To position the proposed method within recent developments in heart disease prediction, this subsection compares the performance of the proposed optimized Random Forest model with several recent studies. The comparison focuses on reported accuracy and ROC–AUC values, as well as key methodological characteristics, to highlight the relative improvements and contributions of the present study. In addition, recent review studies on machine learning for cardiovascular disease prediction indicate that most existing approaches achieve accuracies below 95% and often rely on fixed decision thresholds, which may not adequately address the asymmetric risks of misclassification in clinical diagnosis. This limitation further motivates the adoption of adaptive threshold optimization in the proposed framework.

4.4. Scientific and Clinical Contributions toward Decision Support Systems

The experimental results demonstrate that the proposed enhanced machine learning–based decision support system provides clinically meaningful improvements in heart disease prediction. The optimized Random Forest model achieves high accuracy and ROC–AUC, indicating strong discriminative capability in distinguishing between patients with and without heart disease. From a clinical perspective, this is particularly important because misclassification, especially false negatives, may lead to delayed diagnosis and increased cardiovascular risk.

Despite the promising results obtained in this study, several limitations should be acknowledged. First, the dataset used in this research was obtained from a publicly available repository and may not fully represent the diversity of patient populations encountered in real clinical environments. Variations in demographic characteristics, clinical practices, and data quality across healthcare institutions may influence model performance when deployed in different settings. Therefore, further validation using larger and more heterogeneous datasets from multiple healthcare systems would be necessary to confirm the generalizability and robustness of the proposed model.

5. CONCLUSION

This study proposed an enhanced machine learning–based decision support system for heart disease prediction by integrating ensemble learning, automated hyperparameter optimization, and decision threshold tuning within a unified predictive framework. Unlike many existing studies that rely on fixed model configurations and static classification thresholds, the proposed approach systematically optimizes both model parameters and decision policies to improve diagnostic reliability and clinical

Vol. 7, No. 2, April 2026, Page. 1853-1870
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5684

relevance. Experimental results demonstrate that the optimized Random Forest model, combined with threshold tuning, consistently outperforms baseline machine learning models in terms of accuracy, ROC-AUC, sensitivity, and F1-score. In particular, the improvement in sensitivity and the reduction of false negative cases are clinically significant, as they directly support earlier identification of individuals at high cardiovascular risk and help minimize missed diagnoses. These findings confirm that meaningful performance gains can be achieved not only through the adoption of more complex algorithms, but also through principled optimization strategies applied to well-established machine learning models. From a clinical decision support perspective, the proposed framework aligns with the final stage of the research workflow, namely the Decision Support Output, by providing risk predictions that are more reliable and actionable for clinicians. The integration of optimized prediction models with calibrated decision thresholds enhances the interpretability and practical utility of the system in screening and triage scenarios. Overall, this study contributes a robust and clinically oriented machine learning–based decision support system that offers both methodological advancements and tangible benefits for early cardiovascular disease risk assessment. Future research may further extend this framework by validating the proposed approach using larger and more diverse clinical datasets collected from multiple healthcare institutions. Such validation would help assess the generalizability and robustness of the model across different patient populations and clinical environments. In addition, the proposed methodology, which integrates automated hyperparameter optimization and decision threshold calibration, has the potential to be adapted for other predictive healthcare applications beyond heart disease, such as diabetes risk prediction, stroke detection, or other chronic disease screening tasks. Further studies may also explore the integration of this framework into real clinical decision support systems to evaluate its usability, scalability, and potential impact on healthcare decision-making.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest, whether financial or non-financial, between the authors and any parties related to the research object or the results presented in this article.

REFERENCES

[1] L. Yang, B. Zheng, and Y. Gong, "Global , regional and national burden of ischemic heart disease and its attributable risk factors from 1990 to 2021 : a systematic analysis of the Global Burden of Disease study," *BMC Cardiovascular Disorders*, vol. 25, no. August, pp. 1–15, 2025, doi: 10.1186/s12872-025-05022-x.

[2] X. Wu, J. Sha, Q. Yin, Y. Gu, and X. He, "Global burden of hypertensive heart disease and attributable risk factors , 1990 – 2021 : insights from the global burden of disease study," *Scientific Reports*, vol. 15, no. 14594, pp. 1–26, 2025, doi: 10.1038/s41598-025-99358-1.

[3] M. L. Costa and G. Cecatti, "Urgent and decisive action needed in maternal morbidity and mortality to prevent stagnation in progress," *The Lancet Global Health*, vol. 13, no. 4, pp. e600–e601, 2025, doi: 10.1016/s2214-109x(25)00069-5.

[4] M. S. Khan, I. Shahid, A. Bennis, A. Rakisheva, M. Metra, and J. Butler, "Global epidemiology of heart failure," *Nature Reviews Cardiology*, vol. 21, no. October, pp. 717–734, 2024, doi: 10.1038/s41569-024-01046-6.

[5] F. Rizky, M. Chaq, E. Chaq, Z. Multazam, A. Mustofa, and W. Socha, "The 30 Years of Shifting in The Indonesian Cardiovascular Burden-Analysis of The Global Burden of Disease Study Global Burden of Disease," *Journal of Epidemiology and Global Health*, vol. 14, no. March, pp. 193–212, 2024, doi: 10.1007/s44197-024-00187-8.

[6] B. Chong *et al.*, "Global burden of cardiovascular diseases : projections from 2025 to 2050," *European Journal of Preventive Cardiology*, vol. 32, no. 11, pp. 1001–1015, 2025, doi: 10.1093/eurjpc/zwae281.

[7] G. B. of C. D. and R. 2023 Collaborators, "Global, Regional, and National Burden of Cardiovascular Diseases and Risk Factors in 204 Countries and Territories, 1990-2023," *Journal*

of the American College of Cardiology, vol. 86, no. 22, pp. 2167–2243, 2025, doi: 10.1016/j.jacc.2025.08.015.

[8] P. Huang, Y. Guo, G. Ren, and L. Guo, "Global trends of cardiovascular disease burden attributable to high body mass index from 1990 to 2021 and projections to 2035," *Frontiers in Cardiovascular Medicine*, vol. 12, no. January, p. 1700540, 2026, doi: 10.3389/fcvm.2025.1700540.

[9] B. Chong *et al.*, "The Global Syndemic of Modi fi able Cardiovascular Risk Factors Projected From 2025 to 2050," *Journal of the American College of Cardiology*, vol. 86, no. 3, pp. 165–177, 2025, doi: 10.1016/j.jacc.2025.04.061.

[10] S. Xie, L. Zhang, Y. Zhang, Z. Zheng, J. Li, and S. Zhou, "Global, regional, and national burden of ischemic heart disease and ischemic stroke and their risk factors in youths and young adults aged 15-39 years (1990–2021): a comparative analysis of risk factors from global burden of disease study 2021," *Global Health Action*, vol. 18, no. 1, pp. 1–20, 2025, doi: 10.1080/16549716.2025.2560711.

[11] U. K. Lilhore *et al.*, "A deep learning approach for heart disease detection using a modified multiclass attention mechanism with BiLSTM," *Scientific Reports*, vol. 15, no. July, pp. 1–22, 2025, doi: 10.1038/s41598-025-09594-8.

[12] S. Liu, Y. Cui, and M. Chen, "Heart rate variability : a multidimensional perspective from physiological marker to brain-heart axis disorders prediction," *Frontiers in Cardiovascular Medicine*, vol. 12, no. November, pp. 1–16, 2025, doi: 10.3389/fcvm.2025.1630668.

[13] S. Dhandapani, H. Somasundaram, and T. Angamuthu, "Hybrid deep learning framework for heart disease prediction using ECG signal images," *Scientific Reports*, vol. 15, no. September, pp. 1–17, 2025, doi: 10.1038/s41598-025-10062-6.

[14] H. Bilal and Y. Muhammad, "Identification and diagnosis of chronic heart disease : A deep learning-based hybrid approach," *Alexandria Engineering Journal*, vol. 124, no. March, pp. 470–483, 2025, doi: 10.1016/j.aej.2025.03.025.

[15] V. Idakwo, "From diagnosis to management : unveiling the challenges of artificial intelligence solutions in cardiovascular healthcare," *BMC Digital Health*, vol. 3, no. September, pp. 1–15, 2025, doi: 10.1186/s44247-025-00187-z.

[16] A. Sharma *et al.*, "A Systematic Review on Machine Learning Intelligent Systems for Heart Disease Diagnosis," *Archives of Computational Methods in Engineering*, vol. 32, no. October, pp. 4303–4329, 2025, doi: 10.1007/s11831-025-10271-2.

[17] E. Church, K. Poppe, and S. Wells, "Scoping review of the use of multimorbidity variables in cardiovascular disease risk prediction," *BMC Public Health*, vol. 25, no. March, pp. 1–11, 2025, doi: 10.1186/s12889-025-22169-6.

[18] A. Mansouri *et al.*, "Health system barriers to cardiovascular disease prevention and control in Iran: a qualitative study," *BMC Health Services Research*, vol. 25, no. November, pp. 1–14, 2025, doi: 10.1186/s12913-025-13674-w.

[19] X. Y. Zhu, M. Q. Shi, Z. M. Jiang, X. Li, J. W. Tian, and F. F. Su, "Global, regional, and national burden of cardiovascular diseases attributable to metabolic risks across all age groups from 1990 to 2021: an analysis of the 2021 global burden of disease study data," *BMC Public Health*, vol. 25, no. May, pp. 1–21, 2025, doi: 10.1186/s12889-025-22702-7.

[20] J. Bhatt, S. Jain, and D. D. Bhatia, "Artificial intelligence (Al) in healthcare diagnosis: evidence-based recent advances and clinical implications," *Sensors and Diagnostics*, vol. 4, no. October, pp. 1047–1059, 2025, doi: 10.1039/d5sd00146c.

[21] H. Liu and R. K. Tripathy, "Machine Learning and Deep Learning for Healthcare Data Processing and Analyzing: Towards Data-Driven Decision-Making and Precise Medicine," *Diagnostics*, vol. 15, no. 8, pp. 1–6, 2025, doi: 10.3390/diagnostics15081051.

[22] Z. Bin Akhtar, "Artificial intelligence within medical diagnostics: A multi-disease perspective," *Artificial Intelligence in Health*, vol. 2, no. 3, pp. 44–62, 2025, doi: 10.36922/aih.5173.

[23] T. Y. Lee, "Application of Artificial Intelligence in Healthcare: Trends, Challenges in Disease Diagnosis and Management," *Theoretical and Natural Science*, vol. 87, no. 1, pp. 208–217, 2025, doi: 10.54254/2753-8818/87/2025.20353.

[24] C. M. Bertoncelli *et al.*, "PredictMed-CDSS : Artificial Intelligence-Based Decision Support

Vol. 7, No. 2, April 2026, Page. 1853-1870
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5684

System Predicting the Probability to Develop Neuromuscular Hip Dysplasia," *Bioengineering*, vol. 12, no. 8, pp. 1–19, 2025, doi: 10.3390/bioengineering12080846.

[25] Q. Abbas and W. Jeong, "Explainable AI in Clinical Decision Support Systems: A Meta-Analysis of Methods, Applications, and Usability Challenges," *Healthcare*, vol. 13, no. 17, pp. 1–28, 2025, doi: 10.3390/healthcare13172154.

[26] S. Kierner, P. Kierner, and J. Kucharski, "Combining machine learning models and rule engines in clinical decision systems : Exploring optimal aggregation methods for vaccine hesitancy prediction," *Computers in Biology and Medicine*, vol. 188, no. April, p. 109749, 2025, doi: 10.1016/j.compbiomed.2025.109749.

[27] Y. Kırelli, "Medical Decision Support Systems in Emergency Departments: An NLP and Machine Learning-Based Triage Model," *Arabian Journal for Science and Engineering*, no. December, p. 10924, 2025, doi: 10.1007/s13369-025-10924-1.

[28] Y. Liu, Z. Shi, C. Xiao, and B. Wang, "Computerized Medical Imaging and Graphics A deep learning-based clinical decision support system for glioma grading using ensemble learning and knowledge distillation," *Computerized Medical Imaging and Graphics*, vol. 124, no. July, p. 102602, 2025, doi: 10.1016/j.compmedimag.2025.102602.

[29] J. Miguel, L. Alcaraz, H. Bouma, and N. Strodthoff, "Enhancing clinical decision support with physiological waveforms — A multimodal benchmark in emergency care," *Computers in Biology and Medicine*, vol. 192, no. Part A, p. 110196, 2025, doi: 10.1016/j.compbiomed.2025.110196.

[30] M. Mahdavi, A. Kazemnejad, A. Asosheh, D. Khalili, and A. Tajari, "Development and validation of an artificial intelligence-based model for cardiovascular disease prediction using longitudinal data," *BMC Medical Informatics and Decision Making*, vol. 25, no. December, pp. 1–11, 2025, doi: 10.1186/s12911-025-03280-5.

[31] R. Başar, Ö. Ocak, A. Erturk, and M. de la Roche, "Leveraging Machine Learning Techniques to Predict Cardiovascular Heart Disease," *Information*, vol. 16, no. 8, pp. 1–16, 2025, doi: 10.3390/info16080639.

[32] L. Yu *et al.*, "Interpretable machine learning model for cardiovascular disease risk prediction: a feature decomposition-based study," *BMC Public Health*, vol. 25, no. October, pp. 1–9, 2025, doi: 10.1186/s12889-025-24921-4.

[33] C. S. Parsons, A. Zuiderwijk, N. A. Orchard, J. H. F. Oosterhoff, and M. De Reuver, "Task-Technology Fit of Artificial Intelligence- based clinical decision support systems : a review of qualitative studies," *BMC Medical Informatics and Decision Making*, vol. 25, no. October, pp. 1–16, 2025, doi: 10.1186/s12911-025-03237-8.

[34] N. Newton, A. Bamgboje-ayodele, R. Forsyth, A. Tariq, and M. T. Baysari, "A systematic review of clinicians' acceptance and use of clinical decision support systems over time," *npj Digital Medicine*, vol. 8, no. May, pp. 1–17, 2025, doi: 10.1038/s41746-025-01662-7.

[35] F. Yousefi, R. Dehnavieh, M. Laberge, and M. P. Gagnon, "Opportunities, challenges, and requirements for Artificial Intelligence (AI) implementation in Primary Health Care (PHC): a systematic review," *BMC Primary Care*, vol. 26, no. June, pp. 1–24, 2025, doi: 10.1186/s12875-025-02785-2.

[36] M. D. Teja and G. M. Rayalu, "Optimizing heart disease diagnosis with advanced machine learning models: a comparison of predictive performance," *BMC Cardiovascular Disorders*, vol. 25, no. March, pp. 1–12, 2025, doi: 10.1186/s12872-025-04627-6.

[37] G. Alwakid, F. U. Haq, N. Tariq, M. Humayun, M. Shaheen, and M. Alsadun, "Optimized machine learning framework for cardiovascular disease diagnosis: a novel ethical perspective," *BMC Cardiovascular Disorders*, vol. 25, no. February, pp. 1–28, 2025, doi: 10.1186/s12872-025-04550-w.

[38] T. Banerjee and İ. Paçal, "A systematic review of machine learning in heart disease prediction," *Turkish Journal of Biology*, vol. 49, no. 5, pp. 600–634, 2025, doi: 10.55730/1300-0152.2766.

[39] Z. Jiang *et al.*, "Development of an explainable machine learning model for 3-year cardiovascular risk prediction in new-onset type 2 diabetes using the TyG index and ultrasound features," *BMC Medical Informatics and Decision Making*, vol. 25, no. November, pp. 1–15, 2025, doi: 10.1186/s12911-025-03247-6.

Vol. 7, No. 2, April 2026, Page. 1853-1870
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5684

[40] B. Van Calster *et al.*, "Evaluation of performance measures in predictive artificial intelligence models to support medical decisions: overview and guidance," *The Lancet Digital Health*, vol. 7, no. 12, pp. 1–13, 2025, doi: 10.1016/j.landig.2025.100916.

[41] M. Nasir, N. S. Summerfield, S. Carreiro, D. Berlowitz, and A. Oztekin, "A machine learning approach for diagnostic and prognostic predictions, key risk factors and interactions," *Health Services and Outcomes Research Methodology*, vol. 25, no. 1, pp. 1–28, 2025, doi: 10.1007/s10742-024-00324-7.

[42] B. S. Kumar and L. M. A. Kumar, "Hyperparameter-Tuned Machine Learning Model for Accurate Prediction of Heart Disease," *Journal of Neonatal Surgery*, vol. 14, no. 31, pp. 135–143, 2025, doi: 10.63682/jns.v14i31S.7080.

[43] T. Vu *et al.*, "Machine Learning Model for Predicting Coronary Heart Disease Risk: Development and Validation Using Insights From a Japanese Population–Based Study," *JMIR Cardio*, vol. 9, no. e68066, pp. 1–15, 2025, doi: 10.2196/68066.

[44] S. Saha, M. M. Rahman, T. T. Suki, M. M. Alam, M. S. Alam, and M. A. S. Haque, "Heart Disease Prediction Using Machine Learning Algorithms: Performance Analysis," in *2024 3rd International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)*, IEEE, 2024, pp. 1–6. doi: 10.1109/ICAEEE62219.2024.10561820.

[45] R. Gayathri *et al.*, "Enhancing heart disease prediction with reinforcement learning and data augmentation," *Systems and Soft Computing*, vol. 6, no. August, p. 200129, 2024, doi: 10.1016/j.sasc.2024.200129.

[46] M. N. H. Dipu and M. H. Apu, "From data to diagnosis: Heart disease likelihood prediction using six integrated datasets and rigorous evaluation of some machine learning models for Android app deployment," *In Silico Research in Biomedicine*, vol. 2, no. December 2025, p. 100165, 2026, doi: 10.1016/j.insi.2025.100165.

[47] S. Wan, F. Wan, and X. Dai, "Machine learning approaches for cardiovascular disease prediction: A review," *Archives of Cardiovascular Diseases*, vol. 118, no. 10, pp. 554–562, 2025, doi: 10.1016/j.acvd.2025.04.055.

[48] M. Siddhartha, "Heart Disease Dataset (Comprehensive)," *IEEE Dataport*, no. 1, p. 3, 2025, doi: 10.21227/dz4t-cm36.

[49] T. J. H. Mim, M. Hassan, B. K. Paul, M. Sarwar, and H. Mollah, "Machine Learning Approaches for Cardiovascular Disease Prediction : A Comparative Study," *Biomedical Materials & Devices*, no. November, pp. 1–15, 2025, doi: 10.1007/s44174-025-00564-2.

[50] Y. Rimal, N. Sharma, S. Paudel, and A. Alsadoon, "Comparative analysis of heart disease prediction using logistic regression, SVM, KNN, and random forest with cross-validation for improved accuracy," *Scientific Reports*, vol. 15, no. 1, pp. 1–14, 2025, doi: 10.1038/s41598-025-93675-1 1.

[51] M. Gulhane *et al.*, "Comprehensive review of heart disease prediction: A comparative study from 2019 onwards," *Artificial Intelligence In Medicine*, no. January, p. 103354, 2026, doi: 10.1016/j.artmed.2026.103354.

[52] N. Nasution, F. B. Nasution, and M. A. Hasan, "Predicting Heart Disease Using Machine Learning : An Evaluation of Logistic Regression , Random Forest , SVM , and KNN Models on the UCI Heart Disease Dataset," *IT Journal Research and Development*, vol. 9, no. 2, pp. 140–150, 2025, doi: 10.25299/itjrd.2025.17941.