Journal of Intelligent Decision Support System (IDSS) 8 (2) (2025), pp. 74-84

Published by: Institute of Computer Science (IOCS)

Journal of Intelligent Decision Support System (IDSS)

Journal homepage: www.idss.iocspublisher.org

Comparison of coronary heart disease prediction using basic model and ensemble learning

Rachmat[1], Syamsul Bhahri Iskandar[2], Kasmawaru[3], Suherwin[4]

[1,4] *Teknik Informatika, Universitas Pejuang Republik Indonesia, Makassar, Indonesia*
[2] *Teknik Informatika, STMIK Kharisma Makassar, Makassar, Indonesia*
[3] *Teknik Informatika, Universitas Dipa Makassar, Makassar, Indonesia*

Article Info

Article history:
Received May 28, 2025
Revised Jun 20, 2025
Accepted Jun 30, 2025

Keywords:

Coronary heart disease
Ensemble learning
Machine learning
Model evaluation
Permutation importance

ABSTRACT

Coronary heart disease (CHD) remains one of the leading causes of death worldwide, highlighting the urgent need for accurate early detection. This study aims to compare the performance of various machine learning models—including Decision Tree, K-Nearest Neighbor (KNN), Logistic Regression, Random Forest, XGBoost, and Stacking Ensemble—in predicting CHD using the UCI Heart Disease Dataset. The models were evaluated using five metrics: accuracy, precision, recall, F1-score, and AUC. The results indicate that Stacking and Logistic Regression achieved the highest AUC scores (0.80), while XGBoost obtained the best F1-score (0.40). Simpler models such as Decision Tree and KNN showed relatively lower performance. In addition, feature importance analysis using permutation methods revealed that features like number of major vessels (ca), thalassemia (thal), ST depression (oldpeak), and age play a critical role in prediction accuracy. These findings demonstrate that ensemble learning approaches, especially Stacking and XGBoost, can effectively improve diagnostic performance and offer strong potential for clinical decision support systems (CDSS) in the early detection of coronary heart disease.

Corresponding Author:

Rachmat,
Teknik Informatika,
Universitas Pejuang Republik Indonesia,
Jl. Raya Baruga No.Raya, Antang, Kec. Manggala, Kota Makassar, Sulawesi Selatan 90234, Indonesia
Email: rachmat27udinus@gmail.com

Introduction

Coronary heart disease (CHD) is one of the leading causes of death globally and poses a major challenge to healthcare systems in many countries (Minja et al., 2022; Shao et al., 2020). According to World Health Organization (WHO) reports, CHD accounts for millions of deaths each year, particularly in developing and developed countries, where many cases are not diagnosed early due to limited access and suboptimal detection methods (Organization, 2023). CHD is characterized by narrowing of the coronary blood vessels due to the build-up of atherosclerotic plaque, which can lead to impaired blood flow to the heart muscle and potentially sudden cardiac arrest (Fick & Hillegass, 2022; Mir et al., 2024). Major risk factors that contribute to the development of CHD include hypertension, high cholesterol, obesity, diabetes, smoking and a sedentary lifestyle (Ciumărnean et al., 2021; Pająk et al., 2022). Therefore, accurate early detection is crucial to prevent serious complications and lower the mortality rate from this disease.

In recent decades, the development of artificial intelligence technology, especially machine learning, has opened up great opportunities in improving the accuracy of heart disease diagnosis non-

invasively and effectively. Machine learning can process complex and multivariate medical data to identify patterns that are difficult to detect by traditional methods (An et al., 2023; Shehab et al., 2022). Datasets such as the UCI Heart Disease Dataset are a key source for developing prediction models as they include a number of important clinical features such as age, blood pressure, cholesterol, maximum heart rate, and other variables relevant to the patient's condition. The use of this dataset allows the development of classification algorithms that are able to distinguish between patients with or without heart disease more systematically.

Various machine learning algorithms have been applied for CHD prediction, including Support Vector Machine (SVM), Decision Tree (DT), K-Nearest Neighbor (KNN), and Logistic Regression. For example, Zaki et al., (2020) used Logistic Regression and managed to achieve an accuracy of about 96,1% on the UCI dataset, although the model was less able to handle the complexity of non-linear relationships between features. Secondly,

Spencer et al., (2020) conducted experiments with Random Forest and SVM, resulting in increased accuracy, but their performance is highly dependent on the data preprocessing stage and the model parameters chosen. Research by Lin et al., (2022) introduced ensemble learning techniques such as AdaBoost that improved prediction stability and sensitivity, but the study was limited to only one type of ensemble method. Other studies, such as by Rahman et al., (2025), proposed stacking ensemble methods that combine multiple classifiers to optimize prediction, but the comparative evaluation against other ensemble techniques still lacks depth.

Most of these single models have limitations, especially in the face of imbalanced medical data and non-linearly interacting features. This data imbalance often results in the model being biased towards the majority class thus reducing sensitivity towards the minority class, i.e. patients who actually have heart disease. Therefore, ensemble learning techniques that combine results from multiple learning algorithms are expected to improve the stability, accuracy, and generalization of the model. Methods such as Bagging, Boosting, Voting, and Stacking have their respective advantages in overcoming overfitting and improving more consistent prediction performance (Mahajan et al., 2023). However, a comprehensive and systematic exploration of various ensemble techniques as well as hyperparameter tuning evaluation and feature sensitivity analysis in the context of UCI datasets is still very limited.

This research aims to fill this gap by designing and implementing several ensemble learning methods for coronary heart disease prediction using the UCI Heart Disease Dataset. The focus of the research includes evaluating model performance based on accuracy, precision, sensitivity, F1-score, and AUC score metrics. In addition, the research also applied permutation importance to perform feature sensitivity analysis. This method measures the contribution of each medical attribute by randomizing the feature value and observing its impact on model performance degradation, thus providing a real picture of the influence of each feature on prediction. This approach is essential to improve the interpretability of the model in the clinical domain. Thus, it is hoped that a prediction model can be obtained that is not only technically superior, but also reliable as a clinical decision support system (CDSS) for early detection of CHD accurately and efficiently.

The original contribution of this research lies in the comprehensive approach of exploring various ensemble methods directly which is still rare in the literature. Moreover, the integration of rigorous hyperparameter tuning, the use of multi-metric evaluation, and in-depth feature analysis will provide new insights for the development of predictive models adaptive to the complex and imbalanced characteristics of medical data. Thus, this research not only has strong theoretical value in the field of computer science and data mining, but also provides a significant practical impact on the medical world, especially in supporting efforts to prevent and treat coronary heart disease more effectively.

Method

2.1 Data Collection

In this study, the data used is historical heart disease data taken from the Heart Disease database which consists of 303 rows of data with 13 main features (Janosi et al., 1989). This study uses data from the Cleveland database. Data attributes include age, sex, cp (type of chest pain), trestbps (resting blood pressure), chol (cholesterol level), fbs (fasting blood sugar), restecg (resting

electrocardiography results), thalach (maximum heart rate), exang (exercise-induced angina), and oldpeak (exercise-induced ST depression). The target attribute is heart disease severity with integer values from 0 (no disease) to 4 (high severity). The following research dataset is shown in Table 1

Table 1 Research Dataset

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
63	1	1	145	233	1	2	150	0	2,3	3	0	6	0
67	1	4	160	286	0	2	108	1	1,5	2	3	3	2
67	1	4	120	229	0	2	129	1	2,6	2	2	7	1
37	1	3	130	250	0	0	187	0	3,5	3	0	3	0
41	0	2	130	204	0	2	172	0	1,4	1	0	3	0
.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.
45	1	1	110	264	0	0	132	0	1,2	2	0	7	1
68	1	4	144	193	1	0	141	0	3,4	2	2	7	2
57	1	4	130	131	0	0	115	1	1,2	2	1	7	3
57	0	2	130	236	0	2	174	0	0	2	1	3	1
38	1	3	138	175	0	0	173	0	0	1		3	0

2.2 Data Preprocessing

Data preprocessing is an important step in the data analysis process that involves cleaning and transforming raw data into a format suitable for further analysis. The goal is to improve data quality by addressing issues such as missing values, outliers, and discrepancies (Luengo et al., 2020). Some common techniques used in data preprocessing include:

1. Data Integration: Combining data from multiple sources.
2. Data Transformation: Adjusting data formats, such as normalization and encoding.
3. Data Cleaning: Handles missing values, errors, and outliers.
4. Data Reduction: Simplifying the data by feature selection or techniques such as PCA.

2.3 Split Data

The stage of splitting data into training and testing sets is an important step in the development of machine learning models. The training set, which usually includes about 80% of the data, is used to train the model to recognize patterns in the data. Meanwhile, about 20% of the data is allocated to the testing set to evaluate the model's ability to predict data that has never been seen before. Model performance evaluation is done with metrics such as accuracy, precision, sensitivity, F1-score, and AUC to assess how well the model works as well as identify potential overfitting.

2.4 Base Model

The base model is a machine learning algorithm that is used as an initial comparison before applying advanced models. The three basic models in this research are Decision Tree, K-Nearest Neighbors (KNN), and Logistic Regression.

2.4.1 Decision Tree

Decision Tree divides data based on the most informative attributes using metrics such as Entropy and Information Gain.

Entropy is calculated by :

$$Entropy(S) = \sum_{i=1}^{n} -pi * \log_2 pi \quad (1)$$

The higher the entropy, the more disorganized the data, meaning that the data is evenly spread across the various classes. Conversely, if the entropy is low, the data is more concentrated in one class (Abrori & Fatah, 2025).

While Gain is formulated as:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Gain indicates the effectiveness of attribute A in dividing the data. The larger the Gain value, the better the attribute is at reducing uncertainty. The attribute with the highest Gain is selected as the node in the Decision Tree (Abrori & Fatah, 2025).

2.4.2 K-Nearest Neighbors (KNN)

KNN classifies data based on the k nearest neighbors using a distance metric, the most common of which is Euclidean Distance:

a. Euclidean Distance Formula:

$$d(x, x') = \sqrt{\sum_{i=1}^{n} (x_i - x'_i)^2} \quad (3)$$

b. After calculating the distances of all the training data to the test data, the algorithm selects the k closest data, then determines the class based on the majority of the labels of the k neighbors (Lubis & Lubis, 2020).

2.4.3 Logistic Regression

Logistic Regression uses a sigmoid function to model the probability of the target class based on input features (Mustafa et al., 2023):

a. Sigmoid function:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (4)$$

b. Probability model:

$$P(y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^{n} \beta_i X_i)}} \quad (5)$$

c. Loss Function (binary cross-entropy):

$$L = -\frac{1}{m} \sum_{i=1}^{m} y_i \log(\hat{y}_i) + (1 - \hat{y}_i) \log(\hat{y}_i) + (1 - \hat{y}_i) \log(1 - \hat{y}_i)] \quad (6)$$

This model will estimate the parameters β menggunakan using the Maximum Likelihood Estimation (MLE) method.

2.5 Ensemble Learning

Ensemble Learning is a method in machine learning that combines several models into a single unit to produce more accurate predictions than using a single model (Rayadin et al., 2024). The concept of combining various machine learning approaches is expected to overcome the individual weaknesses of each model and create more accurate and stable predictions. The use of multiple learning models instead of a single model to unravel a given difficult problem has always been an accepted solution. Ensemble Learning methods have theoretically and empirically obtained better performance than weak single learners, especially when dealing with computationally high-dimensional, complex regression and classification problems (Shaikh et al., 2024). Ensemble Learning that can be used are as follows:

2.5. Bagging

Bagging (Bootstrap Aggregating) is one of the techniques in Ensemble Learning that aims to improve the stability and accuracy of prediction models by combining multiple learning models trained in parallel. In this method, each model is built using a subset of training data taken randomly by bootstrap sampling from the original dataset (Ngo et al., 2022).

Random Forest is one of the popular machine learning algorithms and is an application of bagging to decision trees. It builds a number of decision trees independently, where each tree is trained on a random subset of the available data and features. The process of randomly selecting features at each tree node is also a key characteristic of Random Forest that helps reduce correlation between trees (Salman et al., 2024). The main processes in Random Forest are as follows:

a. Bootstrap Sampling i.e., each decision tree is trained using random samples (with returns) from the original dataset:
b. Random Feature Selection i.e. each node selects a random feature subset F_t⊆F to determine the best split.
c. Prediction of the final tree (classification):

$$\hat{y}_i = mode\{h_1(x), h_2(x), \dots, h_T(x)\} \quad (7)$$

1. Prediction of the final tree (regresion):

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \quad (8)$$

2.5.1 Boosting

Boosting is an ensemble method that builds models sequentially with the aim of improving prediction accuracy. Each base model is trained on weighted training data, where these weights are set based on the previous model's error. In this way, boosting transforms weak learners into strong learners by gradually optimizing the loss function (Tyralis & Papacharalampous, 2021).

XGBoost (Extreme Gradient Boosting) is a highly efficient and scalable implementation of decision tree-based boosting. It combines regularization techniques, parallel processing, and missing data handling to improve performance and prevent overfitting. Unlike traditional boosting that only uses the first derivative of the loss function, XGBoost applies a second-order Taylor approximation for more accurate optimization and enables parallel computation using CPU multithreading (Sagi & Rokach, 2021). The final XGBoost model is a combination of K basic decision trees, where the prediction for the input data x is formulated as:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in \mathrm{F} \quad (9)$$

Objective function is written: :

$L = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k)$

For information, l is a loss function that describes the error between the predicted value and the actual value. Then Ω is a function used for regularization to prevent overfitting, as follows:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda\left|\left|w\right|\right|^2 \quad (10)$$

The value of T represents the number of leaves per tree, and w represents the weight of the leaves of each tree. After the second-order Taylor expansion of the objective function and other detailed calculations, the information gain of the objective function after each splitting is :

$$\text{gain} = \frac{1}{2}\left[\frac{\left(\sum_{i-I_L} g_i\right)^2}{\sum_{i-I_L} h_i + \lambda} + \frac{\left(\sum_{i-I_R} g_i\right)^2}{\sum_{i-I_R} h_i + \lambda} + \frac{\left(\sum_{i-I_L} g_i\right)^2}{\sum_{i-I} h_i + \lambda}\right] - \gamma \quad (11)$$

Description :

gain : Information gain describes how much information is gained from a separation.
$\sum_{i-I_L} g_i$: Total gradient of the incoming data to the left node after splitting.
$\sum_{i-I_L} h_i$: Total hessian of the data entering the left node after splitting. Hessian to adjust the gradient and give more weight to more significant observations.
$\sum_{i-I_R} g_i$: Total gradient of the data entering the right node after splitting.
$\sum_{i-I_R} h_i$: Total hessian of the data entering the right node after splitting.
$\sum_{i-I_L} g_i$: Total gradient of the data at the node before splitting.
$\sum_{i-I} h_i$: Total hessian of the data at the node before splitting.

λ : Regularization parameter used to control model complexity and prevent overfitting.

γ : Threshold to control tree growth. Node splitting will only occur if information gain is greater than the threshold value γ.

2.5.2 Stacking

Stacking is an ensemble learning technique that combines multiple models to improve prediction accuracy. The basic idea of stacking is to use a base learner to generate metadata for input and then use a meta-learner which is generally called a level 1 learner to process the metadata (Gu et al., 2022). The base learner is usually called level 0 learner, and the meta learner stacked on the base learner is called stacking.In this research, two models are used as base learners, namely K-Nearest Neighbors (KNN) and Support Vector Machine (SVM), and 1 model for meta learner, namely Logistic Regression.

a. K-Nearest Neighbors (KNN)
   Probability for *c* class:
   $$P(y = c|x) = \frac{1}{k}\sum_{i=1}^{k} 1(y_i = c) \quad (12)$$
   Where *k* adalah is the number of nearest neighbors, and $1(y_i = c)$ is the indicator that the i-th neighbor has label *c*.
b. Support Vector Machine (SVM) (dengan probability=True):
   After training, the SVM will generate a margin score I, then this score is converted into probability with a logistic
   sigmoid function:
   $$P(y - 1|x) = \frac{1}{1 + \exp{(Af(x) + B)}} \quad (13)$$
   Where AAA and BBB are the parameters studied using the Platt scaling method for probability calibration.
c. Logistics Regression
   Logistic Regression is used as a meta-learner that accepts *z* inputs (outputs from base learners).

2.6 Evaluation

At this point, the accuracy, accuracy, recall, and F1 score values are calculated using a confusion matrix. A confusion matrix is a matrix that shows the actual and predicted classification results of LXL size, Where LLL is referred to as the number of labels or classifications, and is interpreted as the number of classes (Rayhan & Setyohadi, 2021). In this study, a 2x2 confusion matrix is used because it contains two (2) lettering/classes The terminology for the 2x2 confusion matrix is shown in Figure 1 below.

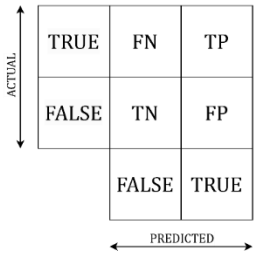


Figure 1 Confusion Matrix 2x2

The values for accuracy, precision, recall, and f1-score can be calculated using the following equations (Maulana et al., 2024) (8) - (11).

$$\text{Accuracy} = \frac{(\mathbf{TP+TN})}{(\mathbf{TP+TN+FP+FN})} \quad (14)$$

$$\text{Precision} = \frac{\mathbf{TP}}{(\mathbf{TP+FP})} \quad (15)$$

$$\text{Recall} = \frac{\mathbf{TP}}{(\mathbf{TP+FN})} \quad (16)$$

$$\text{F1-Score} = \frac{\mathbf{2*Precision*Recall}}{\mathbf{Precision+Recall}} \quad (17)$$

ROC-AUC is used to evaluate the performance of binary classification models. ROC Curve shows the relationship between True Positive Rate (TPR) and False Positive Rate (FPR) at various thresholds (Li, 2024).

$$\text{TPR} = \frac{\mathbf{TP}}{\mathbf{TP-FN}} \quad (18)$$

$$\text{FPR} = \frac{\mathbf{FP}}{\mathbf{FP-TN}} \quad (19)$$

AUC measures the area under the ROC curve. AUC values range from 0 to 1, where values close to 1 indicate excellent model performance. AUC helps measure the model's ability to distinguish between positive and negative classes, regardless of the threshold chosen.

To identify the most influential features, the Permutation Importance graph was used. This method randomizes the value of a feature in the validation data, and then recalculates the model performance. The decrease in performance after randomization indicates the level of influence the feature has on the model's predictions (Demir & Sahin, 2023).

$$PI_j = M_{Base} - M_{Perm(j)} \quad (20)$$

In other words, the greater the performance degradation after features are randomized, the more important they are to the model. This method is model-agnostic and provides a more tangible picture of the contribution of features to predictive ability, which is very useful in model interpretability, especially in clinical applications.

Results and Discussions

This study implements six classification models, namely three basic models and 3 ensemble models for coronary heart disease (CHD) prediction using the UCI Heart Disease dataset. These models are Decision Tree, K-Nearest Neighbor (KNN), Logistic Regression, Random Forest, XGBoost, and Stacking Ensemble. Performance evaluation is carried out using five main metrics namely Accuracy, Precision, Recall, F1-score, and Area Under Curve (AUC). The results of performance comparison between models are presented in Figure 2 and Table 2 below:

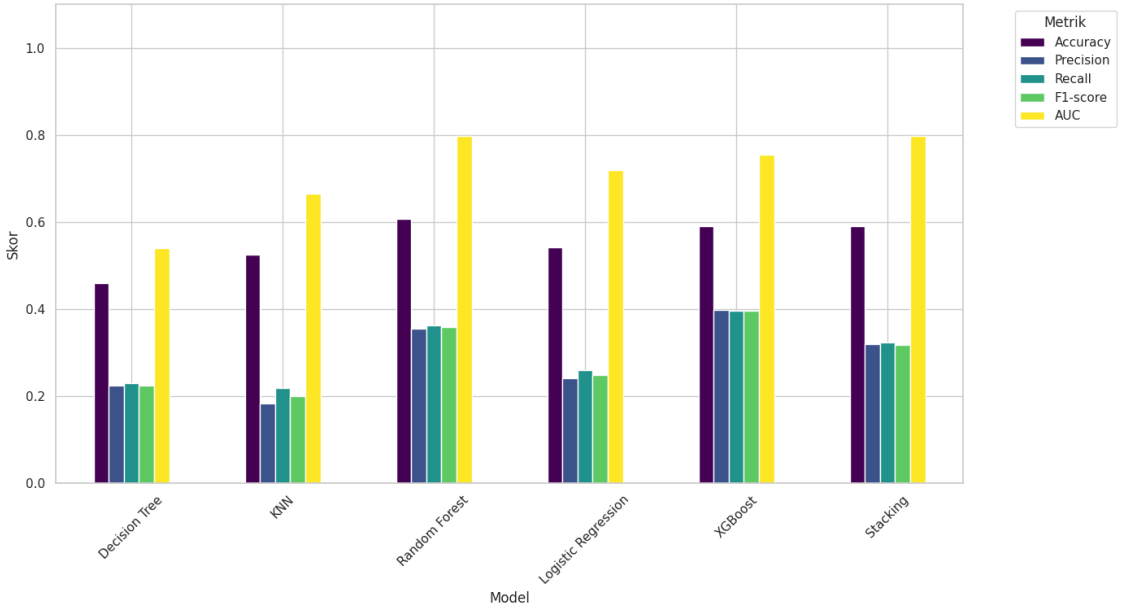


Figure 2 Comparison of Model Performance Visualization

Figure 2 shows that the Stacking and Logistic Regression models have the highest AUC score of 0.80, which indicates the model's ability to optimally distinguish between positive and negative classes. Meanwhile, the XGBoost model shows a balanced performance on all metrics with the highest F1-score of 0.40, followed by Logistic Regression with an F1-score of 0.36. The Decision Tree and KNN models showed low performance on all metrics. For example, Decision Tree only achieved an accuracy of 0.46 and an F1-score of 0.22, while KNN had the lowest precision of 0.18.

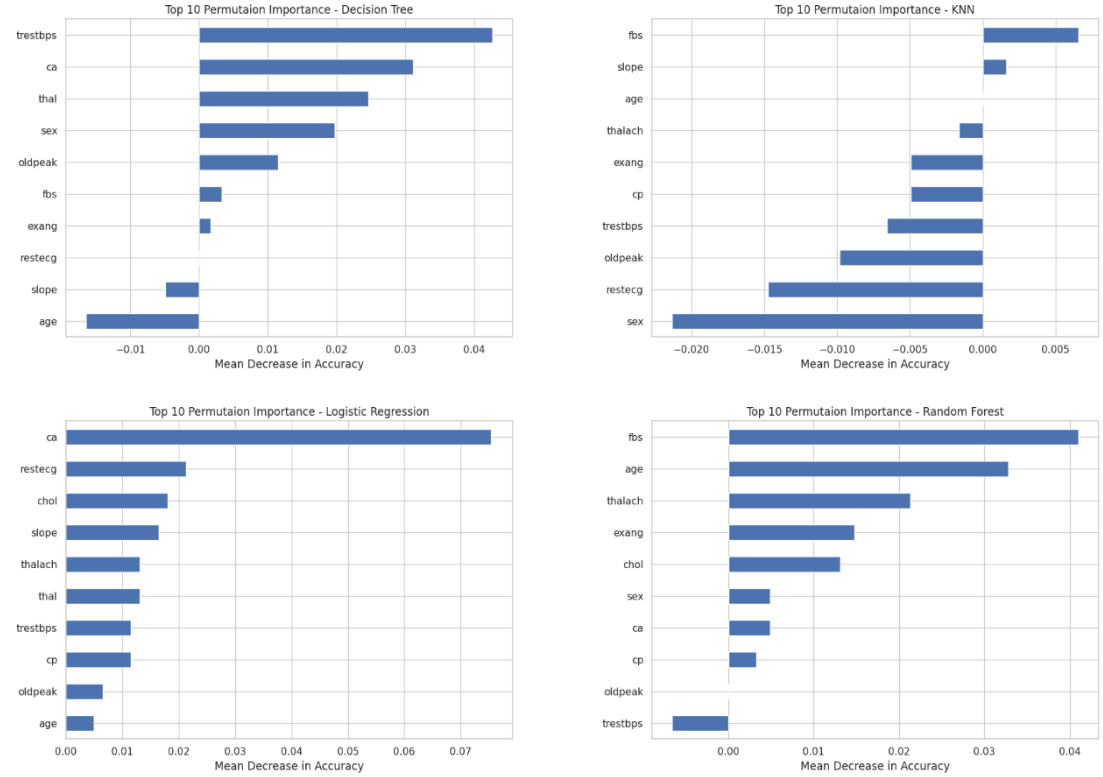
Table 2. Comparison of Model Performance

	Accuracy	Precision	Recall	F1-Score	AUC
Decision Tree	0,46	0,22	0,23	0,22	0,54
KNN	0,52	0,18	0,22	0,20	0,66
Logistic Regression	0,61	0,36	0,36	0,36	0,80
Random Forest	0,54	0,24	0,26	0,25	0,72
XGBoost	0,59	0,40	0,39	0,40	0,76
Stacking	0,59	0,32	0,32	0,32	0,80

From these results it can be concluded that ensemble models such as Stacking and XGBoost generally give better results than individual models. This shows the superiority of the ensemble approach in overcoming the complexity of unbalanced medical data.

Analisis Permutation Feature Importance

To understand the contribution of each feature in heart disease prediction, Permutation Feature Importance analysis was performed on all models. The results are presented in



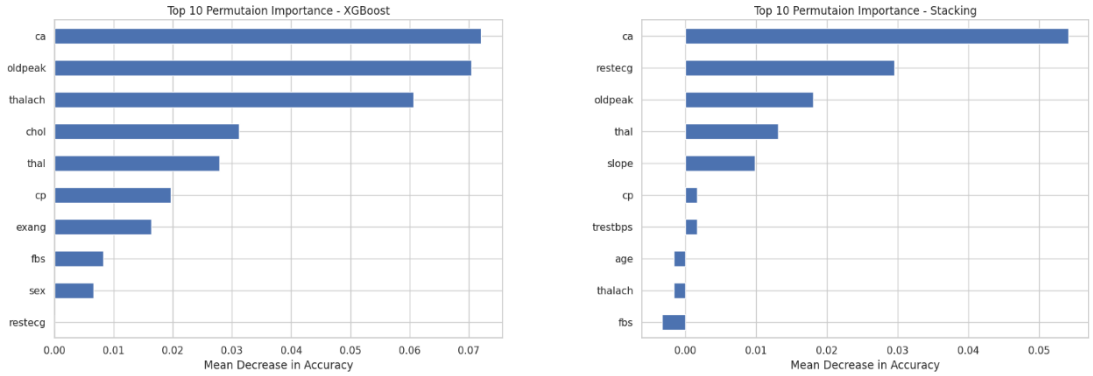


Figure 3 Comparison of Permutation Feature Importance

Figure 3 shows that the features that appear most often as important across models are:
a. ca (number of major blood vessels exposed to fluoroscopy)
b. thal (thalium test result)
c. oldpeak (exercise-related ST depression)
d. age and sex
e. trestbps (resting blood pressure)

Models such as Random Forest and XGBoost considered age, ca, oldpeak, and thalach (maximum heart rate) as features that strongly influenced prediction. In Logistic Regression, the ca feature had a dominant influence with the highest average decrease in accuracy when its value was randomized. These features are clinically associated with heart disease risk, strengthening the interpretive reliability of the model results. This is important in the context of developing a Clinical Decision Support System (CDSS) as it allows clinicians to understand the basis of decisions taken by the model.

Discussion

From the results of the evaluation and analysis of the above features, several important points can be concluded:

1. The ensemble model consistently outperforms the single model in terms of classificationperformance, especially in the AUC and F1-score metrics.
2. Stacking is the most balanced model, with high AUC and stable performance in other metrics, making it a strong candidate for CDSS.
3. Feature analysis showed that the features considered important by the model were mostly consistent with existing medical knowledge, such as ca, thal, and oldpeak.Simple models such as KNN and Decision Tree tend to underperform due to limitations in capturing non-linear interactions and sensitivity to data distribution.

Thus, the results of this study corroborate the findings that the use of ensemble techniques, particularly Stacking and XGBoost, as well as an interpretability approach through permutation importance, can provide robust and relevant prediction models to support early diagnosis of coronary heart disease.

Conclusions

This study compares the performance of several classification algorithms in predicting heart disease risk using the Heart Disease dataset. The analysis showed that the Stacking and Logistic Regression models performed best based on the Area Under Curve (AUC) value of 0.80, which indicates excellent classification ability. In addition, the XGBoost algorithm recorded the highest F1-score value of 0.40, which reflects a balance between precision and recall. Features such as ca, thal, oldpeak and age proved to have the most significant influence on prediction. This finding confirms that ensemble models, especially Stacking and XGBoost, have great potential in supporting early heart disease prediction

accurately. Based on these findings, it is recommended that future research use larger and more diverse datasets to improve the generalizability of the model to a wider population. In addition, deep learning approaches can be explored to improve the accuracy and complexity of predictive models, especially in handling nonlinear and complex medical data. Furthermore, the development of the model into a Clinical Decision Support System (CDSS) should also be considered for integration in clinical practice, so that it can make a real contribution to data-driven medical decision making.

References

Abrori, S., & Fatah, Z. (2025). *Implementasi Metode Decission Tree Dalam Mengklasifikasi Depresi Menggunakan Rapidminer*. *5*(2), 123–132.

An, Q., Rahman, S., Zhou, J., & Kang, J. J. (2023). A comprehensive review on machine learning in healthcare industry: classification, restrictions, opportunities and challenges. *Sensors*, *23*(9), 4178.

Ciumărnean, L., Milaciu, M. V., Negrean, V., Orăşan, O. H., Vesa, S. C., Sălăgean, O., Iluţ, S., & Vlaicu, S. I. (2021). Cardiovascular risk factors and physical activity for the prevention of cardiovascular diseases in the elderly. *International Journal of Environmental Research and Public Health*, *19*(1), 207.

Demir, S., & Sahin, E. K. (2023). An investigation of feature selection methods for soil liquefaction prediction based on tree-based ensemble algorithms using AdaBoost, gradient boosting, and XGBoost. *Neural Computing and Applications*, *35*(4), 3173–3190.

Fick, A., & Hillegass, E. (2022). Ischemic cardiovascular conditions and other vascular pathologies. *Essentials of Cardiopulmonary Physical Therapy-E-Book*, 51.

Gu, J., Liu, S., Zhou, Z., Chalov, S. R., & Zhuang, Q. (2022). A stacking ensemble learning model for monthly rainfall prediction in the Taihu Basin, China. *Water*, *14*(3), 492.

Janosi, A., Steinbrunn, W., Pfisterer, M., & Detrano, R. (1989). Heart Disease [Dataset]. In *UCI Machine Learning Repository*. https://doi.org/10.24432/C52P4X

Li, J. (2024). Area under the ROC Curve has the most consistent evaluation for binary classification. *PLoS One*, *19*(12), e0316019.

Lin, S., Zheng, H., Han, B., Li, Y., Han, C., & Li, W. (2022). Comparative performance of eight ensemble learning approaches for the development of models of slope stability prediction. *Acta Geotechnica*, *17*(4), 1477–1502.

Lubis, A. R., & Lubis, M. (2020). Optimization of distance formula in K-Nearest Neighbor method. *Bulletin of Electrical Engineering and Informatics*, *9*(1), 326–338.

Luengo, J., García-Gil, D., Ramírez-Gallego, S., García, S., & Herrera, F. (2020). Big data preprocessing. *Cham: Springer*, *1*, 1–186.

Mahajan, P., Uddin, S., Hajati, F., & Moni, M. A. (2023). Ensemble learning for disease prediction: A review. *Healthcare*, *11*(12), 1808.

Maulana, R., Narasati, R., Herdiana, R., Hamonangan, R., & Anwar, S. (2024). Komparasi Algoritma Decision Tree Dan Naive Bayes Dalam Klasifikasi Penyakit Diabetes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(6), 3865–3870. https://doi.org/10.36040/jati.v7i6.8265

Minja, N. W., Nakagaayi, D., Aliku, T., Zhang, W., Ssinabulya, I., Nabaale, J., Amutuhaire, W., de Loizaga, S. R., Ndagire, E., & Rwebembera, J. (2022). Cardiovascular diseases in Africa in the twenty-first century: gaps and priorities going forward. *Frontiers in Cardiovascular Medicine*, *9*, 1008335.

Mir, M. A., Dar, M. A., & Qadir, A. (2024). Exploring the Landscape of Coronary Artery Disease: A Comprehensive Review. *Am. J. Biomed. Pharm*, *1*, 9–22.

Mustafa, H., Mohamed, C., Nabil, O., & Noura, A. (2023). Machine learning techniques for diabetes classification: A comparative study. *International Journal of Advanced Computer Science and Applications*, *14*(9).

Ngo, G., Beard, R., & Chandra, R. (2022). Evolutionary bagging for ensemble learning. *Neurocomputing*, *510*, 1–14.

Organization, W. H. (2023). *Global report on hypertension: the race against a silent killer*. World Health Organization. https://doi.org/10.24432/C52P4X.

Pająk, A., Jankowski, P., & Zdrojewski, T. (2022). The burden of cardiovascular disease risk factors: A

current problem. *Polish Heart Journal (Kardiologia Polska)*, *80*(1), 5–15.

Rahman, A., Khan, M. S. I., Eidmum, M. D. Z. A., Shaha, P., Muiz, B., Hasan, N., Debnath, T., Kundu, D., Tamanna, J. T., & Sayduzzaman, M. (2025). Stacked Ensemble Method: An Advanced Machine Learning Approach for Anomaly-based Intrusion Detection System. *Statistics, Optimization & Information Computing*.

Rayadin, M. A., Musaruddin, M., Saputra, R. A., & Isnawaty, I. (2024). Implementasi Ensemble Learning Metode XGBoost dan Random Forest untuk Prediksi Waktu Penggantian Baterai Aki. *BIOS: Jurnal Teknologi Informasi Dan Rekayasa Komputer*, *5*(2), 111–119.

Rayhan, Y., & Setyohadi, D. B. (2021). Classification of grape leaf disease using convolutional neural network (CNN) with pre-trained model VGG16. *2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, 1–5.

Sagi, O., & Rokach, L. (2021). Approximating XGBoost with an interpretable decision tree. *Information Sciences*, *572*, 522–542.

Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning*, *2024*, 69–79.

Shaikh, T. A., Rasool, T., Verma, P., & Mir, W. A. (2024). A fundamental overview of ensemble deep learning models and applications: systematic literature and state of the art. *Annals of Operations Research*, 1–77.

Shao, C., Wang, J., Tian, J., & Tang, Y. (2020). Coronary artery disease: from mechanism to clinical practice. *Coronary Artery Disease: Therapeutics and Drug Discovery*, 1–36.

Shehab, M., Abualigah, L., Shambour, Q., Abu-Hashem, M. A., Shambour, M. K. Y., Alsalibi, A. I., & Gandomi, A. H. (2022). Machine learning in medical applications: A review of state-of-the-art methods. *Computers in Biology and Medicine*, *145*, 105458.

Spencer, R., Thabtah, F., Abdelhamid, N., & Thompson, M. (2020). Exploring feature selection and classification methods for predicting heart disease. *Digital Health*, *6*, 2055207620914777.

Tyralis, H., & Papacharalampous, G. (2021). Boosting algorithms in energy research: a systematic review. *Neural Computing and Applications*, *33*(21), 14101–14117.

Zaki, Z., Shah, M. A., Wakil, K., & Sher, F. (2020). Logistic regression based human activities recognition. *J. Mech. Contin. Math. Sci*, *15*(4), 228–246.