



Performance Comparison of Supervised Learning Algorithms in Heart Disease Risk Classification

Fatimah Azzahra¹, Muhammad Rafiq Pohan^{2*},
Ainul Mardhiah Binti Mohammed Rafiq³, Imran Hazim Bin Abdullah Salim⁴,
Azwa Nurnisya Binti Ayub⁵, Nuralya Medina Binti Mohammad Nizam⁶

^{1,2}Department of Information System, Faculty of Science and Technology,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

^{3,4}Department of Computer Science, Faculty of Information and Communication Technology,
International Islamic University Malaysia, Malaysia

^{5,6}Department of Information System, Faculty of Information and Communication Technology,
International Islamic University Malaysia, Malaysia

E-Mail: ¹12250324568@students.uin-suska.ac.id, ²12250311460@students.uin-suska.ac.id,
³mrdhainul06@gmail.com, ⁴imranhazim02@gmail.com,
⁵Azwannurnisya612@gmail.com, ⁶medinamohdnizam@gmail.com

Received Jun 12th 2025; Revised Jan 15th 2026; Accepted Jan 27th 2026; Available Online Jan 31th 2026
Corresponding Author: Muhammad Rafiq Pohan
Copyright © 2026 by Authors, Published by Institute of Research and Publication Indonesia (IRPI)

Abstract

Heart disease is one of the leading causes of death worldwide, so early diagnosis is essential for effective treatment and prevention. This study evaluates the performance of five machine learning algorithms, namely Decision Tree, Naive Bayes, K-Nearest Neighbors (K-NN), Random Forest, and Support Vector Machine (SVM), using a heart disease dataset obtained from Kaggle. The dataset comprises 14 variables, including 13 attributes and 1 target variable. The models were tested on data split ratios of 70:30, 80:20, and 90:10, with performance measured by accuracy, precision, recall, and F1-score. The results showed that Random Forest performed best, achieving the highest accuracy of 98.54% at the 80:20 ratio. Decision Trees followed, yielding similar results, while K-NN performed the best at the 90:10 ratio, with a precision of 100%. In contrast, Naive Bayes performed lower due to high false positives. These results are in line with previous studies, confirming the effectiveness of Random Forest and Decision Trees in predicting heart disease risk. This research contributes to the development of reliable machine-learning models for early diagnosis and risk assessment of heart disease.

Keywords: Heart Disease, Machine Learning, Model Evaluation, Random Forest, Supervised Learning

1. INTRODUCTION

Data from the WHO show that heart disease remains a major global health challenge, with an estimated 19.8 million deaths in 2022, representing approximately 32% of all global deaths. Of these, about 85% were due to heart attacks and strokes, with more than three-quarters occurring in low- and middle-income countries. Although conventional diagnostic methods such as Electrocardiogram (ECG), Magnetic Resonance Imaging (MRI), and angiography are available, the need for rapid, accurate, and affordable early detection systems is urgent to reduce premature mortality, which reaches 38% among non-communicable disease patients under the age of 70. Heart disease is a condition involving disorders of the heart and its blood flow [1]. These diseases range from blood vessel problems to arrhythmias [2]. The diagnosis of heart disease is usually based on the patient's medical history, blood pressure, blood sugar, and cholesterol tests, as well as tests such as an ECG, stress test, X-ray, echocardiogram, coronary angiography, and MRI and CT scans [3]. Several risk factors play a major role in the development of heart disease, including smoking, excessive alcohol consumption, obesity, inactive lifestyle, chronic stress, and genetic diseases. In addition, people between the ages of 40 and 50 have an increased risk of heart disease [4], [5]. Heart disease is responsible for 32% of deaths worldwide each year, killing about 17 million people each year [6]. Therefore, early diagnosis is essential for effective treatment [7].

Supervised learning is particularly suitable for this domain because it uses labeled datasets in which the 'target' variable, representing the presence or absence of heart disease, serves as the ground truth. By training on these historical labels, the algorithms can learn a mapping function to predict outcomes for new, unseen patient data, effectively mimicking the clinical diagnostic process. This approach serves as a reliable decision-

support system, providing clinicians with data-driven insights to clarify diagnoses and improve treatment efficacy. In the medical field, data analysis is increasingly used to clarify diagnoses, improve research techniques, and plan equipment accordingly. Artificial intelligence-based software is now available to analyze data and predict optimal outcomes [8]. a study by Ali et al. [9] found that the use of supervised learning and data mining for predicting and detecting heart disease has promising clinical potential, although their development remains challenging. Early prediction provides important information to doctors, enabling them to diagnose and treat patients more effectively [10]. In this study, several supervised machine learning algorithms, such as K-Nearest Neighbors (K-NN), Decision Trees, and Random Forests, were evaluated to find diagnostic accuracy [9].

In previous research by Sumwiza et al. [11], the Random Forest-based ensemble method, combined with feature selection techniques, achieved 99% accuracy. This is compared to other models such as Support Vector Machine (SVM), K-NN, and Logistic Regression. The Random Forest algorithm is very effective because it can handle complex and imbalanced data, provides reliable results through ensemble methods, and has flexibility [12]. K-NN also shows superiority in predicting heart disease, achieving the highest accuracy of 99.71%, which can reach 100% after feature extraction using only four attributes. K-NN is the best choice for predictive analysis based on Health datasets because the kappa statistic of 0.9941 indicates a very reliable classification [1]. The K-NN algorithm is very useful for dynamic exercise monitoring datasets because it is very easy to use, flexible, and can recognize local patterns in the data [13].

The SVM achieves a maximum accuracy of 93.3%, which is higher than that of other methods reported by Support et al. [14]. SVM can also effectively reduce error in heart disease prediction and is preferred for its good generalization, strong theoretical basis, and resilience to dimensionality [15]. Research comparing the performance of Decision Tree and Naïve Bayes algorithms for predicting heart disease shows that Naïve Bayes is superior to Decision Tree because it is better suited to numerical data and exhibits better generalization [16]. In contrast, Decision Tree tends to have lower performance because it is prone to overfitting, especially if its depth is not set, and is sensitive to data distribution and outliers [17].

However, as shown by previous research [18], a Decision Tree can achieve competitive results, with 99.0% accuracy on heart disease datasets. This success is supported by several factors, such as effective data preprocessing [19], model parameter tuning [20], and dataset characteristics relevant to rule-based patterns that the Decision Tree can capture [18]. Based on these studies, this study aims to compare the performance of supervised learning algorithms in heart disease classification, namely Random Forest, K-NN, Naïve Bayes, SVM, and Decision Tree, with heart disease datasets [10], [12], [14], [16], [18]. It is hoped that the results of this study can provide insight into the development of more accurate and effective methods for diagnosing heart disease.

2. MATERIAL AND METHOD

This research methodology is designed to compare the performance of classification algorithms (C4.5, K-Nearest Neighbor, Naive Bayes, Random Forest, and SVM) in predicting heart disease. These five algorithms were selected for their proven efficacy and widespread use in medical diagnostic research. Random Forest and Decision Tree were chosen for their ability to handle complex medical datasets and capture rule-based patterns. K-NN was included for its reliability in classifying health data and its flexibility in capturing local patterns. SVM provides a robust solution for binary classification in multidimensional spaces while avoiding overfitting. Lastly, Naive Bayes was selected for its efficiency in handling numerical data and its strong generalization capabilities. By comparing these diverse approaches ranging from tree-based ensembles to probabilistic and distance-based models, this study aims to identify the most stable architecture for heart disease risk assessment. The stages of the research methodology are described in Figure 1.

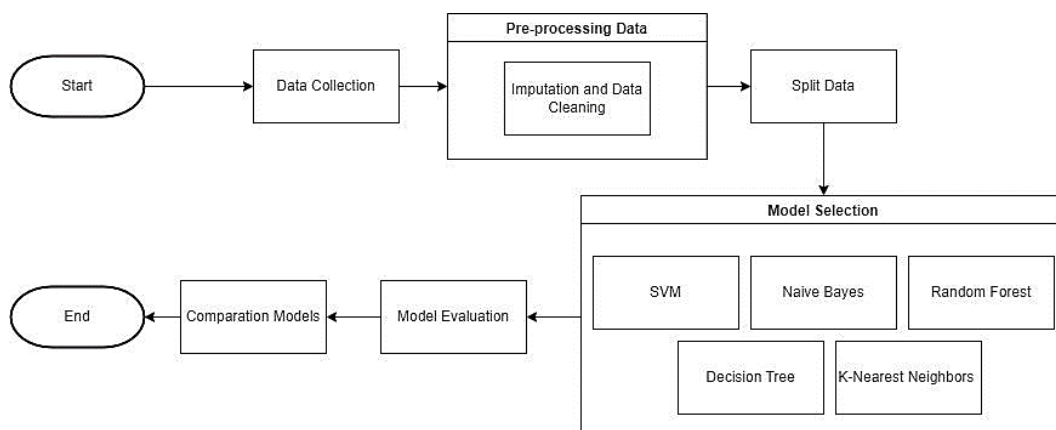


Figure 1. Research Methodology

2.1. Data collecting

The dataset used comes from the Kaggle platform, which contains data on patients with risky heart disease. This dataset contains 14 variables with a total of 1025, each with 13 characteristics, and one class used to predict heart disease. Demographic information, health test results, and medical history are entered into the patient data records. Heart disease is the target variable, indicating whether the patient has heart disease (1) or not (0). This dataset has been used in several studies. One of them is a study by Mohammed Amine Bouqentar et al. (2024), who used this dataset to train various machine learning algorithms, including Decision Trees, RandomForestst, and Support VectoMachinesne. This study aimed to predict heart disease early with an accuracy of 92 percent through ten-fold cross-validation. The results showed that this dataset can help in the development of heart disease prediction models [3].

Tabel 1. DataSet

No	Attribute	Type	Description
1	Age	Integer	The patient's age in years.
2	Sex	Category	Gender of the patient (1 = male; 0 = female).
3	Chest Pain Type	Category	Type of chest pain (1: typical angina, 2: atypical angina, 3: non-anginal pain, 4: asymptomatic).
4	Resting Blood Pressure	Integer	Resting blood pressure in mm Hg upon hospital admission.
5	Cholestoral	Integer	Serum cholestoral level measured in mg/dl.
6	Fasting Blood Sugar	Category	Fasting blood sugar > 120 mg/dl (1 = true; 0 = false).
7	Rest ECG	Category	Resting electrocardiographic results (0: normal, 1: ST-T wave abnormality, 2: left ventricular hypertrophy).
8	Max Heart Rate	Integer	Maximum heart rate achieved (thalach).
9	Exercise Induced Angina	Category	Presence of chest pain/angina induced by exercise (1 = yes; 0 = no).
10	Oldpeak	float64	ST depression induced by exercise relative to rest.
11	Slope	Category	The slope of the peak exercise ST segment (1: upsloping, 2: flat, 3: downsloping).
12	Vessels Colored by Fluorescein	Category	Number of major vessels (0-3) colored by fluoroscopy.
13	Thalassemia	Category	Status of thalassemia blood disorder (3 = normal; 6 = fixed defect; 7 = reversible defect).
14	Target	Integer	Diagnosis of heart disease (0 = No Heart Disease; 1 = Heart Disease).

2.2. Pre-processing Data

Data preprocessing is a part of machine learning that transforms raw data into a more organized and understandable format [21]. This data usually comes from various sources and is not ready for direct analysis, so it needs to be checked for missing features, outliers, or incorrect information, and then cleaned from the data set [11]. Before performing analysis on the dataset, it is very important to fill in the missing information [22].

2.3. Splitting data

Data splitting is the process of dividing a dataset into several parts in order to evaluate the performance of a particular model or algorithm. Generally, a dataset is separated into training data and testing data, where the training data are used to train the algorithm, while the testing data are used to assess the model's performance [23]. In this study, the dataset was divided into three training–testing scenarios with ratios of 70:30, 80:20, and 90:10. Using multiple data split ratios aims to evaluate the stability and consistency of model performance across different proportions of training data.

The 70:30 ratio was used as a standard split to balance training and testing. The 80:20 ratio was selected to observe potential performance improvements as the amount of training data increased. Meanwhile, the 90:10 ratio was used to analyze model behavior when trained on the maximum available data and to examine the potential risk of overfitting. Comparing model performance across these three split scenarios allows for a more comprehensive evaluation of the model's generalization capability.

2.4. Decision tree

Decision Tree is a popular and useful algorithm for predicting and classifying medical data[24]. A decision tree has a tree-like structure for regression and classification. Each internal node is depicted by a rectangle, and leaf nodes by ovals, with branches representing choices and leaves representing outcomes. A decision tree is built using a top-down approach with the divide-and-conquer method. This algorithm is popular because it is easy to understand and use [25]. To divide the data into the most similar subgroups based on the

outcome variable, attributes at the points are selected in the process [26]. The method uses entropy calculations and information gain [27].

1. Determine the attribute to be used as the root by determining the lowest entropy value and the highest gain value. Determine the entropy value using Equation 1 [27] .

$$\text{Entropy}(y) = \sum_{i=1}^n - P_i \log_2 p_i \quad (1)$$

2. Determine the gain value using Equation 2 [27] .

$$\text{Gain}(y, A) = \text{Entropy}(y) - \sum_e \text{enilai}(A) \frac{y_e}{y} \text{entropy}(y_e) \quad (2)$$

3. Create a branch for each value.
4. Repeat the process for each branch, so that all cases have the same class.

2.5. Naïve Bayes Classifier

A naive Bayes (NB) classifier is a classification algorithm based on the Bayes theorem, with the assumption that each predictor is independent of the others. Since this algorithm is probability-based, it is very important to study the sample distribution and the type of features used [28]. In the Naive Bayes structure, each attribute in the training data is conditionally independent and has the same role in determining the class [29]. Equation 3 shows the relationship between these Naive Bayes prediction equations [30].

$$Q(C|X) = \frac{Q(X) \cdot Q(C)}{Q(X)} \quad (3)$$

The Naive Bayes algorithm consists of several main parts. First, based on the given input, X denotes the case criteria. Next, C denotes a particular label or class in the dataset, and i denotes the number of available class labels. If it is known that the class label is C, the probability of the occurrence of class label C based on the input criteria X is expressed as Q(C|X). On the other hand, Q(X|C) denotes the probability of the input criteria X occurring if the class label is C. Finally, Q(C) denotes the probability of the class label C appearing first or earlier in the dataset. Using Bayes' theory, the combination of these elements is used to calculate the posterior probability.

2.6. K-Nearest Neighbors (K-NN)

The K-Nearest Neighbors (K-NN) algorithm is a robust and supervised machine learning algorithm [31]. It works by measuring the distance between similar data points and then clustering them based on their proximity. One of the main goals of the K-NN algorithm is to create a model f that can predict the class label y for an unknown pattern x [32]. It assigns the majority class label to the closest patterns in the data space and assigns those patterns to the target x. To determine the similarity in the data space, it uses the Hamming distance as a metric [3].

$$d_{\text{Euclidian}} = \sqrt{\sum_{i=1}^n (x_{i2} - x_{i1})^2} \quad (4)$$

In the K-Nearest Neighbors (K-NN) algorithm, the Euclidean distance is used to calculate the distance between the sample data (X i1) and the test data (X i2), based on n attributes. This distance helps determine the distance between the sample data and the test data, allowing the test data to be classified based on the majority class among the nearest neighbors. Furthermore, the k parameter in the K-Nearest Neighbors (K-NN) algorithm was determined using a Grid Search to obtain the optimal configuration. Several values of k were evaluated on the training dataset, namely k = 1, 3, 5, 7, and 9. The performance of each k value was then assessed based on accuracy and F1-score metrics to ensure a balanced trade-off between precision and recall [1].

2.7. Random Forest

One of the most effective batch-learning methods for large and imbalanced classification problems is the random forest classifier [33]. To predict new, unseen instances, the random forest algorithm uses the average of the base models' predictions [34]. Random forest can handle both categorical and continuous variables and, compared to a single DT model, is less prone to overfitting. Given an instance x with class F, the final prediction of a batch of N trees can be calculated using the following equation [35]:

$$H(N(X)) = \arg \max (y) - \sum_{k=1}^k 1(H_k(x) = j), \text{ untuk } j = 1, \dots, c \quad (5)$$

In the Random Forest algorithm, $H(N(x))$ is the final decision for input x , which is based on the results of K decision trees. Each k th tree produces a decision $h_k(x)$, and the majority of decisions from these trees determine the final class. If the condition is true, the indicator function 1 is 1 , and C is the total number of possible classes.

2.8. Support Vector Machine

Support vector machines have been very good in disease prediction in recent years [36]. Used for prediction, regression, and classification of linear and non-linear data, the SVM algorithm is a recent solution to the binary classification problem on multidimensional data [4]. This technique is used in classification, prediction, and regression, both linear and non-linear [37]. Its ability to maximize the classification margin, thereby improving accuracy and generalization, is the main advantage of SVM. In addition, SVM avoids the problems of overfitting and excessive dimensionality, which can increase computational complexity and prevent getting stuck in local extremes [38]. SVM shows mathematical equality [36].

$$f(x) = W^T X + b \quad (6)$$

In these equations, x is a vector point, and w is a vector of weights. Therefore, the separated data in Equation (1) must always be greater than zero, and the separated data in Equation (2) must always be less than zero. SVM selects the hyperplane that maximizes the distance to all available ones.

2.9. Model Evaluation

An important part of building a successful machine learning (ML) model is evaluating the model's performance. Predictive models are evaluated using the confusion matrix performance metric [39]. This matrix is usually used to evaluate classification performance and determine the performance of the machine learning model [40]. The classification performance of each method is calculated using the Confusion Matrix to obtain accuracy, precision, and recall results. This matrix is also used to evaluate the performance of the machine learning model. Figure 2 shows a diagram of the confusion matrix [41].

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

Figure 2. Confusion Matrix Diagram [42]

Accuracy, precision, recall, specificity, and F1 score are some of the matrices that will be calculated in the confusion matrix. The formula will be obtained during the calculation [41].

3. RESULTS AND DISCUSSION

To select the best algorithm, five supervised learning methods were applied to a dataset of patients with heart disease risk obtained from Kaggle. After going through the pre-processing process, the algorithms were evaluated using five different performance measures. The experimental results are summarized in tables and graphs to facilitate analysis, and compared with previous studies to determine the best configuration for each algorithm.

3.1. Preprocessing Data

The first step in data preprocessing is data cleaning. This process includes removing missing values, outlier detection, outlier handling, model training, and selecting the right model. After loading the dataset and importing the necessary libraries, the data is divided into three scenarios: 70% training data and 30% test data, 80% with 20%, and 90% with 10%. All these steps are done before applying the classification algorithm to ensure optimal data. The best model is selected based on evaluation using various performance metrics, and only that model is used to test the dataset [3].

The last step of the preprocessing process is to apply performance metrics to evaluate the prediction results. This evaluation determines whether the patient has heart disease or not, as well as assessing the effectiveness of the proposed computer-aided diagnosis system.

Table 2. Data

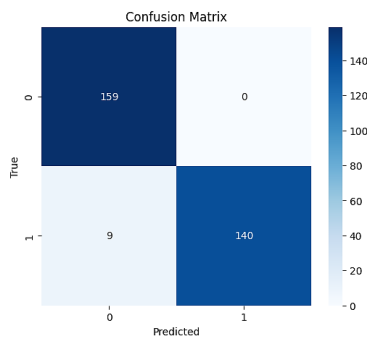
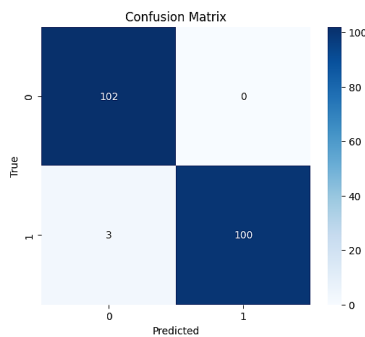
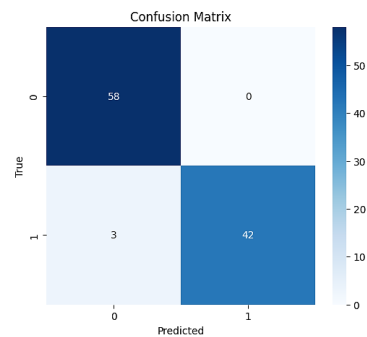
Age	Sex	Chest Pain Type	Resting Blood Pressure	...	Vessels Colored By Flourosopy	Thalassemia	Target
52	Male	Typical angina	125	...	Two	Reversable Defect	0
53	Male	Typical angina	140	...	Zero	Reversable Defect	0
70	Male	Typical angina	145	...	Zero	Reversable Defect	0
61	Male	Typical angina	148	...	One	Reversable Defect	0
62	Female	Typical angina	138	...	Three	Fixed Defect	0
...	...	...	...	...	...	...	...
59	Male	Atypical angina	140	...	Zero	Fixed Defect	1
60	Male	Typical angina	125	...	One	Reversable Defect	0
47	Male	Typical angina	110	...	One	Fixed Defect	0

3.2. Decision tree

In the Decision Tree algorithm, the model performance shows high accuracy at various data split ratios, namely 70:30, 80:20, and 90:10, with the best results at the 80:20 ratio, achieving 98.54% accuracy, 98.57% precision, and 98.54% recall. The absence of false positives (FP) at all ratios indicates the model's ability to consistently avoid false positive prediction errors. At the 70:30 ratio, the model achieves 97.08% accuracy, 97.32% precision, and 96.98% recall, while at the 90:10 ratio, accuracy remains high at 97.09% with 97.54% precision and 96.67% recall. These results, as shown in Figures 3, 4, and 5, indicate that the Decision Tree can provide good and consistent prediction results on the heart disease risk dataset.

Table 3. Decision Tree Evaluation Results

Algorithm	Split	Evaluation	Result (%)
C45	70% : 30%	Accuracy	97,08
		Precision	97,32
		Recall	96,98
		F1-Score	97,07
	80% : 20%	Accuracy	98,54
		Precision	98,57
		Recall	98,54
		F1-Score	98,54
	90% : 10%	Accuracy	97,09
		Precision	97,54
		Recall	96,67
		F1-Score	97,02

**Figure 3.** Confusion Matrix with Ratio 70:30**Figure 4.** Confusion Matrix with Ratio 80:20**Figure 5.** Confusion Matrix with Ratio 90:10

3.3. Naive Bayes

In the Naive Bayes algorithm, the model performance shows a consistent level of accuracy at various data-sharing ratios, namely 70:30, 80:20, and 90:10. The best results were obtained at the 90:10 ratio with an

accuracy of 79.61%, a precision of 72.22%, and a recall of 86.67%. At the 70:30 ratio, the model achieved 79.55% accuracy, 76.54% precision, and 83.22% recall, although there were quite high false positives (FPs) of 38, indicating incorrect positive predictions. For the 80:20 ratio, accuracy was 78.05%, precision 76.36%, and recall 81.55%. The confusion matrix results for each ratio, as shown in Figures 6, 7, and 8, indicate that Naive Bayes can provide fairly good predictions, although it is more susceptible to false positives than other algorithms. Overall, this model shows balanced and reliable performance for the prediction case on the heart disease risk dataset.

Table 4. Naive Bayes Evaluation Results

Algorithm	Split	Evaluation	Result (%)
Naïve Bayes	70% : 30%	Accuracy	79,55
		Precision	76,54
		Recall	83,22
		F1-Score	79,74
	80% : 20%	Accuracy	78,05
		Precision	76,36
		Recall	81,55
		F1-Score	78,87
	90% : 10%	Accuracy	79,61
		Precision	72,22
		Recall	86,67
		F1-Score	78,79

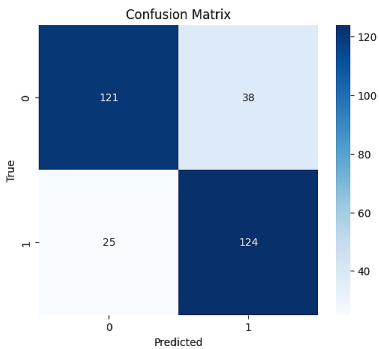


Figure 6. Confusion Matrix with Ratio 70:30

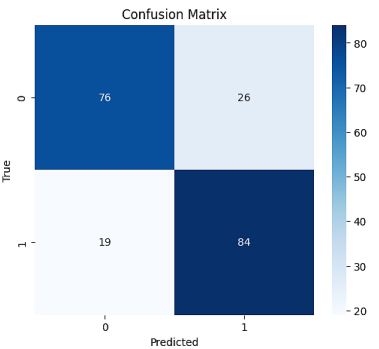


Figure 7. Confusion Matrix with Ratio 80:20

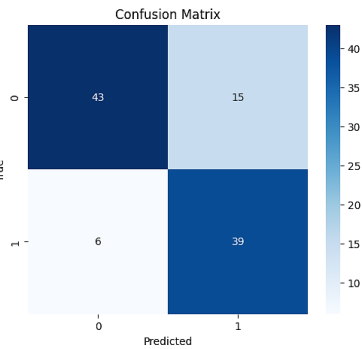


Figure 8. Confusion Matrix with Ratio 90:10

3.4. K-Nearest Neighbors (K-NN)

In the K-Nearest Neighbors algorithm with the best parameter value (best k) of 3, the model performance shows very good results in various data sharing ratios: 70:30, 80:20, and 90:10. At the 70:30 ratio, the model achieves an accuracy of 89.29% with a precision of 89.19% and a recall of 88.59%. Although there are 16 false positives (FP), these results reflect the model's ability to provide reliable predictions. At the 80:20 ratio, accuracy increases to 90.73%, with precision 90.38% and recall 91.26%. The high F1 Score of 90.82% indicates a balance between precision and recall, making the model effective at classifying test data.

The best results were obtained at a 90:10 ratio, with an accuracy of 97.09%, a precision of 100.00%, and a recall of 93.33%. The absence of false positives (FP) at this ratio confirms the model's ability to avoid false positive prediction errors. The confusion matrix results illustrating the model's performance for each data-split scenario are presented in Figures 9, 10, and 11. Overall, the K-NN algorithm exhibits consistent and excellent performance in predicting heart disease risk. Furthermore, the parameter evaluation results indicate that $k = 3$ yields the most stable and optimal performance across different data-splitting scenarios, particularly in improving precision and F1-score. Therefore, $k = 3$ was selected as the optimal parameter and used across all K-NN experiments in this study, consistent with previous research showing that performance-based selection of k enhances the reliability and generalization capability of K-NN models for medical data analysis.

Table 5. K-Nearest Neighbors Evaluation Results

Algorithm	Split	Evaluation	Result (%)
K-NN	70% : 30%	Accuracy	89,29
		Precision	89,19
		Recall	88,59
		F1-Score	88,89

Algorithm	Split	Evaluation	Result (%)
	80% : 20%	Accuracy	90,73
		Precision	90,38
		Recall	91,26
		F1-Score	90,82
	90% : 10%	Accuracy	97,09
		Precision	100
		Recall	93,33
		F1-Score	96,55

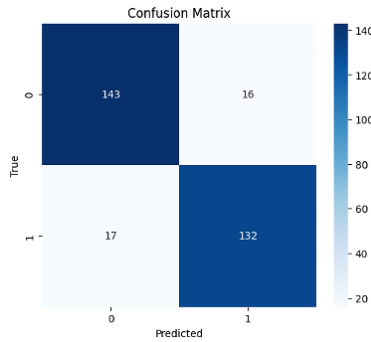


Figure 9. Confusion Matrix with Ratio 70:30

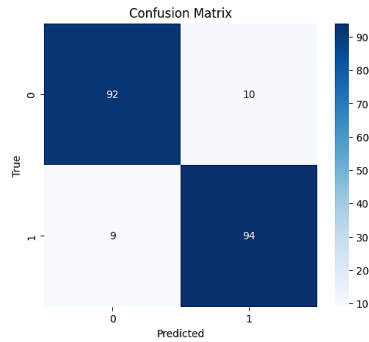


Figure 10. Confusion Matrix with Ratio 80:20

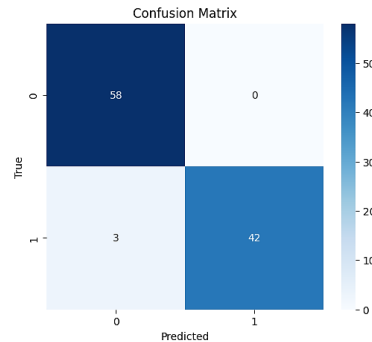


Figure 11. Confusion Matrix with Ratio 90:10

3.5. Random Forest

In the Random Forest algorithm, the model performance shows excellent results in various ratios. With a ratio of 70:30, the model achieves an accuracy of 98.05% with, a precision of 98.12%, and a recall of 98.05%. The absence of false positives (FP) in all ratios confirms the model's ability to produce accurate predictions with minimal errors, as shown in Figures 12, 13, and 14.

Table 6. Random Forest Evaluation Results

Algorithm	Split	Evaluation	Result (%)
Random Forest	70% : 30%	Accuracy	98,05
		Precision	98,12
		Recall	98,05
		F1-Score	98,05
	80% : 20%	Accuracy	98,54
		Precision	98,58
		Recall	98,54
		F1-Score	98,54
	90% : 10%	Accuracy	97,09
		Precision	97,23
		Recall	97,09
		F1-Score	97,07

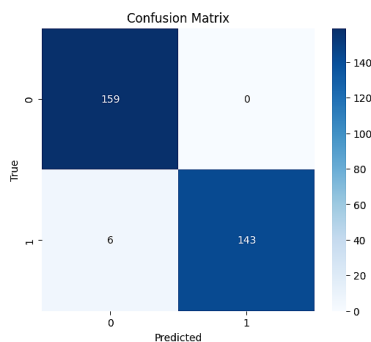


Figure 12. Confusion Matrix with Ratio 70:30

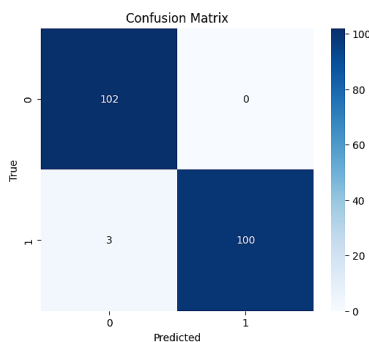


Figure 13. Confusion Matrix with Ratio 80:20

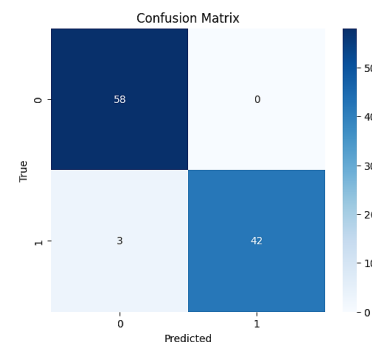


Figure 14. Confusion Matrix with Ratio 90:10

3.6. Support Vector Machine (SVM)

In the Support Vector Machine (SVM) algorithm, with a 70:30 split, the model achieves an accuracy of 91.56%, precision of 91.93%, and recall of 91.56%, although there are 20 false positives (FP). The confusion matrix results in Figures 15, 16, and 17 show that the SVM algorithm can provide reliable and consistent predictions.

Table 7. SVM Evaluation Results

Algorithm	Split	Evaluation	Result (%)
Support Vector Machine (SVM)	70% : 30%	Accuracy	91,56
		Precision	91,93
		Recall	91,56
		F1-Score	91,55
	80% : 20%	Accuracy	88,78
		Precision	89,23
		Recall	88,78
		F1-Score	88,75
	90% : 10%	Accuracy	91,26
		Precision	91,76
		Recall	91,26
		F1-Score	91,3

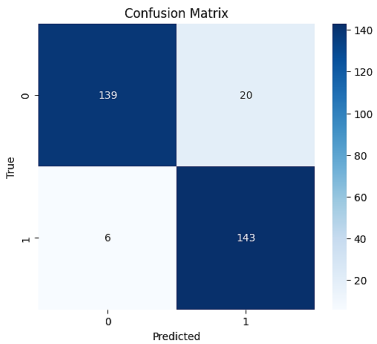


Figure 15. Confusion Matrix with Ratio 70:30

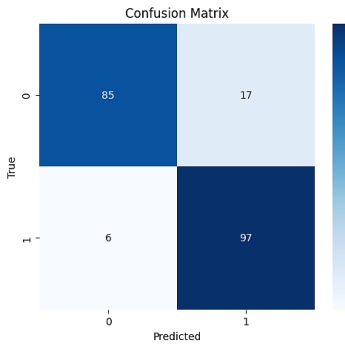


Figure 16. Confusion Matrix with Ratio 80:20

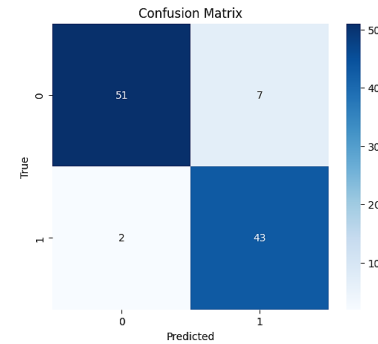


Figure 17. Confusion Matrix with Ratio 90:10

3.7. Algorithm comparison

The results of the comparison of these five algorithms show that Decision Tree has the best performance in all data-sharing ratios, with the highest accuracy at the 80:20 ratio (98.54%) and a balance between precision (98.57%) and recall (98.54%). Naïve Bayes yields lower results, especially at the 70:30 and 80:20 ratios, with accuracies of 79.55% and 78.05%, but improves at the 90:10 ratio to 79.61%. K-NN shows very good results at all ratios, with the best accuracy at the 90:10 ratio (97.09%) and a precision of 100%. Random Forest also shows excellent performance, achieving over 98% accuracy at the 70:30 and 80:20 ratios and 97.09% at the 90:10 ratio, with very consistent precision and recall. SVM achieved good results, with 91.56% accuracy at the 70:30 ratio and 91.26% at the 90:10 ratio, although there was a slight decrease at the 80:20 ratio (88.78%). Overall, Random Forest and Decision Tree proved to be the most effective algorithms for heart disease prediction on this dataset, as shown in Figure 18.

3.8. Discussion

This study demonstrates that both the Random Forest (RF) and Decision Tree (DT) algorithms are effective for heart disease detection, with the RF consistently achieving superior classification performance. The experimental results indicate that the ensemble learning mechanism employed by Random Forest enables the model to capture more complex decision boundaries while reducing the risk of overfitting compared with a single Decision Tree. This finding suggests that combining multiple decision trees improves model robustness, particularly when handling medical datasets that contain complex relationships among clinical variables. In a study conducted by Reddy et al. [16], although Naïve Bayes achieved the highest accuracy, Decision Tree remained a competitive classifier with an accuracy of 82%, confirming its capability as an interpretable prediction model. Similarly, Zriqat et al. [18] reported that Random Forest slightly outperformed Decision Tree while maintaining consistently high predictive performance. These findings collectively indicate that tree-based algorithms remain reliable and practical solutions for heart disease prediction, as they offer high predictive accuracy and transparent decision-making.

The superior performance of Random Forest can be attributed to its ensemble learning strategy, in which multiple decision trees are generated from different bootstrap samples and combined via majority voting. This mechanism effectively reduces model variance and minimizes the influence of noisy observations, leading to better generalization on unseen patient data. Furthermore, Random Forest can model nonlinear relationships among clinical variables without extensive data preprocessing or feature transformations. Research conducted by Minnoora and Bath [12] supports these findings, reporting that Random Forest achieved an accuracy of 99.30% while significantly reducing false-negative predictions, an essential factor in medical diagnosis because undetected heart disease may delay treatment and increase patient risk. The current study also indicates that Random Forest performs well even with relatively small datasets, making it a practical option for healthcare institutions with limited patient records. Therefore, the algorithm offers not only high predictive performance but also operational advantages for real-world clinical implementation.

Another important finding is the influence of feature selection on classification performance. Clinical datasets often contain redundant or less informative variables that may reduce model efficiency and prediction accuracy. By selecting only the most relevant features, machine learning algorithms can focus on significant clinical indicators while minimizing unnecessary computational complexity. This observation is consistent with the study by Sumwiza et al. [11], which demonstrated that Random Forest achieved approximately 99% accuracy after applying an effective feature selection strategy. These results emphasize that the quality of input features is equally important as the choice of classification algorithm. Consequently, integrating feature selection techniques into the machine learning pipeline can substantially improve prediction reliability while reducing computational costs.

Overall, the findings confirm that Random Forest is the most suitable algorithm among the evaluated models due to its high accuracy, stability, and ability to reduce false-negative classifications. Although Decision Tree remains an effective and highly interpretable model, its predictive performance is generally lower because it relies on a single-tree structure that is more sensitive to data variation. The comparison with previous studies further strengthens the conclusion that ensemble learning methods provide more reliable performance than conventional single-tree approaches in medical diagnosis. From a practical perspective, implementing Random Forest in clinical decision-support systems could assist healthcare professionals in identifying high-risk patients more accurately and consistently. Nevertheless, this study is limited by the relatively small dataset and the absence of external validation using patient data from different healthcare institutions. Future research should therefore evaluate the proposed model on larger multicenter datasets, investigate advanced hyperparameter optimization techniques, and explore explainable artificial intelligence (XAI) methods to enhance both predictive performance and clinical interpretability.

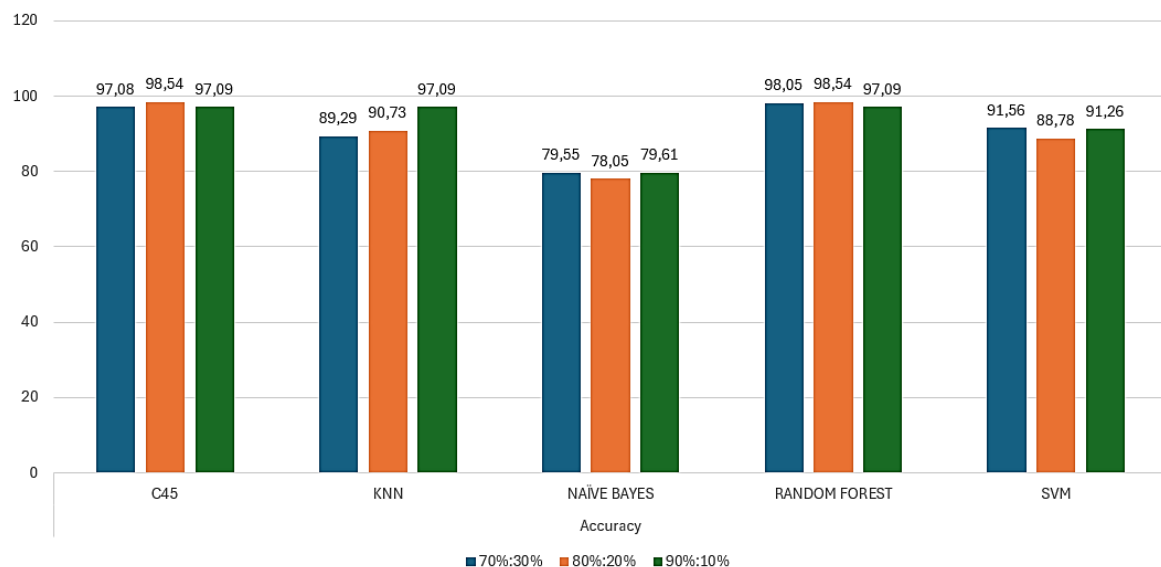


Figure 18. Result of Comparison Models

4. CONCLUSION

Based on the evaluation results of five supervised learning algorithms, namely Decision Tree, Naive Bayes, K-Nearest Neighbors (K-NN), Random Forest, and Support Vector Machine (SVM), it can be concluded that Decision Tree and Random Forest provide the best performance in predicting the risk of heart disease in this dataset. Decision Tree shows the highest accuracy with a very good balance of precision and recall, while Random Forest also provides consistent results with an accuracy of more than 98%. K-NN provides good results at a ratio of 90:10 with an accuracy of 97.09%, but decreases at other ratios, while SVM

shows good performance despite a slight decrease at a ratio of 80:20. Naive Bayes provides lower results, especially at ratios of 70:30 and 80:20. Overall, Decision Tree and Random Forest are the best choices based on accuracy and balance of evaluation metrics.

REFERENCES

- [1] K. M. Almustafa, "Prediction of heart disease and classifiers ' sensitivity analysis," pp. 1–18, 2020.
- [2] M. Ozcan and S. Peker, "Healthcare Analytics A classification and regression tree algorithm for heart disease modeling and prediction," *Healthc. Anal.*, vol. 3, no. July 2022, p. 100130, 2023, doi: 10.1016/j.health.2022.100130.
- [3] M. Amine et al., "Heliyon Early heart disease prediction using feature engineering and machine learning algorithms," *Heliyon*, vol. 10, no. 19, p. e38731, 2024, doi: 10.1016/j.heliyon.2024.e38731.
- [4] G. Manikandan, B. Pragadeesh, V. Manojkumar, A. L. Karthikeyan, and R. Manikandan, "Informatics in Medicine Unlocked Classification models combined with Boruta feature selection for heart disease prediction," *Informatics Med. Unlocked*, vol. 44, no. August 2023, p. 101442, 2024, doi: 10.1016/j.imu.2023.101442.
- [5] T. A. Gaziano, A. Bitton, S. Anand, S. Abrahams-gessel, and A. Murphy, "Growing Epidemic of Coronary Heart Disease in Low- and Middle-Income Countries," *YMCD*, vol. 35, no. 2, pp. 72–115, 2010, doi: 10.1016/j.cpcardiol.2009.10.002.
- [6] L. Mohammed, S. Zafar, L. Mohammed, S. Akoosh, L. Mohammed, and S. Akoosh, "ScienceDirect Anticipating the Nearness of Coronary Heart Infection Anticipating the Nearness of Coronary Heart Infection Anticipating the Nearness of Coronary Heart Infection Utilizing Machine Learning Classifiers Utilizing Machine Learning Classifiers Utilizing Machine Learning Classifiers Utilizing Learning Classifiers Salem Machine," *Procedia Comput. Sci.*, vol. 235, pp. 2619–2629, 2024, doi: 10.1016/j.procs.2024.04.247.
- [7] M. G. El-shafiey and A. Hagag, "A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest," pp. 18155–18179, 2022.
- [8] Z. Ying, G. Zhang, Z. Pan, C. Chu, and X. Liu, "FedECG : A federated semi-supervised learning framework for electrocardiogram abnormalities prediction," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 6, p. 101568, 2023, doi: 10.1016/j.jksuci.2023.101568.
- [9] M. Ali et al., "Heart disease prediction using supervised machine learning algorithms : Performance analysis and comparison," *Comput. Biol. Med.*, vol. 136, no. July, p. 104672, 2021, doi: 10.1016/j.combiomed.2021.104672.
- [10] V. Chang, V. Rupa, A. Qianwen, and M. A. Hossain, "Healthcare Analytics An artificial intelligence model for heart disease detection using machine learning algorithms," *Healthc. Anal.*, vol. 2, no. September 2021, p. 100016, 2022, doi: 10.1016/j.health.2022.100016.
- [11] K. Sumwiza, C. Twizere, G. Rushingabigwi, and P. Bakunzibake, "Informatics in Medicine Unlocked Enhanced cardiovascular disease prediction model using random forest algorithm," *Informatics Med. Unlocked*, vol. 41, no. August, p. 101316, 2023, doi: 10.1016/j.imu.2023.101316.
- [12] M. Minnoor and V. Baths, "ScienceDirect ScienceDirect Diagnosis of Breast Cancer Using Random Forests Diagnosis of Breast Cancer Using Random Forests," *Procedia Comput. Sci.*, vol. 218, no. 2022, pp. 429–437, 2023, doi: 10.1016/j.procs.2023.01.025.
- [13] A. Bilal et al., "Classification and monitoring of arm exercises using machine learning and wrist-worn band," vol. 27, no. September, 2024, doi: 10.1016/j.eij.2024.100534.
- [14] B. O. Support, "Informatics in Medicine Unlocked learning prospective," vol. 26, 2021, doi: 10.1016/j.imu.2021.100696.
- [15] N. V. S, D. Sripada, V. Sugumaran, and M. Aghaei, "Heliyon Detection of visual faults in photovoltaic modules using a stacking ensemble approach," *Heliyon*, vol. 10, no. 6, p. e27894, 2024, doi: 10.1016/j.heliyon.2024.e27894.
- [16] C. Series, "Prediction on Cardiovascular disease using Decision tree and Naïve Bayes classifiers Prediction on Cardiovascular disease using Decision tree and Naïve Bayes classifiers", doi: 10.1088/1742-6596/2161/1/012015.
- [17] E. Heart, D. Prediction, U. Machine, and L. Techniques, "Learning Techniques," 2023.
- [18] A. Zriqat, A. M. Altamimi, and M. Azzeh, "A Comparative Study for Predicting Heart Diseases Using Data Mining Classification Methods," vol. 14, no. 12, pp. 868–879, 2016.
- [19] S. Srinivasan, S. Gunasekaran, S. K. Mathivanan, B. A. M. M. B, P. Jayagopal, and G. T. Dalu, "An active learning machine technique based prediction of cardiovascular heart disease from UCI - repository database," *Sci. Rep.*, pp. 1–19, 2023, doi: 10.1038/s41598-023-40717-1.
- [20] U. Boosting and E. Techniques, "Using Boosting Ensemble Techniques," pp. 1–22, 2024.
- [21] J. B. Raja and S. C. Pandian, "Computer Methods and Programs in Biomedicine PSO-FCM based data mining model to predict diabetic disease," vol. 196, 2020, doi: 10.1016/j.cmpb.2020.105659.
- [22] T. Islam, M. Raihan, F. Farzana, N. Aktar, P. Ghosh, and S. Kabiraj, "Typical and Non-Typical Diabetes

- Disease Prediction using Random Forest Algorithm,” pp. 1–6, 2020.
- [23] A. Putri et al., “Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 20–26, 2023, doi: 10.57152/malcom.v3i1.610.
 - [24] J. M. Gonzalez-cava, J. A. Reboso, J. L. Casteleiro-roca, J. L. Calvo-rolle, J. Albino, and M. Pérez, “A Novel Fuzzy Algorithm to Introduce New Variables in the Drug Supply Decision-Making Process in Medicine,” vol. 2018, 2018, doi: 10.1155/2018/9012720.
 - [25] A. Sinha, B. Rout, S. Mohanty, S. R. Mishra, H. Mohapatra, and S. Dey, “ScienceDirect Exploring Exploring Sentiments Sentiments in in the the Russia-Ukraine Russia-Ukraine Conflict : Conflict : A Comparative Analysis of K-NN , Decision Tree And Logistic Comparative Analysis of K-NN , Decision Tree And Logistic Regression Machine Machine Learning Learning Classifiers .,” *Procedia Comput. Sci.*, vol. 235, pp. 1068–1076, 2024, doi: 10.1016/j.procs.2024.04.101.
 - [26] D. Colledani, P. Anselmi, and E. Robusto, “Machine learning-decision tree classifiers in psychiatric assessment : An application to the diagnosis of major depressive disorder,” *Psychiatry Res.*, vol. 322, no. October 2023, p. 115127, 2023, doi: 10.1016/j.psychres.2023.115127.
 - [27] A. F. Lubis, H. Z. Haq, I. Lestari, and M. Iltizam, “Classification of Diabetes Mellitus Sufferers Eating Patterns Using K-Nearest Neighbors , Naïve Bayes and Decission Tree,” vol. 2, no. July, pp. 44–51, 2024.
 - [28] O. Peretz, M. Koren, and O. Koren, “Engineering Applications of Artificial Intelligence Naive Bayes classifier – An ensemble procedure for recall and precision enrichment,” *Eng. Appl. Artif. Intell.*, vol. 136, no. PB, p. 108972, 2024, doi: 10.1016/j.engappai.2024.108972.
 - [29] T. Kim and J. Lee, “Exponential Loss Minimization for Learning Weighted Naive Bayes Classifiers,” *IEEE Access*, vol. 10, pp. 22724–22736, 2022, doi: 10.1109/ACCESS.2022.3155231.
 - [30] S. Farhana, “ScienceDirect ScienceDirect Classification of Academic Performance for University Research Evaluation by Implementing Modified Naive Bayes Algorithm,” *Procedia Comput. Sci.*, vol. 194, pp. 224–228, 2021, doi: 10.1016/j.procs.2021.10.077.
 - [31] R. Ehsani and F. Drabløs, “Robust Distance Measures for k NN Classification of Cancer Data,” 2020, doi: 10.1177/1176935120965542.
 - [32] O. Bouattane, “Atherosclerosis disease prediction using Supervised Machine Learning Techniques,” 2020.
 - [33] S. Asadi, S. Roshan, and M. W. Kattan, “Random forest swarm optimization-based for heart diseases diagnosis,” *J. Biomed. Inform.*, vol. 115, no. January, p. 103690, 2021, doi: 10.1016/j.jbi.2021.103690.
 - [34] G. Obaido et al., “Machine Learning with Applications Supervised machine learning in drug discovery and development : Algorithms , applications , challenges , and prospects,” *Mach. Learn. with Appl.*, vol. 17, no. July, p. 100576, 2024, doi: 10.1016/j.mlwa.2024.100576.
 - [35] R. T. Aruleba et al., “COVID-19 Diagnosis : A Review of Rapid Antigen , RT-PCR and Artificial Intelligence Methods,” pp. 1–18, 2022.
 - [36] S. I. Ayon, M. Islam, R. Hossain, M. Islam, and R. Hossain, “Coronary Artery Heart Disease Prediction : A Comparative Study of Coronary Artery Heart Disease Prediction : A Comparative Study of Computational Intelligence Techniques,” *IETE J. Res.*, vol. 0, no. 0, pp. 1–20, 2020, doi: 10.1080/03772063.2020.1713916.
 - [37] N. Mohammadi, M. Esteki, and J. Simal-gandara, “Machine learning for authentication of black tea from narrow-geographic origins : Combination of PCA and PLS with LDA and SVM classifiers,” *LWT*, vol. 203, no. June, p. 116401, 2024, doi: 10.1016/j.lwt.2024.116401.
 - [38] M. Libnao, M. Misula, C. Andres, J. Mariñas, and J. Mariñas, “ScienceDirect Procedia ScienceDirect Traffic incident prediction and classification system using naïve Traffic incident prediction and algorithm classification system using naïve bayes bayes algorithm,” *Procedia Comput. Sci.*, vol. 227, pp. 316–325, 2023, doi: 10.1016/j.procs.2023.10.530.
 - [39] K. Moulaci, M. Shanbehzadeh, Z. M. Taghiabad, and H. K. Arpanahi, “Comparing machine learning algorithms for predicting COVID - 19 mortality,” *BMC Med. Inform. Decis. Mak.*, vol. 0, pp. 1–12, 2022, doi: 10.1186/s12911-021-01742-0.
 - [40] P. F. Pratama, D. Rahmadani, R. S. Nahampun, D. Harmutika, A. Rahmadeyan, and M. F. Evizal, “Random Forest Optimization Using Particle Swarm Optimization for Diabetes Classification,” *Public Res. J. Eng. Data Technol. Comput. Sci.*, vol. 1, no. 1, pp. 41–46, 2023, doi: 10.57152/predatecs.v1i1.809.
 - [41] E. Helmud, E. Helmud, and P. Romadiana, “Classification Comparison Performance of Supervised Machine Learning Random Forest and Decision Tree Algorithms Using Confusion Matrix,” vol. 13, pp. 92–97, 2024.
 - [42] Z. C. Dwinnie, L. Khairani, M. A. M. Putri, J. Adhiva, and M. I. F. Tsamarah, “Application of the Supervised Learning Algorithm for Classification of Pregnancy Risk Levels,” *Public Res. J. Eng. Data Technol. Comput. Sci.*, vol. 1, no. 1, pp. 26–33, 2023, doi: 10.57152/predatecs.v1i1.806.

- [43] F. Duran et al., “Perbandingan Kinerja Algoritma Random Florest Classifier,” vol. 3A, no. 2, pp. 98–103, 2024.
- [44] A. P. Jawalkar, P. Swetcha, N. Manasvi, P. Sreekala, and S. Aishwarya, “Early prediction of heart disease with data analysis using supervised learning with stochastic gradient boosting,” J. Eng. Appl. Sci., pp. 1–18, 2023, doi: 10.1186/s44147-023-00280-y.
- [45] R. Ehsani and F. Drabløs, “Robust Distance Measures for k NN Classification of Cancer Data,” 2020, doi: 10.1177/1176935120965542.