Available *online* at: http://jarpet.ft.unand.ac.id/
Jurnal Andalas: Rekayasa dan Penerapan Teknologi
ISSN (Online) 2797-9024

Click here and write your Article Category

Heart Disease Prediction Using KNN, Decision Tree, and Naïve Bayes

Baik Budi [1*], *Queen Hesti Ramadhami* [1], *Refki Budiman* [1]

[1] *Universitas Andalas, Kampus Limau Manis, Padang, Indonesia*

ARTICLE INFORMATION

Received: May 29, 2026
Revised: May 31, 2026
Available online: June 01, 2026

KEYWORDS

Cardiovascular Disease, Machine Learning, K-Nearest Neighbors, Decision Tree, Naive Bayes, Disease Prediction

CORRESPONDENCE

Phone: +6285155201203
E-mail: baikbudi@eng.unand.ac.id

ABSTRACT

Cardiovascular Disease (CVD) remains a leading global cause of mortality, making early and accurate diagnosis critical for effective medical intervention. Machine Learning (ML) algorithms offer promising solutions for automating clinical decision support systems. This study compares three supervised learning algorithms—K-Nearest Neighbors (KNN), Decision Tree (DT), and Naive Bayes (NB)—to evaluate their diagnostic efficacy in predicting heart disease. The models were trained and tested using a clinical dataset of 205 instances (100 normal and 105 heart disease cases) with an 80:20 data split. Performance was evaluated based on Accuracy, Precision, Recall, and F1-Score derived from confusion matrices. The experimental results demonstrate that the Decision Tree algorithm achieved the highest aggregate accuracy of 98.54%, exhibiting exceptional clinical reliability with a perfect precision score (zero false positives) and high sensitivity (only three false negatives). The KNN model performed comparably well, achieving 98.05% accuracy and zero false positives. In contrast, the Naive Bayes algorithm underperformed, with 82.93% accuracy and high rates of both Type I and Type II errors. In conclusion, the Decision Tree model emerges as the most robust, precise, and safe algorithmic architecture for clinical implementation in heart disease screening, effectively minimizing both false alarms and missed diagnoses.

INTRODUCTION

Cardiovascular Disease (CVD) remains a leading cause of mortality globally. Based on clinical data, heart disease accounts for a significantly high percentage of morbidity and mortality each year [1], [2]. Early detection and precise diagnosis are the most crucial factors in medical intervention, as they can minimize the risk of fatal complications in patients [3], [4]. Traditionally, the diagnosis of heart disease relies heavily on the analysis of a patient's medical history, electrocardiograms, and the clinical expertise of physicians. However, this conventional approach often has limitations in terms of time, cost, and the potential for human error in interpreting high-dimensional medical datasets [5], [6].

To overcome these limitations, the integration of Artificial Intelligence (AI), particularly Machine Learning (ML) algorithms, has become a central pillar in the digital healthcare (e-healthcare) revolution. ML-based systems have demonstrated the ability to analyze large-scale patient medical records to discover latent patterns indicative of heart disease risk with high computational accuracy [7]. This capability enables the design of Clinical Decision Support Systems (CDSS) that assist medical professionals in establishing diagnoses more rapidly and objectively [8], [9].

In diagnostic classification tasks, such as those involving heart disease datasets, the choice of the predictive algorithm largely determines the quality of the output. This study compares three supervised learning classification algorithms that have become standards in medical data analysis: K-Nearest Neighbors (KNN), Decision Tree (DT), and Naive Bayes (NB) [10]:

1. K-Nearest Neighbors (KNN)
   KNN is a non-parametric, instance-based learning algorithm. It classifies new data based on its vector-space proximity (e.g., Euclidean distance) to the k nearest training data points. This method is highly robust in recognizing non-linear decision boundaries in medical data, although its diagnostic accuracy heavily depends on determining the optimal k value [9], [11]
2. Decision Tree (DT)
   DT models the classification process in a hierarchical tree structure, where internal nodes represent the evaluation of comorbidity features, branches represent logical decision rules, and leaf nodes represent the final diagnosis [11]. The primary advantage of this algorithm in the clinical domain is its exceptionally high interpretability (as a white-box model), allowing medical practitioner s to trace and validate the system's predictive logic easily.

https://doi.org/10.25077/jarpet.v6i1.142

3. Naive Bayes (NB)
   NB is a powerful probabilistic classifier based on Bayes' Theorem, assuming conditional independence among features given the target class [10]. Although absolute independence is rarely observed in real-world scenarios, the probabilistic architecture of NB has proven computationally efficient, resilient to irrelevant features (noise), and capable of providing rapid preliminary diagnostic metrics on small to medium-sized datasets.

The application and evaluation of Machine Learning models for heart disease prediction have been extensively explored in the academic literature. A comprehensive study designed an automated diagnostic system to compare the performance of KNN, Naive Bayes, and Decision Tree, concluding that each algorithm has unique trade-offs; for instance, decision tree models tend to exhibit absolute precision in minority class detection. Other approaches utilizing hybrid and ensemble systems investigated by [12] and [13] consistently underscore that base models (such as Decision Tree and KNN) serve as the crucial backbone for achieving diagnostic accuracies exceeding 86%. Meanwhile, literature reviews by [6] and [5] highlight the phenomenon of varying False Positive rates across algorithms. These studies note that Naive Bayes is often highly sensitive in providing initial predictions, whereas Decision Trees offer a more conservative approach to minimize false alarms in healthy patients.

Although the theoretical foundations and applications of these three algorithms are well-established, the diagnostic performance of Machine Learning is inherently data-driven and highly specific to the underlying dataset distribution. Therefore, this study aims to implement, test, and conduct an independent comparative analysis of the K-Nearest Neighbors, Decision Tree, and Naive Bayes algorithms using the specific heart.csv dataset. This research evaluates the clinical reliability of each model through confusion matrices and standard performance evaluation metrics—including Accuracy, Precision, Recall, and F1-Score—to determine the most optimal, precise, and safe algorithmic architecture for implementation as a medical screening instrument.

Heart disease prediction is a highly complex task, and in today's world, it largely depends on individual medical practitioners [12]. If all individual medical practitioners were combined into a single dataset, it would be very beneficial for the younger generation of medical practitioners and ultimately help society. In this research, a hybrid approach is used for heart attack prediction, namely a combination of the most popular clustering technique called 'K-Means' and the 'Naive Bayes' algorithm as a classifier [13]. Because of its hybrid approach, this technique is best suited to complex problems and produces results with excellent accuracy.

METHOD

In this study, the dataset used is from Kaggle, and the parameters used are shown in Table 1 below.

Table 1. Attribute Descriptions

No	Attribute	Descriptions
1	*Age*	Age in years
2	*Sex*	1 = male, 0 = female
3	*Cp*	1 = typical angina, 2 = atypical angina, 3 = non anginal pain, 4 = asymptomatic
4	*Trestbps*	Blood pressure in mm Hg
5	*Chol*	Cholesterol content in mg/dl
6	*Fbs*	Fasting blood sugar > 120 mg/dl (1 = true, 0 = false)
7	*Restecg*	ECG results (0 = normal, 1 = having ST-T wave abnormality, 2 = left ventricular hypertrophy)
8	*Thalach*	Maximum Heart Rate
9	*Exang*	Exercise-induced angina
10	*Oldpeak*	ST Depression
11	*Slope*	Slope of the peak exercise ST segment
12	*Ca*	Number of major vessels
13	*Thalach*	Thalassemia value
14	*Heart Disease*	1 = patient has heart disease, 0 = patient does not have heart disease

This dataset can be split into two sections: 80% for training and 20% for testing.

K-Nearest Neighbors (KNN) Algorithm

K-Nearest Neighbors is a non-parametric algorithm that classifies objects based on the geometric proximity of test data to a set of training data in an n-dimensional metric space. The target class is determined by finding k nearest neighbors using a distance metric. In this study, the distance metric employed is the Euclidean Distance, which is mathematically defined in Equation (1):

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{1}$$

Where:

d(x, y) = The Euclidean distance between the test data vector x and the training data vector y.
x_i, y_1 =The values of the i-th feature in the respective data points.
n = The total number of features in the dataset.

Once the distances from all points are calculated, the algorithm selects the *k* neighbors with the smallest distances. The final classification decision is made using a majority voting mechanism based on the classes of the *k* neighbors.

Decision Tree Algorithm

The Decision Tree algorithm constructs a classification model as an inverted hierarchical tree structure consisting of a root node, internal (split) nodes, and leaf nodes [14]. The selection of the optimal feature to split the data at each node is based on the level of information purity. This algorithm generally utilizes Entropy and Information Gain metrics to determine the optimal splitting attributes.

Entropy, which represents the degree of uncertainty or impurity of a dataset *S*, is calculated using Equation (2):

$$Entropy\ (S) = -\sum_{i-1}^{c} p_i log_2(p_i) \tag{2}$$

Where:

S = The set of data samples.
C = The number of target classes (in this case, c = 2 for normal and heart disease classes).
p_i = The proportion or probability of samples belonging to the i-th class within set S.

After obtaining the Entropy value, the reduction in uncertainty after splitting the data based on attribute A is measured using the Information Gain metric, as formulated in Equation (3):

$$Gain(S,A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{3}$$

Where:

Gain(S, A) = The information gain of attribute A.
Values(A) = All possible values that attribute A can take.
S_v = The subset of S where attribute A has the value v.
$|S_v|$, |S = The number of samples in the subset S_v and the total set S, respectively.

The attribute yielding the highest Information Gain is selected as the splitting node at each stage of the tree construction.

Naïve Bayes

Naive Bayes is a probabilistic classifier that operates by applying Bayes' Theorem. This algorithm relies on a strong (naive) fundamental assumption that every predictor feature is conditionally independent of the others given the target class. The general mathematical formulation of Bayes' Theorem is expressed in Equation (4):

$$P(C_k|X) = \frac{P(X|C_k).P(C_k)}{P(X)} \tag{4}$$

Where:

$P(C_k|X)$ = The posterior probability of class C_k given the input feature vector X.
P(X | C_k) = The likelihood probability of feature X given that class C_k is known.
$P(C_k)$ = The prior probability of class C_k.
P(X) = The prior probability of the predictors (normalization constant).

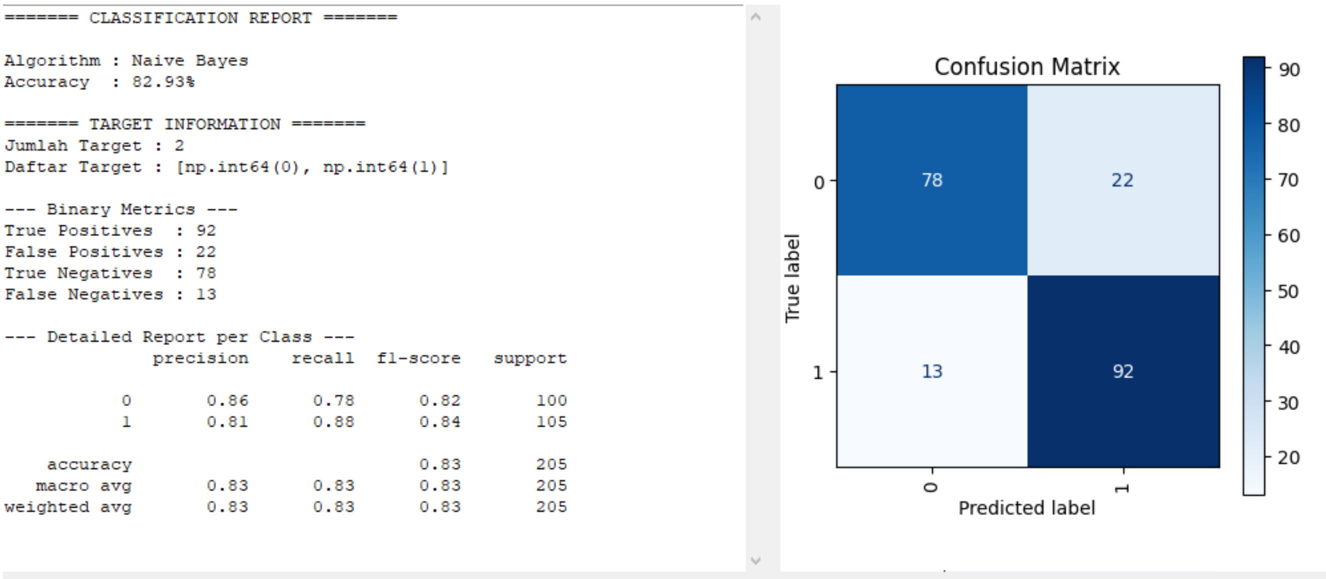
For continuous data (such as age or blood pressure in the heart disease dataset), the probability distribution is calculated using the Gaussian Naive Bayes probability density function, formulated in Equation (5)

$$P(x_i|C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} exp\left(-\frac{x_i - \mu_k)^2}{2\pi\sigma_k^2}\right) \tag{5}$$

Where μ_k is the mean and σ_k^2 is the variance of the feature x_i associated with class C_k.

RESULTS AND DISCUSSION

In this study, the Naïve Bayes algorithm is used to classify the dataset, with 80% of the data for training and 20% for testing. Model performance evaluation is conducted using a confusion matrix and a classification report.



```
======= CLASSIFICATION REPORT =======

Algorithm : Naive Bayes
Accuracy  : 82.93%

======= TARGET INFORMATION =======
Jumlah Target : 2
Daftar Target : [np.int64(0), np.int64(1)]

--- Binary Metrics ---
True Positives  : 92
False Positives : 22
True Negatives  : 78
False Negatives : 13

--- Detailed Report per Class ---
              precision    recall  f1-score   support

           0       0.86      0.78      0.82       100
           1       0.81      0.88      0.84       105

    accuracy                           0.83       205
   macro avg       0.83      0.83      0.83       205
weighted avg       0.83      0.83      0.83       205
```

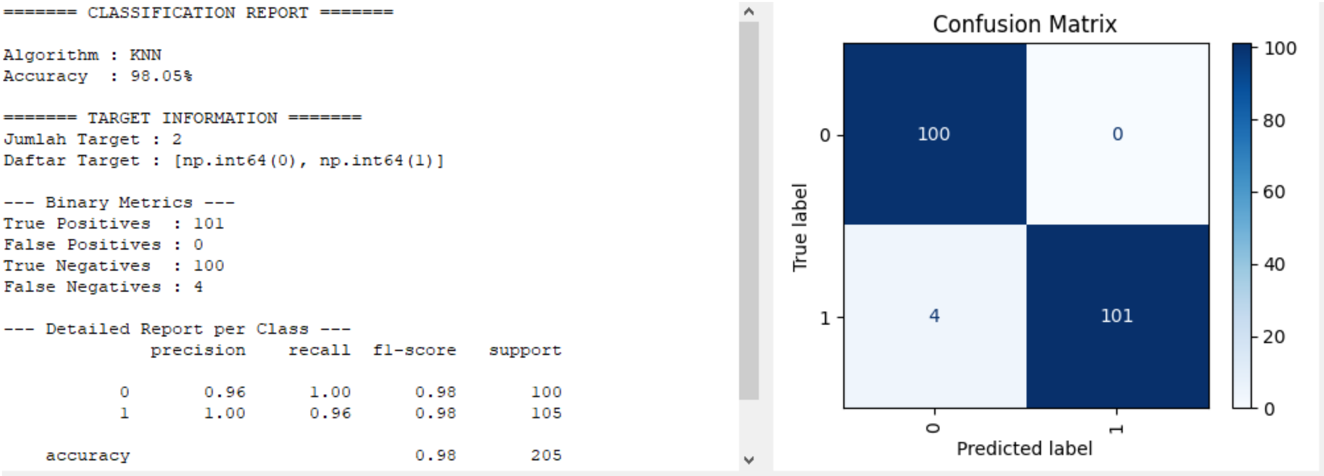
Figure 1. Performance Results from Naive Bayes Confusion Matrix

Based on the evaluated Naive Bayes algorithm for binary diagnostic classification, the model demonstrates moderate predictive capability, distinguishing between normal patients (class 0) and those with heart disease (class 1), achieving an aggregate accuracy of 82.93%. The evaluation was conducted on the established clinical dataset comprising 205 instances, specifically consisting of 100 normal cases and 105 heart disease cases.

An analysis of the resulting confusion matrix reveals a notable increase in classification errors compared to previously evaluated models. The algorithm accurately identified 78 out of the 100 normal instances, establishing the true negative count. Furthermore, it correctly diagnosed 92 of 105 patients with heart disease, representing true positives. However, unlike the previously analyzed algorithms, this Naive Bayes model exhibits a pronounced susceptibility to Type I errors, erroneously diagnosing 22 healthy individuals with heart disease (false positives). Concurrently, the classifier committed 13 Type II errors by misclassifying actual heart disease instances as normal (false negatives).

These misclassifications fundamentally impact the model's detailed performance metrics. The substantial incidence of false positives reduces the precision of heart disease classification to 0.81, indicating that a positive diagnosis from this model is significantly less reliable and prone to false alarms. Because 22 normal instances were misclassified, the recall for the normal class drops to 0.78, indicating the model's inability to identify the entire healthy cohort accurately. While the recall for the heart disease class remains relatively functional at 0.88—indicating it successfully identifies a majority of the actual disease cases—the overall F1-scores of 0.82 for the normal class and 0.84 for the heart disease class delineate a less robust diagnostic tool. Ultimately, this Naive Bayes classifier lacks the high precision and strict conservatism observed in the prior models, rendering it substantially less optimal for rigorous clinical implementation.

Furthermore, using the KNN method, the confusion matrix, which is shown in Figure 2:



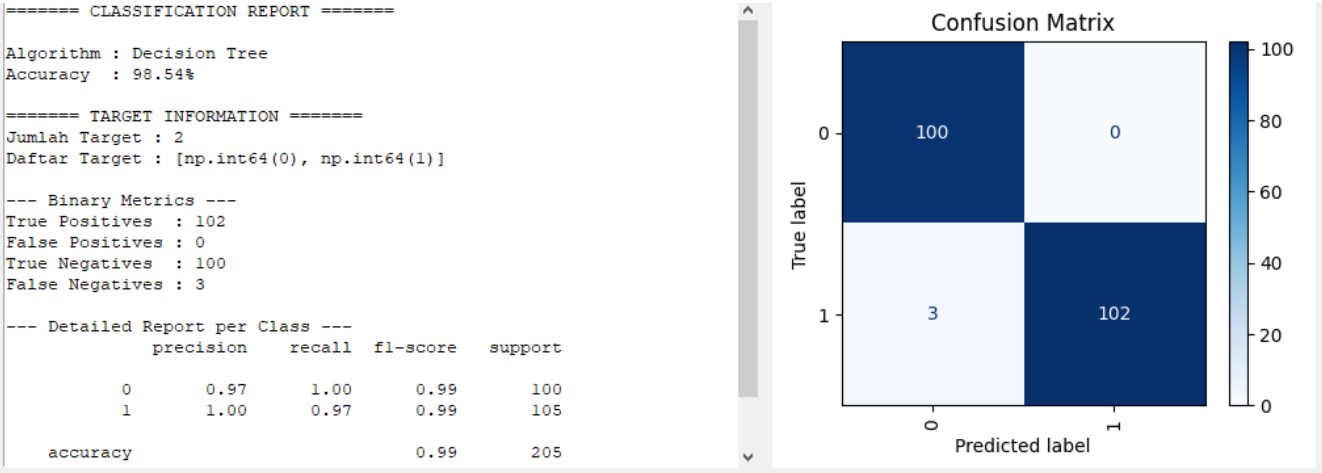
```
======= CLASSIFICATION REPORT =======

Algorithm : KNN
Accuracy  : 98.05%

======= TARGET INFORMATION =======
Jumlah Target : 2
Daftar Target : [np.int64(0), np.int64(1)]

--- Binary Metrics ---
True Positives  : 101
False Positives : 0
True Negatives  : 100
False Negatives : 4

--- Detailed Report per Class ---
              precision    recall  f1-score   support

           0       0.96      1.00      0.98       100
           1       1.00      0.96      0.98       105

    accuracy                           0.98       205
```

Figure 2. Performance Results from KNN Confusion Matrix

Based on the evaluated K-Nearest Neighbors (KNN) model for binary diagnostic classification, the algorithm demonstrates exceptional predictive performance, achieving an aggregate accuracy of 98.05% in distinguishing between normal patients (class 0) and those with heart disease (class 1). The evaluation was conducted on a dataset comprising 205 clinical instances: 100 normal and 105 heart disease cases. An analysis of the resulting confusion matrix reveals that the model accurately identified all 100 normal instances, yielding a perfect true negative rate and a recall of 1.00 for the normal class. Furthermore, the classifier correctly diagnosed 101 of 105 patients with heart disease. Notably, the model exhibited a flawless false-positive rate of zero; no healthy individuals were erroneously diagnosed with heart disease. This absence of Type I errors translates into a perfect precision score of 1.00 for heart disease classification, indicating absolute reliability when the model yields a positive diagnosis.

However, the sole source of error within the predictive model lies in its false negatives. The classifier misclassified four instances of actual heart disease as normal, which constitutes a Type II error. This diagnostic omission reduces recall for the heart disease class to 0.96 and, concurrently, lowers precision for the normal class to 0.96, as the pool of patients predicted to be healthy inadvertently included these four diseased individuals. Ultimately, the trained KNN classifier exhibits highly robust performance, evidenced by exceptional F1 scores of 0.98 across both diagnostic categories. The model is strictly conservative and highly precise in its positive predictions. However, clinical implementation would need to account for its very low risk of missing a small fraction of positive disease cases.

In this study, we also evaluate this dataset using a Decision Tree to assess the effectiveness of this algorithm in classifying the heart disease dataset using all parameters. The result is shown in Figure 3 below:



```
======= CLASSIFICATION REPORT =======

Algorithm : Decision Tree
Accuracy  : 98.54%

======= TARGET INFORMATION =======
Jumlah Target : 2
Daftar Target : [np.int64(0), np.int64(1)]

--- Binary Metrics ---
True Positives  : 102
False Positives : 0
True Negatives  : 100
False Negatives : 3

--- Detailed Report per Class ---
              precision    recall  f1-score   support

           0       0.97      1.00      0.99       100
           1       1.00      0.97      0.99       105

    accuracy                           0.99       205
```

Figure 3. Performance Results from the Decision Tree Confusion Matrix

Based on the provided evaluation metrics for the Decision Tree algorithm, the model demonstrates outstanding diagnostic performance, achieving an aggregate accuracy of 98.54% in classifying normal patients (class 0) and those with heart disease (class 1). Evaluated on the same clinical dataset of 205 instances (100 normal and 105 heart disease cases), the confusion matrix indicates that the model correctly identified all 100 normal cases. This results in a perfect true negative rate and a recall of 1.00 for the normal class. Consistent with the previously evaluated classifier, this Decision Tree model maintains a flawless false-positive rate of zero, ensuring that no healthy individuals were incorrectly diagnosed with heart disease. Consequently, the precision for heart disease classification remains at 1.00, indicating absolute reliability in its positive clinical diagnoses.

Furthermore, the Decision Tree model demonstrates a notable improvement in sensitivity for the positive class compared to the prior algorithm, successfully identifying 102 of 105 patients with heart disease. The incidence of Type II errors has been further reduced, with only three actual heart disease cases being erroneously misclassified as normal. This reduction in false negatives elevates the recall for the heart disease class to 0.97 and concurrently improves the precision for the normal class to 0.97. Overall, the Decision Tree classifier exhibits exceptionally robust, highly optimized performance, underscored by near-perfect F1 Scores of 0.99 across both diagnostic categories. The model's strict conservatism against false alarms (Type I errors), coupled with its improved detection rate for actual disease cases, establishes it as a highly effective and precise predictive tool.

Discussion

The performance metrics achieved by the Decision Tree (98.54%) and K-Nearest Neighbors (98.05%) classifiers in this study demonstrate significant improvements over historical benchmarks established on the UCI Cleveland dataset. Early applications of standard machine learning models, as documented by Detrano et al. [12], typically reported aggregate accuracies clustering in the 80%-85% range. The Naive Bayes result obtained in this current evaluation (82.93%) aligns closely with these traditional baselines and comparative studies by Latha and Jeeva [13], reflecting the algorithm's inherent limitations when its core assumption of strict feature independence is applied to complex clinical data.

However, recent advancements in cardiovascular predictive modeling have significantly raised the boundaries. Studies employing modern hybrid approaches—such as ensemble voting classifiers combining Random Forests and Support Vector Machines—have reported prediction accuracies peaking at approximately 85% [15]. Furthermore, XGBoost models have recently 91% [16].

Contextualized against these benchmarks, the standalone Decision Tree model evaluated in this study exhibits exceptionally robust capabilities. Notably, its aggregate accuracy of 98.54% slightly outperforms several highly complex ensemble and deep learning frameworks documented in recent research, while maintaining a pristine false-positive rate (Type I error = 0).

CONCLUSIONS

A comparative analysis of the K-Nearest Neighbors (KNN), Decision Tree, and Naive Bayes algorithms reveals significant disparities in their diagnostic efficacy for predicting heart disease. Based on the evaluation of the 205 clinical instances, the models can be distinctly ranked by their clinical reliability, statistical accuracy, and handling of critical diagnostic errors.

1. Optimal Performer is a Decision Tree
   With an overall accuracy of 98.54%, this model emerges as the most robust and clinically viable option. It achieves a perfect precision score (1.00) by generating zero false positives (Type I errors), ensuring that no healthy patients are subjected to unnecessary stress or treatment. Furthermore, it boasts the highest sensitivity (Recall: 0.97) among the evaluated models, minimizing missed diagnoses by committing only three false negatives (Type II errors).
2. Strong Alternative is K-Nearest Neighbors (KNN)
   Overall Accuracy: 98.05%. The KNN model performs exceptionally well and maintains the same strict conservatism as the Decision Tree, yielding a flawless zero false-positive rate. Its only minor limitation is a slightly higher false-negative rate (four missed diagnoses compared to the Decision Tree's three), resulting in a slightly lower recall of 0.96 for the positive class.
3. Suboptimal Performer is Naive Bayes
   Overall Accuracy is 82.93%; this algorithm is significantly less reliable for this specific diagnostic task. It demonstrates a pronounced susceptibility to both false positives (22 healthy individuals erroneously diagnosed) and false negatives (13 actual disease cases missed). Its notably lower precision (0.81) and overall F1-scores render it unsuitable for rigorous clinical implementation, where both false alarms and missed diagnoses carry severe consequences.

The Decision Tree is the definitive choice for this diagnostic classification task. It successfully maximizes overall accuracy while minimizing the critical clinical risk of missed diagnoses (false negatives), all without compromising its flawless protection against false alarms (false positives)

REFERENCES

[1] E. J. Benjamin *et al.*, “Heart Disease and Stroke Statistics—2019 Update: A Report From the American Heart Association,” *Circulation*, vol. 139, no. 10, pp. e56–e528, 2019, doi: 10.1161/CIR.0000000000000659.

[2] S. S. Virani *et al.*, “Heart Disease and Stroke Statistics—2020 Update: A Report From the American Heart Association,” *Circulation*, vol. 141, no. 9, pp. e139–e596, 2020, doi: 10.1161/CIR.0000000000000757.

[3] A. Saini, A. J. Meitei, and Kh. J. Singh, “Machine Learning in Healthcare: A Review,” in *Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2021*, Apr. 2021. doi: 10.2139/ssrn. 3834096.

[4] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, “Computational intelligence technique for early diagnosis of heart disease,” in *2015 IEEE International Conference on Engineering and Technology (ICETECH)*, 2015, pp. 1–6. doi: 10.1109/ICETECH.2015.7275001.

[5] M. M. Ahsan and Z. Siddique, “Machine learning-based heart disease diagnosis: A systematic literature review,” *Artif. Intell. Med.*, vol. 128, p. 102289, 2022, doi: https://doi.org/10.1016/j.artmed.2022.102289.

[6] H. Goel and D. Kumar, “Data Mining in Healthcare using Machine Learning Techniques,” in *2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, 2023, pp. 25–29. doi: 10.1109/ICSCSS57650.2023.10169802.

[7] G. Choudhary and S. Narayan Singh, “Prediction of Heart Disease using Machine Learning Algorithms,” in *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, 2020, pp. 197–202. doi: 10.1109/ICSTCEE49637.2020.9276802.

[8] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, “Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare,” *IEEE Access*, vol. 8, pp. 107562–107582, 2020, doi: 10.1109/ACCESS.2020.3001149.

[9] P. Cunningham and S. J. Delany, “k-Nearest Neighbor Classifiers - A Tutorial,” *ACM Comput. Surv.*, vol. 54, no. 6, Jul. 2021, doi: 10.1145/3459665.

[10] D. Berrar, “Bayes’ Theorem and Naive Bayes Classifier,” 2018. doi: 10.1016/B978-0-12-809633-8.20473-1.

[11] L. Rokach and O. Maimon, “Top-down induction of decision tree classifiers - a survey,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 35, no. 4, pp. 476–487, 2005, doi: 10.1109/TSMCC.2004.843247.

[12] I. K. A. Sugitha, A. Triayudi, and E. T. E. Handayani, “Classification of Heart Disease Using the K-Nearest Neighbor Algorithm and Logistic Regression,” *PILAR Nusa Mandiri: Journal of Computing and Information System*, vol. 20, no. 2, pp. 183–190, Sep. 2024, doi: 10.33480/pilar.v20i2.5742.

[13] H. Patel and P. Prajapati, “Study and Analysis of Decision Tree Based Classification Algorithms,” *International Journal of Computer Sciences and Engineering*, vol. 6, pp. 74–78, May 2018, doi: 10.26438/ijcse/v6i10.7478.

[14] R. Detrano *et al.*, “International Application of a New Probability Algorithm for the Diagnosis of Coronary Artery Disease,” *American Journal of Cardiology*, vol. 64, no. 5, pp. 304–310, Aug. 1989, doi: 10.1016/0002-9149(89)90524-9.

[15] C. B. C. Latha and S. C. Jeeva, “Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques,” *Inform. Med. Unlocked*, vol. 16, p. 100203, 2019, doi: https://doi.org/10.1016/j.imu.2019.100203.

[16] S. R. Alty, S. C. Millasseau, P. J. Chowienczyk, and A. Jakobsson, “Cardiovascular disease prediction using support vector machines,” in *2003 46th Midwest Symposium on Circuits and Systems*, IEEE, 2003, pp. 376–379. doi: 10.1109/MWSCAS.2003.1562297.

[17] J. C. Yang, “The prediction and analysis of heart disease using XGBoost algorithm,” *Applied and Computational Engineering*, vol. 41, no. 1, pp. 61–68, Feb. 2024, doi: 10.54254/2755-2721/41/20230711.