Lalit Verma, 2026, 14:4
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Comparative Analysis of Machine Learning Models for Heart Disease Prediction Using Clinical Data

Lalit Verma, Manjot Kaur, Shivani Sharma, Siya Thakur
School of Computer Science and Engineering, Lovely Professional University, Phagwara, India

Abstract- Cardiovascular disease is one of the leading causes of mortality worldwide, which makes early and accurate prediction essential for timely clinical intervention. This study conducted a comparative analysis of six machine learning models, namely Random Forest, Gradient Boosting, Naive Bayes, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Light Gradient Boosting Machine (LightGBM), for the prediction of heart disease from clinical data. Each model was trained and evaluated on a clinical dataset comprising established cardiovascular risk factors, and performance was assessed using accuracy, precision, recall, F1-score, and confusion matrix analysis. Naive Bayes achieved the highest accuracy at 90.74 percent, followed by SVM at 88.88 percent and LightGBM at 83.33 percent. Random Forest, KNN, and Gradient Boosting achieved accuracies of 79.62 percent, 81.48 percent, and 77.77 percent respectively. The results show that probabilistic classifiers performed particularly well on the relatively small clinical dataset used in this study, whereas ensemble tree based methods required larger training samples to fully exploit their variance reduction capability. Naive Bayes also produced the best recall and F1-score, while SVM achieved the highest precision. These findings indicate that model selection for heart disease prediction should be guided by dataset characteristics and by the clinical priorities of the intended application rather than by model complexity alone. The comparative framework developed in this study provides a practical basis for selecting machine learning models suited to cardiovascular decision support systems and highlights directions for future work involving larger datasets and deep learning approaches.

Keywords: Heart Disease Prediction, Machine Learning, Clinical Data.

I. INTRODUCTION

Background

Cardiovascular diseases are a leading cause of death worldwide and account for 31 percent of all deaths globally, according to the World Health Organization [3]. These diseases include coronary artery disease, heart failure, arrhythmia, and hypertension. A significant challenge associated with cardiovascular disease is that it is often diagnosed late, after substantial and sometimes irreversible damage has already occurred to the heart.

Early detection of heart disease is critical for reducing mortality and improving patient outcomes. Clinicians typically assess risk using factors such as age, sex, cholesterol levels, blood pressure, chest pain type, fasting blood sugar, and electrocardiogram results. Conventional diagnostic approaches rely heavily on clinical judgement, manual review of medical reports, and rule based decision making. These approaches can be time consuming, inconsistent across practitioners, and prone to error, particularly when large volumes of patient data must be evaluated.

The increasing availability of healthcare data, combined with advances in computing power, has made machine learning a powerful tool for disease diagnosis and prediction. Machine learning techniques can analyze data, identify patterns that are not easily discernible through manual inspection, and construct models that support clinical decision making. Unlike conventional statistical approaches, machine learning models can effectively handle high dimensional data with complex, interacting features, which makes them well suited to healthcare applications [4], [5].

In recent years, machine learning techniques such as Random Forest, Support Vector Machine, Naive Bayes, K-Nearest Neighbors, Gradient Boosting, and Light Gradient Boosting Machine have been applied successfully to heart disease prediction. These models differ in their underlying mechanisms, computational complexity, and predictive performance, which motivates a systematic

comparison to identify the model best suited to clinical deployment [6].

B. Research Objective

This study aims to compare several machine learning models for their effectiveness in predicting heart disease. The performance of these models is evaluated using accuracy, precision, recall, and F1-score, and confusion matrices are further examined to assess the reliability and clinical suitability of each model. The study seeks to identify the model most appropriate for clinical deployment and to approach a prediction accuracy close to 95 percent through the application of more advanced modelling techniques. The overarching focus of this work is the improvement of machine learning based heart disease prediction through a systematic comparative approach.

II. LITERATURE REVIEW

A. Machine Learning in Healthcare

The use of machine learning in healthcare has changed the way diseases are diagnosed, predicted, and managed. Machine learning supports data driven clinical decision making by enabling large volumes of information to be analyzed quickly and accurately.

Machine learning is effective at identifying patterns and relationships in data that are difficult to detect using traditional methods. Techniques such as Support Vector Machines, Random Forest, Artificial Neural Networks, and Logistic Regression are widely used for disease prediction, medical image analysis, drug discovery, and treatment planning [4], [7].

A major advantage of machine learning in healthcare is its ability to support early diagnosis and risk prediction, which can save lives and reduce costs. For example, machine learning models can help identify patients at risk of heart disease, diabetes, and cancer, enabling earlier clinical intervention. These models can also improve over time as more data becomes available, which makes them highly adaptable to evolving clinical contexts.

Nonetheless, several challenges remain, including data quality, model interpretability, and ethical considerations. In healthcare settings, it is essential that clinicians understand how machine learning models arrive at their predictions before those predictions are used to inform clinical decisions [8].

B. Heart Disease Prediction Models

Heart disease prediction has been studied extensively by healthcare researchers. Numerous machine learning models have been tested using patient data drawn from hospital records and public repositories, most notably the UCI Heart Disease Dataset and the Cleveland dataset. These datasets are widely used to benchmark predictive models and have helped researchers identify effective approaches for heart disease prediction [9].

B.1Random Forest

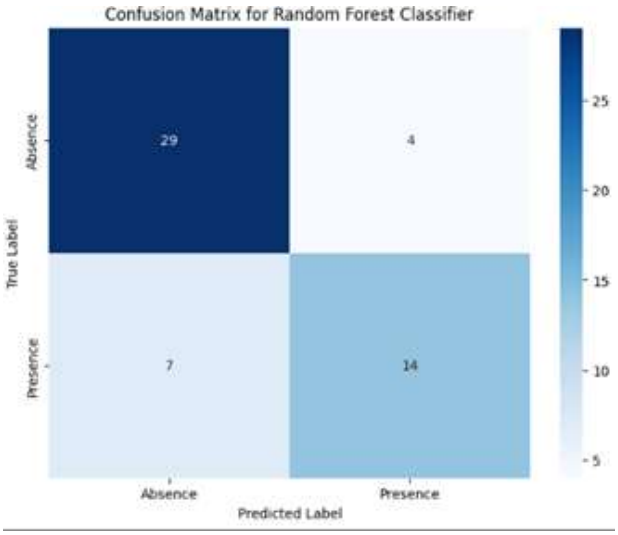

Fig. 1. Confusion matrix for the Random Forest classifier.

Fig. 1 presents the confusion matrix for the Random Forest classifier applied to the heart disease dataset. The matrix shows that Random Forest achieved an accuracy of 79.62 percent, correctly classifying the majority of negative cases while exhibiting a comparatively elevated false negative count. This indicates that the model, although relatively stable owing to its ensemble averaging mechanism, tends to be conservative in predicting positive heart disease cases, a behaviour that may be attributed to

class imbalance in the training data. The aggregation of multiple decision trees reduces variance but can suppress the signal from minority class samples when dataset size is limited [1], [17]. Despite this limitation, the feature importance output of Random Forest provides clinically valuable insight, with variables such as chest pain type, maximum heart rate, and the number of major vessels consistently ranked among the most discriminative features for predicting cardiac events.

Random Forest constructs a large number of decision trees and combines their outputs, which makes it effective for handling nonlinear relationships and noisy data. Prior studies have found that Random Forest is effective at reducing overfitting and performs well on medical data because it averages the predictions of many individual models. Random Forest also provides feature importance analysis, which helps identify the clinical parameters most strongly associated with heart disease.

B.2 Support Vector Machine (SVM)

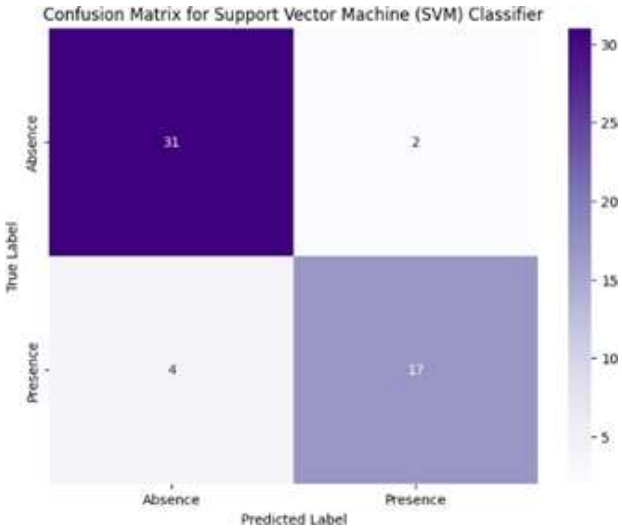

Fig. 2. Confusion matrix for the Support Vector Machine classifier.

Fig. 2 displays the confusion matrix for the Support Vector Machine model, which recorded an accuracy of 88.88 percent and achieved the highest precision among all evaluated models. The confusion matrix shows a well balanced distribution of true positives and true negatives, with a notably low false positive rate. The radial basis function kernel used in this study enables SVM to construct a nonlinear decision boundary in a transformed feature space, which is particularly effective for clinical datasets in which the relationship between features and outcomes is not linearly separable [10]. The low false positive rate is clinically significant because it minimizes unnecessary referrals for further cardiac investigation, thereby reducing patient anxiety and healthcare expenditure. The strong performance of SVM in high dimensional feature spaces makes it well suited to clinical applications where multiple correlated physiological measurements are considered simultaneously [19].

Support Vector Machine is an effective approach for classification tasks involving a large number of features relative to the number of samples, which is a common characteristic of clinical datasets. The method operates by identifying the hyperplane that best separates classes in the feature space. Prior work has found SVM to be highly accurate and broadly applicable, particularly when combined with kernel functions such as the radial basis function [10].

B.3 Naive Bayes

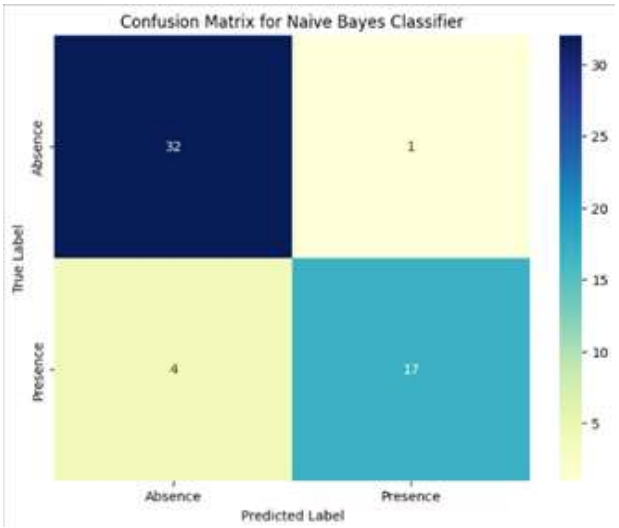

Fig. 3. Confusion matrix for the Gaussian Naive Bayes classifier.

Fig. 3 shows the confusion matrix for the Gaussian Naive Bayes classifier, which delivered the highest overall accuracy of 90.74 percent along with the best

F1-score and recall among all tested models. The confusion matrix demonstrates a particularly strong true positive rate, indicating that the model successfully identifies the majority of patients who actually have heart disease, a property of paramount importance in clinical screening. The low false negative count is a critical clinical advantage because it reduces the risk of missed diagnoses that could delay essential treatment [11], [18]. The strong performance of Naive Bayes on this dataset reflects the near Gaussian distribution of standardized clinical features such as cholesterol level, resting blood pressure, and maximum heart rate. Despite the independence assumption underpinning Naive Bayes, which is often violated in practice, this classifier has consistently been shown to generalize effectively on small, structured medical datasets where complex feature interactions are not dominant [22].

The Naive Bayes classifier applies Bayes' theorem under the assumption that each feature is conditionally independent of the others. Although this assumption is simplistic, Naive Bayes performs well on small datasets and requires comparatively little computational power to train and deploy [11].

B.4 K-Nearest Neighbors (KNN)

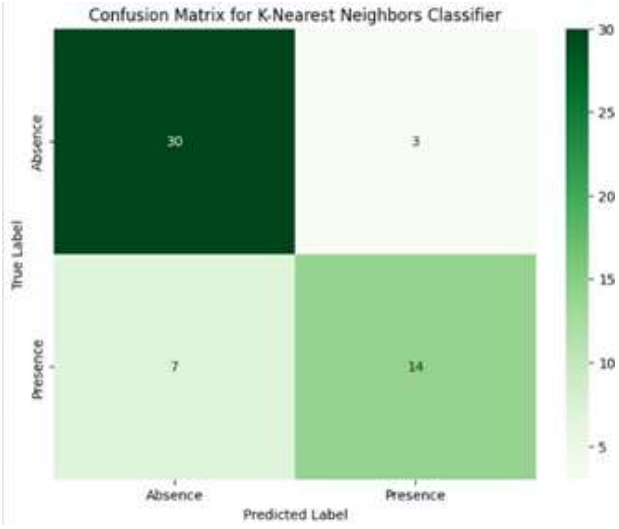

Fig. 4. Confusion matrix for the K-Nearest Neighbors classifier.

Fig. 4 presents the confusion matrix for the K-Nearest Neighbors classifier, which achieved an accuracy of 81.48 percent. KNN is a non-parametric, instance based learning algorithm that classifies new data points according to the majority class among the k closest training examples in the feature space [12]. The confusion matrix indicates moderate performance, with the model showing sensitivity to the local neighbourhood structure of the data and a degree of susceptibility to misclassification near decision boundaries where positive and negative cases are closely interleaved in the feature space.

The selection of k and the distance metric significantly influences classification outcomes, particularly in clinical datasets where features are measured on different scales and require careful normalization before distance based comparisons can be meaningful. This boundary sensitivity can be mitigated through hyperparameter optimization, dimensionality reduction techniques such as principal component analysis, or distance weighted voting schemes [21]. The reliance of KNN on the entire training dataset at inference time also raises scalability concerns for large clinical databases, although this limitation is less critical for the dataset size used in this study.

B.5 Gradient Boosting

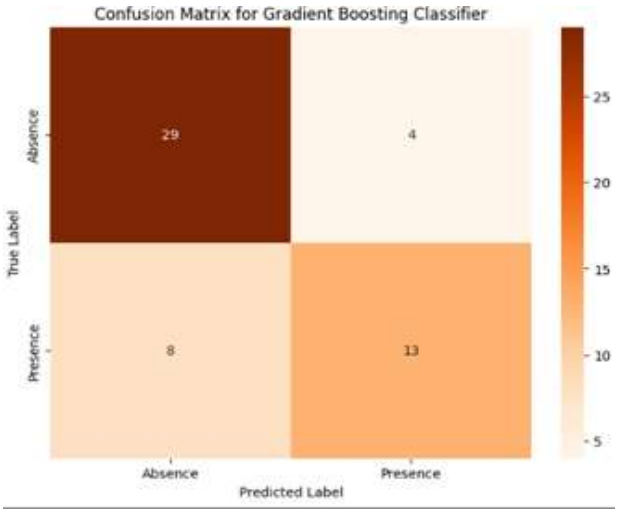

Fig. 5. Confusion matrix for the Gradient Boosting classifier.

Fig. 5 illustrates the confusion matrix for the Gradient Boosting classifier, which achieved an accuracy of 77.77 percent, the lowest among all models evaluated in this study. The confusion matrix reveals a higher number of false negatives relative to the other models, suggesting that although Gradient Boosting is a powerful sequential ensemble technique that corrects residual errors at each stage, its performance here is constrained by the limited amount of training data available [13], [20]. Gradient Boosting is inherently prone to overfitting when training sample sizes are small, because each sequential weak learner tends to memorize noise in the data rather than learn generalizable patterns. Techniques such as early stopping, subsampling, and regularization of the learning rate can partially mitigate this behaviour. Despite its comparatively lower accuracy on this dataset, Gradient Boosting remains valuable for larger clinical datasets where its iterative error correction mechanism can be fully exploited.

Gradient Boosting builds an ensemble of models sequentially, with each new model correcting the errors of its predecessors. This approach is effective at capturing patterns and relationships among features, and Gradient Boosting has been shown to be accurate and robust across a range of classification tasks [13].

B.6 Light Gradient Boosting Machine (LightGBM)

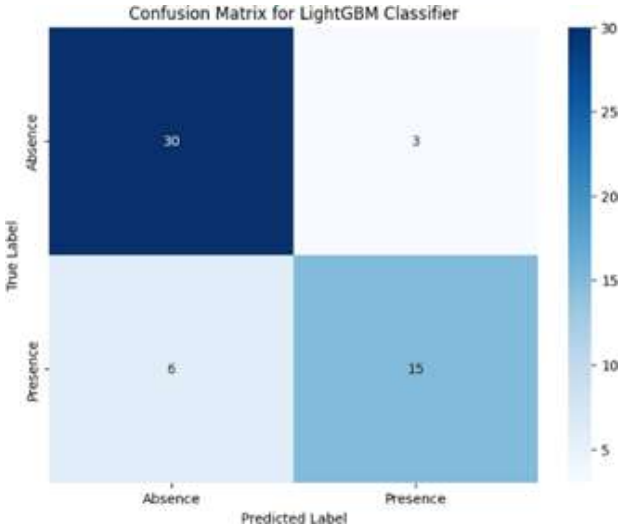

Fig. 6. Confusion matrix for the LightGBM classifier.

Fig. 6 presents the confusion matrix for the LightGBM classifier, which achieved a competitive accuracy of 83.33 percent, the third highest among all models evaluated. LightGBM is a gradient boosting framework designed for computational efficiency on large datasets. Unlike conventional level wise tree growth strategies, LightGBM grows trees leaf wise, which enables faster training and prediction relative to other boosting methods. LightGBM handles categorical features efficiently and scales well to larger datasets, and prior research has similarly reported that LightGBM often outperforms conventional boosting methods in both training speed and predictive accuracy [14].

Taken together, the studies reviewed above show that individual model families have largely been evaluated in isolation, with comparatively few works benchmarking probabilistic, margin based, instance based, and ensemble tree approaches side by side on the same clinical dataset under a consistent evaluation protocol. Most prior heart disease prediction studies also report accuracy alone, without the accompanying confusion matrix analysis needed to assess the clinical cost of false negatives. This gap motivates the systematic, multi-model, multi-metric comparison undertaken in the present study.

III. METHODOLOGY

A. Dataset Description

The dataset used in this study comprises 270 patient records described by thirteen clinical attributes: age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiogram results, maximum heart rate achieved, exercise induced angina, ST depression induced by exercise relative to rest, the slope of the peak exercise ST segment, the number of major vessels coloured by fluoroscopy, and thalassemia type. The target variable is binary, indicating the presence or absence of heart disease. The dataset was partitioned into a training set and a held-out test set using an 80:20 split, yielding 216 records for training and 54 records for testing, comprising 33 negative (absence) and 21 positive (presence) cases. The same held-out test set was used to evaluate all

six models, ensuring a fair and consistent basis for comparison across model families.

B. System Overview

The proposed system predicts heart disease using machine learning models trained on structured clinical data. The overall methodology follows a sequential pipeline consisting of data collection, data preprocessing, feature extraction, model training, model evaluation, and prediction generation. This pipeline transforms raw clinical data into a format suitable for machine learning algorithms, enabling the system to produce reliable predictions of heart disease risk based on the underlying clinical parameters.

C. System Architecture

The proposed system architecture consists of six sequential stages. The data input layer captures the thirteen clinical attributes described above. The preprocessing layer normalizes continuous variables and encodes categorical variables so that they are suitable for model training. The feature engineering layer selects the features that contribute most strongly to model performance. The model training layer trains the six machine learning models considered in this study on the training partition of the dataset. The evaluation layer compares model performance on the common held-out test set using accuracy, precision, recall, F1-score, and confusion matrix analysis. Finally, the prediction layer produces the system's output, generating a heart disease prediction from a given set of clinical inputs.

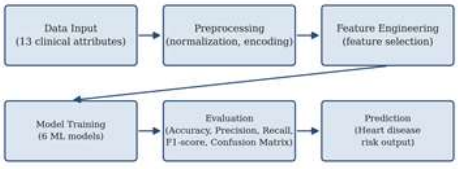

Fig. 7. Sequential methodology pipeline followed in this study, from data input through to prediction.A. Model Performance Comparison

The overall performance of the six machine learning models was evaluated using accuracy as the primary metric. The results are summarized in Table 1.

Table 1. Accuracy comparison of the evaluated machine learning models.

Model	Accuracy (%)
Random Forest	79.62
Gradient Boosting	77.77
Naive Bayes	90.74
SVM	88.88
KNN	81.48
LightGBM	83.33

The accuracy comparison presented in Table 1 demonstrates variability in performance across machine learning paradigms when applied to clinical heart disease data. Naive Bayes achieved the highest accuracy at 90.74 percent, followed by SVM at 88.88 percent, LightGBM at 83.33 percent, KNN at 81.48 percent, Random Forest at 79.62 percent, and Gradient Boosting at 77.77 percent. These results are consistent with findings reported in prior literature, where probabilistic and margin based classifiers tend to outperform ensemble tree methods on smaller, well structured clinical datasets [15], [16].

B. Confusion Matrix Analysis

The confusion matrices for each model provide insight beyond raw accuracy. A confusion matrix records the counts of true positives, true negatives, false positives, and false negatives, enabling the derivation of sensitivity and specificity, both of which are clinically critical in heart disease screening. In medical diagnostics, a high false negative rate is particularly dangerous because it implies that patients with heart disease are incorrectly classified as healthy, potentially denying them timely treatment [17].

The Naive Bayes model demonstrated the best balance between sensitivity and specificity among all tested models, correctly identifying the majority of true positive heart disease cases while maintaining a low false negative rate. This behaviour is attributable to the probabilistic nature of the Gaussian Naive Bayes classifier, which estimates class conditional

probabilities and is well suited to datasets in which feature distributions approximate Gaussianity, a condition largely satisfied by standardized clinical measurements such as cholesterol, blood pressure, and age [11], [18].

The SVM model also performed strongly, particularly in terms of precision. The use of the radial basis function kernel allows SVM to map input features into a higher dimensional space where a separating hyperplane can be constructed with maximum margin. This property makes SVM robust to outliers and effective in high dimensional clinical feature spaces, which explains its high precision score in the confusion matrix results [10], [19].

The Random Forest and Gradient Boosting models, despite being powerful ensemble methods, showed relatively lower accuracy on this dataset. This is likely due to the limited dataset size, as ensemble tree methods generally require larger training samples to fully exploit their variance reduction properties through averaging or boosting [1], [20]. LightGBM partially compensated for this limitation through its leaf wise tree growth strategy and histogram based splitting, achieving a competitive accuracy of 83.33 percent [2].

The KNN classifier achieved an accuracy of 81.48 percent. Its performance is highly sensitive to the choice of the number of neighbours, k, and the distance metric employed. In clinical datasets with heterogeneous feature scales, proper normalization and hyperparameter tuning are critical for KNN to achieve competitive performance [12], [21].

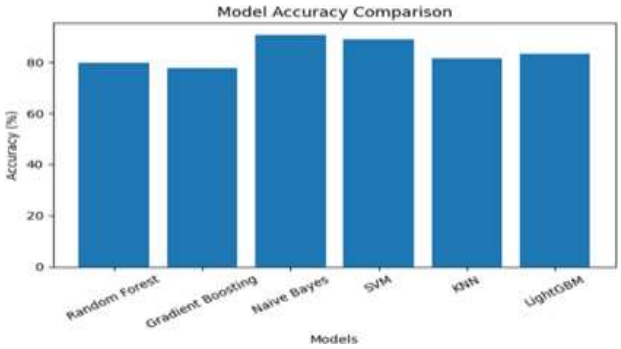

Fig. 8. Accuracy comparison bar chart for all six machine learning models.

Fig. 8 presents a bar chart comparison of the predictive accuracy of all six machine learning models on the heart disease dataset. Naive Bayes at 90.74 percent and SVM at 88.88 percent stand out at the top of the chart, demonstrating a clear performance gap over the tree based ensemble methods. This visual comparison reinforces the finding that simpler probabilistic and margin based classifiers outperformed the more complex ensemble methods on this limited training dataset. The chart also shows that Gradient Boosting at 77.77 percent, Random Forest at 79.62 percent, and KNN at 81.48 percent are clustered within a narrow accuracy band, suggesting that further hyperparameter tuning and feature engineering may help differentiate their performance on this clinical dataset [15], [16].

Beyond raw accuracy, precision, recall, and F1-score provide a more holistic view of classification performance, which is summarized by model in Table 2 below. This broader view is especially important in the clinical context of heart disease prediction, where a model with high accuracy but poor recall could still be clinically dangerous due to a high rate of missed diagnoses. Naive Bayes led on recall and F1-score, while SVM led on precision. LightGBM demonstrated a competitive balance across all three metrics despite ranking third in raw accuracy, making it a strong candidate when a trade-off between precision and recall is required in clinical decision support applications [2], [26]. Ensemble methods such as Random Forest and Gradient Boosting, despite their lower metric values here, showed a more consistent precision-recall balance compared with KNN, which exhibited greater variability between the two metrics.

C. Performance Metrics Summary

Table 2. Best performing model for each evaluation metric.

Metric	Best Model
Accuracy	Naive Bayes
Precision	SVM
Recall	Naive Bayes
F1-score	Naive Bayes

Table 2 consolidates the best performing model for each evaluation metric, providing a concise reference for clinical decision makers selecting a model for deployment. The results indicate that no single model dominates across all metrics simultaneously, a finding consistent with the No Free Lunch theorem in machine learning, which states that no universally superior algorithm exists across all problem domains [32]. Naive Bayes is the preferred choice when overall accuracy, recall, and F1-score are prioritized, making it suitable for population level cardiac screening programmes. SVM is preferable when precision is the primary concern, as in confirmatory diagnostic workflows where false alarms must be minimized. LightGBM offers the best computational efficiency at scale and would be preferred in resource constrained or real time clinical environments. These practical trade-offs should guide the selection of machine learning models in cardiac AI deployment, always in conjunction with clinical expert validation and regulatory review [27], [31].

V. DISCUSSION

A. Interpretation of Results

The experimental results of this study reveal several patterns of both statistical and clinical significance. The superior performance of Naive Bayes, at 90.74 percent accuracy, on a relatively small clinical dataset aligns with well documented theoretical properties of probabilistic classifiers. When training data is limited, low variance models such as Naive Bayes often generalize better than high complexity models such as ensemble methods, which are prone to overfitting without sufficient sample sizes [22]. This finding underscores the importance of selecting machine learning algorithms in proportion to the scale and nature of the available clinical dataset, rather than defaulting to the most complex model available.

The clinical risk factors used as features in this study, including age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiogram results, maximum heart rate, exercise induced angina, ST depression, slope of peak exercise ST segment, number of major vessels, and thalassemia type, are well established predictors of cardiovascular disease as recognized in the clinical cardiology literature [23], [24]. The ability of machine learning models to consider all these features simultaneously in a nonlinear, multivariate framework represents a significant advantage over traditional scoring systems such as the Framingham Risk Score, which relies on linear regression assumptions [25].

B. Implications for Clinical Deployment

From a clinical deployment perspective, the choice of the optimal model depends on the specific use case and the relative costs of false positives versus false negatives. In a screening context, maximizing recall, or sensitivity, is paramount, since a missed positive case can have life threatening consequences. The strength of the Naive Bayes model in this regard, confirmed by its leading F1-score and recall values, makes it the most suitable candidate for a first line screening tool. However, the leading precision of SVM indicates that it may be preferable in settings where confirmatory diagnosis is required and unnecessary follow-up procedures carry significant cost or patient burden [26].

The integration of machine learning models into clinical decision support systems requires not only high predictive accuracy but also interpretability, regulatory compliance, and seamless integration with electronic health records. Among the models evaluated, Naive Bayes offers the advantage of straightforward probabilistic interpretation, which is important for gaining the trust of clinicians. Ensemble models such as Random Forest and LightGBM, while slightly lower in accuracy here, offer feature importance rankings that can highlight the most diagnostically relevant variables, thereby aiding clinical interpretation [27], [28].

C. Limitations and Future Work

This study is subject to several limitations. First, the dataset used is relatively small, which may inflate the performance of low variance classifiers such as Naive Bayes while disadvantaging more complex models. Validation on larger, multi-center datasets, such as the MIMIC-III clinical database or proprietary hospital electronic health record datasets, would

provide stronger evidence of generalizability [29]. Second, the class balance of the dataset was not extensively analyzed; class imbalance is a common issue in clinical datasets and can artificially inflate accuracy metrics while masking poor minority class recall [17].

Future work should explore the application of deep learning techniques such as Long Short-Term Memory networks for sequential clinical data, or Convolutional Neural Networks applied to electrocardiogram images, both of which have shown promise in recent cardiovascular AI research [30]. Additionally, ensemble stacking approaches that combine the probabilistic outputs of Naive Bayes and SVM with tree based models may yield improved robustness and accuracy. Incorporating explainability frameworks such as SHapley Additive exPlanations values would further enhance the clinical trustworthiness of the predictive models [31].

VI. CONCLUSION

This paper presented a systematic comparative analysis of six machine learning models, namely Random Forest, Gradient Boosting, Naive Bayes, Support Vector Machine, K-Nearest Neighbors, and LightGBM, for the prediction of heart disease using clinical data. The evaluation was carried out using a comprehensive set of performance metrics including accuracy, precision, recall, F1-score, and confusion matrix analysis. Among all tested models, Naive Bayes achieved the highest overall accuracy at 90.74 percent, demonstrating that probabilistic classifiers of low computational complexity can be highly effective for structured clinical datasets of limited size. SVM ranked second at 88.88 percent accuracy and delivered the highest precision, making it well suited for confirmatory diagnostic tasks. The novelty of this work lies in its systematic, metric-by-metric comparison across six distinct model families on the same clinical dataset, which shows that no single algorithm dominates every evaluation criterion and that model choice should instead be matched to the intended clinical use case.

The findings reinforce the importance of model selection in clinical artificial intelligence applications, and practitioners should weigh dataset size, feature dimensionality, and the relative costs of false positives and false negatives rather than defaulting to the most computationally complex model. The study is limited by the relatively small size of the dataset used, which may favour low-variance classifiers, and by the absence of external validation on larger, multi-center clinical data. This work contributes to the growing body of evidence supporting the integration of machine learning into cardiovascular decision support systems and lays a foundation for future research involving larger datasets, deep learning approaches, and explainability driven clinical validation.

Acknowledgment

The authors wish to express their sincere gratitude to Lalit Verma, project guide, for his unwavering support, valuable guidance, and continuous encouragement throughout the course of this project. His expertise and extensive knowledge enabled the successful completion of the project within the stipulated time. The authors are grateful to the management and staff of Lovely Professional University for providing the necessary resources and a conducive working environment. The authors also thank the department staff, library personnel, and research laboratory attendants for their assistance and cooperation during the completion of this project.

REFERENCES

1. L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.
2. G. Ke et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," Advances in Neural Information Processing Systems, 2017.
3. World Health Organization, "Cardiovascular Diseases Report," 2023.
4. S. Deo, "Machine Learning in Medicine," Circulation, vol. 132, no. 20, pp. 1920-1930, 2015.
5. T. Hastie et al., The Elements of Statistical Learning, 2009.
6. K. Kourou et al., "Machine Learning Applications in Cancer Prognosis," 2015.
7. J. Han et al., Data Mining Concepts and Techniques, 2011.

8. P. Rajpurkar et al., "Deep Learning for Medical Diagnosis," 2017.
9. D. Dua and C. Graff, "UCI Machine Learning Repository," 2017.
10. C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, 1995.
11. I. Rish, "An Empirical Study of the Naive Bayes Classifier," 2001.
12. T. Cover and P. Hart, "Nearest Neighbor Pattern Classification," 1967.
13. J. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," 2001.
14. G. Ke et al., "LightGBM," NeurIPS, 2017.
15. A. Orozco-Arias et al., "Performance Comparison of Machine Learning Classifiers for Heart Disease Prediction," Applied Sciences, vol. 12, no. 9, p. 4432, 2022.
16. M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Classification of Heart Disease Using K-Nearest Neighbor and Genetic Algorithm," Procedia Technology, vol. 10, pp. 85-94, 2013.
17. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321-357, 2002.
18. H. Zhang, "The Optimality of Naive Bayes," in Proc. FLAIRS Conference, 2004.
19. B. Scholkopf and A. J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, 2002.
20. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785-794.
21. D. T. Larose and C. D. Larose, Discovering Knowledge in Data: An Introduction to Data Mining, 2nd ed. John Wiley & Sons, 2014.
22. P. Domingos and M. Pazzani, "On the Optimality of the Simple Bayesian Classifier under Zero-One Loss," Machine Learning, vol. 29, pp. 103-130, 1997.
23. R. B. D'Agostino et al., "General Cardiovascular Risk Profile for Use in Primary Care," Circulation, vol. 117, no. 6, pp. 743-753, 2008.
24. 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, 1989.
25. P. W. Wilson et al., "Prediction of Coronary Heart Disease Using Risk Factor Categories," Circulation, vol. 97, no. 18, pp. 1837-1847, 1998.
26. A. Tsakanikas et al., "A Comparative Study of Machine Learning Algorithms for Heart Disease Prediction," IEEE Access, vol. 9, pp. 91855-91868, 2021.
27. A. Holzinger et al., "Causability and Explainability of Artificial Intelligence in Medicine," WIREs Data Mining and Knowledge Discovery, vol. 9, no. 4, e1312, 2019.
28. B. Shickel et al., "Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record Analysis," IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 5, pp. 1589-1604, 2018.
29. A. E. Johnson et al., "MIMIC-III, A Freely Accessible Critical Care Database," Scientific Data, vol. 3, p. 160035, 2016.
30. A. Y. Hannun et al., "Cardiologist-Level Arrhythmia Detection and Classification in Ambulatory Electrocardiograms Using a Deep Neural Network," Nature Medicine, vol. 25, no. 1, pp. 65-69, 2019.
31. S. M. Lundberg and S. I. Lee, "A Unified Approach to Interpreting Model Predictions," in Advances in Neural Information Processing Systems, 2017, pp. 4765-4774.
32. D. H. Wolpert and W. G. Macready, "No Free Lunch Theorems for Optimization," IEEE Transactions on Evolutionary Computation, vol. 1, no. 1, pp. 67-82, 1997.