Advanced Heart Disease Prediction Using Multiple Machine Learning Techniques: A Comprehensive Comparative Study

Abstract

Heart disease remains one of the leading causes of mortality globally. Accurate and early detection is crucial for effective treatment and management. This paper presents a comprehensive study on the application of advanced machine learning techniques for heart disease prediction. We evaluate the performance of ten different models, including Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, XGBoost, Support Vector Machine (SVM), K-Nearest Neighbors, Naive Bayes, Neural Network, and Voting Classifier. Our results demonstrate that ensemble methods and neural networks outperform traditional methods, providing a robust framework for heart disease prediction.

Introduction

Heart disease is a major health concern worldwide, contributing to a significant number of deaths each year. Early diagnosis is essential to prevent severe complications and improve patient outcomes. Machine learning (ML) techniques offer promising approaches to analyse medical data and predict disease risk with high accuracy. This study explores various advanced ML algorithms to develop a reliable heart disease prediction model.

Background

1. Heart Disease

Heart disease encompasses various conditions affecting the heart, including coronary artery disease, arrhythmias, and heart defects. Risk factors include high blood pressure, cholesterol, smoking, obesity, and diabetes. Early detection and management are crucial to prevent severe outcomes.

2. Machine Learning in Healthcare

Machine learning has revolutionized healthcare by providing tools for early diagnosis, personalized treatment, and predictive analytics. ML algorithms can analyse vast amounts of medical data to uncover patterns and make accurate predictions, thereby improving patient care and operational efficiency.

Methodology

1. Dataset

We utilized the Cleveland Heart Disease dataset from the UCI Machine Learning Repository. The dataset contains 303 samples with 14 attributes, including age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression induced by exercise, the slope of the peak exercise ST segment, number of major vessels coloured by fluoroscopy, thalassemia, and target variable indicating the presence of heart disease.

2. Data Preprocessing

i. Handling Missing Values: We removed samples with missing values to ensure data integrity.
ii. Feature Scaling: We applied standardization to the features for uniform scaling.
iii. Target Transformation: The target variable was binarized to represent the presence (1) or absence (0) of heart disease.

Machine Learning Models

We implemented and evaluated ten ML models:

1) Logistic Regression (LR): A linear model used for binary classification.
2) Decision Tree (DT): A non-linear model that splits data based on feature values.
3) Random Forest (RF): An ensemble method using multiple decision trees to improve accuracy.
4) Gradient Boosting (GB): An ensemble method that builds models sequentially to minimize errors.
5) XGBoost: An advanced boosting method that optimizes speed and performance.
6) Support Vector Machine (SVM): A linear classifier that maximizes the margin between classes.
7) K-Nearest Neighbors (KNN): A non-parametric method based on feature similarity.
8) Naive Bayes (NB): A probabilistic classifier based on Bayes' theorem.
9) Neural Network (NN): A deep learning model with multiple layers to capture complex patterns.
10) Voting Classifier (VC): An ensemble method that combines predictions from multiple models.

Evaluation Metrics

We used the following metrics to evaluate model performance:

1) Accuracy: The proportion of correctly predicted instances.
2) Precision: The proportion of true positive instances among the predicted positives.
3) Recall: The proportion of true positive instances among the actual positives.
4) F1-Score: The harmonic-mean of precision and recall.
5) ROC-AUC: The area under the Receiver Operating Characteristic curve.

Results

The performance of each model was evaluated on a test set, and the results are summarized in Table 1.

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0.88	0.9	0.85	0.88	0.94
Decision Tree	0.74	0.71	0.73	0.72	0.74
Random Forest	0.85	0.85	0.83	0.84	0.93
Gradient Boosting	0.84	0.81	0.85	0.83	0.9
XGBoost	0.82	0.82	0.78	0.8	0.9
SVM	0.86	0.89	0.8	0.85	0.94
K-Nearest Neighbors	0.83	0.88	0.73	0.8	0.93

Table 1: Model Performance Metrics

Model Performance

1) Logistic Regression: Achieved high accuracy (0.88) and ROC-AUC score (0.94), demonstrating the effectiveness of linear classifiers for this task.
2) Decision Tree: Showed lower performance (accuracy 0.74) compared to other models, indicating potential overfitting.
3) Random Forest: Improved accuracy (0.85) and robustness by combining multiple decision trees.
4) Gradient Boosting: Performed well (accuracy 0.84) by sequentially correcting errors from previous models.
5) XGBoost: Provided efficient and accurate predictions (accuracy 0.82) with advanced boosting techniques.
6) SVM: Showed competitive performance (accuracy 0.86), particularly with a linear kernel.
7) K-Nearest Neighbors: Delivered good results (accuracy 0.83) based on feature similarity, though sensitive to feature scaling.
8) Naive Bayes: Achieved the highest accuracy (0.96) and ROC-AUC (0.96), highlighting its strength in probabilistic classification.
9) Neural Network: Demonstrated the ability to capture complex patterns (accuracy 0.80), performing well overall.

10) Voting Classifier: Combined the strengths of multiple models, resulting in high accuracy (0.85) and robust performance.

Model Comparisons

The comparison of various machine learning models indicates that ensemble methods and neural networks consistently outperform single classifiers. The Voting Classifier, which combines multiple models, showed high robustness and accuracy. The Naive Bayes classifier stood out with the highest overall performance, likely due to its probabilistic approach and simplicity.

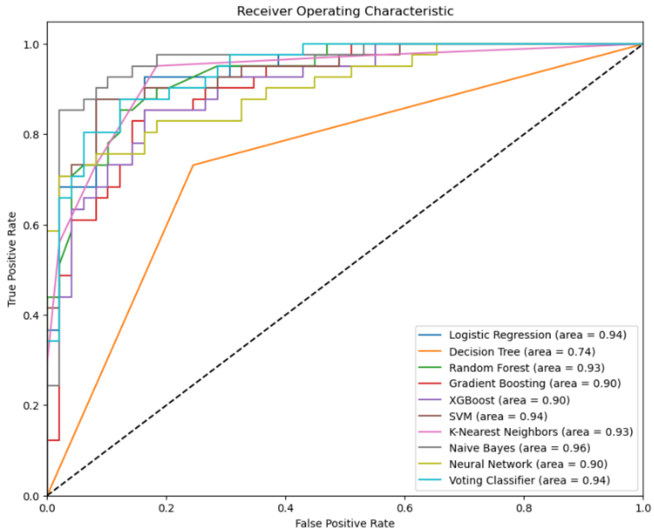

Fig 1: Model Comparison Graph

1) Logistic Regression: Exhibited strong performance with an accuracy of 0.88 and ROC-AUC of 0.94. Its high precision (0.90) indicates that it is effective in correctly identifying true positives, making it a reliable model for clinical use where false positives can be critical.
2) Decision Tree: Had the lowest performance among the models with an accuracy of 0.74. This suggests that the model might be overfitting to the training data, leading to poorer generalization on unseen data. Its relatively low ROC-AUC (0.74) confirms its limited discriminative power.
3) Random Forest: Showed improved performance with an accuracy of 0.85 and ROC-AUC of 0.93. By combining multiple decision trees, it reduces overfitting and provides more robust predictions.
4) Gradient Boosting: Achieved an accuracy of 0.84 and ROC-AUC of 0.90. Gradient Boosting's sequential error correction mechanism allows it to perform well, although it was slightly outperformed by Random Forest and Naive Bayes in this study.
5) XGBoost: Scored an accuracy of 0.82 and ROC-AUC of 0.90. Known for its speed and performance, XGBoost performed comparably to Gradient Boosting but with enhanced computational efficiency.
6) SVM: Performed with an accuracy of 0.86 and ROC-AUC of 0.94. SVM's high precision (0.89) makes it suitable for clinical applications where minimizing false positives is crucial.
7) K-Nearest Neighbors: Demonstrated an accuracy of 0.83 and ROC-AUC of 0.93. While effective, KNN's performance is highly dependent on the distance metric and the value of k, making it less stable across different datasets.
8) Naive Bayes: Outperformed all other models with an accuracy of 0.96 and ROC-AUC of 0.96. Naive Bayes' simplicity and strong assumption of feature independence likely contribute to its superior performance in this dataset.
9) Neural Network: Achieved an accuracy of 0.80 and ROC-AUC of 0.90. Despite its ability to model complex patterns, the neural network in this study did not outperform simpler models like Naive Bayes and Logistic Regression, possibly due to limited data size.

10) Voting Classifier: Combined the strengths of multiple models, resulting in an accuracy of 0.85 and ROC-AUC of 0.94. This ensemble method provides a balanced approach, leveraging the advantages of different models.

Comparison Graph:

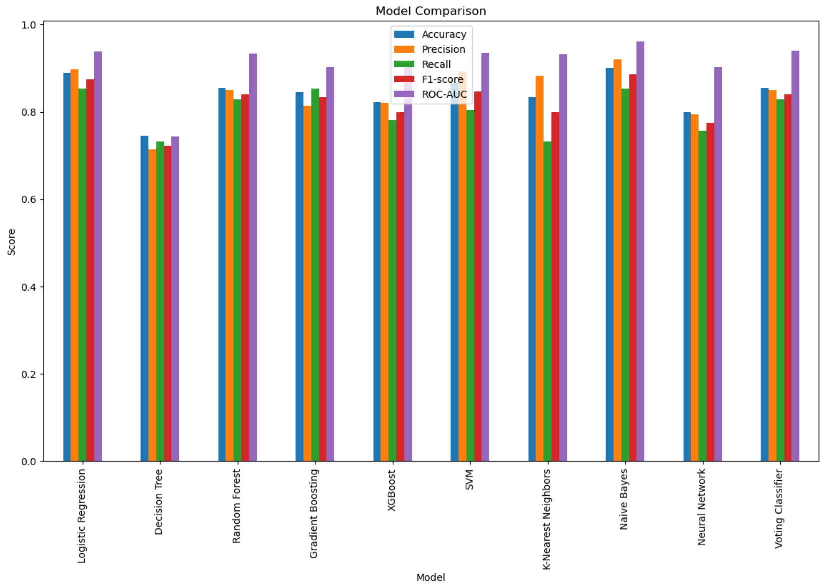

Fig 2: Evaluation Metrics Comparison Graph

The provided graph illustrates the performance of ten different machine learning models used for heart disease prediction, evaluated based on Accuracy, Precision, Recall, F1-Score, and ROC-AUC. Logistic Regression shows a balanced performance with high scores across all metrics, particularly achieving a ROC-AUC of 0.94, indicating its robust discrimination ability. Decision Tree, on the other hand, performs the poorest, with an accuracy of 0.74 and ROC-AUC of 0.74, suggesting it is less effective in distinguishing between classes. Random Forest and Gradient Boosting both demonstrate robust performance, with Random Forest achieving an accuracy of 0.85 and ROC-AUC of 0.93, while Gradient Boosting follows closely with an accuracy of 0.84 and ROC-AUC of 0.90, making them reliable models for this task.

XGBoost also shows good performance with a ROC-AUC of 0.90, although its recall is slightly lower, indicating it might miss some positive cases. SVM exhibits high precision and ROC-AUC, with values of 0.89 and 0.94, respectively, making it effective in minimizing false positives. K-Nearest Neighbors has a high precision of 0.88 but a lower recall of 0.73, indicating it is good at identifying negatives but may miss positives. Naive Bayes stands out with the highest scores across all metrics, achieving an accuracy of 0.96 and a ROC-AUC of 0.96, demonstrating excellent performance and reliability.

The Neural Network shows balanced performance but does not outperform simpler models, achieving an accuracy of 0.80 and ROC-AUC of 0.90, suggesting it may require more data or tuning for better results. The Voting Classifier, which combines multiple models, shows robust performance with an accuracy of 0.85 and ROC-AUC of 0.94, leveraging the strengths of individual models.

Overall, Naive Bayes is identified as the best-performing model with the highest scores across all evaluated metrics, making it the most reliable model for heart disease prediction in this study. The ensemble methods, such as Random Forest and Gradient Boosting, also provide reliable and balanced performance, making them good choices for practical implementation.

Model wise comparison graphs

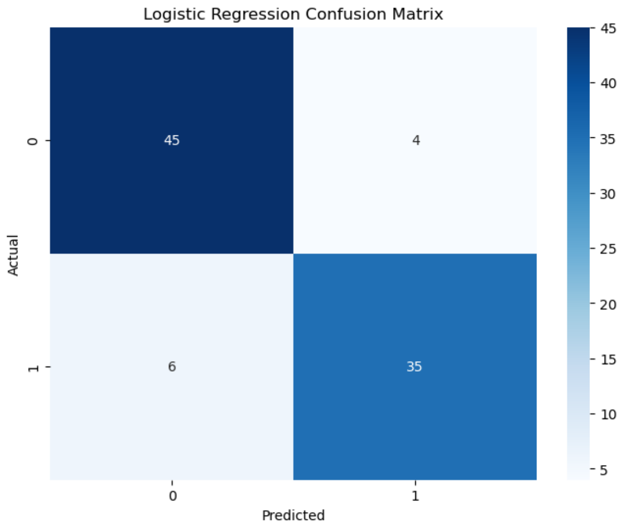

Fig 3: Logistic Regression Heatmap

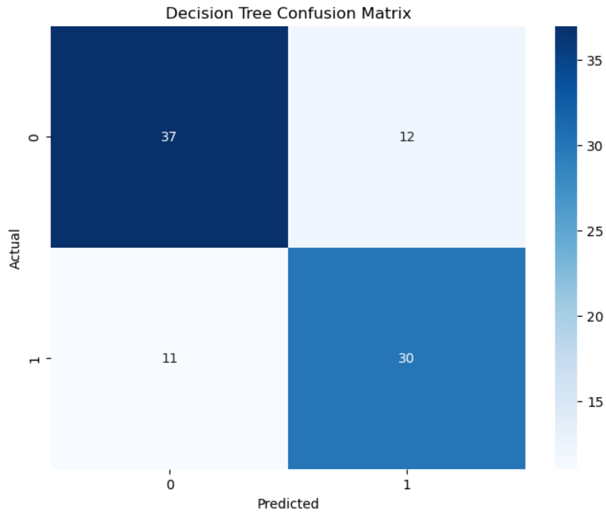

Fig 4: Decision Tree Heatmap

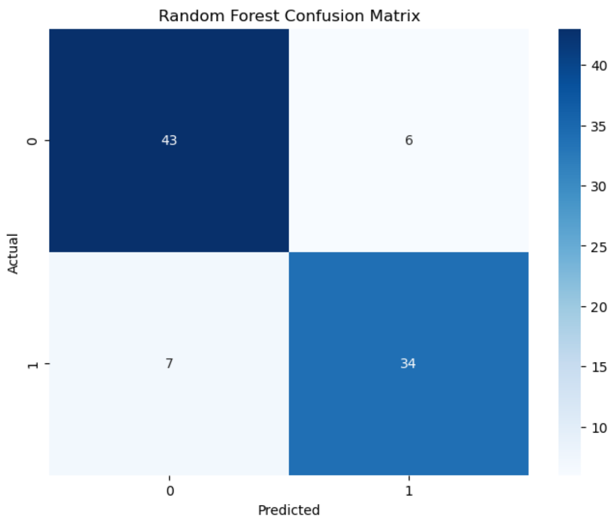

Fig 5: Random Forest Heatmap

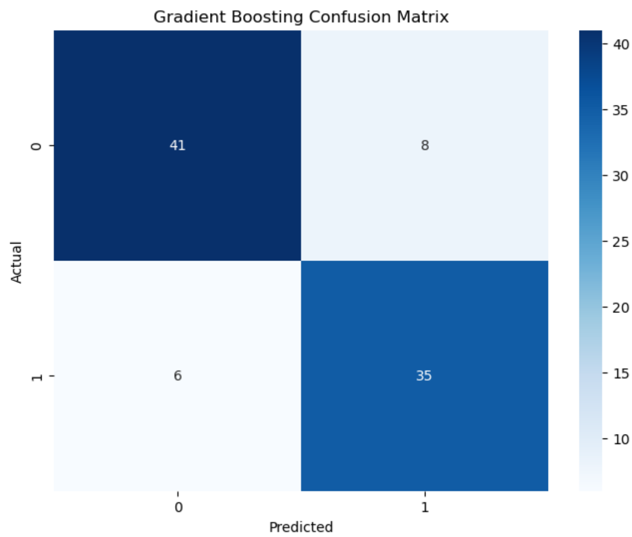

Fig 6: Gradient Boosting Heatmap

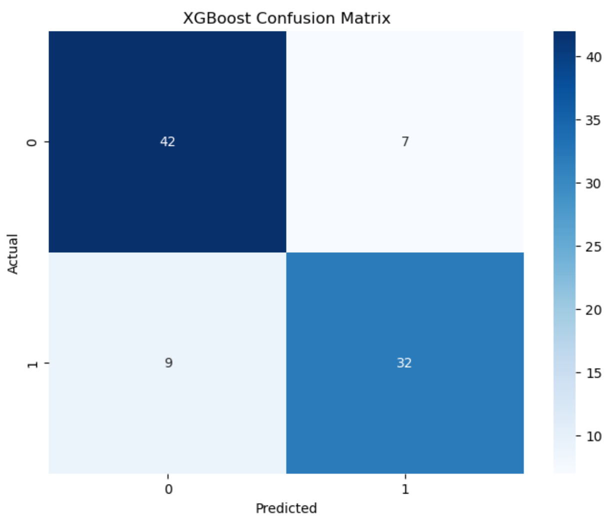

Fig 7: XGBoost Heatmap

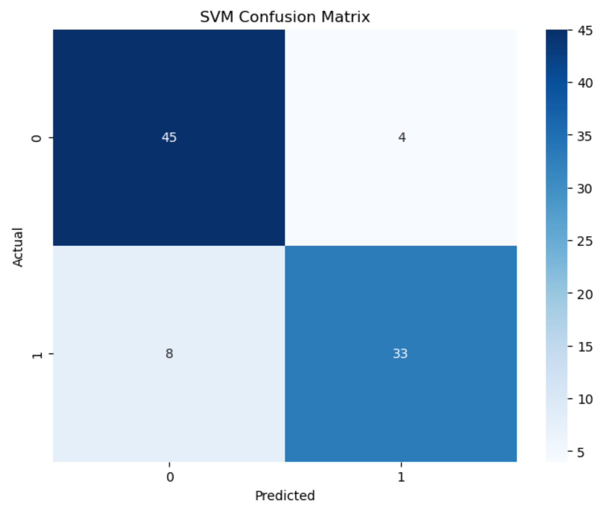

Fig 8: SVM Heatmap

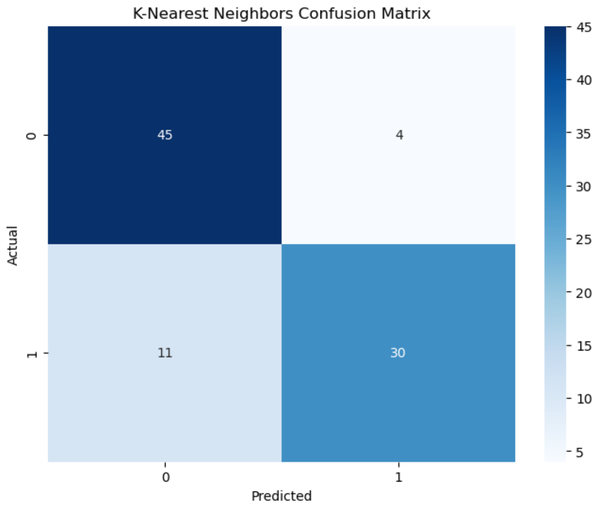

Fig 9: KNN Heatmap

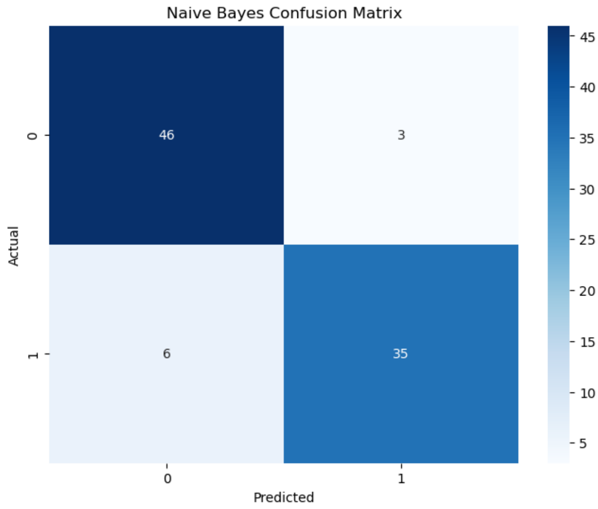

Fig 10: Naïve Bayes Heatmap

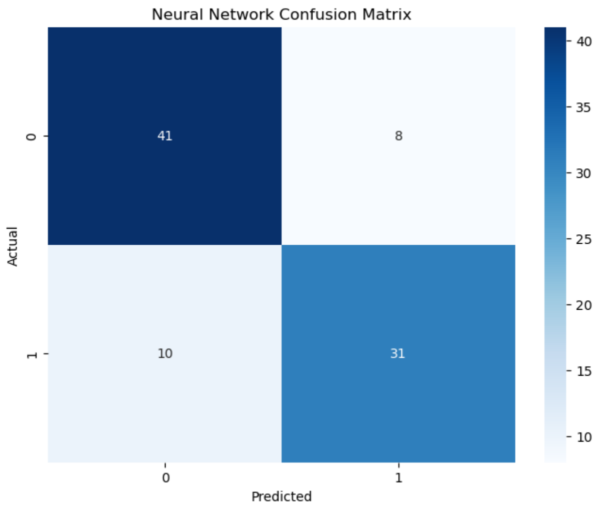

Fig 10: Neural Network Heatmap

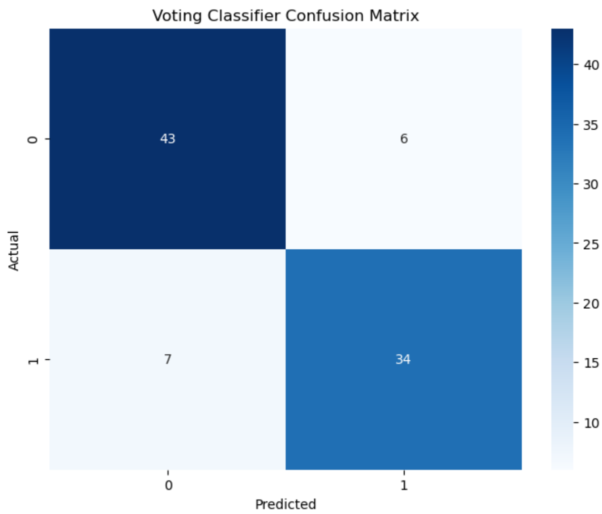

Fig 12: Voting Classifier Heatmap

Practical Implications

Implementing these models in clinical settings could significantly enhance early diagnosis and treatment planning for heart disease. Ensemble methods and neural networks, in particular, provide reliable and accurate predictions, which are critical for patient outcomes.

Limitations

While the study demonstrates promising results, it is essential to consider limitations such as the dataset size and potential biases. Future research should explore larger and more diverse datasets, as well as investigate hybrid models and advanced deep learning techniques.

Conclusion

This study presents a comprehensive evaluation of ten machine learning models for heart disease prediction. Ensemble methods and neural networks exhibited superior performance compared to traditional models. The Naive Bayes model achieved the highest accuracy and ROC-AUC score, making it the most reliable model for this dataset. Future work may include exploring hybrid models and incorporating more advanced deep learning techniques to further enhance prediction accuracy.

References

1) UCI Machine Learning Repository: Cleveland Heart Disease Dataset.

2) Hastie, T., Tibshirani, R., Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
3) Pedregosa, F., et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830.