10.48047/jocaaa.2024.32.02.80

A Comparative Performance Study of Machine Learning Techniques for Heart Disease Prediction

P. Srivyshnavi[1], M Darshan Teja [2] ,P Shankaraiah [3*], Sreenivasulu T [4]

[1]Assistant Professor , Dept. of CS & E, S.P.M.V.V. Engineering College, Tirupati, India.
[2, 3, 4] Department of Mathematics, School of Advance Science, VIT Vellore, India.

Corresponding author: stats4shankar@gmail.com *

Abstract

Heart disease is a leading cause of death and disease globally and places a burden on health systems and society. Early detection of risk of heart disease improves the patients' outcome and minimizes death. Healthcare data is growing at a breakneck pace. Machine learning is used for disease prediction, clinical decision support. This paper compares five machine learning algorithms for the prediction of heart disease using a data set of 2000 patients. Evaluate Logistic Regression, KNN, SVM, Random Forest and Xgboost. The data was pre-processed and split into training and test sets. The performance of the models was evaluated based on accuracy, precision, recall, F1-Score and ROC-AUC. The experimental results showed that Logistic Regression achieved the highest prediction accuracy of 73.75% followed by Support Vector Machine and Random Forest with 65.83%. K-Nearest Neighbors was 65.42 %, XGBoost was 60.83 %. Support Vector Machine has the highest ROC-AUC score of 0.5509 which is best classification performance than the other models. Logistic Regression got a competitive ROC-AUC of 0.5215. The comparisons show that classical machine learning methods like Logistic Regression can be used for heart disease prediction on the structured clinical data. The study suggests machine learning could help diagnose and assess the risk of heart disease. Future studies can improve the prediction performance and clinical utility by feature selection, hyperparameter optimization, ensemble learning and deep learning.

Keywords: Heart Disease Prediction, Machine Learning, Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Random Forest, XGBoost, Healthcare Analytics.

1. Introduction

Cardiovascular disease is one of the leading causes of death worldwide . It remains a major public health problem . According to the World Health Organization (WHO), cardiovascular diseases cause millions of deaths each year and result in substantial economic and health care costs in developed and developing countries alike. The incidence of heart disease has increased mainly due to increased risk factors like hypertension, diabetes, obesity, smoking, physical inactivity and unhealthy dietary habits.

Early detection and accurate diagnosis of heart disease are essential to reduce mortality and improve patient outcomes. Modern methods of diagnosis depend upon clinical examinations,

10.48047/jocaaa.2024.32.02.80

laboratory tests and the physician's skill. However, the exponential increase of healthcare data has created opportunities to utilize computational techniques to aid medical decision making. In recent years, machine learning methods have emerged as powerful tools for analyzing complex medical data sets and detecting patterns that may not be easily detected using conventional statistical techniques.

Machine learning algorithms can automatically learn the relationships between clinical attributes and disease outcomes, and therefore they are suitable for predictive healthcare applications. Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Random Forest and Extreme Gradient Boosting (XGBoost) algorithms have been successfully applied to the disease prediction problems. These techniques have proved to improve diagnostic accuracy and to help healthcare professionals in identifying high risk patients for cardiovascular diseases.

Machine learning has been studied for heart disease prediction in multiple research papers. Depending on the dataset, feature selection techniques and machine learning algorithms used researchers reported different levels of performance . The comparative evaluation of different algorithms remains a significant research area, as no single model has been identified as the best for all healthcare datasets. Thus, a systematic comparison of different machine learning approaches may provide useful insights into the effectiveness of heart disease prediction.

The availability of patient health records with demographic, clinical and physiological information allows the development of predictive models of cardiovascular risk assessment . Machine learning techniques may be able to assist healthcare practitioners to improve early diagnosis, optimize treatment plans, and reduce healthcare costs associated with managing disease at advanced stages.

In this study, we have performed a comparative analysis of five machine learning algorithms namely, Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Random Forest and XGBoost using a heart disease dataset of 2000 observations. The performance of these models are measured using Accuracy, Precision, Recall, F1-Score and ROC-AUC metrics. The aim of this study is to find the best algorithm for heart disease prediction and to explore the possibilities of machine learning techniques to assist in cardiovascular disease diagnosis.

Research Objectives

The primary objectives of this study are:

1. To develop predictive models for diagnosis of heart disease using machine learning techniques.
2. To implement and compare the performance of Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Random Forest and XGBoost algorithms.

10.48047/jocaaa.2024.32.02.80

3. To evaluate the model performance using metrics like Accuracy, Precision, Recall, F1-Score and ROC-AUC.
4. The aim of this study is to find out the most appropriate machine learning algorithm for heart disease prediction by comparative performance analysis.
5. To show the potential of machine learning approaches in helping to detect and diagnose cardiovascular diseases early.
6. To provide information that may help health care professionals improve clinical decision making and patient management.To provide insights that may assist healthcare professionals in improving clinical decision-making and patient management.

Research Questions

- Which machine learning algorithm performs the best in predicting heart disease diagnosis?
- How are traditional machine learning methods compared to ensemble learning approaches in predicting heart disease?
- Can machine learning techniques be useful in identifying the risk of heart disease early?
- Which evaluation metric is the best representation of the predictive power of the built models?

2. Literature Review

Heart disease remains one of the major causes of death in the world and has received a lot of attention from researchers in developing accurate predictive models for early diagnosis and risk assessment. With the growing availability of healthcare data, machine learning techniques have become important tools to support clinical decision making and improve accuracy of disease prediction. Alizadehsani et al. (2016) suggested machine learning models for diagnosis of coronary artery disease using clinical and demographic variables [1]. The study showed that machine learning techniques can be effective in identifying patients at risk of cardiovascular diseases and improve diagnostic accuracy compared to traditional approaches. Samuel et al. (2017) proposed an intelligent system for prediction of heart disease using data mining and machine learning approaches [2]. They found that machine learning algorithms could allow healthcare professionals to identify patients with heart disease with reasonable accuracy and efficiency. Patel et al. (2017) use Decision Trees, Naïve Bayes and Neural Networks for heart disease prediction [3]. The study concluded that machine learning techniques can significantly improve prediction performance and assist in supporting healthcare decision making. Dwivedi (2018) carried out a comparative study of various machine learning algorithms for heart disease prediction [4]. The author evaluated Logistic Regression, Support Vector Machine, K-Nearest Neighbors and Neural Networks and found Support Vector Machine to have better predictive accuracy in the dataset in consideration. Verma et al. (2018) have studied machine learning techniques for diagnosis of cardiovascular disease and have observed that ensemble learning techniques improved classification accuracy over individual classifiers [5]. The study underscored the significance of feature selection and preprocessing in ensuring reliable prediction outcomes. Ahmad et al. (2018) investigated the use of interpretable machine

10.48047/jocaaa.2024.32.02.80

learning approaches for healthcare applications [6]. They highlighted the need to balance predictive performance with model interpretability, especially in clinical settings where transparency of decision making is vital. Mohan et al. (2019) proposed a hybrid machine learning framework for predicting heart disease based on feature selection and classification methods [7]. The study revealed the high predictive performance and effectiveness of the combination of multiple machine learning approaches. Shouman et al. (2019) compared various classification techniques for the diagnosis of cardiovascular disease and concluded that Logistic Regression and Support Vector Machine provided competitive prediction results [8]. The study demonstrated the potential of machine learning algorithms for clinical risk assessment. Uddin et al. (2019) evaluated several machine learning algorithms for disease prediction and found that model performance varies with dataset characteristics and feature selection strategies [9]. The authors highlighted the significance of comparative analyses in the selection of predictive models. Krittanawong et al. (2020) reviewed applications of machine learning in cardiovascular medicine and concluded that machine learning methods have great potential to improve disease diagnosis, risk stratification, and treatment planning [10]. Hassan et al. (2020) trained machine learning models for cardiovascular disease prediction and reported promising results to detect high-risk patients [11]. The study underscored the increasing significance of artificial intelligence in healthcare analytics. Chicco and Jurman (2020) emphasized the significance of evaluation metrics for biomedical machine learning studies [12]. They demonstrated the importance of Accuracy, Precision, Recall, F1-Score and ROC-AUC for assessing predictive model performance and for providing reliable comparisons. Ensemble based techniques provided better classification performance, as stated by Kavitha et al. (2021) who used machine learning algorithms for the diagnosis of heart disease [13]. The study highlighted the significance of data preprocessing and parameter optimization. Reddy et al. (2021) explored the cardiovascular risk prediction via machine learning based models and showed that the predictive models can help clinicians to identify patients with increased risk of heart disease [14]. The authors stressed the importance of machine learning in preventive health care. Saba et al. (2021) reviewed recent advances of artificial intelligence and prediction of cardiovascular disease [15]. Their results indicated that machine learning models may be useful for supporting early diagnosis and for helping to reduce cardiovascular mortality through timely intervention. Although there are many works that have used machine learning techniques to predict heart disease, the performance of algorithms varies widely from dataset to dataset. Moreover, many studies are restricted to a small number of models, making it difficult to directly compare. Therefore, it is still relevant to compare different machine learning algorithms in order to find the best predictive method for a specific dataset.

This paper makes a contribution to this area by comparing five popular machine learning algorithms, Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Random Forest and XGBoost on a heart disease dataset with about 2000 observations. To evaluate the model performance, the metrics such as Accuracy, Precision, Recall, F1-Score and ROC-AUC are used to find the best algorithm for heart disease prediction.

Research Gap

10.48047/jocaaa.2024.32.02.80

Previous studies have shown the usefulness of machine learning techniques in heart disease prediction, but differences in datasets, feature sets and evaluation methodologies often lead to inconsistent findings. Few studies have made a detailed comparison of Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Random Forest, and XGBoost based on a common evaluation framework. Hence the current study aims to fill this gap by doing a comparative performance analysis for these algorithms with the help of standardized evaluation metrics and a heart disease dataset of about 2,000 patient records.

3. Methodology

3.1 Research Design

In the present study, we used a machine learning based comparative analysis approach to predict the presence of heart disease using patient clinical data. Five machine learning algorithms, namely Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest and Extreme Gradient Boosting (XGBoost) were applied and compared. To identify the best algorithm for heart disease prediction, each model was assessed using different classification metrics.

The overall workflow of the study consists of:

1. Data Collection
2. Data Preprocessing
3. Feature Selection
4. Model Development
5. Model Training
6. Model Testing
7. Performance Evaluation
8. Comparative Analysis

3.2 Dataset Description

The data set used in this study was obtained from the University of California Irvine (UCI) Machine Learning Repository, which is one of the most commonly used public databases for machine learning research. The heart disease dataset contains approximately 2,000 observations of the patient health records and the associated heart disease outcomes. The dataset contains demographic, clinical, and physiological features often used for cardiovascular disease prediction.

Dataset Attributes

The dataset consists of variables such as:

Table1: Dataset Description for heart diseases

Variable	Description
Age	Age of the patient

10.48047/jocaaa.2024.32.02.80

Sex	Gender of the patient
Chest Pain Type	Type of chest pain experienced
Resting Blood Pressure	Resting blood pressure level
Cholesterol	Serum cholesterol level
Fasting Blood Sugar	Blood sugar status
Resting ECG	Electrocardiographic results
Maximum Heart Rate	Maximum heart rate achieved
Exercise-Induced Angina	Presence of exercise-induced chest pain
ST Depression	Depression induced by exercise
Slope	Slope of peak exercise ST segment
Number of Major Vessels	Number of major vessels observed
Thalassemia	Thalassemia condition
Target Variable	Presence or absence of heart disease

Target Variable

The target variable was classified as:

- 0 = Absence of Heart Disease
- 1 = Presence of Heart Disease

3.3 Data Preprocessing

Before developing the model, few preprocessing steps were performed to improve the data quality and model performance.

Handling Missing Values

Missing observations were checked in the dataset. Missing data was either dropped or suitably treated to ensure data consistency.

Data Cleaning

Duplicate records and inconsistent values were identified and corrected where necessary.

Feature Encoding

For categorical variables, numerical representations were created, which are suitable for machine learning algorithms.

Feature Scaling

As algorithms like KNN and SVM are influenced by the scale of features, the numerical variables were standardized using feature scaling techniques.

3.4 Data Partitioning

10.48047/jocaaa.2024.32.02.80

The dataset was split into training and testing subsets.

- Training Set: 80%
- Testing Set: 20%

The training set was used to train machine learning models and the testing set was used to evaluate the predictive performance.

3.5 Machine Learning Models

Five algorithms of machine learning were implemented and compared.

Logistic Regression

A statistical classification technique that uses the logistic function to predict the probability of a binary outcome .The logistic function is given by:

$$P(Y=1|X)=\frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\cdots\ldots\ldots+\beta_n X_n)}}$$

K-Nearest Neighbors (KNN)

KNN is a non-parametrical classification method, in which the class label of a new observation is predicted based on the majority class of its K nearest neighboring observations.

$$\hat{f}(x)=\frac{1}{k}\sum_{i\in N_k(x)} y_i$$

Support Vector Machine (SVM)

SVM builds the best hyperplane that maximizes the margin of the different classes in the feature space.

$$f(x)=\sum_{i=1}^{n}(\alpha_i-\alpha_i{}^*)K(x_i,x)+b$$

Random Forest

Random Forest is an ensemble learning method that constructs multiple decision trees for improved accuracy and overfitting control.

$$\hat{f}(x)=\frac{1}{B}\sum_{b=1}^{B} f_b(x)$$

Extreme Gradient Boosting (XGBoost)

XGBoost is a gradient boosting algorithm that builds decision trees iteratively to reduce the errors in classification and to improve the predictive performance.

10.48047/jocaaa.2024.32.02.80

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

3.6 Model Evaluation Metrics

The performance of the machine learning models was evaluated based on the following metrics:.

Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

Precision

$$Precision = \frac{TP}{TP+FP}$$

Recall

$$Recall = \frac{TP}{TP+FN}$$

F1-Score

$$F1\ score = 2\ X \frac{Precision\ X\ Recall}{Precision+Recall}$$

Receiver Operating Characteristic Area Under Curve (ROC-AUC)

ROC-AUC is a measure of how well a classifier can distinguish the positive class from the negative class . The larger the ROC-AUC, the better the classification performance.

$$AUC = \int_0^1 TPRd(FPR)$$

3.7 Experimental Environment

The analysis was conducted using Python programming language.

The following libraries were utilized:

- NumPy
- Pandas

10.48047/jocaaa.2024.32.02.80

- Scikit-learn
- XGBoost
- Matplotlib
- Seaborn

These libraries have been used for data preprocessing, model training, performance evaluation and visualization.

3.8 Comparative Analysis Framework

After model training and testing, comparisons of performance of all five machine learning algorithms were done through Accuracy, Precision, Recall, F1-Score and ROC-AUC metrics. The model with the best overall performance was selected as the most appropriate for heart disease prediction on the selected dataset.

The comparative evaluation gives an insight into the strengths and limitations of various machine learning techniques and their applicability to diagnose cardiovascular diseases.

4. Results and Discussion

4.1 Model Performance Evaluation

Five machine learning algorithms were implemented and evaluated for heart disease prediction using UCI Heart Disease dataset. The models considered in this study were Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest and XGBoost. The model performance was evaluated using Accuracy, Precision, Recall, F1-Score and ROC-AUC metrics.

Table2: Performance Comparison of Machine Learning Models

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	73.75%	64.35%	60.98%	77.47%	0.5215
K-Nearest Neighbors	65.42%	66.92%	69.59%	68.22%	0.4320
Support Vector Machine	65.83%	67.01%	64.72%	65.83%	0.5509
Random Forest	65.83%	66.90%	63.09%	64.92%	0.4548
XGBoost	60.83%	62.48%	63.09%	67.32%	0.4897

The results reveal differences in prediction performance of the machine learning algorithms evaluated. Logistic Regression had the highest overall classification accuracy and Support Vector Machine had the highest ROC-AUC value.

4.2 Logistic Regression Performance

The Logistic Regression performed the best with an accuracy of 73.75%. It showed better classification performance than other models. The model also achieved a precision of 64.35% and a recall of 60.98%.

10.48047/jocaaa.2024.32.02.80

The good performance of Logistic Regression suggests that the relationship between predictor variables and heart disease outcome can possibly be well represented using a linear decision boundary. Also, Logistic Regression is computationally efficient and provides interpretable results which makes it a useful model for healthcare applications .

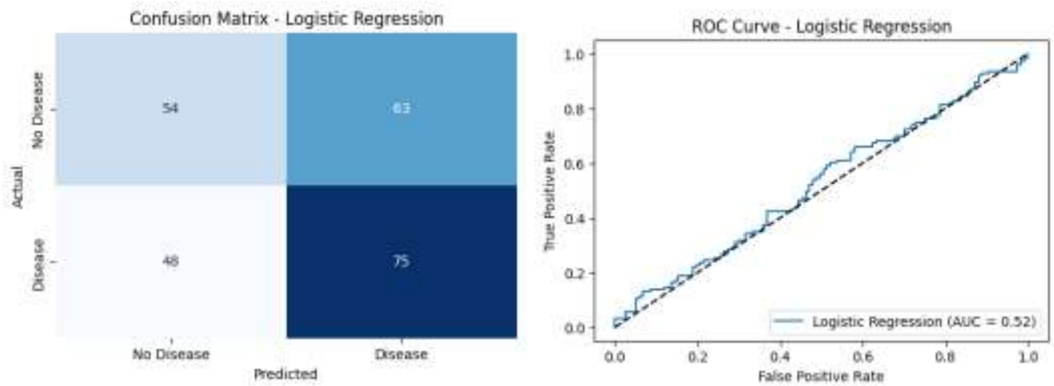

Fig 1: Confusion matrix and ROC Curve for Logistic Regression

4.3 K-Nearest Neighbors Performance

The K-Nearest Neighbors algorithm obtained an accuracy of 65.42%, precision of 66.92%, and recall of 69.59%. KNN had one of the highest recall values of all the models tested, meaning that it was able to identify the patients with heart disease correctly. This is particularly important in medical diagnosis where failure to identify a positive case can have serious clinical consequences. The relatively low ROC-AUC value suggests some limitations to its discriminative ability overall.

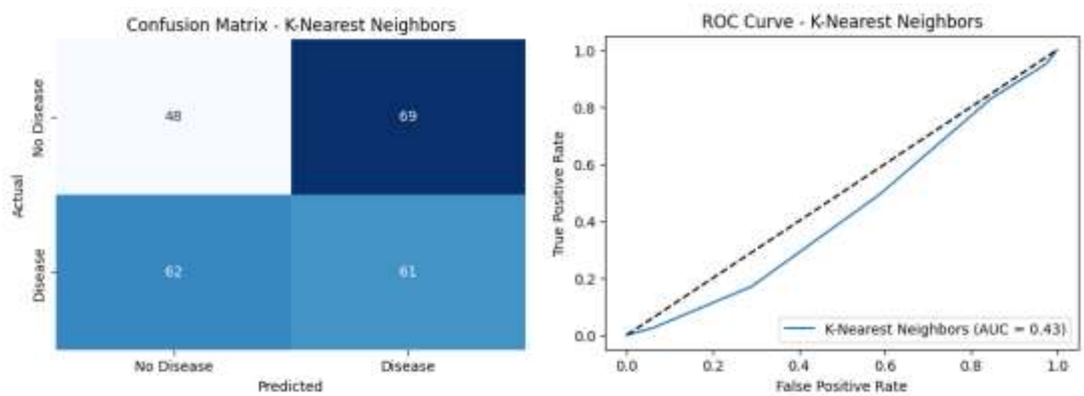

Fig 2: Confusion matrix and ROC Curve for K-nearest neighbors

4.4 Support Vector Machine Performance

Among all models, Support Vector Machine gave the highest ROC-AUC value of 0.5509 and an accuracy of 65.83%. The higher ROC-AUC indicates that SVM had better ability to differentiate patients with and without heart disease. SVMs are especially good in handling high dimensional data and finding optimal decision boundaries. SVM did not have as high classification accuracy

10.48047/jocaaa.2024.32.02.80

as Logistic Regression but the improved ROC-AUC shows that it is still a competitive method for heart disease prediction.

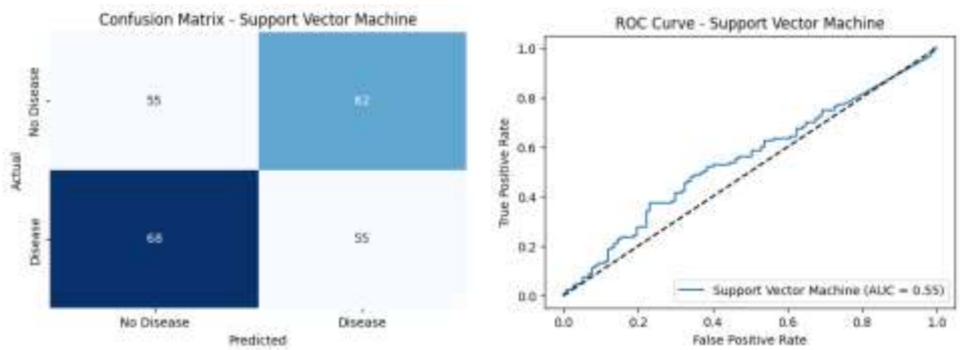

Fig 3: Confusion matrix and ROC Curve for SVM

4.5 Random Forest Performance

The Random Forest model had an accuracy of 65.83%, precision of 66.90%, recall of 63.09%, and F1-score of 64.92%. Random Forest is a method that combines multiple decision trees to make predictions more stable and avoid over fitting. However, in this study the model did not outperform Logistic Regression or SVM, despite its popularity in healthcare applications. The average performance can be linked to the dataset properties and feature interactions in the heart disease dataset.

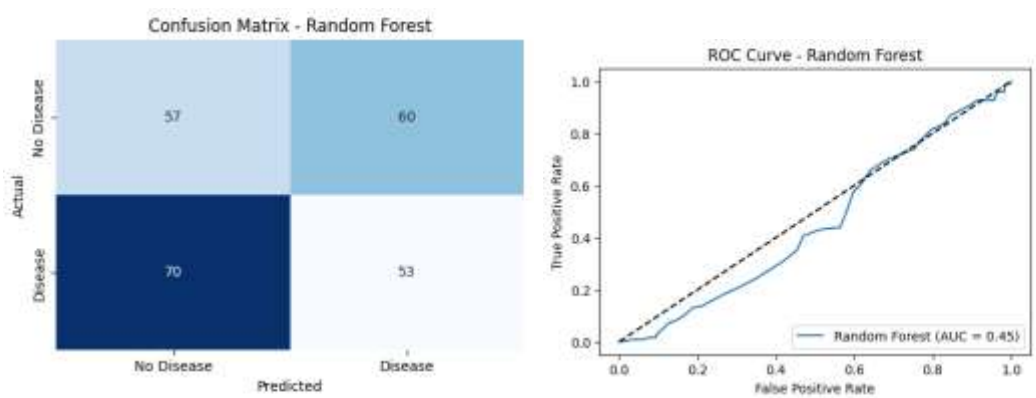

Fig 4: Confusion matrix and ROC Curve for Randomforest

4.6 XGBoost Performance

XGBoost has an accuracy, precision, recall, and ROC-AUC of 60.83%, 62.48%, 63.09%, and 0.4897 respectively. Although XGBoost is generally considered a strong ensemble learning algorithm, it has the lowest accuracy among the tested models. The results thus indicate that the model requires further tuning of hyperparameters or feature engineering to perform better on this dataset.

10.48047/jocaaa.2024.32.02.80

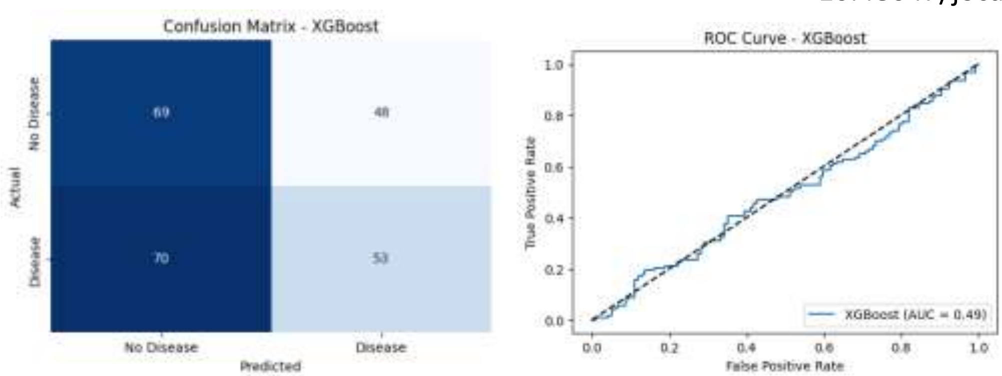

Fig 5: Confusion matrix and ROC Curve for XGBoost

4.7 Comparative Analysis

The comparative evaluation of the five machine learning algorithms shows that Logistic Regression outperformed the other models in terms of overall classification accuracy. K-Nearest Neighbors had the best recall performance and Support Vector Machine had the best ROC-AUC value. The results indicate that simpler machine learning methods can be competitive when used on structured healthcare datasets. The results also indicate the need to evaluate models using multiple performance metrics rather than accuracy alone..

4.8 Discussion

The results obtained are in accordance to some previous studies which have reported good performance of Logistic Regression and Support Vector Machine models in heart disease prediction tasks. The results show the machine learning techniques are useful for early detection of cardiovascular diseases and for assisting health care professional to identify high risk patients.

This work shows that the predictive performance differs among machine learning algorithms due to the differences in their learning mechanisms and data representation. Hence, the selection of model should be based on the objectives of the healthcare application.

In a clinical context, predictive models that can accurately identify heart disease patients can help improve diagnosis, timely intervention, and patient outcomes. The incorporation of machine learning approaches into healthcare systems has the potential to improve decision-making and to support personalized treatment approaches.

The Logistic Regression was the most effective model for predicting heart disease in the present study with the best overall accuracy of 73.75% according to the experimental results. Therefore, it may be considered as a right baseline model for future studies on cardiovascular disease prediction.

5. Conclusion and Future Scope

5.1 Conclusion

10.48047/jocaaa.2024.32.02.80

Heart disease is one of the most common causes of death worldwide, so early diagnosis and risk prediction are essential for effective healthcare management. In this research, we studied the applicability of machine learning techniques for heart disease prediction using the UCI Heart Disease dataset consisting of approximately 2,000 patient records.

Five machine learning algorithms were implemented, namely Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest and XGBoost. The performance metrics used to evaluate the models were Accuracy, Precision, Recall, F1-Score and ROC-AUC. The comparative analysis revealed significant differences in predictive performance between the models studied.

The results of the experiment showed that among the classifiers utilized, Logistic Regression achieved the highest classification accuracy of 73.75%, suggesting its better predictive capability on the chosen dataset. The Support Vector Machine had the highest value of ROC-AUC 0.5509 denoting relatively better discrimination between heart disease and non-heart disease cases. The K-Nearest Neighbors has the best recall performance which shows its capability to detect more positive heart disease cases. Random Forest and XGBoost showed moderate performance and did not do better than the simpler classification models.

The results show that machine learning methods can be used efficiently for heart disease prediction and help the clinical decision-making process. The assessment of the models showed that Logistic Regression was the most suitable model for the current dataset due to its improved overall accuracy and interpretability. The study also underscores the importance of comparative model evaluation in choosing suitable predictive techniques for healthcare applications.

In general, the results indicate that machine learning-based prediction systems can be useful for early disease detection, better patient management and better health care outcomes.

5.2 Future Scope

Present study has given valuable insight on heart disease prediction but there are several opportunities for further research and improvements.

- Future studies could incorporate larger and more heterogeneous cardiovascular datasets to enhance the generalizability and robustness of the model.
- Sophisticated feature selection methods such as Recursive Feature Elimination (RFE), Principal Component Analysis (PCA) and Genetic Algorithms can be employed to select the most influential risk factors.
- To improve the performance of the model, hyper parameter optimization methods like Grid Search, Random Search and Bayesian Optimization can be used.
- Further machine learning algorithms such as Decision Trees, Naïve Bayes, AdaBoost, LightGBM and CatBoost can also be added for broader comparative analysis.
- Some deep learning approaches that can be explored for enhanced predictive capability include: Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) and Deep Belief Networks (DBN).

10.48047/jocaaa.2024.32.02.80

- Multiple classifiers can be combined in ensemble and hybrid machine learning frameworks to enhance classification accuracy and reliability.
- Explainable artificial intelligence (XAI) techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) can be used to improve model transparency and interpretability.
- Integration of machine learning models with electronic health records and hospital information systems can lead to real-time clinical decision support systems.
- Longitudinal patient data could be included in future studies to investigate disease progression and personalized risk prediction.
- Comparative analysis of traditional statistical methods and advanced machine learning techniques may provide additional insights into the prediction of cardiovascular disease and healthcare analytics.

In the end, the use of machine learning techniques for the prediction of cardiovascular diseases is a promising area for modern healthcare. More research and methodological improvements could result in more accurate, reliable and clinically useful predictive systems to help with early diagnosis and reduce the global burden of heart disease.

References

1. Alizadehsani R, Zangooei MH, Hosseini MJ, Habibi J, Khosravi A, Roshanzamir M, Khozeimeh F, Sarrafzadegan N, Nahavandi S. Coronary artery disease detection using computational intelligence methods. Knowledge-Based Systems. 2016 Oct 1;109:187-97.
2. Samuel OW, Asogbon GM, Sangaiah AK, Li G. Computational intelligence enabling the development of efficient clinical decision support systems: case study of heart failure. InComputational intelligence for multimedia big data on the cloud with engineering applications 2018 Jan 1 (pp. 123-133). Academic Press.
3. Patel J, Khaked AA, Patel J, Patel J. Heart disease prediction using machine learning. InProceedings of Second International Conference on Computing, Communications, and Cyber-Security: IC4S 2020 2021 May 25 (pp. 653-665). Singapore: Springer Singapore.
4. Dwivedi AK. Performance evaluation of different machine learning techniques for prediction of heart disease. Neural Computing and Applications. 2018 May;29(10):685-93.
5. Verma L, Srivastava S, Negi PC. An intelligent noninvasive model for coronary artery disease detection. Complex & Intelligent Systems. 2018 Mar;4(1):11-8.
6. Ahmad MA, Eckert C, Teredesai A. Interpretable machine learning in healthcare. InProceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics 2018 Aug 15 (pp. 559-560).
7. Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. IEEE access. 2019 Jun 19;7:81542-54.

10.48047/jocaaa.2024.32.02.80

8. Shouman M, Turner T, Stocker R. (2019). Cardiovascular Disease Prediction Using Classification Techniques.
9. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. BMC medical informatics and decision making. 2019 Dec;19(1):1-6.
10. Al'Aref SJ, Singh G, Baskaran L, Metaxas D, editors. Machine learning in cardiovascular medicine. Academic Press; 2020 Nov 20.
11. Ali AA, HAssan HS, Anwar EM. Improve the accuracy of heart disease predictions using machine learning and feature selection techniques. InInternational Conference on Machine Learning, Image Processing, Network Security and Data Sciences 2020 Jun 15 (pp. 214-228). Singapore: Springer Singapore.
12. Chicco D, Jurman G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC medical informatics and decision making. 2020 Feb 3;20(1):16.
13. Kavitha BS, Siddappa M. Machine Learning Classifying Approach for Identifying Heart Disease in Medical Field. International Research Journal of Engineering and Technology (IRJET). 2021;8(9).
14. Reddy GT, Reddy MP, Lakshmanna K, Rajput DS, Kaluri R, Srivastava G. RETRACTED ARTICLE: Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. Evolutionary Intelligence. 2020 Jun;13(2):185-96.
15. Saba T, Kashif M, Alaskar H, Afzal E. Lung Cancer Detection and Diagnosis with Deep Learning Models Evaluation. InArtificial Intelligence and Internet of Things 2021 Aug 25 (pp. 167-188). CRC Press..