Heart disease detection using different machine learning algorithms

Baikani Akhila

Department of Artificial Intelligence and Data Science

Vardhaman College of Engineering

Hyderabad, India

baikaniakhila2004@gmail.com

Jaitaram Thrisha Reddy

Department of Artificial Intelligence and Data Science

Vardhaman College of Engineering

Hyderabad, India

thrishajaitharam@gmail.com

Yarram Sai Ruthik

Department of Artificial Intelligence and Data Science

Vardhaman College of Engineering

Hyderabad, India

sairuthik2004@gmail.com

Shanmugasundaram Hariharan

Department of Artificial Intelligence and Data Science

Vardhaman College of Engineering

Hyderabad, India

mailtos.hariharan@gmail.com

Vinay Kukreja  
Centre for Research Impact & Outcome, Chitkara University,

Rajpura,Punjab, India

[vinay.kukreja@chitkara.edu.in](mailto:vinay.kukreja@chitkara.edu.in)

Murugaperumal Krishnamoorthy  
*Department of Electrical and Electronics Engineering,*   
*Vardhaman College of Engineering*Hyderabad, India  
[murugaee@gmail.com](mailto:murugaee@gmail.com)

*Abstract*—Now a days,Heart Disease had became an significant issue world wide.Early prediction of heart disease can be done using Machine learning algorithms. The main objective of writing this paper is to develop a quick and easy method for predicting the heart disease using machine learning algorithms . ML can be used to improve the prognosis of heart disease.This article examines the results of various ML algorithms using different performance evaluation metrics.In this paper,we used six(6) machine learning algorithms i.e,Support Vector Machine (SVM),Decision tree, Naive Bayes, Logistic Regression,Random forest and also K-Nearest Neighbors (KNN). In comparison to other ML algorithms, the Random forest classifier approach attained a very high accuracy of 95.00%, according to the data obtained from the classification report.Machine learning contain the effective techniques for making decisions. The medical industry makes extensive use of machine learning techniques. This study offers the machine learning methods for predicing heart disease.The data set used in this is taken from the kaggle. We developed an approach to predict heart disease using actual patient data. This work uses different classification algorithms. This mixed technique is implemented using Google Colab.

*Keywords*— Heart disease, Predictive models, Supervised learning , Random Forest, Support Vector Machine, Cardiovascular issues, Mortality rate.

# Introduction

According to reports from the World Health Organization (WHO), heart disease is a global health concern. Heart disease risk factors include advanced age, high blood pressure, abnormal ECG results, male gender, exercise-induced chest discomfort, lower maximal heart rate, higher cholesterol, old peak with physical activity, and uneven ST slope [10]. Predictive or machine learning models use these parameters, along with the type of chest discomfort and fasting blood sugar, to assess the risk of heart disease.A significant issue that results in many deaths is a lack of early-stage diagnosis. As a result, reliable and efficient prediction models are essential for early identification of cardiac disease.Recent statistics indicate that over 17 million individuals on a globe suffer from heart failure [5]. The causes of heart failure may be separated into two groups: those associated with structural damage, such as a prior heart attack, and those associated with functional dysfunctions, such as high blood pressure.Some symptoms include fatigue,swelling of abdomen,coughing ,shortness of breath,loss of appetite. Treatment options include prescription medication,lifestyle modifications, and, in some cases, surgery.. Research suggests that quick recognition and appropriate treatment of heart failure can improve overall well-being and increase longevity. The purpose of this work is to create a ML model for improved patient outcomes and heart failure management.Early detection of these heart disease may help in saving lots of human lives [6].

According to the reports emphasized in the American Heart Association 2024 depicts the following results.In 1950 there were 60% of the heart failure deaths are reduced due to the advancements in the medical technology.But in 20th century due to the several factors there is the rise in heart disease rates.The factors include age,sex,chest pain type,resting BP s,cholesterol,fasting blood sugar,resting ECG,max heart rate,exercise angina,old peak,ST slope etc.43% of the females and more than 50% of the males are effected by these heart disease due to high blood pressure.Above 45% of the people of over the age 20 are effected due to cardiac insufficiency.Obesity is one of the factor which contribute a larger part for the heart failure.71% of the American adults are overweight and it had lead to heart failures on a large scale. In 2020, there was a significant rise in deaths related to cardiovascular issues due to the COVID-19 outbreak, especially affecting Asian, Black, and Hispanic communities.This study assists in identifying heart disease through various factors. Our data set includes key information such as age,max heart rate,chest pain type,resting BP,fasting blood sugar,resting ECG,sex,exercise angina,cholesterol,oldpeak,ST slope. The data set also includes a target variable called "target," which is a binary variable that indicates whether or not heart disease exists, with values of 0 (no heart disease) and 1 (heart illness)[1].

Analyzing this data set allows us to explore the relationships between different variables. By identifying significant risk factors, our goal is to improve the understanding of heart disease causes and recommend preventive measures and personalized health care interventions.Heart is the vital organ in the human body for survival.Heart diseases can be fatal. Factors such as Unhealthy diet(diets high in saturated fats), Smoking(damage blood vessels), Excessive alcohol consumption(can raise blood pressure), High cholesterol can lead to heart disease [7].The heart is an essential organ that pumps blood throughout the body, provides oxygen, and removes toxins.Heart Disease prediction is crucial for improving patient predicting and reducing mortality rate. The mortality rate from heart disease continues to rise across all age groups. The primary objective is to evaluate the effectiveness of these algorithms in early heart disease prediction which improves the patient survival rates [8]. Machine learning revolutionizes healthcare by facilitating early disease detection through robust data analysis techniques [2].

Machine learning is the subset of Artificial Intelligence.Machine learning can be defined as a machine is able to take its decision on its own without any human interference and perform actions based on the past trained data. Machine learning is the study of making a machine think like human.This paper's primary goal is to give doctors a tool for making first heart disease diagnoses. By doing this, adverse effects can be prevented and patients can receive successful treatment . When it comes to identifying various underlying patterns and subsequently analyzing the given data, machine learning is crucial. Following data analysis, machine learning approaches aid in the early detection and diagnosis of cardiac disease. This research objective is to select the best performing model that fits the data set by analyzing the efficiency of several machine learning techniques.Machine learning algorithms play an important role in analyzing the data and making good decisions for heart disease prediction in real time.Nowadays, machine learning algorithms play a crucial role in health care. Here, we introduce some Machine learning models that can help predict heart disease based on its features and symptoms.For improving the heart disease patient health it is essential to predict it in the early stage[4].

Machine learning models will give us many advantages for predicting the heart disease.We use the predictive models in machine learning which take the humans past data who is suffering from the heart disease and predict the person who may have the chance of getting the heart disease based on his/her symptoms.ML models have the high capability of predicting the accurate data.The model can we evaluated its performance by finding its accuracy.This study utilizes supervised machine learning algorithms, including Support Vector Machine (SVM),Decision tree, Naive Bayes, Logistic Regression,Random forest and also K-Nearest Neighbors (KNN), to predict the heart disease. Among the tested models, Random Forest Classifier has the highest accuracy , achieving an accuracy rate of 95% . This model achieves highest performance and evaluated based on precision ,recall ,accuracy and area under the curve.Machine learning is highly involved in medical diagnoses and the health care industry. In the medical industry, machine learning is used for numerous things, such as drug development, diagnosis using medical imaging, epidemic prediction, and heart failure prediction. Large medical data sets can be used to identify patterns and do predictive analysis using machine learning techniques. When compared to traditional medical techniques, machine learning offers numerous benefits, including reduced expenses and time, which enhances diagnosis [10].

In this research we considered the data set which contain the features such as age,sex,chest pain type,resting BP,cholesterol,fasting blood sugar,resting ECG,max heart rate,exercise angina,oldpeak,ST slope [4]. We trained the different models using the 80% of the data in the data set.The models include Support Vector Machine (SVM),Decision tree classifier, Naive Bayes, Logistic Regression,Random forest and also K-Nearest Neighbors (KNN) [7].The accuracy of the models are Support vector machine - 58% ,Decision tree classifier - 89% ,Naive Bayes or Gaussian NB - 86% ,Logistic Regression - 83% ,Random forest classifier - 95% and K-Nearest Neighbors(KNN) - 72%.In order to improve the convergence of the models preprocess the data and scale it.The study's data set is accessible to the general public via Kaggle. Python was used for all of the computation, preprocessing, and visualization on Google Colab. Up to 95% of heart disease cases have been predicted accurately in the past with machine learning approaches . These studies, however, frequently used tiny sample sizes, so it's possible that the findings can't be applied to wider groups [9]. By employing a bigger and more varied data set, our study seeks to overcome this constraint, which should improve the results generalizability [3].

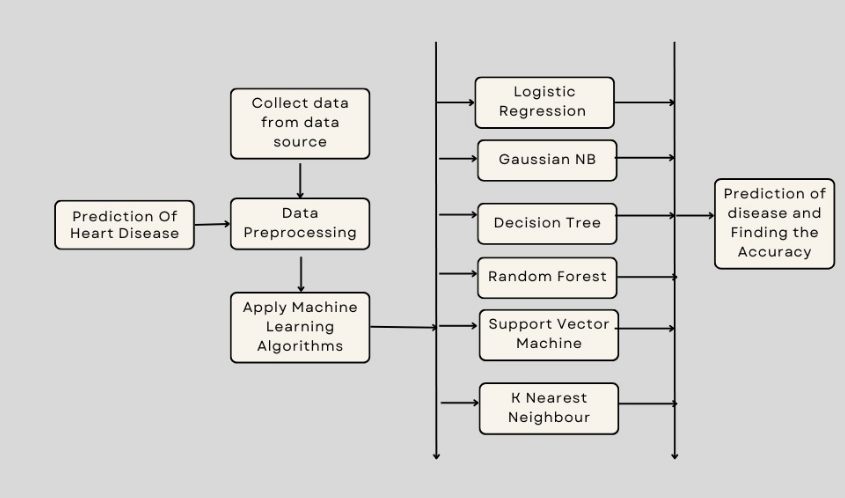
# Related Work

Python programming language can be used effectively to predict heart disease by analyzing data from various healthcare facilities and patients. This data-driven approach helps doctors provide better treatment options and improve the overall healthcare delivery system. The focus is on using Python to detect heart diseases. This application relies on a dataset containing patient information, such as age, sex, cholesterol levels, and other relevant factors. To build this application, several libraries like Matplotlib, NumPy, and Pandas are imported, enhancing Python’s capabilities for analyzing patient data. Importantly, the application complies with HIPAA regulations to ensure patient data security. It also discusses how machine learning techniques are used to detect heart disease, specifically focusing on various algorithms. The project primarily employs the Random Forest algorithm, achieving an accuracy of about 83% with the training data. This algorithm is chosen for its effectiveness in analyzing data and providing reliable predictions. Data analysis plays a key role in this process, especially with categorical variable. The application also considers medical factors such as age, gender, blood pressure and cholesterol to predict heart disease risk. Overall, using machine learning in Python is a branch of Artificial Intelligence, and libraries like SKLEARN are essential for making these predictions. The Random Forest algorithm is highlighted as one of the best methods for accurately predicting heart disease outcome.[11]

This presents a model for detecting heart disease using machine learning algorithms. Four algorithms were employed to train and analyze a dataset containing test results from various patients. The accuracy of these algorithms was evaluated, and a graph was plotted using Matplotlib to visualize the results. The findings showed that the Decision Tree, Random Forest, and Support Vector Machine algorithms achieved the highest accuracy, around 98.83%, while the K-Nearest Neighbors algorithm had an accuracy of approximately 97.4%.The Decision Tree model was successfully integrated into a web application using an API called Flask. It performed well when tested five times online without any errors. The research suggests that this model could be further developed into a real-time system using Deep Learning, allowing users to upload their test results as images for analysis.After comparing the results, the Decision Tree Classifier was selected for making predictions due to its high accuracy. This model was then saved and integrated into a web application using Flask.[12]

This study aimed to classify heart disease using different models, focusing on a real-world dataset of 70,000 patients from Kaggle. Researchers applied the k-modes clustering algorithm to predict heart disease, preprocessing the data by converting age into 5-year intervals and blood pressure into 10 intervals. They also split the dataset by gender to address differences in heart disease progression.The elbow curve method helped determine the best number of clusters for men and women. The multilayer perceptron (MLP) model achieved the highest accuracy at 87.23%, with other algorithms also performing well, all above 86%. However, the study had limitations: it relied on a single dataset, considered few demographic factors, and did not assess model performance on a test dataset. The ability to interpret the results and understand the clusters formed was also not evaluated. Future research should compare k-modes clustering with other algorithms like k-means and hierarchical clustering, explore the effects of missing data and outliers, and evaluate model performance on new data. This could improve understanding of the clusters and enhance decision-making in heart disease diagnosis and treatment[13].

This research focused on creating algorithms to detect and classify ventricular cardiac diseases. The first step involved a detection algorithm that identifies if a heart signal is diseased or normal, using normal signals from the MIT-BIH Normal Sinus Rhythm database. If a signal is identified as diseased, it is then analyzed by a classifier to determine the specific disease. The study concentrated on two diseases: ventricular tachycardia and premature ventricular contraction, using data from different MIT-BIH databases. The heart signals were divided into two sets: one for training the detection and classification algorithms and another for testing their performance.



A key finding from the evaluation was that removing noise and baseline shifts from the signals is unnecessary when transforming them into the spectral domain. This is because the added noise doesn't affect the important frequency values needed for analysis. Additionally, the researchers noted that the same results could be achieved even if the sampling rates of the signals were adjusted. They observed that normal heart signals have higher frequency components compared to diseased signals, indicating that these components are reduced in the presence of disease[14].

The heart is essential for human health, but heart disease is becoming increasingly common worldwide. To address this issue, we need a reliable model to predict early signs of heart disease. This study aims to create an accurate machine learning model to help diagnose heart conditions at a lower cost. The focus is on predicting heart disease using various machine learning algorithms and evaluating their performance through accuracy rates derived from a confusion matrix. After comparing five different algorithms, the Logistic Regression model was found to be the most accurate, achieving an impressive 95% accuracy rate. It also demonstrated strong performance in other metrics, with anF1-score, recall, and precision all at 95%. These results indicate that machine learning can effectively aid in predicting heart disease, suggesting that such algorithms could become valuable tools in healthcare. The approach used in this study could also be extended to diagnose other diseases, such as cardiovascular conditions, diabetes, breast cancer, and tumors. Future research may involve analyzing historical data and combining different machine learning techniques to enhance the accuracy and effectiveness of disease predictions[15]. It explores how machine learning techniques can classify heart diseases effectively. Classifiers play a vital role in healthcare, as their results help determine the best treatments for patients. The study examines and compares existing techniques to find the most efficient and accurate systems. Machine learning can significantly improve the accuracy of predicting cardiovascular risks, allowing for earlier identification of patients and enabling preventive treatments.

Fig. 1. Proposed heart prediction work flow diagram

Overall, there is great potential for machine learning algorithms in predicting heart-related diseases. While some algorithms perform very well in certain situations, they may not be as effective in others[16].

# PROPOSED WORK

Fig. 1 presents the proposed diagram for hear disease prediction. The prediction of heart disease is done by using various models and algorithms. The steps are detailed down in this section.

Step 1:Collection data from data source

Collecting data from various types of medical fields such as Hospitals and Other medical related fields, The data should be related to the persons who are suffering with the Heart disease. This is one of the major tasks in the process. The data which is collected should be in the form of either CSV files(Comma separated values) , JSON format or it can be XML file.

Step 2: Data Preprocessing

The Data should be Preprocessed before going to apply the machine learning algorithms. The data should be prepared in such a way that their should be no errors or missing values in the given input dataset. This data Preprocessing includes the Data cleaning which is used the remove the empty values or the null values in the given dataset. Handling error values includes, if given data contains any errors or unknown values which are not used for applying the Algorithms they will be removed. The data should be divided in to training and testing models before going to apply the algorithms to the models. If the data contains any numerical values , the values should be normalized.

Step 3:Applying Machine Learning Algorithms

Applying various machine learning algorithms once after the Data is preprocessed. The algorithms we have used in the detection of heart disease are :

1)Logistic Regression: Logistic regression is an analysis which give the Binary outputs such as “True” or “False” values. The outcome of this type of model will be based on minimum one independent variable.

2)Gaussian NB: This algorithm is used for classification of tasks which are continuous. And Mainly we use this Gaussian Nb algorithm for describing of data in which the output graph is in bell-shaped normal curve. For the prediction of Heart Disease we have used Gaussian Naïve Bayes algorithm for the Classification of the given Dataset.

3)Decision Tree Classifier : Decision Tree algorithm Divides the given Dataset into subsets and further classification or regression is performed. For the heart disease detection we used this algorithm in order to make the given data to be divided into sub sets of the data which helps and maintain the accuracy of the data.

4)Support Vector Machine Algorithm: This is a Supervised Algorithm Which is used to classify the data and used to find thehyperplane between the Support vectors Which gives us the accurate information about the model.

5)Random Forest Algorithm :It is an another method as same as Decision tree Algorithm. This algorithm consists of various Decision trees in which the result from each tree Is combined the get the accurate result. In Heart Disease Detection Model the Random Forest Algorithm helps to get the accurate results.

1. K Nearest neighbor Algorithm(KNN): K nearest neighbor algorithm is a supervised algorithm which helps the data to find out the nearest neighbor which depends on the value of k. The k value is depend on the model accuracy. Mainly these algorithms are used on the big datasets like one related to medical fields.

Step 4: After all these algorithms performed on these data set we check the results and accuracy of the given data and graphs

# Results and Discussions

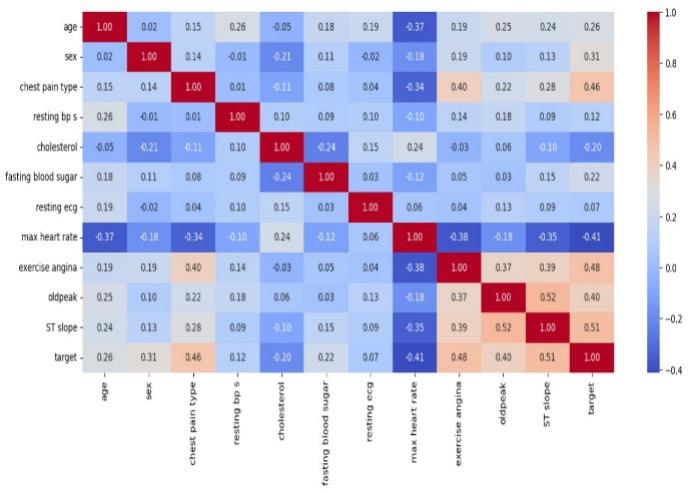
Fig. 2 presents the heatmap whcichy analyzes features related to the heart disease like exercise-induced angina, ST slope, and chest pain type have positive correlation. It is also observed that there is strong likelihood associativity with heart disease. On other hand, maximum heart rate has negative correlation, meaning they have lower chance of having heart disease. The evaluation of the algorithms is done using well known metrics namely precision, recall and F1-score (presented using Eqn. 1-3).

Fig. 2: Heatmap for heart disease diagnosis.

(1)

(2)

(3)

Table 1: Algorithm comparison as compared with different metrics (No Heart diesase)

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Precision** | **Recall** | **F1-Score** |
| LogisticRegression | 0.82 | 0.80 | 0.81 |
| GaussianNB | 0.85 | 0.83 | 0.84 |
| DecisionTree | 0.84 | 0.92 | 0.88 |
| RandomForest | 0.96 | 0.93 | 0.95 |
| SVM | 0.54 | 0.54 | 0.54 |
| KNN | 0.68 | 0.71 | 0.69 |

Table 2: Algorithm comparison as compared with different metrics (With Heart disease)

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Precision** | **Recall** | **F1-Score** |
| LogisticRegression | 0.84 | 0.85 | 0.85 |
| GaussianNB | 0.86 | 0.88 | 0.87 |
| DecisionTree | 0.93 | 0.86 | 0.89 |
| RandomForest | 0.95 | 0.97 | 0.96 |
| SVM | 0.62 | 0.62 | 0.62 |
| KNN | 0.75 | 0.73 | 0.74 |

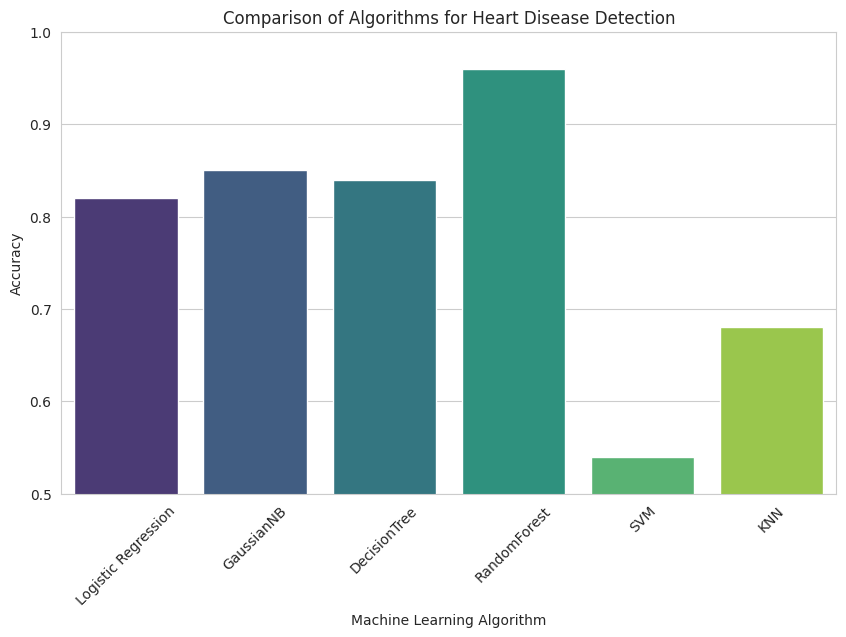


Fig.3: Graph about Training accuracy

Fig.3 presents the comparison of several algorithm. Also it is found the random forest outperformed better as compared to other machine learning algorithms. Fig.4 provides valuable understanding of most influential features in this model’s decision making process. ROC (Receiver Operating Characteristic) curve plots are used for evaluating the performance of various machine learning algorithms for certain tasks. Fig.5 shows that RandomForestClassifier and XGBClassifier are the top-performing models. K Neighbors Classifier performs poor compared to other models. Finally, As most have AUC values significantly above 0.9, All the models perform well in predicting heart disease.



Fig.4: Graph about training accuracy

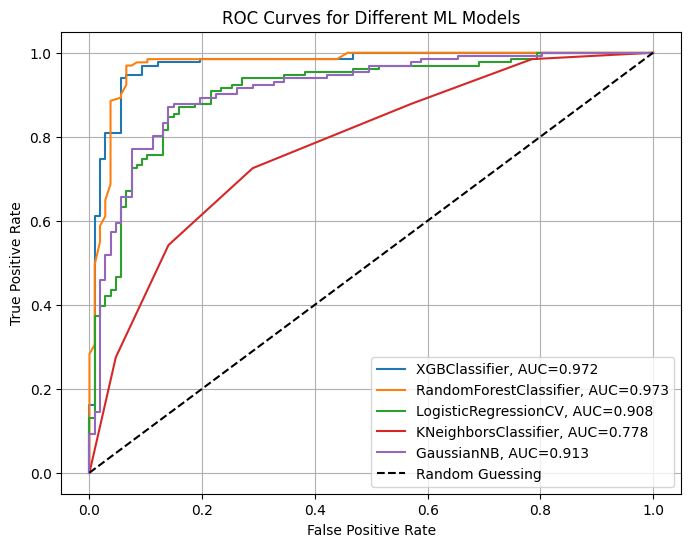


Fig.5: ROC curves for machine learning models.

# Conclusion

In the conclusion, with the help of Machine learning algorithms we can detect the disease which is related to heart. Various Machine learning algorithms give various outputs in the form of graphs, In which we can detect disease .In future these machine learning algorithms helps to improve the medical diagnostics which improves the accuracy in the results. Some supervised machine algorithms like Decision tree , support vector machines will give accurate results and classify the patients who have high risk of heart disease. Based upon on previous papers , this paper helps to improve the models accuracy , precision. This paper aims to give advance features in detecting the heart disease by using Machine learning algorithms. At the end , implementing the machine leaning algorithms no only in detecting the heart diseases it also reduce the humans failures or errors. The experimental findings suggest that the proposed learning system can assist clinicians in better detecting heart failure.

##### References

[1] B. U. Rindhe, et al., "Heart disease prediction using machine learning,"Heart Disease, vol. 5, no. 1, 2021..

[2] M. A. Naser, A. A. Majeed, M. Alsabah, T. R. Al-Shaikhli, and K. M. Kaky, "A review of machine learning’s role in cardiovascular disease prediction: Recent advances and future challenges," Algorithms, vol. 17, no. 2, p. 78, 2024.

[3] N. Kahouadji, "On the generalizability of machine learning classification algorithms and their application to the Framingham Heart Study," Information, vol. 15, no. 5, p. 252, 2024.

[4] Md. J. Rahaman, "Evaluate the predictive performance of supervised machine learning algorithms in diabetes dataset," International Journal of Science and Research (IJSR), vol. 13, pp. 538-543, 2024. DOI: 10.21275/SR24708194212.

[5] M. Ahmed and I. Husien, "Heart disease prediction using hybrid machine learning: A brief review," Journal of Robotics and Control (JRC), vol. 5, no. 3, pp. 884-892, 2024.

[6] A. M. Qadri, A. Raza, K. Munir, and M. S. Almutairi, "Effective feature engineering technique for heart disease prediction with machine learning," IEEE Access, vol. 11, pp. 56214-56224, 2023.

[7] J. Miah, D. M. Ca, M. A. Sayed, E. R. Lipu, F. Mahmud, and S. Y. Arafat, "Improving cardiovascular disease prediction through comparative analysis of machine learning models: A case study on myocardial infarction," in 2023 15th International Conference on Innovations in Information Technology (IIT), 2023, pp. 49-54.

[8] R. Ahmed, M. Bibi, and S. Syed, "Improving heart disease prediction accuracy using a hybrid machine learning approach: A comparative study of SVM and KNN algorithms," International Journal of Computations, Information and Manufacturing (IJCIM), vol. 3, no. 1, pp. 49-54, 2023.

[9] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective heart disease prediction using machine learning techniques," Algorithms, vol. 16, no. 2, p. 88, 2023.

[10] N. Chandrasekhar and S. Peddakrishna, "Enhancing heart disease prediction accuracy through machine learning techniques and optimization," Processes, vol. 11, no. 4, p. 1210, 2023.

[11] V. Chang, et al., "An artificial intelligence model for heart disease detection using machine learning algorithms,"Healthcare Analytics, vol. 2, 2022, Art. no. 100016.

[12] O. E. Taylor, P. S. Ezekiel, and F. B. Deedam-Okuchaba, "A model to detect heart disease using machine learning algorithm,"International Journal of Computer Sciences and Engineering, vol. 7, no. 11, pp. 1-5, 2019.

[13] C. M. Bhatt, et al., "Effective heart disease prediction using machine learning techniques,"Algorithms, vol. 16, no. 2, p. 88, 2023.

[14] K. M. Aamir, et al., "Automatic heart disease detection by classification of ventricular arrhythmias on ECG using machine learning,"Computers, Materials & Continua, vol. 71, no. 1, 2022.

[15] M. K. Hossen, "Heart disease prediction using machine learning techniques,"American Journal of Computer Science and Technology, vol. 5, no. 3, pp. 146-154, 2022.