The Role of Machine Learning in Heart Disease Prediction

Shruti Jha
IT-Student
Indira Gandhi Delhi Technical University For Women

Tisha
CSE-Student
Indira Gandhi Delhi Technical University For Women

Sonam
CSE-Student
Indira Gandhi Delhi Technical University For Women

Abstract

Heart disease is a leading cause of global morbidity and mortality, underscoring the urgent need for effective prevention and early intervention strategies. This research explores the application and effectiveness of various machine learning algorithms in the context of heart disease prediction.

We referred a large dataset with clinical and diagnostic information from a variety of patient populations. A variety of machine learning techniques, including gradient booster, bragging and boosting, decision trees, random forests, neural networks, KNN models, and KNN models were incorporated. The models were evaluated using performance criteria like accuracy and precision.

I. INTRODUCTION

Heart disease is one of the prevalent disease that can lead to reduce the lifespan of human beings nowadays. Each year 17.5 million people are dying due to heart disease [1].Cardiovascular diseases remain the most important global health problem and account for a significant proportion of morbidity and mortality worldwide. Timely identification of people at risk of heart disease is paramount for effective prevention and treatment. Machine learning, which can distinguish complex patterns from complex data sets, has become a powerful tool for cardiac risk assessment and prediction. Machine learning (ML) has emerged as a promising tool for enhancing heart disease prediction, offering the potential to revolutionize how healthcare providers identify and manage this debilitating condition.

Early detection of heart failure risk allows for timely intervention and preventive measures. Heart failure can be anticipated in advance, allowing for more efficient care and treatment. High-risk patients can receive specialized care plans made just for them, improving health outcomes, lowering hospital stays, and improving quality of life.

Common attributes used for heart disease are Age,Sex, Fasting Blood Pressure, Chest Pain type, Resting ECG(test that measures the electrical activity of the heart),Number of major vessels colored by fluoroscopy, Threst Blood Pressure (high blood pressure), Serum Cholesterol (determine the risk for developing heart disease), Thalach (maximum heart rate achieved), ST depression (finding on an electrocardiogram, trace in the ST segment is abnormally low below the baseline), painloc (chest pain location (sub sternal=1, otherwise=0)), Fasting blood sugar, Exang.(exercise included angina), smoke, Hypertension, Food habits, weight, height and obesity[2]

II. LITERATURE REVIEW

There are numerous works has been done related to disease prediction systems using different techniques and machine learning algorithms in medical centres
.
K. Polaraju et al, [3] proposed Prediction of Heart Disease using Multiple Regression Model and it proves that Multiple Linear Regression is appropriate for predicting heart disease chance. The work is performed using training data set consists of 3000 instances with 13 different attributes which has mentioned earlier. The data set is divided into two parts that is 70% of the data are used for training and 30% used for testing. Based on the results, it is clear that the classification accuracy of Regression algorithm is better compared to other algorithms

Ashok Kumar Dwivedi et al, [4] recommended different algorithms like Naive Bayes, Classification Tree, KNN,Logistic Regression, SVM and ANN. The Logistic Regression gives better accuracy compared to other algorithms.

Sairabi H.Mujawar et al, [5] used k-means and naïve bayes to predict heart disease. This paper is to build the system using historical heart database that gives diagnosis. 13 attributes have considered for building the system. To extract knowledge from database, data mining techniques such as clustering, classification methods can be used. 13 attributes total of 300 records were used from the Cleveland Heart Database. This model is to predict whether the patient have heart disease or not based on the values of 13 attributes.

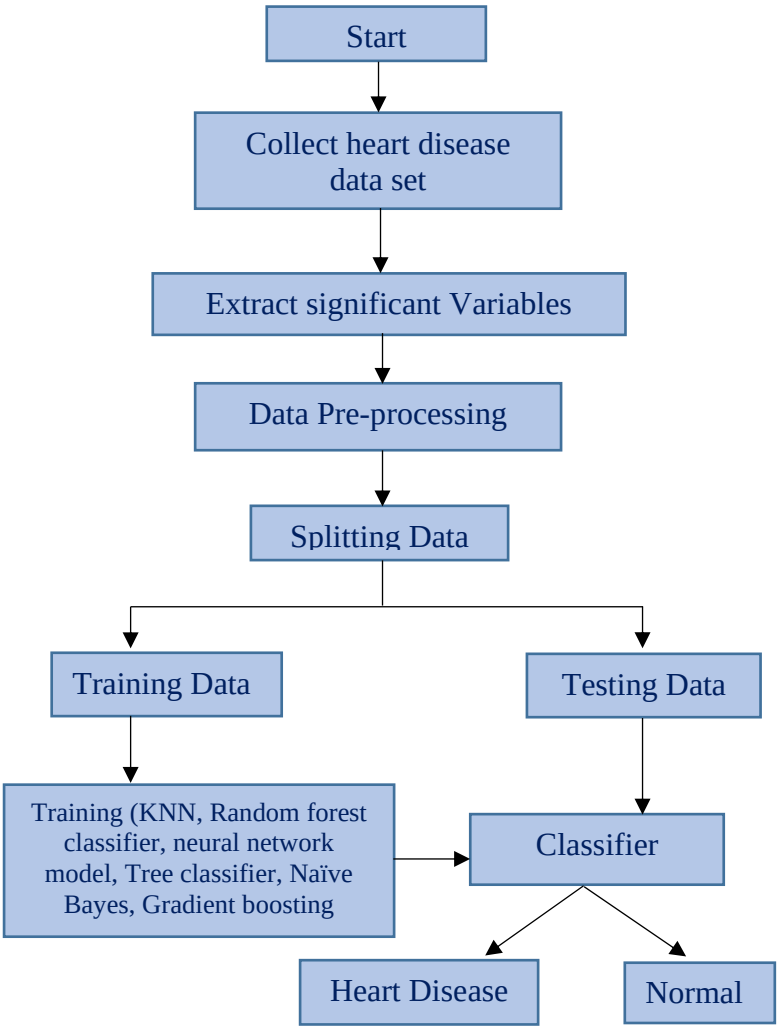
These studies collectively explore different approaches for predicting heart disease. They focus on the application of a Multiple Regression Model and highlights the suitability of Multiple Linear Regression for predicting heart disease with superior classification accuracy. Also previous research employ k-means clustering and Naive Bayes classification techniques on a historical heart database with 13 attributes, aiming to predict the presence of heart disease in patients, although specific accuracy metrics are not provided in the available information. Together, these studies underscore the diversity of approaches in the field of heart disease prediction, offering valuable insights for future research endeavors.

III. METHODOLOGY

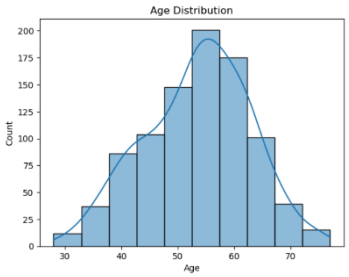
In this paper, we have referenced from dataset which was curated from 5 heart datasets combined over 11 common features that makes it largest available dataset on heart attack prediction so far for research purposes containing 918 observations. Then we pre-processed the dataset which deals with missing values, cleaning and normalization .In next step, exploratory data analysis provided with significant observations. We used the results obtained to then extract significant features, split the dataset into train and test in 80:20.

This paper contains the analysis of different machine learning model which are Neural networks, Ensemble models like bagging -Random Forest classifier and boosting-Gradient Booster, Decision- tree Classifier, KNN,K-Means and Naïve-Bayes which will help to determine the heart diseases in early stages and prevent further consequences ,It will ultimately help in the field of medical studies.

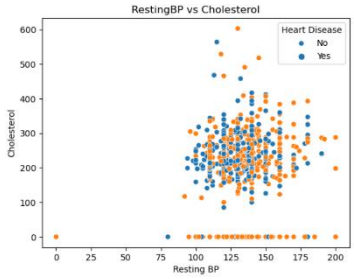
After application of different model ,when we evaluated our proposed model on the basis of evaluation metrics. Hence, this heart disease prediction system can give better results in the future work if future scope is included during research



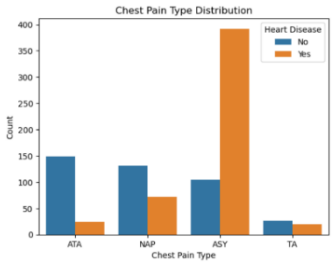
IV. RESULT



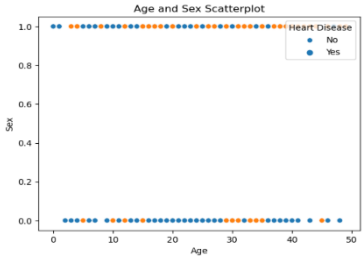
A. This graph shows the Age Distribution between different groups.



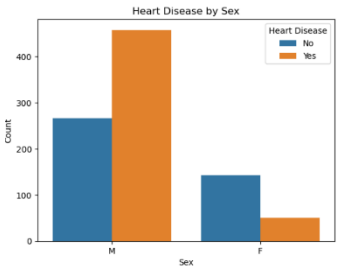
B. This graph shows RestingBP vs. Cholesterol



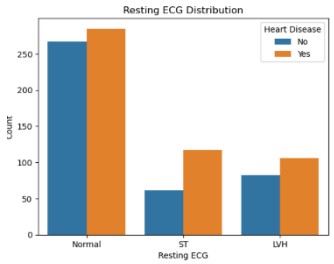
C. This graph shows the Chest Pain Distribution.



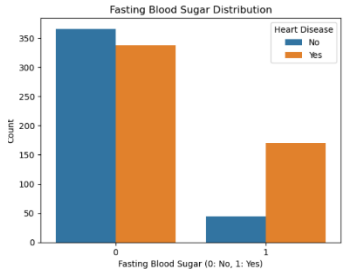
D. This graph shows Age & Sex scatterpiot .



E. This graph shows Heart Disease by Sex.



F. This graph shows Resting ECG Distribution.



G. This graph shows Fasting Blood Sugar Distribution.

V. DISCUSSION

From the analysis of the dataset we concluded that heart disease rate. The chances of having heart failure due to heart diseases are greater in men than in women. On an average the risks of having severe effects due to heart diseases has sudden increases males of age 40-50. From previous record of heart patients, it is seen that patients having heart pain type ASY are more likely to have heart diseases and chances of heart failure in future. While the ATA pain type seems to have less chances of having a heart failure risk in future.

In this research paper, the accuracy with which a model can predict a heart disease can help determine the early stages of a disease and start the treatment in early stages only which will prevent the chances of heart failure and can save many lives.

Our limitations for this research was small amount of reliable dataset which can be found for this research work in which the records of patients were very less ,which makes it difficult to interpret it for the proper results and analysis .In future research on huge amount of dataset and applying the techniques used in following research paper can help found a more reliable method for the prediction of the chances of heart failure due to heart disease .This method can be used for predicting

the chances of kidney and lungs failure also in future given that a reliable dataset is available .

VI. CONCLUSION

A cardiovascular disease detection model has been developed using six ML classification modelling techniques. This project predicts people with cardiovascular disease by extracting the patient medical history that leads to a fatal heart disease from a dataset that includes patients' medical history such as chest pain, sugar level, blood pressure, etc. The algorithms used in building the given model are Random Forest Classifier and KNN etc. The accuracy of our model is 99%. By using these, computer aided techniques we can predict the patient fast and better and the cost can be reduced very much. There are a number of medical databases that we can work on as these Machine learning techniques are better and they can predict better than a human being which helps the patient as well as the doctors.

Maximum accuracy of the neural network model has reached up to 83%. The KNN model has given a maximum accuracy of 73%. The decision Tree classifier model has given an accuracy of 100% but we can't consider it because dataset is very small. The Naïve Bayes model gave accuracy of maximum 63%. Random forest classifier gave maximum accuracy of 88%. Last but not the least gradient boosting gave maximum score of 99% which is highest accuracy of our model.

VII. REFERNCES

[1] Animesh Hazra, Arkomita Mukherjee, Amit Gupta,Asmita Mukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review", Research Gate Publications, July 2017, pp.2137-2159.

[2] T.Mythili, Dev Mukherji, Nikita Padaila and Abhiram Naidu, "A Heart Disease Prediction Model using SVM-Decision Trees- Logistic Regression (SDL)", International Journal of Computer Applications, vol. 68,16 April 2013.

[3] K. Polaraju, D. Durga Prasad, "Prediction of Heart Disease using Multiple Linear Regression Model",International Journal of Engineering Development andResearch Development, ISSN:2321-9939, 2017.

[4] Ashok kumar Dwivedi, "Evaluate the performance of different machine learning techniques for prediction of heart disease using ten-fold cross-validation", Springer,17 September 2016.

[5] Sairabi H.Mujawar, P.R.Devale, "Prediction of Heart Disease using Modified K-means and by using Naïve Bayes", International Journal of Innovative research in Computer and Communication Engineering, vol.3,October 2015, pp.10265-10273.