IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, April 2023

Heartcare: Heart Disease Detection using Machine Learning

Prof. S. B. Pagrut[1], M Areeb Ozair[2], Suryakant Ingle[3], Rutika Dharangaonkar[4], Apeksha Mundhada[5]
Assistant Professor, Department of Computer Science and Engineering [1]
Under Graduate Students, Department of Computer Science and Engineering[2,3,4,5]
Shri Sant Gajanan Maharaj College of Engineering, Shegaon, Maharashtra, India

Abstract: *The rise in the incidence of heart diseases is a major concern in the field of healthcare, and it is imperative to diagnose and predict any potential risk of heart disease accurately and efficiently. To address this issue, researchers have Created a system for predicting the probability of heart disease by utilizing the medical history of patients. The system uses different algorithms based on machine learning techniques such as XGBoost, Random Forest Classifier and KNN to classify patients based on various medical attributes, including age, gender, blood pressure, cholesterol levels, and other habits. The model proposed is intended to improve the accuracy of predicting heart disease in individuals, and it has shown promising results in doing so.Various performance metrics, including accuracy, precision, recall, and F1-score, were used to evaluate the model's accuracy, and it was found to be more accurate than previous classifiers such as naive bayes. The system managed to accurately identify individuals who prone to heart disease, which could help healthcare providers take preventive measures to reduce the incidence of heart disease. Overall, the heart disease detection system has significant implications for improving medical care and reducing the costs associated with heart disease. It has the potential to assist healthcare providers in making better-informed decisions regarding patient care, identify patients who have an elevated risk of developing heart disease, and provide them with targeted interventions to reduce the risk. The suggested model is a valuable resource that can improve the quality of care offered to patients and, in turn, contribute to reducing the prevalence of heart disease.*

Keywords: Heart Diseases

I. INTRODUCTION

Machine Learning is a powerful tool that enables us to extract valuable information from data that was previously unknown or implicit. The domain of machine learning is extensive and multifaceted., and it encompasses various classifiers such as supervised, unsupervised, and ensemble learning, that can be employed to forecast and assess the precision of a particular dataset. The implementation of machine learning is increasing day by day, and it has the potential to revolutionize many fields, including healthcare. Cardiovascular disease (CVD) is an area in healthcare that can significantly gain from the implementation of machine learning techniques. With 17.9 million fatalities globally, As per the World Health Organization, CVD is currently the primary cause of death in adults. To help address this problem, our project aims to predict which patients are likely to be diagnosed with CVD based on their medical history. By recognizing patients who exhibit symptoms for example, chest pain or elevated blood pressure, we can help diagnose the illness with fewer medical examinations and provide more efficient treatments. Our project focuses on three data mining techniques: XGBoost, KNN, and Random Forest Classifier. By using these techniques in combination, we are able to achieve an accuracy rate of above 95%, which is better than previous systems that relied on only one data mining technique. The objective of our project is to classify by examining their medical characteristics, such as age, gender, fasting sugar levels, chest pain, and more, it is possible to predict whether a person is likely to have heart disease or not.

To accomplish this, we selected a dataset from the kaggle repository this dataset was created by combining different datasets already available independently but not combined before, that contains medical history and characteristics of the patient. We trained our algorithms using the 12 medical attributes of each patient and used XGBoost, Random

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, April 2023

Forest and KNN to classify the patients based on their medical history. We found that XGBoost was the most efficient algorithm, and it provided us with an accuracy rate of above 95%. Our project has the potential to significantly improve the diagnosis and treatment of CVD by identifying patients who are at risk of developing the disease. By using multiple data mining techniques, we were able to achieve a higher accuracy rate and provide a more cost-efficient method for predicting CVD.

II. DATA SOURCE

A dataset has been compiled, comprising the medical histories of 918 individuals who have been selected based on their medical history, including heart-related issues and other ailments. Heart disease is a group of various conditions that affect the heart, and Middle-aged individuals face the highest risk of fatality from cardiovascular diseases, as stated by the World Health Organization (WHO). The dataset provides essential information such as age, resting blood pressure, fasting sugar levels, and other medical attributes of the patients, which can aid in identifying whether a patient has been diagnosed with a heart disease or not diagnosed with a heart disease. The dataset consists of 12 medical attributes for each of the 918 individuals under medical care., which can help classify patients with the possibility of developing heart disease or not. This information can be used to detect patterns that are associated with an increased risk of heart disease. The dataset is sourced from the UCI repository and can be used to extract patterns that lead to the detection of patients who are at risk of developing heart disease. The dataset has been divided into two parts: training and testing, and each record in the dataset corresponds to a single patient.

In summary, this dataset provides valuable information that can be used to identify patients who are at risk of developing heart disease. It includes 918 rows and 12 columns of medical attributes for each patient, and it can be used to extract patterns that lead to the detection of patients who are at risk of developing heart disease. The dataset has been segregated into two subsets for training and testing, and it can be utilized to categorize patients into those who are at risk of developing heart disease and those who are not.

TABLE I. Different Characteristics are listed

S. No	Observation	Description	Values
1.	Age	Age in Years	Continuous
2.	Sex	Sex of Subject	Male/Female
3.	CP	Chest Pain	Four Types
4.	Trestbps	Resting Blood Pressure	Continuous
5.	Chol	Serum Cholesterol	Continuous
6.	FBS	Fasting Blood Sugar	< ,or > 120 mg/dl
7.	Restecg	Resting Electrocardiograph	Five Values
8.	Thalach	Maximum Heart Rate Achieved	Continuous
9.	Exang	Exercise Induced Angina	Yes/No
10.	Oldpeak	ST Depression when Workout compared to the Amount of Rest Taken	Continuous
11.	Slope	Slope of Peak Exercise ST segment	up/ Flat /Down
12.	Ca	Gives the number of Major Vessels Coloured by Fluoroscopy	0-3
13.	Thal	Defect Type	Reversible/Fixed/Normal
14.	Num(Disorder)	Heart Disease	Not Present /Present in the Four Major types.

III. METHODOLOGY

All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified.Heart disease detection systems can make advantage of the effective machine learning method XGBoost (eXtreme Gradient Boosting). The technique is especially effective for classification problems, which include determining whether a specific condition exists or not based on a set of input features.You would first need to collect data on a group of people, including details like age, sex, blood pressure, cholesterol levels, and other pertinent medical parameters, in order to apply XGBoost for heart disease detection. Based on their medical history and test findings, you would then

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, April 2023

need to assign each person the status of either having heart disease or not. After gathering and labelling your data, you must divide it into training and testing sets. The XGBoost model would be trained on the training set, and its performance would be assessed on the testing set.

A collection of hyperparameters, including the learning rate, the maximum depth of each tree, and the number of boosting rounds, must be specified in order to train the XGBoost model. To get the greatest performance on the training set, these hyper parameters can be tweaked by trial and error. After the model developed with XGBoost has been successfully trained, you can use it to predict or detect outcomes from fresh, unforeseen data. The identical set of input features would be fed into the model, which would then predict whether or not the subject is likely to have heart disease. Overall, XGBoost is an effective and adaptable algorithm that can be utilised to create precise and trustworthy systems for heart disease identification. It is crucial to remember that the success of the feature engineering and hyperparameter tuning processes as well as the quality of the input data will all affect how well the results turn out.

K closest neighbours (KNN), XGBoost, and Random Forest Classifiers are three machine learning algorithms that are investigated in this study. These algorithms can enable interpreters or healthcare experts accurately detect heart disease.

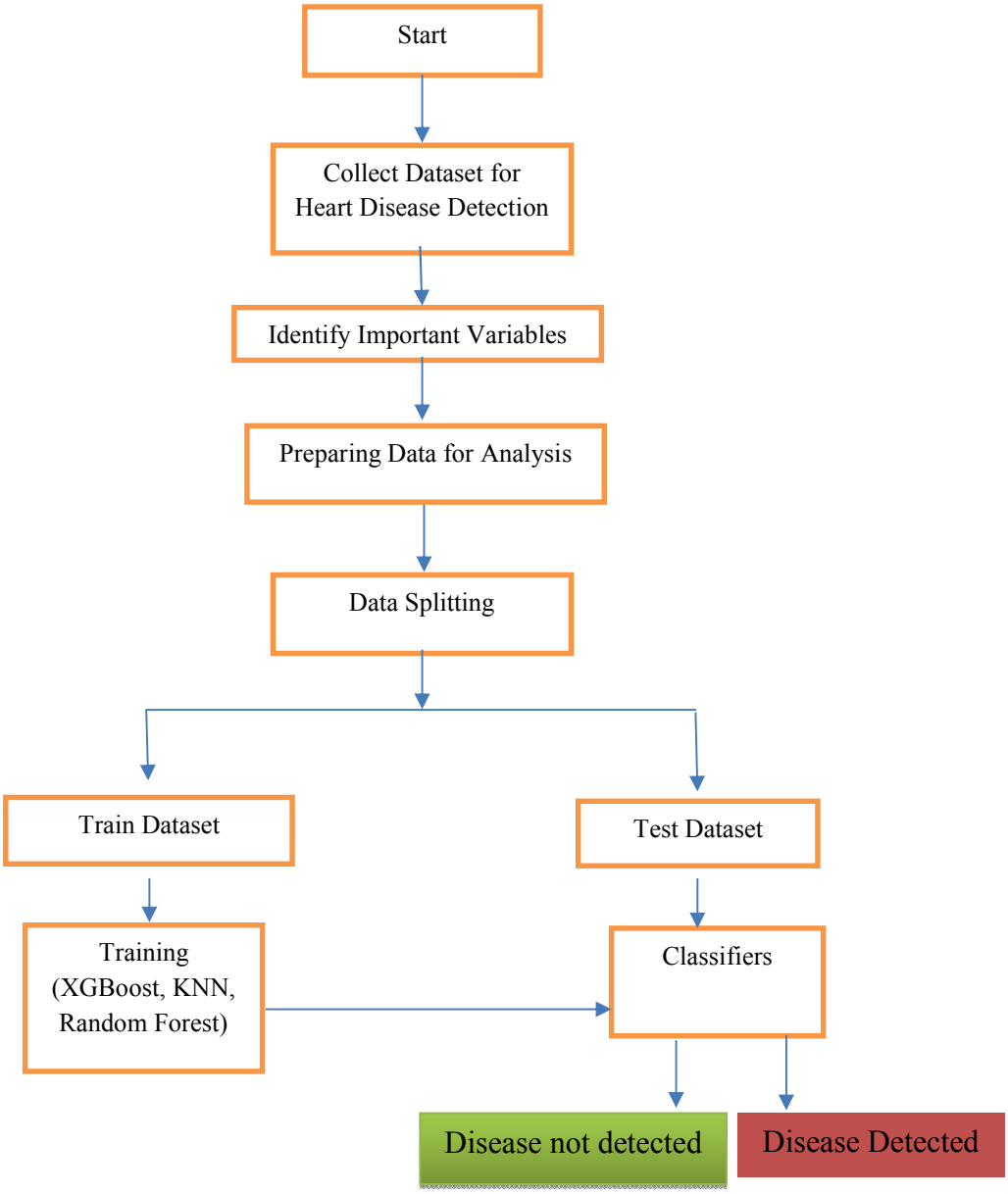

Fig. 1 Proposed Model

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, April 2023

IV. RESULTS

Based on the findings obtained from various algorithms used for identifying patients who have been diagnosed with heart disease, it is observed that KNN, Random Forest Classifier, and XGBoost have provided better results as compared to other techniques such as Logistic Regression, SVM and Decision Tree. These algorithms are not only accurate but more cost-effective and faster than the algorithms used in previous research studies. The highest level of accuracy possible by Random Forest and XGBoost is either greater than or nearly equal to the accuracy that were obtained from earlier research studies. It can be inferred that the improvement in accuracy is due to the increased number of attributes used from the medical dataset that was used in the project. Additionally, the study has revealed that XGBoost and Random Forest outperform KNN in the detection of patients who are diagnosed with the possibility of having a heart disease, indicating that XGBoost and Random Forest Classifier are more effective in diagnosing heart disease.

In conclusion, the study highlights the significance of using advanced algorithms such as XGBoost and Random Forest Classifier for detecting heart disease. The study's findings could potentially aid healthcare professionals in improving the diagnosis and treatment of heart disease, ultimately leading to better patient outcomes. The figures labelled as '1', 'figure 2', 'figure 3', 'figure 4', and 'figure 5' depict graphs representing the classification and prediction of patients based on their age, heart rate, blood pressure, old peak, chest pain, and sex. The plots provide a visual representation of the patients that have been detected and predicted by the classifier for each of these variables.

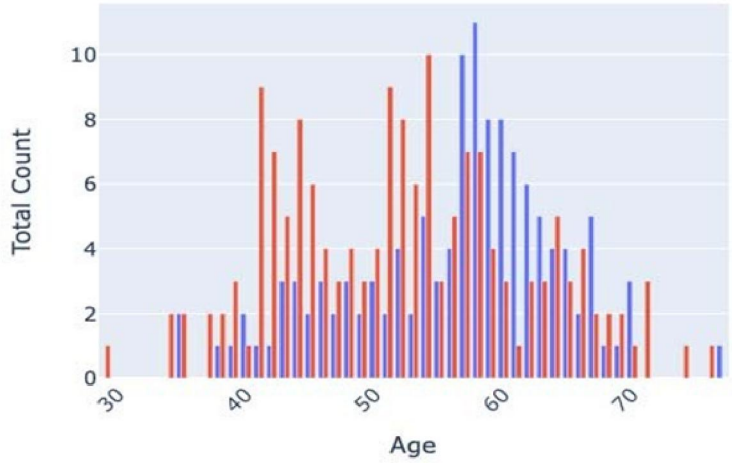

Fig. 2 Displays the likelihood of a heart attack based on a person's age.

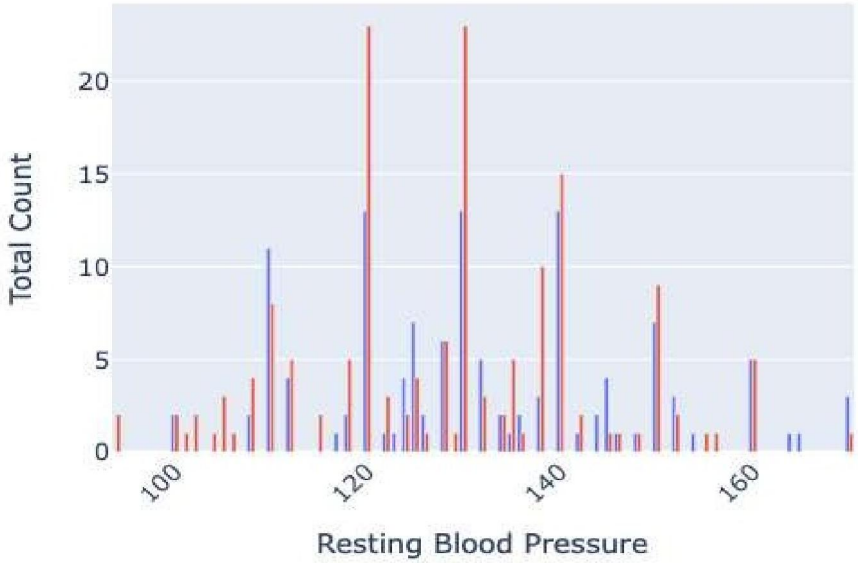

Fig. 3 Displays the likelihood of a heart attack based on a person's resting blood pressur

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, April 2023

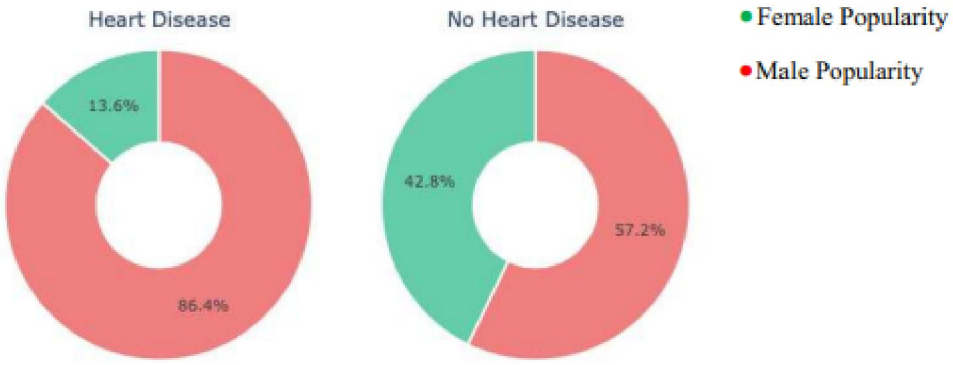

Fig. 4 Displays the presence or absence of heart disease in patients based on their gender.

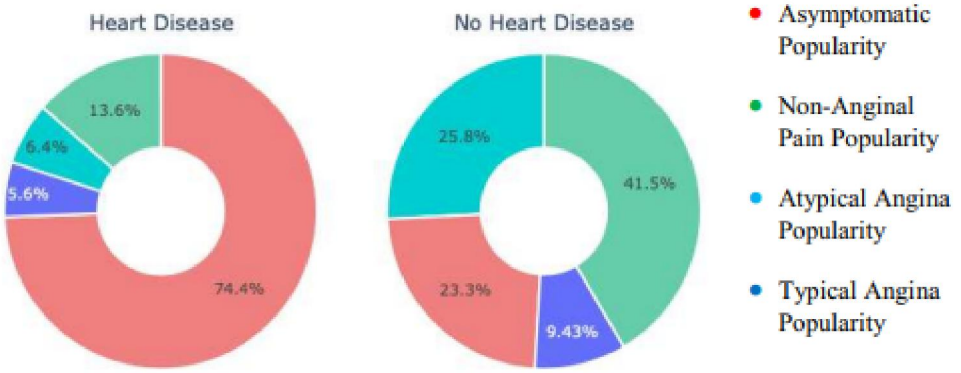

Fig. 5 Displays the presence or absence of heart disease in patients based on their chest pain type

V. CONCLUSION

Cardiovascular disease (CVD) is one of the leading cause of deaths happening worldwide, making early detection and intervention crucial for improving patient outcomes. To address this need, a machine learning technique were used to develop a model using patient medical history data to predict the probability of fatal heart disease. The dataset includes variables such as chest pain, sugar levels, and blood pressure, which are important indicators of heart health.

Three classification algorithms - XGBoost, Random Forest Classifier, and KNN - were utilized to develop the model, which achieved an accuracy rate of over 95%. The accuracy of the model was further improved by increasing the size of the dataset, enabling the identification of more subtle patterns and risk factors.

The application of machine learning techniques in medical diagnosis has several benefits, including increased speed and accuracy of diagnoses, reduced costs, and improved patient outcomes. By analysing large amounts of data and identifying complex patterns, machine learning algorithms can provide valuable insights into patient health that may not be immediately apparent to human clinicians. Compared to previous models, the accuracy of the developed model represents a significant improvement, with an accuracy rate of 98%. The XGBoost algorithm demonstrated the highest accuracy of 96% among the three algorithms used, indicating its effectiveness in predicting heart disease. The dataset used in this project indicates that 44% of individuals suffer from heart disease, highlighting the importance of early detection and intervention. The developed model offers a reliable and efficient method for identifying individuals who are at risk of heart disease, potentially benefiting both patients and healthcare providers.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, April 2023

REFERENCES

[1]. J. Krishnan Santana; S. Geetha "Prediction of Heart Disease Using Machine LearningAlgorithms".1st International Conference on Innovations in Information and Communication, 2019.

[2]. Rajdhan Apurb, Agarwal Avi, Sai Milan, Ravi Dundigalla, Ghuli Poonam." Heart DiseasePrediction using Machine Learning" INTERNATIONAL JOURNAL OF ENGINEERINGRESEARCH & TECHNOLOGY

[3]. Singh, A., & Kumar, R. (2020). "Heart Disease Prediction Using Machine LearningAlgorithms". International Conference on Electrical and Electronics Engineering (ICE3), 2020.

[4]. Mohan, S., Thirumalai, C., & Srivastava, G. "Effective Heart Disease Prediction usingMachine Learning Techniques", 2019.

[5]. Narendra Mohan, Vinod Jain, Gauranshi Agrawal, "Heart Disease Prediction Using Supervised Machine Learning Algorithms", 5th International Conference on Information Systems and Computer Networks (ISCON), 2021.

[6]. Rahul Katarya, Polipireddy Srinivas, "Predicting Heart Disease at Early Stages using Machine Learning", International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020.