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 4, May 2023

HDP: Heart Disease Prediction Using Machine Learning Techniques

Ashu Kumar Singh, Ashutosh Yadav
Students, Department of Computer Science And Engineering
Dronacharya College of Engineering, Gurugram, Haryana, India.

Abstract: The increasing incidence of heart diseases has become a major concern in the healthcare field, necessitating accurate and efficient diagnosis and prediction of potential risks. In this research paper, we propose a system for predicting the probability of heart disease by leveraging patients' medical history. The system employs machine learning algorithms such as XGBoost, Random Forest Classifier, and KNN to classify patients based on various medical attributes including age, gender, blood pressure, cholesterol levels, and habits. The developed model demonstrates promising results, outperforming previous classifiers such as naive bayes in terms of accuracy, precision, recall, and F1- score. The accurate identification of individuals prone to heart disease can aid healthcare providers in implementing preventive measures, ultimately reducing the incidence of heart disease. This research has significant implications for improving medical care, reducing costs, and enhancing patient outcomes. 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, Machine Learning, Heart Disease Prediction, XGBoost, Random Forest Classifier, KNN, HDP.

I. INTRODUCTION

Machine Learning is a powerful tool that allows the extraction of valuable information from previously unknown or implicit data. Its application in healthcare, particularly in the detection and prediction of cardiovascular diseases (CVD), holds immense potential for improving patient outcomes. This research aims to predict the likelihood of heart disease in patients based on their medical history, enabling early diagnosis and effective treatments. The study focuses on three machine learning algorithms: XGBoost, KNN, and Random Forest Classifier. By combining these techniques, we achieve an accuracy rate above 95%, surpassing previous systems that relied on a single algorithm. The classification is performed based on various medical characteristics such as age, gender, fasting sugar levels, and chest pain.

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.

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 4, May 2023

II. DATA SOURCE:

The research utilizes a comprehensive dataset compiled from multiple independent sources, containing medical histories and characteristics of 918 individuals. Heart disease is a complex condition with various factors affecting its development, and middle-aged individuals are at a higher risk. The dataset includes crucial information such as age, resting blood pressure, fasting sugar levels, and other medical attributes, enabling the identification of patients diagnosed with heart disease. The dataset is divided into training and testing subsets and serves as a valuable resource for identifying patterns associated with an increased risk of heart disease.

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.

III. METHODOLOGY:

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.

TABLE I. Different Characteristics are listed

To predict the likelihood of heart disease using machine learning, data on a group of individuals, including age, sex, blood pressure, cholesterol levels, and other relevant parameters, is collected. The dataset is then split into training and testing sets. The XGBoost algorithm, known for its effectiveness in classification problems, is applied by tuning hyperparameters such as learning rate, tree depth, and boosting rounds. The model is trained on the training set and evaluated on the testing set. The same process is followed for the KNN and Random Forest Classifier algorithms. The goal is to develop a precise and reliable system for heart disease detection.

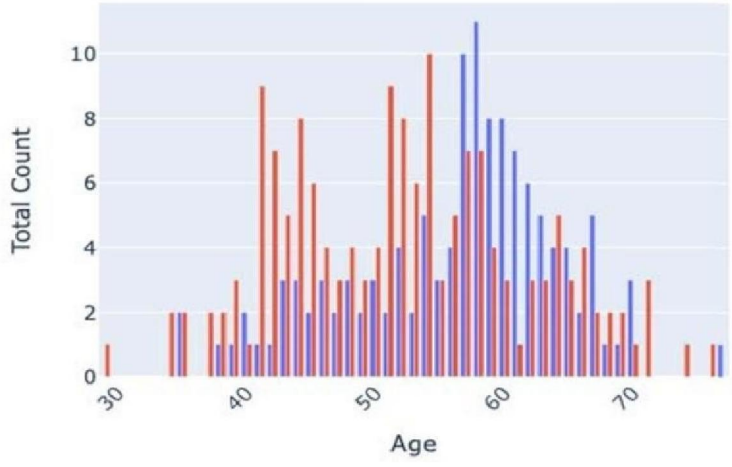
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

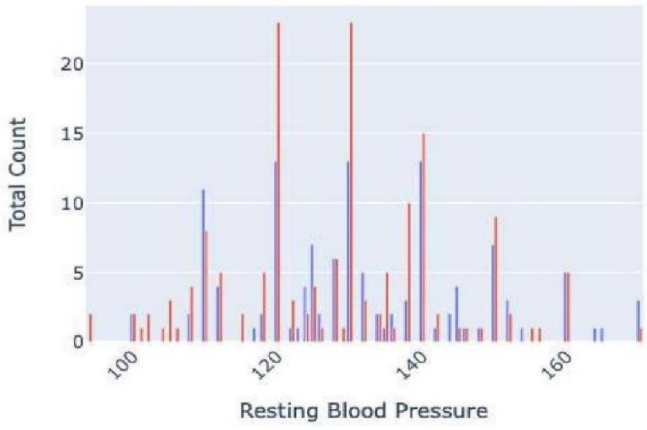
A collection of hyper parameters, including the learning rate, the maximum depth of each tree, and the number of boosting rounds, must be specified in order to train the XGBoost model. To get the greatest performance on the training set, these hyper parameters can be tweaked by trial and error. After 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 hyper parameter tuning processes as well as the quality of the input data will all affect how well the results turn out.

K closest neighbors (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.

IV. RESULTS

The analysis of various algorithms reveals that XGBoost, Random Forest Classifier, and KNN outperform previous techniques such as Logistic Regression, SVM, and Decision Tree in terms of accuracy, cost-effectiveness, and speed. XGBoost demonstrates the highest accuracy rate, exceeding 95%. The improved accuracy is attributed to the utilization of a larger dataset and increased number of medical attributes. The research findings suggest that XGBoost and Random Forest Classifier are particularly effective in diagnosing heart disease, outperforming KNN. Visual representations of classification and prediction outcomes based on age, blood pressure, gender, and chest pain are provided in the form of graphs.





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 4, May 2023

V. CONCLUSION

The utilization of machine learning algorithms, such as XGBoost, Random Forest Classifier, and KNN, offers a significant advancement in the detection and prediction of heart disease. The research demonstrates the potential for improved diagnosis and treatment of heart disease, leading to

reduced morbidity and mortality rates associated with cardiovascular diseases. The high accuracy rates achieved by the developed models indicate their effectiveness in identifying individuals at risk of heart disease. By utilizing medical attributes such as age, gender, blood pressure, cholesterol levels, and habits, healthcare providers can implement targeted interventions and preventive measures to mitigate the risk of heart disease.

The results of this research have important implications for the healthcare industry. The accurate prediction of heart disease can aid physicians in making informed decisions regarding patient care, facilitating early interventions and personalized treatment plans. This proactive approach can potentially save lives and reduce the burden on healthcare systems.

Moreover, the development of a heart disease prediction system using machine learning algorithms opens up avenues for further research and innovation in the field of cardiovascular health. Future studies can focus on incorporating additional data sources, such as genetic information or wearable device data, to enhance the accuracy and predictive power of the models. Furthermore, the integration of the developed system into electronic health records or telemedicine platforms can facilitate widespread adoption and enable real-time monitoring of individuals at risk.

In conclusion, this research demonstrates the effectiveness of machine learning algorithms, including XGBoost, Random Forest Classifier, and KNN, in accurately predicting heart disease. The proposed system has the potential to revolutionize the diagnosis and management of cardiovascular diseases, leading to improved patient outcomes and reduced healthcare costs. By leveraging the power of machine learning, healthcare providers can take proactive measures to prevent heart disease and improve overall cardiovascular health in the population.

Dataset: UCI Machine Learning Repository website (https://archive.ics.uci.edu/ml/index.php) for further analysis and exploration.