IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, July 2022
Impact Factor: 6.252

Heart Disease Prediction using Web Application and Random Forest Algorithm

Mr. Yashas J[1], Anagha A V[2], Savantha V J[3], Sanjay K N[4], Rakesh N V[5]
Associate Professor, Department of Electronics and Communication Engineering[1]
Students, Department of Electronics and Communication Engineering[2,3,4,5]
Malnad College of Engineering, Hassan, Karnataka, India
yashasjgowda@gmail.com, anaghaav20000@gmail.com, savanthgowda333@gmail.com

Abstract: *Heart is the main organ of our body. In today's world heart is one of the major fatal diseases and its prediction is quite difficult. This paper proposes a model to predict the heart disease using machine learning techniques. The proposed system involves data collection, data Preprocessing, data modeling and model deployment. The dataset collected has 14 columns and 421 rows corresponding to 14 medical attributes of 421 patients. This dataset is grouped for training and testing. Various machine learning algorithms were used to check the accuracy out of which the Random Forest Algorithm is best suited with an accuracy of 84%. The model is trained using random forest algorithm. Web app python code is used to load the model. The user can enter the details in the web application which is created using FLASK framework. The result is displays whether a person has heart disease or not based on the user input. If the person is predicted with heart disease then an alert message will be sent.*

Keywords: Heart Disease Prediction

I. INTRODUCTION

As we all know heart is the main organ of our body. If the functionality of the heart fails, pumping of the blood stops which in turn stops the working of other body parts. In today's modern world heart disease is one of the major fatal diseases. The difficulty behind these diseases is their prediction. Consequently, location of cardiovascular anomalies at the beginning phase and devices for the expectation of heart illnesses can save a ton of life and help specialists to plan a compelling treatment plan which eventually lessens the death rate because of cardiovascular sicknesses.ML assumes a truly significant part to recognize the covered up discrete examples and accordingly dissect the given information. After examination of information ML strategies help in heart condition forecast and early finding. This paper presents execution examination of different ML strategies like Decision Tree, K-Nearest Neighbors and Random Forest for predicting heart condition at an early stage. The Random Forest Machine Learning Algorithm is incorporated with the Flask Web system for prediction of Heart Disease is proposed.

II. LITERATURE REVIEW

Cardiovascular Disease is the primary justification behind death inside the world throughout the past ten years. Almost one individual passes on from heart condition concerning every moment inside the U. S. alone. To scale back the quantities of passings from heart sicknesses there should be a quick and prudent discovery strategy double-dealing information handling. By examining the exploratory outcomes, it's finished that the J48 tree procedure clads to be best classifier for heart condition forecast because of it contains a ton of precision and the most un-allout chance to make.[1]
A Random Forest Machine Learning Algorithm is incorporated with the Flask Web structure for anticipating of Heart Disease.Artery Blockage demonstrates the presence of heart disease. The higher the blockage, higher is the phase of heart disease.The Data expected for the prediction contains parameters, for example, Age, Sex, Blood Pressure, Sugar levels. Trial results say that predictions by utilizing the Random Forest Machine Learning Algorithm which is integrated with the Flask Web system is reliably better than those acquired utilizing different strategies. [2]
For the accurate detection of the coronary illness, a proficient ML strategy ought to be utilized which had been gotten from a distinctive examination among several machine learning algorithms in a Java Based Open Access Data Mining Platform, WEKA. to screen the heart disease patient nonstop by his/her caretaker/doctor, a constant patient observing

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, July 2022
Impact Factor: 6.252

framework was created and introduced by utilizing Arduino. An accuracy level of 97.53% accuracy was found from the SVM algorithm along with sensitivity and specificity of 97.50% and 94.94% respectively. [3]

The probabilities of heart condition and classification of patient's risk level by implementing different data processing techniques like Naive Bayes, Decision Tree, Logistic Regression and Random Forest was performed by making use of heart condition dataset available in UCI machine learning repository. Random Forest algorithm has achieved the highest accuracy of 90.16% compared to other ML algorithms executed. [4]

In order to analyse the classification performance of two machine learning algorithms: K-Nearest Neighbour and Support Vector Machine, classifiers were applied on same data and results were compared on the basis of misclassification and correct classification rate and according to experimental results, it was presumed that KNN with principle component analysis is best as compared to SVM. [5]

III. METHODOLOGY

3.1 Flow Chart

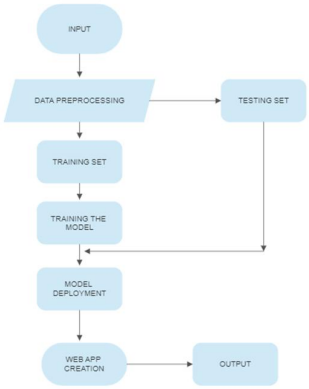

Fig. 1. Flow chart

IV. DATA COLLECTION

A data set is an assortment of data. In tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set. Here we have collected data from Kaggle.com. The dataset collected has 14 columns and 421 rows corresponding to 14 medical attributes of 421 patients.

Sl. No.	Attributes	Range	Type	Description
1.	Age		Numerical	Age in years
2.	Sex	1=Male, 0=Female	Numerical	Describes the gender.
3.	ChestPain	0=Typical angina 1=atypical angina 2=non-anginal pain 3=symptomatic	Numerical	The information about chest pain of a person.
4.	RestBP		Numerical	Resting blood pressure in mm Hg
5.	Chol		Numerical	Serum cholesterol in mg/dl
6.	Fbs	1 = True 0 = False	Numerical	Fasting blood sugar > 120 mg/dl
7.	RestECG	0=Normal 1=having ST-T wave abnormality 2=Probable or definite left ventricular hypertrophy	Numerical	Resting electrocardiographic results

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, July 2022
Impact Factor: 6.252

8.	MaxHR		Numerical	Describes max heart rate
9.	ExAng	1 = Yes 0 = No	Numerical	Describe the information about exercise induced angina
10.	Oldpeak		Numerical	ST depression induced by exercise relative to rest.
11.	Slope	0=Up Sloping 1=Flat 2=Down Sloping	Numerical	The slope of the peak exercise ST segment
12.	Ca		Numerical	Number of major vessels (0-3) coloured by fluoroscopy.
13.	Thal	3 = Normal 6 = Fixed defect 7 = Reversable defect	Numerical	Thalassemia /Blood disorder levels
14.	AHD	0 = No 1 = Yes	Numerical	Heart disease

Table 1: Attributes Description

V. DATA PREPROCESSING

From the main look, the dataset contains 14 columns, 5 of them contain numerical values and 9 of them contain categorical values. The dataset is clean and contains all the information required for each variable. By using info(), describe(), isnull() functions, no errors, missing values and irregularities values are detected.

By checking the percentage of the persons with and without heart diseases, it was found that 56% of the persons in the dataset have heart disease. So, in this way, the dataset is generally adjusted.

The conclusion after data preprocessing is as follows:

- Age and Maximum Heart Rate- Heart disease is emerging frequently in older people, and the max heart rates are lower for elderly people with heart disease.
- Chest Pain and Exercise Induced Angina- There are four sorts of chest pain: typical angina, atypical angina, non-anginal pain and asymptomatic. Majority of the heart disease patients are found to have asymptomatic chest pain. Individuals who have exercise induced angina; they usually experience the ill effects of asymptomatic chest pain, and they are bound to have heart disease.
- Thalassemia - Individuals with reversible defect are likely to have heart disease.
- ST depression and the Slope of the Peak Exercise ST Segment – Individuals who have down sloping ST segment have higher values of ST depression and more chance to be tainted with heart disease. The more prominent the ST depression, the greater the chance of disease.
- Age and Number of Major Vessels (0–3) Coloured by Fluoroscopy - A large portion of the heart disease patients are old and they have at least one major vessel coloured by Fluoroscopy.

VI. DATA MODELING

6.1 Data Splitting

The splitting of data is grouped into training and testing of data. Training is applied to the 80% of the data set and Testing is applied to the 20% of the dataset in all the cases except in case of random forest which is 70% for training and 30% for testing. Testing the data is utilized to assess the execution of the model utilizing a couple of algorithms. Based on the training data and testing data the best model is selected. The training data is not quite the same as testing data; the acquired data is applied to the algorithm.

6.2 Machine Learning Model

Various classifiers and regression models have been utilized namely Decision Tree,K-Nearest Neighbors and Random Forest.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, July 2022
Impact Factor: 6.252

- Decision Tree: In this algorithm, preprocessing made at first by dividing information into training and test data. Feature scaling should be possible in view of normalizing the qualities before prediction. The accuracy achieved with this algorithm is 68.4%.
- K-Nearest Neighbors: KNN is classifier and addressing classification is utilizedissue and it can indicate a new data point by seeing as the Euclidean Distance and majority vote count. The accuracy acquired with this algorithm is63.4%.
- Random Forest: Random Forest is a supervised classification algorithm. it comprises of various Decision Trees and we can easily get the accuracy. Here, various examples of the training data are thought of and models are created for each data. While a prediction for any data is required, each model gives a prediction and the average of these predictions are found to get a better assessment of the real output value. The accuracy acquired with this algorithm is84%.

VII. MODEL DEPLOYMENT

After training, testing and checking for accuracy of the three different algorithms utilized in training the model, the model with the highest percentage of accuracy i.e., random forest algorithm was used to make prediction. This trained model is saved and afterward loaded into a web. Flask web application framework is used to deploy and build the web application. Web app python code is used to load the model. An HTML layout is utilized for the front end to allow the users to input heart disease symptoms of the patient. These inputs are being passed to the trained model to check and make prediction of a patient and display if the patient has heart disease or not. An alert message will be sent if the person is predicted to have disease.

VIII. CONCLUSION

In this paper, a Machine learning model was being prepared in order to predict whether a person has a Heart disease or not. This model purposes a dataset which have 14 attributes directed on various people. The dataset was split for training and testing. Various machine learning algorithms were utilised to check the accuracy. Among all the algorithms Random Forest is the best suited algorithm with the accuracy of 84%. Therefore, the model was trained using random forest algorithm. Flask web application framework was utilised to create the web application where the user can enter the details. Based on the user inputs the result will be displayed. If the person is predicted to have heart disease, the alert message will be sent.

REFERENCES

[1]. Jaymin Patel, Prof.Tejal Upadhyay, Dr. Samir Patel, "Heart Disease Prediction Using Machine Learning and Data Mining Technique", In International Journal of Computer Science & Communication (IJCSC), Volume-7, Number 1 Sept 2015- March 2016 Page No.129 – 137. Shi, D.-P.,Wu, C. "The influence of infrared temperature measurementbased on reflection temperature compensation and incident temperature compensation" Electron. Measur. Technol. 08, 2321–2326 (2015)

[2]. Akram Ahmed Mohammed, Rajkumar Basa, Anirudh Kumar Kuchuru, Shiva Prasad Nandigama ,Maneeshwar Gangolla, "Random Forest Machine Learning technique to predict Heart disease" European Journal of Molecular & Clinical Medicine ISSN 2515-8260 Volume 7, Issue 4, 2020

[3]. Shadman Nashif, Md. Rakib Raihan, Md. Rasedul Islam, Mohammad Hasan Imam, "Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System" World Journal of Engineering and Technology, 2018, 6, 854-873

[4]. RachitMisra , Pulkit Gupta , Prashuk Jain, "Rachit Misra1 , Pulkit Gupta2 , Prashuk Jain", July 2021| IJIRT | Volume 8 Issue 2 | ISSN: 2349-6002

[5]. Rahul Chaurasia, Saksham Gupta and Shipra Singh Siddhu, "Prediction Of Heart Disease Using Machine Learning Algorithm" 2018 IJCRT | Volume 6, Issue 2 April 2018

[6]. O.E. Taylor1, P. S. Ezekiel, F.B. Deedam-Okuchaba, "A Model to Detect Heart Disease using Machine Learning Algorithm" International Journal of Computer Sciences and Engineering Vol.-7, Issue-11, Nov 2019

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, July 2022
Impact Factor: 6.252

[7]. Rahul Chaurasia, Saksham Gupta and Shipra Singh Siddhu, "Prediction Of Heart Disease Using Machine Learning Algorithm" 2018 IJCRT | Volume 6, Issue 2 April 2018

[8]. KillanaSowjanya, Dr. G. Krishna Mohan, "Predicting Heart Disease Using Machine Learning Classification Algorithms and Along With TPOT (AUTOML)" International Journal Of Scientific & Technology Research Volume 9, Issue 04, April 2020