International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 09 | Sep 2021 www.irjet.net p-ISSN: 2395-0072

Heart Disease Prediction Using Different ML Algorithms

Apoorva Shete[1], Nikket Chandwani[2], Prof. Anushree Gupta[3]

[1]Department of Electronics and Telecommunication Engineering, Thadomal Shahani Engineering College, Mumbai, India
[2]Department of Computer Science Engineering, Thadomal Shahani Engineering College, Mumbai, India
[3]Prof., Department of Electronics and Telecommunication Engineering, Thadomal Shahani Engineering College, Mumbai, India

---

Abstract - *Heart plays a significant role in living organisms. Deaths due to heart disease make up for the most number of deaths around the world every year. Diagnosis and prediction of heart diseases or alternatively known as cardiovascular diseases is very important. So the prediction involved requires very high precision and correctness. The correct prediction of heart diseases can prevent lives. The dataset involved here uses 13 main attributes used for performing analysis. The dataset has some irrelevant values and features which have been handled using normalization. Machine Learning algorithms are used for prediction and diagnosis. These algorithms include K-Nearest Neighbour(KNN), Support Vector Classifier, Logistic Regression, Naive Bayes, Random Forest Classifier, etc. Various results are verified using accuracy and confusion matrix. We achieved the highest accuracy on this dataset using Logistic Regression and Naive Bayes.*

Key Words: Heart Disease Prediction, Logistic Regression, Naïve Bayes, Decision Tree and KNN

1.INTRODUCTION

Heart is the one of the most vital organs of the human body so the care and maintenance of the heart is essential. Heart disease describes a range of various conditions that affect the heart. Over the last decade, heart disease or cardiovascular diseases remains the primary basis of death as per the estimates by the World Health Organization. As per their estimates, about 17.9 million deaths occur every year because of some or the other cardiovascular disease. Primarily of them being coronary artery disease and cerebral stroke. The diagnosis and discovery of heart diseases at the earliest is important. Various unhealthy activities increase the chances of heart diseases. These could include high cholesterol, obesity, hypertension, smoking, etc. Symptoms shown by individuals may be varied. They often have back pain, jaw pain, neck pain, breath weakness, chest pain, etc.

Even with heart disease being the primary cause of death in the world in recent years, these can be controlled and managed effectively. The accuracy here lies on the proper time of detection of the disease. Large set of medical records created by medical experts are available for analysis and extraction. Data can be mined using data mining techniques and then this data can be effectively handled by the Machine Learning(ML) algorithms which can be further used for diagnosis, detection and prediction. ML plays an important role to detect the hidden discrete patterns in large datasets which help us analyze the given data.

Multiple algorithms were used on the dataset and highest accuracy of 86.89% was achieved using Logistic Regression and Naive Bayes.

2. LITERATURE REVIEW

The previously implemented models for heart disease prediction and the work done in this field till date is reviewed in this section.

Paper[1] uses various attributes related to heart disease for prediction and a model based on supervised learning algorithms as Naïve Bayes, decision tree, K-nearest neighbor, and Random Forest algorithm is presented in the paper. In this research paper the highest accuracy has been achieved using KNN algorithm. The paper[2] uses a combination of ML algorithms and deep learning for the prediction of cardiovascular diseases. ML models of algorithms such as Random Forest, KNN, SVM, LR and XGBoost were implemented. IN the architecture of deep learning, three dense layers were used. It was concluded that all the ML algorithms performed better for prediction of heart diseases. Paper[3] too discusses a comparative study of different ML algorithms namely SVM, LR, KNN and Decision Tree for the prediction of heart diseases. Here, the best accuracy was achieved by SVM followed by KNN algorithm. The paper[4] presents a comparative study by analysing the performance of different machine learning algorithms. The trial results verified that the Random Forest algorithm has achieved the highest accuracy of 90.16% as compared to other ML algorithms implemented.

In paper[5], a detailed study of the past research in this field has been presented efficiently. This paper has also presented a data classification system using DT, KNN, K-Means and ADABoost algorithms for the prediction and classification of cardiovascular diseases. In paper[6], an artificial neural network algorithm was developed for classifying heart disease based on certain clinical features and attributes. The accuracy achieved in this research was nearly 80%. The proposed model in paper[7] uses

weighted voting of Logistic Regression, Random Forest, K Nearest Neighbour, Gaussian Naive Bayes, and Artificial Neural Network activated with ReLU function algorithms.

In paper[8], the proposed work uses Random Forest algorithm, K Nearest Neighbour algorithm and Logistic regression algorithm to order the dataset. The accuracy of the ensemble model with logistic regression is 95.06% and without logistic regression is 98.77%.

3. DATASET DESCRIPTION

The first and foremost step in this project was the collection of appropriate data along with data processing and cleaning. The dataset used in this paper is the Heart Disease prediction UCI dataset [9]. It provides details about several attributes of a heart disease patient. It has different factors which individually and collectively affect the heart conditions of a person. This dataset has 303 entries and 14 attributes. Out of the 303 entries in the dataset, 165 entries were those of male patients and 138 were that of female patients. This is shown below in Fig.1. Here, '0' represents the number of females whereas '1' represents the number of males.

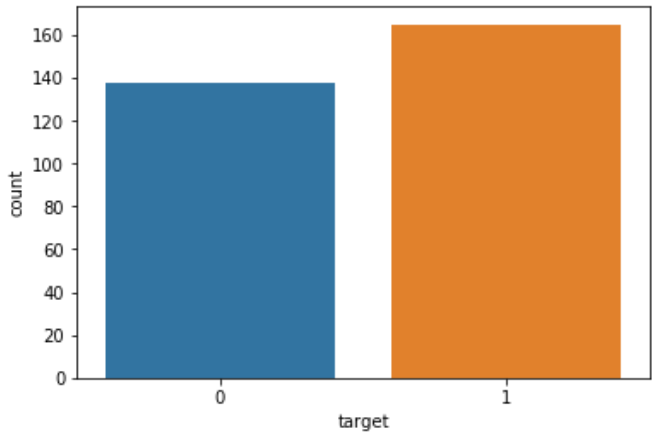

Fig.1. Distribution of patients as per sex in the dataset

From the above mentioned features, only a few important ones were selected for training and testing of the model.

After this, the data was divided into, 80:20 ratio for model testing and training.

4. METHODOLOGY

The first step in this research project was data collection that is appropriate for the prediction and analysis of heart disease. Previous section discusses the detailed description of the dataset used for the purpose of this research. The data was first cleaned and exploratory data analysis was carried out on the dataset.

The heart disease frequency for ages for both the male and female genders was plotted. In this dataset, '0' represents the Female patients' count, whereas '1' represents the male patients' count. This has been shown in Fig.2 below. It can be seen that the most frequent age for heart disease for females is 58 whereas for male it is 54.

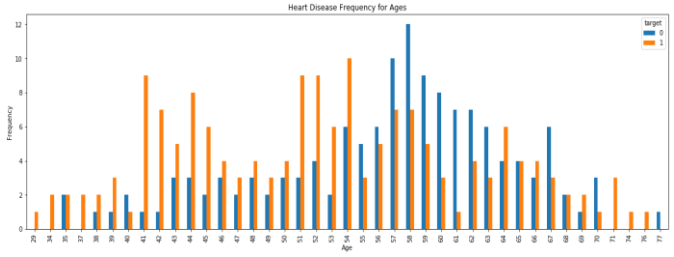

Fig.2. Heart Disease Frequency for Ages

After this the heart disease frequncy for sex was plotted. This is shown in Fig.3 below.

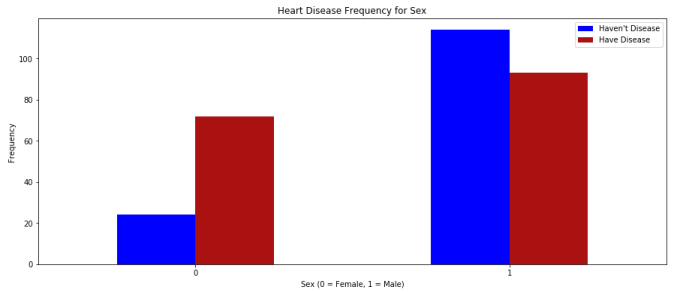

Fig.3. Heart Disease Frequency for Sex

Fig.4 and Fig.5 represent the Thalassemia v/s Cholesterol and Thalassemia v/s Resting Blood Pressure, respectively.

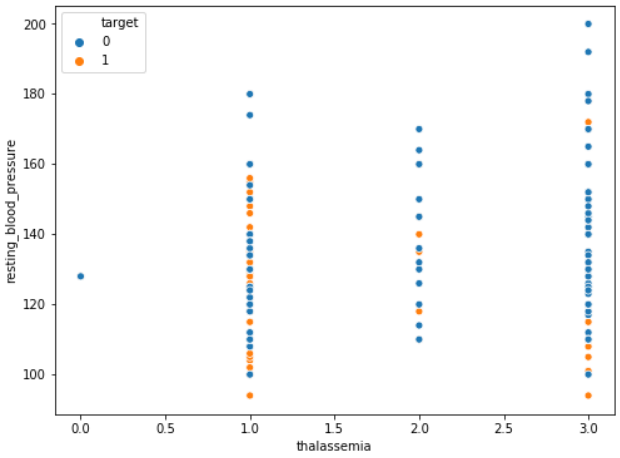

Fig. 4. Thalassemia v/s Cholesterol

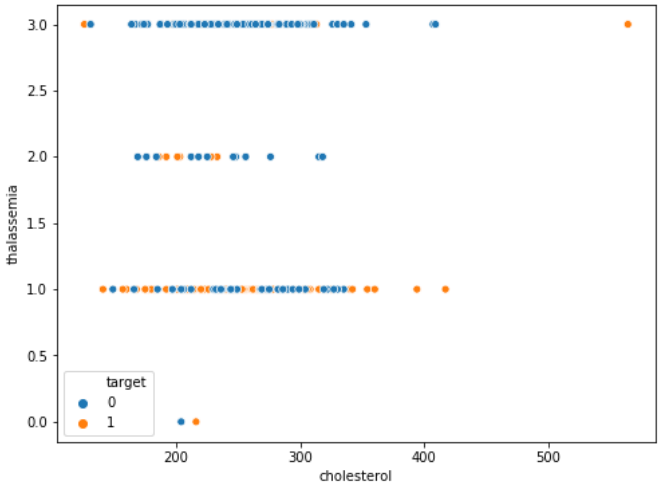

Fig.5. Thalassemia v/s Resting Blood Pressure

Fig. 6 shows the age vs maximum heart disease rate plot.

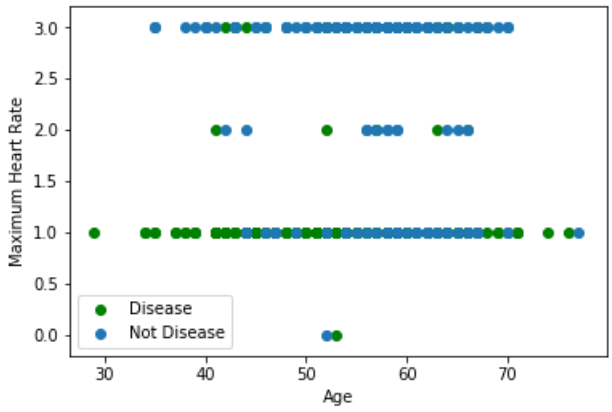

Fig.6. Age v/s Maximum Heart Disease Rate Plot

The fasting blood sugar data has been plotted for both the genders. This is shown below in Fig.7.

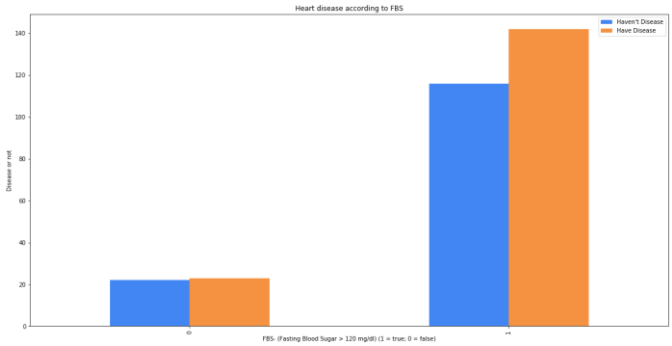

Fig.7. Heart Disease Rate as per Fasting Blood Sugar

After this, from all the features in the dataset, only the important ones were selected. The features selected were: 'age', 'resting_blood_pressure', 'cholesterol', 'max_heart_rate_achieved', 'st_depression', 'num_major_vessels'. A correlation matrix of the selected features was plotted. It is as shown below in Fig.8.

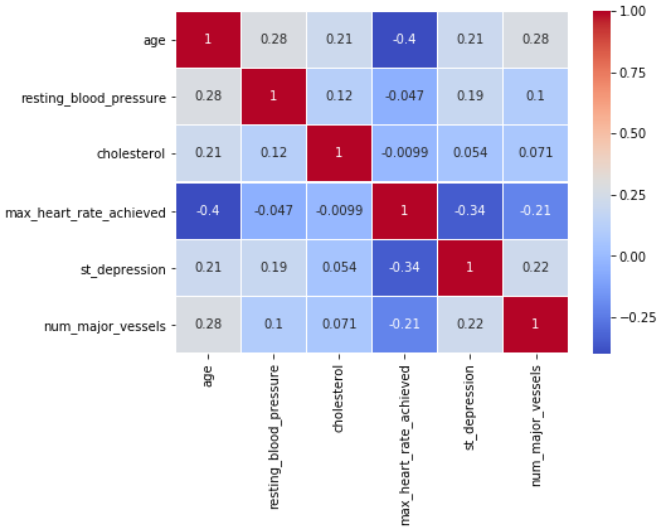

Fig.8. Correlation Matrix

The machine learning algorithms chosen for the purpose of this research were LR, NB, Decision Tree and KNN. A diagram explaining the flow of the procedure followed for the application of this model has been given below in Fig.9.

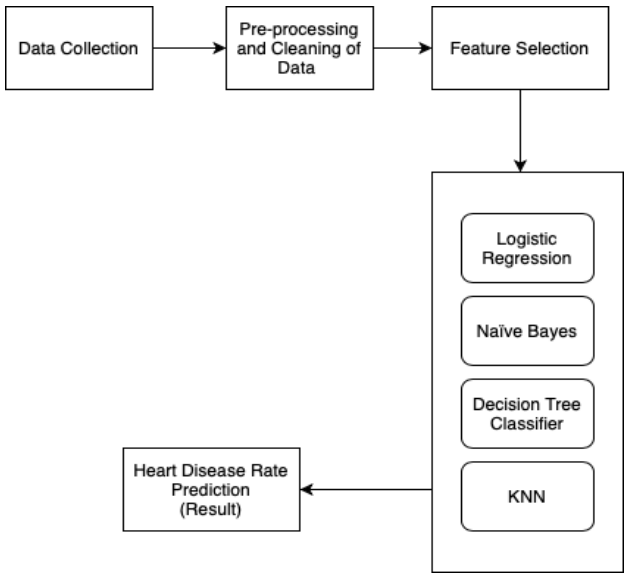

Fig.9. Workflow of procedure followed for the application of this model

5. RESULTS AND DISCUSSIONS

The accuracy obtained by each of the algorithms has been given in the Table. 1 below along with a comparative graph shown below in Fig.10.

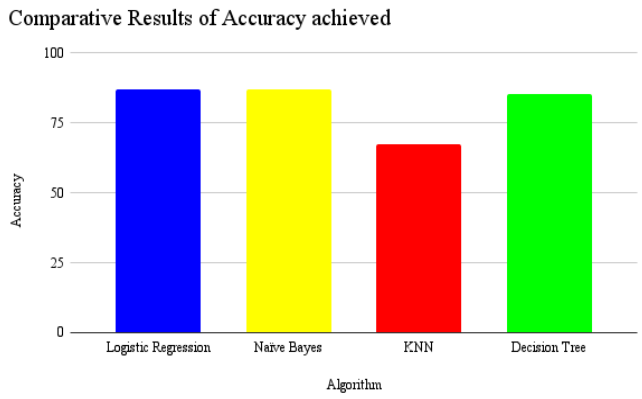

Fig.10. Comparative results of the accuracy achieved

From the results, it is evident that the best accuracy is achieved using the Logistic Regression algorithm, followed by the Naïve Bayes classifier. The accuracy obtained for the LR algorithm was **86.89%,** the best as compared to others. Whereas, the least accuracy was obtained by the KNN classifier, **67.21%**.

6. CONCLUSIONS

Heart or cardiovascular diseases are the most common diseases among humans. If heart diseases can be predicted in humans from early stages itself, it can be very beneficial and helpful to save a lot of lives in the society. This research therefore profoundly examines the different features that are essential in prediction of heart diseases.

In the future, the accuracy can be improved by using a larger dataset with more attributes. This will help the model in achieving more accurate results. Advanced machine learning algorithms can also be used with this larger data.

Hence, there are multiple methods in which this model can be further improved to give the best accuracy, results and insights to the user and help in preventing cardiovascular diseases.

REFERENCES

[1] Shah, D., Patel, S. & Bharti, S.K. Heart Disease Prediction using Machine Learning Techniques. SN COMPUT. SCI. 1, 345 (2020). https://doi.org/10.1007/s42979-020-00365-y

[2] https://www.hindawi.com/journals/cin/2021/8387680/

[3] A. Singh and R. Kumar, "Heart Disease Prediction Using Machine Learning Algorithms," 2020 International Conference on Electrical and Electronics Engineering (ICE3), 2020, pp. 452-457, doi: 10.1109/ICE348803.2020.9122958.

[4] Apurb Rajdhan , Avi Agarwal , Milan Sai , Dundigalla Ravi, Dr. Poonam Ghuli, 2020, Heart Disease Prediction using Machine Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 04 (April 2020)

[5] https://www.ijrte.org/wp-content/uploads/papers/v8i1s4/A11740681S419.pdf

[6] A. H. Chen, S. Y. Huang, P. S. Hong, C. H. Cheng and E. J. Lin, "HDPS: Heart disease prediction system," 2011 Computing in Cardiology, 2011, pp. 557-560

[7] http://www.jcreview.com/fulltext/197-1595424461.pdf

[8] https://www.ijrte.org/wp-content/uploads/papers/v9i1/F9780038620.pdf

[9] "Dataset", https://www.kaggle.com/ronitf/heart-disease-uci