**Heart Disease Prediction System**

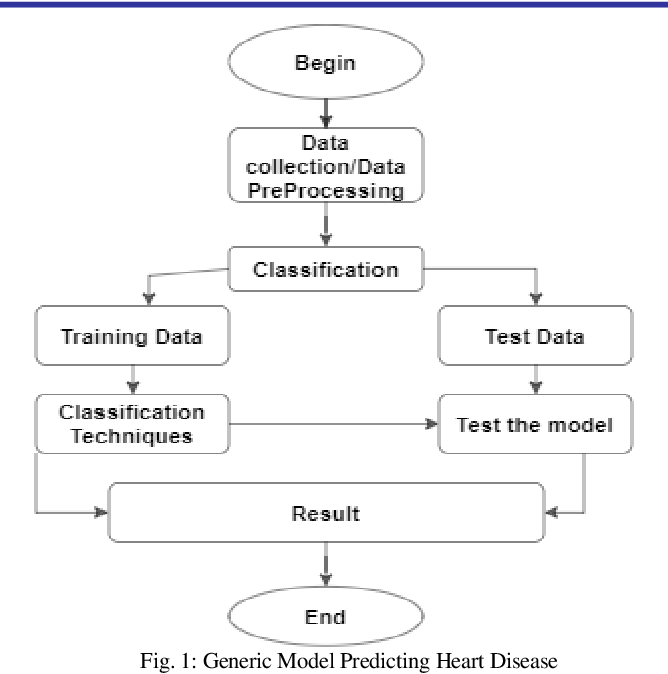
**Introduction**

The cases of heart diseases are increasing with every passing day at an alarming rate. In such a situation it is quite important to predict such diseases beforehand. Diagnosing them might be a difficult task. The project mainly focuses on which patient is more likely to have a heart disease based on various medical attributes.

We have prepared a heart disease prediction system from our end to predict whether the patient is likely to be diagnosed with a heart disease or not using their medical history. By using different ***machine- learning*** algorithms such as ***logistic regression***, ***Naïve Bayes*** and ***Random Forest Classifier*** to predict the patient's disease. *A very useful approach of the above mentioned was to regulate the model's use to improve the accuracy of prediction of heart attack in any individual.*

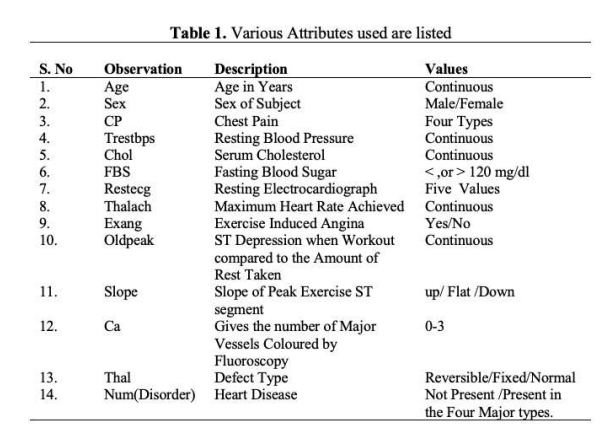
The strength of the proposed model was quiet satisfying being able to predict a heart disease in a particular individual by using *Random Forest* which showed a good accuracy in comparison to the previously used classifier such as Naive Bayes and Logistic Regression. The System enhances medical care and reduces the cost. This project gives us significant knowledge that can help us predict the patients with heart disease. It is implemented in the .pynb format.

**Steps in Heart Disease Prediction**



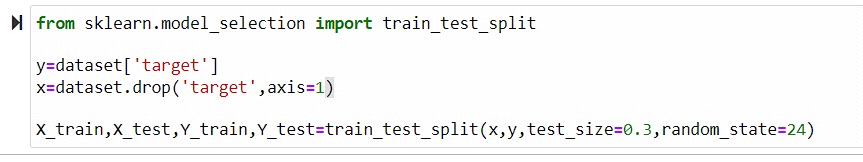
**Data Source**

We take a data source comprising of medical history of 1025 different patients of various age groups. This dataset gives us the much-needed information namely medical attributes (age, resting blood pressure, fasting sugar level etc. ) of the patient.



**Data pre-processing**

Data preprocessing deals with the missing values, cleaning of data and normalization depending on algorithms used. Typically, the data is split into a training set (used to train the model) and a testing set (used to evaluate the model's performance). According to this data set, the pattern which leads to the detection of patient prone to getting a heart disease is extracted. These records are split into two parts: Training and Testing. This data set contains 1025 rows and 14 columns, where each row corresponds to a single record.



**Model Evaluation**

After pre-processing of data, classifier is used to classify the pre-processed data the classifier used in the proposed model are Logistic Regression, Naïve Bayes, Random Forest. Finally, the proposed model is undertaken, where we evaluated our model on the basis of accuracy and performance using various performance metrics has been developed using different classifiers.

**Results & Discussions**

The following ‘figure 1’, ‘figure 2’, ‘figure 3’, ‘figure 4’ shows a plot of the number of patients that are been segregated and predicted by the classifier depending upon the age group, Sex, Chest Pain, Resting Blood Pressure:

Figure 1. Shows the Risk of Heart Attack on the basis of their age.

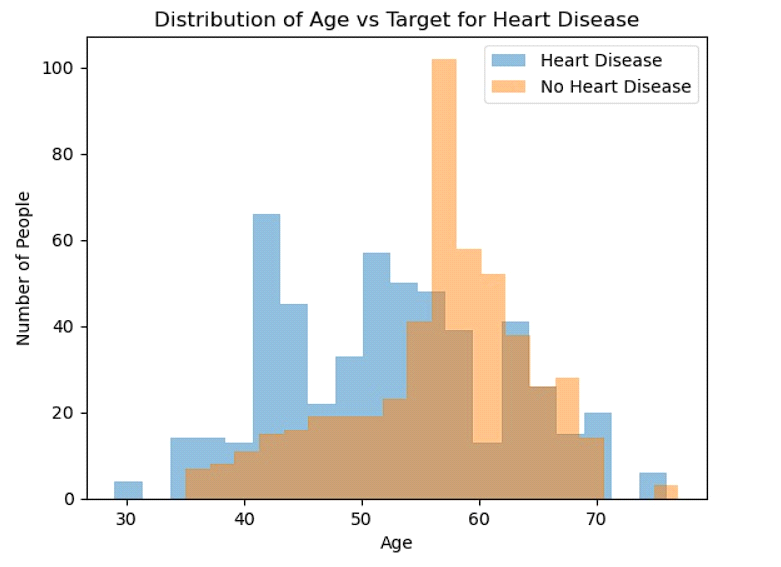


Figure 2. Shows the patients having or not having Heart Disease on the basis of Sex.

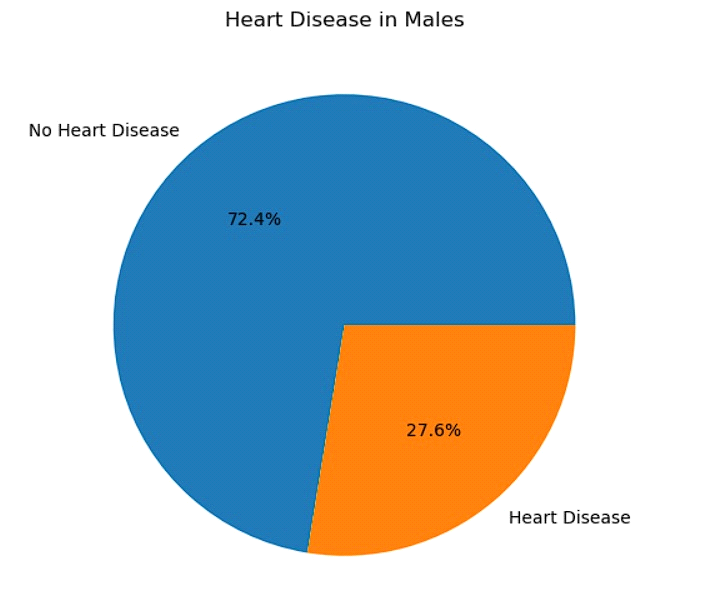
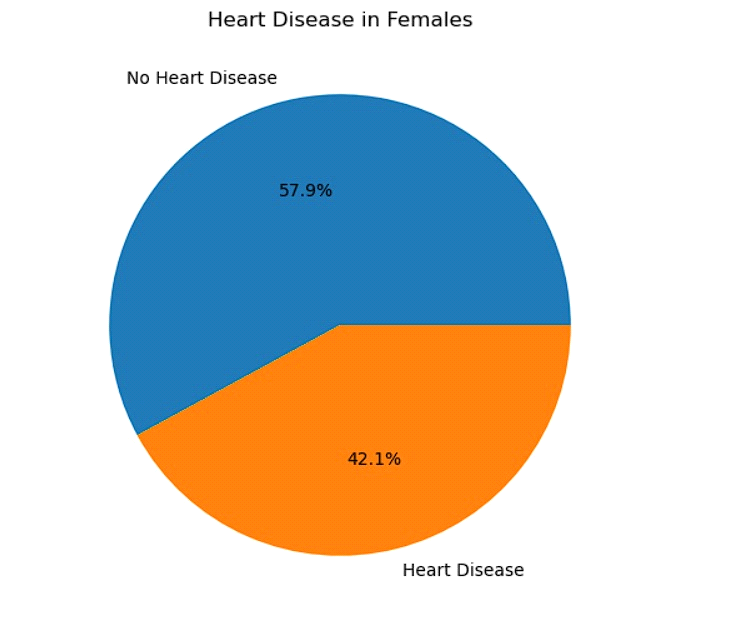


Figure 3. Shows the patients having or not having Heart Disease on the basis of type of Chest Pain.

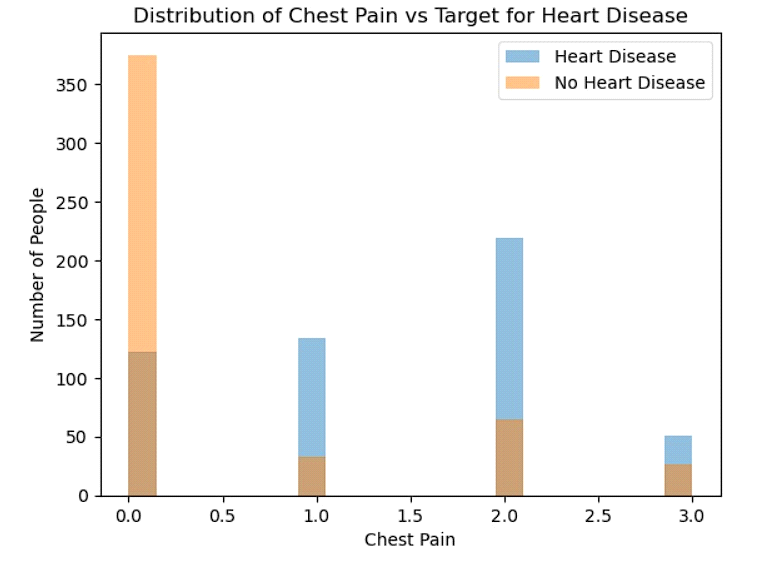
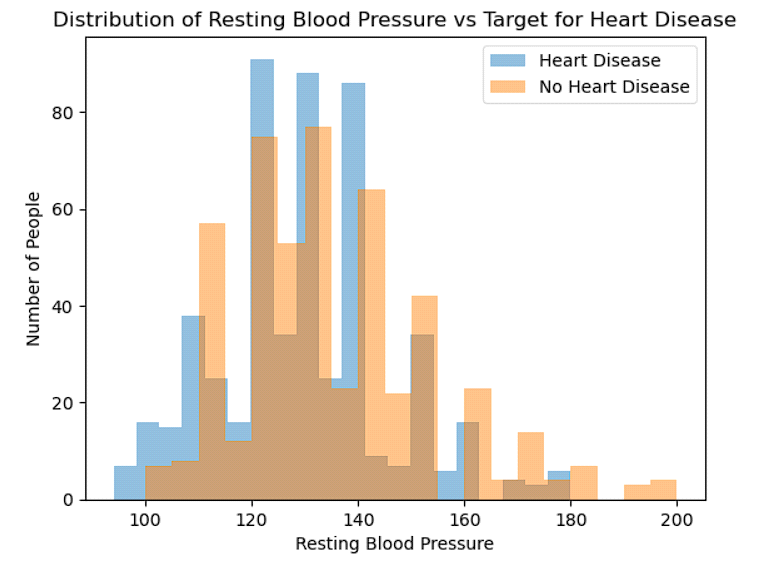


Figure 4. Shows the patients having or not having Heart Disease on the basis of Resting Blood Pressure.



**Conclusion**

The ***Heart Disease Prediction System*** offers a valuable tool for health care professionals in identifying individuals at risk of heart disease. By leveraging patient data and machine learning techniques, this system contributes to early detection, risk assessment, and improved health care outcomes. The algorithm that we used i.e. Random Forest Classifier is more accurate than Logistic Regression and Naive Bayes. Moreover, the maximum accuracy obtained by Random Forest Classifier is equal to 96.75%.