**Abstract –**

Heart disease causes a high mortality rate around the world and has become a very significant health threat for many people. Early prediction can save many lives; detecting cardiovascular disease early can help prevent it from getting worse .Heart diseases and other related diseases are growing at an alarming rate. Machine learning can help provide a healthier future by making accurate predictions about the development of said diseases. The medical industry should not reject machine learning because it will save people in the long run techniques. In the proposed work, a novel machine learning approach is proposed to predict heart disease.The study proposed using the Cleveland heart disease dataset and applied machine learning techniques such as regression, classification, Random Forest and Decision Tree. This study presents a novel technique of machine learning which has not been performed previously. 3 Algorithms like 1. Random Forest, 2. Decision Tree and 3. Hybrid models (Hybrid of random forest and decision tree) are used in machine learning models to create accuracy levels of 88.7% through heart disease prediction models, for instance The interface is designed to take the user's input parameter, which is then analyzed using a hybrid model of Decision Tree and Random Forest.

**Key Words**: Kaggle Heart Disease Database, Decision Trees, Random forest, Hybrid algorithm, Machine learning

**1. INTRODUCTION**

Machine Learning is helpful for analyzing and understanding massive amounts of information. It is used to extract data and make decisions about subsequent applications. Clustering, association rule mining, and classifications are the most common data mining approaches. These data mining approaches can be implemented using a variety of algorithms. Though simulation tools such as Weka are accessible, Python programming is gaining traction with these methods implemented using scikit-learn packages. As a result, the implementation of data mining principles in real time is more trustworthy than ever before.

Machine learning is becoming increasingly popular in the medical diagnosis business, where computer analysis may reduce manual error and enhance accuracy. Machine learning algorithms make disease diagnosis more accurate. Machine learning techniques are used to forecast diseases such as heart disease, liver disease, diabetes, and tumor. In the medical industry, classification techniques such as decision trees, naive bayes, and SVM (Support Vector Machine) were employed; similarly, regression algorithms such as Random forest, lasso, and logistic regressions were used. Deep learning algorithms are widely employed in the medical diagnosis area for most tumor forecasts.

According to surveys, almost 17 million people die each year as a result of cardiovascular disorders (CVD). Many lives could be saved if sickness is detected early, and mortality can be minimized if patients receive therapy on time .Cardiovascular diseases cover a wide range of hazards, including heart disease, stroke, and other conditions. These disorders are becoming more widespread even in younger age groups as a result of a lack of physical exercise caused by lifestyle changes. The primary causes of heart disease are smoking, lack of physical activity, high cholesterol foods, junk food, and poor living practices.

**2. EXISTING SYSTEM**

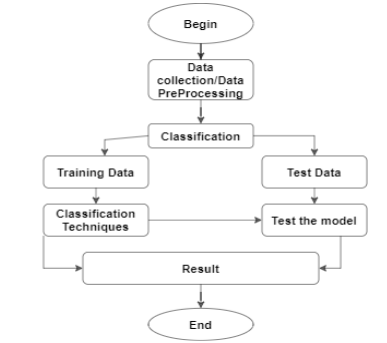
The patient provides the input details for this system. The cardiac illness is then assessed using machine learning algorithms based on the user inputs. The generated findings are now compared to those of current models in the same domain and found to be superior. Patterns are discovered using DT, Support Vector Machines SVM, and Naive Bayes using data from heart disease patients obtained at the UCI laboratory. With these algorithms, the performance and accuracy of the outcomes are compared. In comparison to other current methods, the suggested hybrid method produces results of 87 percent for F-measure.

**2.1 DISADVANTAGE**

1. It is impossible to predict the outcome of cardiovascular disease. 2. Data mining techniques are ineffective in assisting with smart decision-making.

**3. PROPOSED SYSTEM**

We used python and pandas operations to do heart disease classification using data collected from the UCI repository after reviewing the results from previous approaches. It gives a simple visual depiction of the dataset, working environment, and predictive analytics construction. The machine learning process begins with data preprocessing, followed by feature selection based on data cleaning, classification, and evaluation of modeling performance. To improve the accuracy of the outcome, the random forest technique is applied**.**

****

**FIGURE 1: Experiment workflow with Dataset**

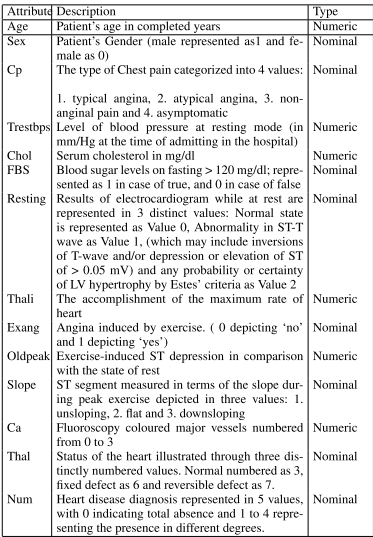
**3.1 ADVANTAGES**

1. Improved accuracy in the diagnosis of cardiac disease. 2. Random forest technique and feature selection are used to handle the largest amount of data. 3. Doctors' time complexity should be reduced. 4. Patients will save money.

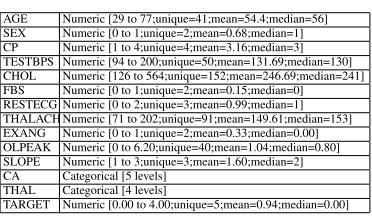
**4. APPROACH**

**4.1 Data Pre- Processing**

Dataset containing attributes sex denotes the patient's gender, age denotes the patient's age, trestbps denotes the patient's resting blood pressure, cp denotes chest pain, fbs denotes fast blood sugar, chol denotes cholesterol, thalach denotes the maximum heart rate achieved, restecg denotes the resting electroc. result (1 anomaly), oldpeak denotes the ST depression induced. ex, exang denotes exercise The slope of peak exercise is indicated by the word slope. The thalassemia is indicated by ST, pred attribute, thal**.**

****

**Table1. UCI Dataset attributes detailed information**

****

**Table 2. UCI dataset range and data types**

**4.2 Feature Selection and Reduction**

Two qualities related to age and sex are utilized to identify the patient's personal information among the 13 attributes in the data set. The remaining qualities are significant because they provide critical clinical information. Clinical data are essential for determining the degree of cardiac disease and diagnosing it.

