

# Cardio Vascular Disease Prediction Using Big Data

Dr. N. Thulasi<sup>1</sup>, Dr. V. Janakiraman<sup>2</sup>, Kural Varman<sup>3</sup>

<sup>1</sup>Professor, Dhanalakshmi Srinivasan College of Engineering and Technology.

<sup>2</sup>Professor, Dhanalakshmi Srinivasan College of Engineering and Technology.

<sup>3</sup>A V B, Student, Dhanalakshmi Srinivasan College of Engineering and Technology.

---

**ABSTRACT-** We proposed as checking the whole patient heart disease using naïve Bayes classification in machine learning. So in that, we will take results of how much percentage patients get disease as a positive information and negative information. Big data is difficult to work with using most relational database management systems and desktop statistics and visualization packages. So we can use machine learning. The proposed shows a machine learning processing model, from the data mining perspective. Using classifiers, we are processing heart percentage and values are showing as a confusion matrix. We proposed anew classification scheme which can effectively improve the classification performance in the situation that training dataset is available. Stent diagnosis of heart disease. Furthermore, the resulting model has a high specificity rate which makes it a handy tool for junior cardiologists to screen out patients who have a high probability of having the disease and transfer those patients to senior cardiologists for further analysis.

**key terms:** statistics, matrix, cardiologists, package

---

## 1. INTRODUCTION

Machine learning is one of the most disruptive technologies of this generation. It is a part of data science where in the computer systems are made to learn from the different data sets on the basis of patterns generated from the datasets. It basically enables a machine to learn on its own based on some useful algorithms specifically developed for it. Machine learning is being heavily used in all the sectors including manufacturing, healthcare, Research and Development etc. In this work we are predicting the heart disease occurrence in a patient based on some important characteristics which are best suited based on our data set that we have collected.

The heart is one of the main organs of the human body. It pumps blood through the blood vessels of the circulatory system. The circulatory system is extremely important because it transports blood, oxygen and other materials to the different organs of the body. Heart plays the most crucial role in circulatory system. If the heart does not function properly then it will lead to serious health conditions including death.

### A. Cardiovascular Diseases

Heart diseases or cardiovascular diseases (CVD) are a class of diseases that involve the heart and blood vessels.

Cardiovascular disease includes coronary artery diseases (CAD) like angina and myocardial infarction (commonly known as a heart attack). There is another heart disease, called coronary heart disease (CHD), in which a waxy substance called plaque develops inside the coronary arteries. These are the arteries which supply oxygen-rich blood to heart muscle. When plaque begins to build up in these arteries,

the condition is called atherosclerosis.

The development of plaque occurs over many years. With the passage of time, this plaque can harden or rupture (break open). Hardened plaque eventually narrows the coronary arteries which in turn reduces the flow of oxygen-rich blood to the heart. If this plaque ruptures, a blood clot can form on its surface. A large blood clot can most of the time completely block blood flow through a coronary artery. Over time, the ruptured plaque also hardens and narrows the coronary arteries. If the stopped blood flow isn't restored quickly, the section of heart muscle begins to die. Without quick treatment, a heart attack can lead to serious health problems and even death.

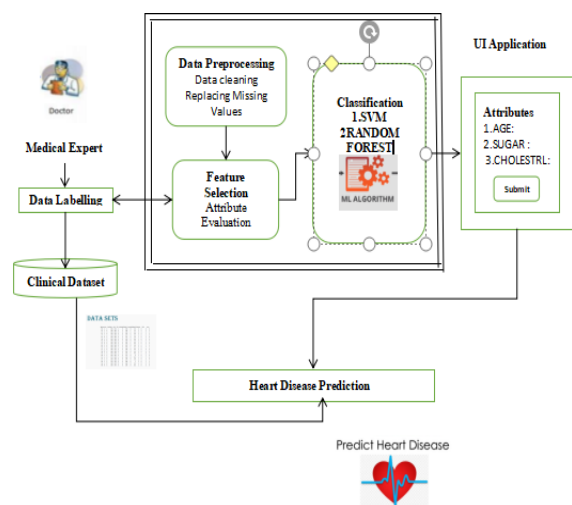
### B. Project Algorithm

The Random Forest algorithm tries to solve the problem, by using tree representation. Each internal node of the tree corresponds to an attribute, and each leaf node corresponds to a class label. It creates a forest of trees where each tree is formed by a random selection of features from the total features. Here, we can vary the number of trees that will be used to predict the class. We are also implementing the support vector machine to classify the datasets based on the class label. The support vector machine works on kernel and uses hyper plane to group datasets.

The logistic regression is also known as sigmoid function which helps in the easy representation in graphs. It also provides high accuracy. In this algorithm first the data should

be imported and then trained. By using equation the logistic regression algorithm the training data estimate the best and approximate coefficient and representation. By implementing these algorithms is used and the health care data which predicts the patient whether they are having heart disease or not.

## 2. MATERIALS AND METHODS



### **A. Data Collection**

In this phase the data is collected to analyze the best proposed method. dataset is used in our research for identifying the performance of proposed methods.

**Age:** displays the age of the individual.

**Sex:** displays the gender of the individual.

**Chest-pain type:** displays the type of chest-pain experienced by the individual.

**Resting Blood Pressure:** displays the resting blood pressure value of an individual in mmHg(unit).

**a. Serum Cholesterol:** displays the serum cholesterol in mg/dl(unit).

**b. Fasting Blood Sugar:** compares the fasting blood sugar value of an individual with 120mg/dl.

**c. Resting ECG :** displays resting electrocardiographic

**d. Max heart rate achieved:** displays the max heart rate achieved by an individual.

### **3. .RESULTSAND DISCUSSIONS**

#### **A. Data Preprocessing**

It is a technique that is used to convert the raw data into a clean dataset. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

Data goes through a series of steps during preprocessing:

**(i) Data Cleaning:** Data is cleansed through processes such as filling in missing values, smoothing the noisy data, or resolving the inconsistencies in the data.

**(ii) Data Integration :**Data with different representations are put together and conflicts with in the data are resolved.

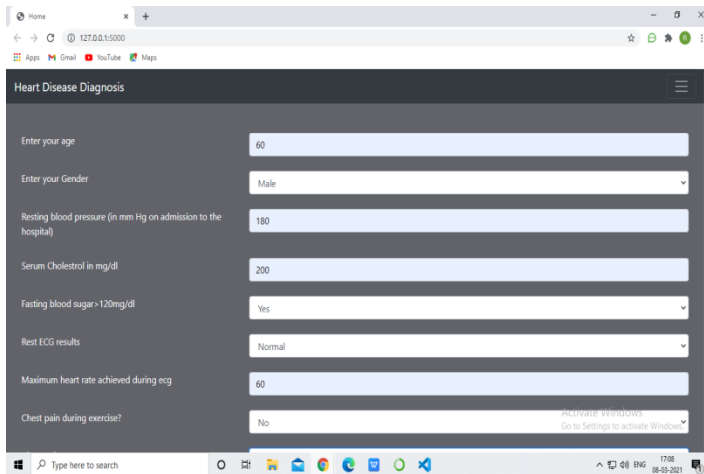
**(iii) Data Transformation:** Data is normalized, aggregated and generalized.

**(iv) Data Reduction:** This step aims to present a reduced representation of the data in a data warehouse.

#### **B. Classification**

Classification is a classic Machine learning technique based on machine learning. Basically classification is used to classify each item in a set of data in to one of predefined set of classes or groups. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics. Classification is a classic data mining technique based on machine learning. Basically classification is used to classify each item in a set of data in to one of predefined set of classes or groups. Classification method makes use of mathematical techniques such as Random Forest algorithm.

The Random Forest algorithm tries to solve the problem, by using tree representation. Each internal node of the tree corresponds to an attribute, and each leaf node corresponds to a class label. It creates a forest of trees where each tree is formed by a random selection of features from the total features. Here, we can vary the number of trees that will be used to predict the class. We are also implementing the support vector machine to classify the datasets based on the class label. The support vector machine works on kernel and uses hyper plane to group datasets.



**Fig 4.1 Screenshot of the result**

Heart diseases when aggravated spiral way beyond control. Heart diseases are complicated and take away lots of lives every year. When the early symptoms of heart diseases are ignored, the patient might end up with drastic consequences in a short span of time. Sedentary lifestyle and excessive stress in today's world have worsened the situation. If the disease is detected early then it can be kept under control. However, it is always advisable to exercise daily and discard unhealthy habits at the earliest. Here we have studied various classification algorithms that can be used for classification of heart disease databases also we have seen different techniques that can be used for classification and the accuracy obtained by them. This investigation tell us about dissimilar technologies that are used in dissimilar papers with dissimilar count of attributes with different accuracies depending on the tools designed for execution.

### 3. CONCLUSION

Heart diseases or cardiovascular diseases (CVD) are a class of diseases that involve the heart and blood vessels. Cardiovascular disease includes coronary artery diseases (CAD) like angina and myocardial infarction (commonly known as a heart attack). There is another heart disease, called coronary heart disease (CHD), in which a waxy substance called plaque develops inside the coronary arteries. These are the arteries which supply oxygen-rich blood to heart muscle. When plaque begins to build up in these arteries, the condition is called atherosclerosis.

### REFERENCES:

1. Prediction of Heart Disease Using Machine Learning, Aditi Gavhane; Gouthami Kukkula; Isha Pandya; Prof. Kailas Devadkar, 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA).
2. Heart Disease Prediction using Evolutionary Rule Learning, Aakash Chauhan; Aditya Jain; Purushottam Sharma;

- Vikas Deep, 2018 4th International Conference on Computational Intelligence & Communication Technology(CICT).
3. Blood Viscosity based Heart Disease Risk Prediction Model in Edge/Fog Computing, R Latha; P Vetrivelan, 2019 11th International Conference on Communication Systems & Networks(COMSNETS).
  4. An Investigation of Heart Disease Prediction Techniques, G.Shanmugasundaram; V. Malar Selvam; R. Saravanan; S. Balaji, 2018 IEEE International Conference on System, Computation, Automation and Networking(ICSCAN).
  5. Newfangled Approach for Early Detection and Prevention of Ischemic Heart Disease using Data Mining, Dhara B. Mehta; Nirali C. Varnagar, 2019 35th International Conference on Trends in Electronics and Informatics (ICOEI).
  6. Real-time machine learning for early detection of heart disease using big data approach Abderrahmane Ed-Daoudy; Khalil Maalimi, 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems(WITS).