Automatic Segmentation and Classification on Cardiovascular Disease Detection using Naïve Bayes from Data Mining Techniques

Aarthi K[1], Anitha Sweety M[1], Lydia Sharon Rose G[1], Balasundaram S[2]

[1]Students, B.E. Electronics and Communication Engineering
[2]Assistant Professor, Department of Electronics and Communication Engineering
[1,2]Meenakshi Sundararajan Engineering College, Kodambakkam, Chennai, Tamil Nadu, India.

---***---

Abstract - *Naïve Bayes Technique is a simple, yet efficacious and commonly utilized machine learning classifier. Enormous data generated by healthcare transactions are complex and voluminous. It needs to be processed and analyzed by different traditional methods. Data mining provides the methodology and technology to transform these volumes of data into useful information for decision making. In today's modern world cardiovascular disease is the most harmful one. Diagnosis of heart disease is a momentous and tedious task in medicine. The detection of heart disease from symptoms is a multidimensional issue. This research paper investigates three different classification models of Data Mining Techniques for detection of cardiovascular disease to facilitate experts in the healthcare domain and highlights the performance of classifications models on cardiovascular disease detection and the same has been justified with the results of different experiments conducted using WEKA machine learning software. Using Naïve Bayes and Linear Regression, the cardiovascular disease are analyzed, predicated complications, medical data are emerged with the aim of providing innovative and efficient ways to avail patients in their daily lives by monitoring vital functions and disease patterns, detecting falls or changes in motion patterns or routines or apprising health care professionals of patients' health status. It is utilized in clinical advances and in hospitals in measuring and promoting patient's health.*

1. INTRODUCTION

Now a days in many countries Technology has being improved a lot but there are slight variations in using the technologies based on their development.

Heart disease is the major cause of mortality rate in the United States, England, Canada and Wales. It was reported that heart disease kills one person every 34 seconds in the United States and the mortality rate due to Cardiovascular disease was decreased drastically by 41% in the UN from 1990 to 2016. But, in India, mortality rate increased by 34% from 115.7 to 209, one death per 1 lakh population and 54.5 million people were dead in 2016. Nearly 14% of the people in the world have been affected by heart attack. In Hong Kong males have the lowest death rate for cardiovascular disease because the Russian federation rate was about six times greater, for females the lowest death rate are found in France, Hong Kong and Japan. Turkmenistan has the highest mortality rate from cardiovascular disease in 2012, with 72 deaths per 1, 00,000 people. Kazakhstan had the highest mortality rate, with 635 deaths per 100,000 people.

There has been recent improvement in congested heart failure treatment. Researchers say that people with this disease is still unfavorable, with about 50% having an average life hope of less than 5 years. Nearly 90% of people die within 1 year due to heart failure.

Heart attacks and strokes are mainly caused due to blockage of blood from flowing to heart or brain. The main reason for this heart attack is the buildup of fatty deposits on the inner walls of blood vessels wherein the strokes are caused by bleeding from the blood vessels in the brain or by blood clots. The main reasons of heart attack and strokes is due to the usage of tobacco, unhealthy diet, obesity, physical inactivity, harmful use of alcohol, hypertension, diabetics and hyperlipidemia.

The heart disease can be reduced with so many technologies and with proper healthy diet but before reducing the heart disease, one could at least predict whether the person is having heart disease or not according to the symptoms the person is going through. The cardiovascular disease is deducted using various data mining techniques. From various data mining, one of the techniques used here is Naïve Bayes Technique.

2. LITERATURE SURVEY:

In the below survey, we will see various data mining that have been used to classify the cardiovascular disease.

In the year 2016,Hyeongsoo Kim 1, Musa Ibrahim M. Ishag 1, Minghao Piao 2,Taeil Kwon 3 and Keun Ho Ryu have tested with several machine learning techniques namely Neural Networks, Support Vector Machine(SVM), Classification based on Multiple Association Rule(CMAR), Decision tree indication and Bayesian classifiers.

In the year 2017, MIN CHEN1, YIXUE HAO1, KAI HWANG2, LU WANG1, AND LIN WANG said that they proposed new Convolutional Neural Networks (CNN) which is based on multimodal disease risk prediction algorithm using structured and unstructured data

compared with several typical prediction algorithms, there prediction accuracy reached about 94.8% with a convergence speed, which is quicker than the CNN based unimodal disease risk prediction algorithm.

In the year 2019, panel C. Beulah Christalin LathaS. CarolinJeeva said about machine learning which involves artificial intelligence being used in solving many problems in data science. Ensemble classification is a stronger machine learning technique that is commonly used for prediction. A maximum increase of 7% accuracy for weaker classifier was got with the help of ensemble classification.

In the year 2019, Avinash Golande, Pavan Kumar T combined said that they used different data mining techniques such as decision tree, KNN, adaboost, k means clustering.

In the year 2020, Chunyan guo, Jiabing Zhang, Yang liu, Yaying Xie, Zhiqiang Han, and JIanshe Yu said that they proposed recursion enhanced random forest (RERF-ILM) with an improved linear model to check whether the heart disease is present or not.

3. DATASET AND MODEL DESCRIPTION:

From this project we have categorized the patients detail into three processes. Firstly, the patient's information regarding ECG signal are being collected. Secondly, the patient's age and symptoms are being collected and updated in the dataset.

Finally, the dataset is being created with the help of hospitals, clinical and website. As per the existing system, one of the best ways to diagnose the heart disease is by using the echocardiography. This echocardiography is a painless test which uses sound waves to create the picture of a heart and it also gives the information regarding size and shape of the heart and how good the heart chambers and values are working. With existing system, only patients ECG signal is taken into consideration thereby one will know whether the patient is undergoing cardiovascular disease or not. Only with this information one cannot easily predict or reduce the heart disease.

From the above existing system, the cardiovascular disease cannot be detected effectively. Hence, this project proposes a data mining technique to make an effective way. The data mining technique which is used is Naïve Bayes Technique. In the first step to process the ECG Input signal, discrete wavelet transform (DWT) issued. DWT is a linear signal processing technique and

it is closely related to discrete Fourier transform (DFT). DWT involves sine and cosine signal. Here DWT is used in order to differentiate the PQRST signal.
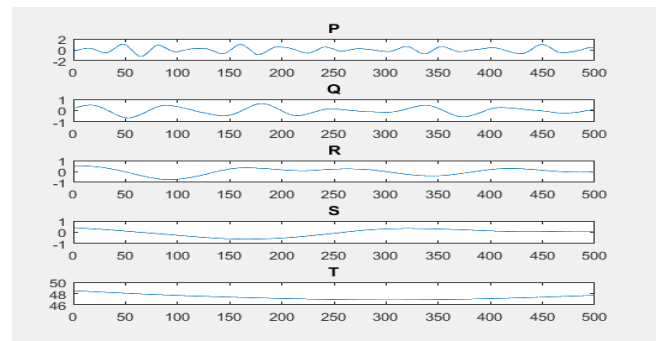

Fig – 1: Representation of PQRST spectrum

In the second step, the patient's age and symptoms are given into the Naïve Bayes Technique, which is used to classify the algorithm and classify the disease based on the symptoms of the patients.

In the final step, the ECG spectrum results, Naïve Bayes results along with data base are given to a linear regression technique.

Linear Regression explains the relationship between a dependent variable and extraordinary variable in a straight line. This linear regression technique is the one which compares the three given signal results linearly by constructing the matched output of the result of the patient.
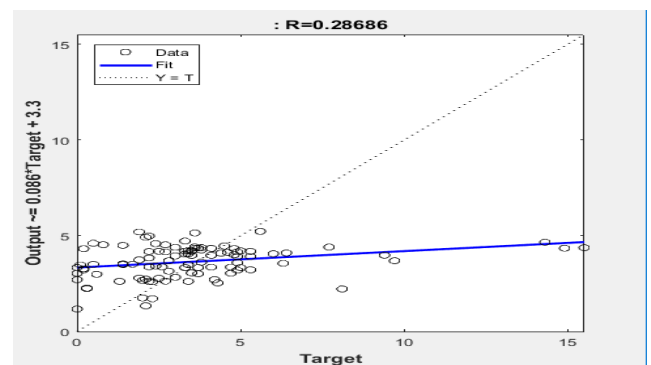

Fig – 2: Representation of Linear Regression model

From this linear regression technique the data's are analyzed and the required results are produced in a straight line with necessary accuracy and precision.
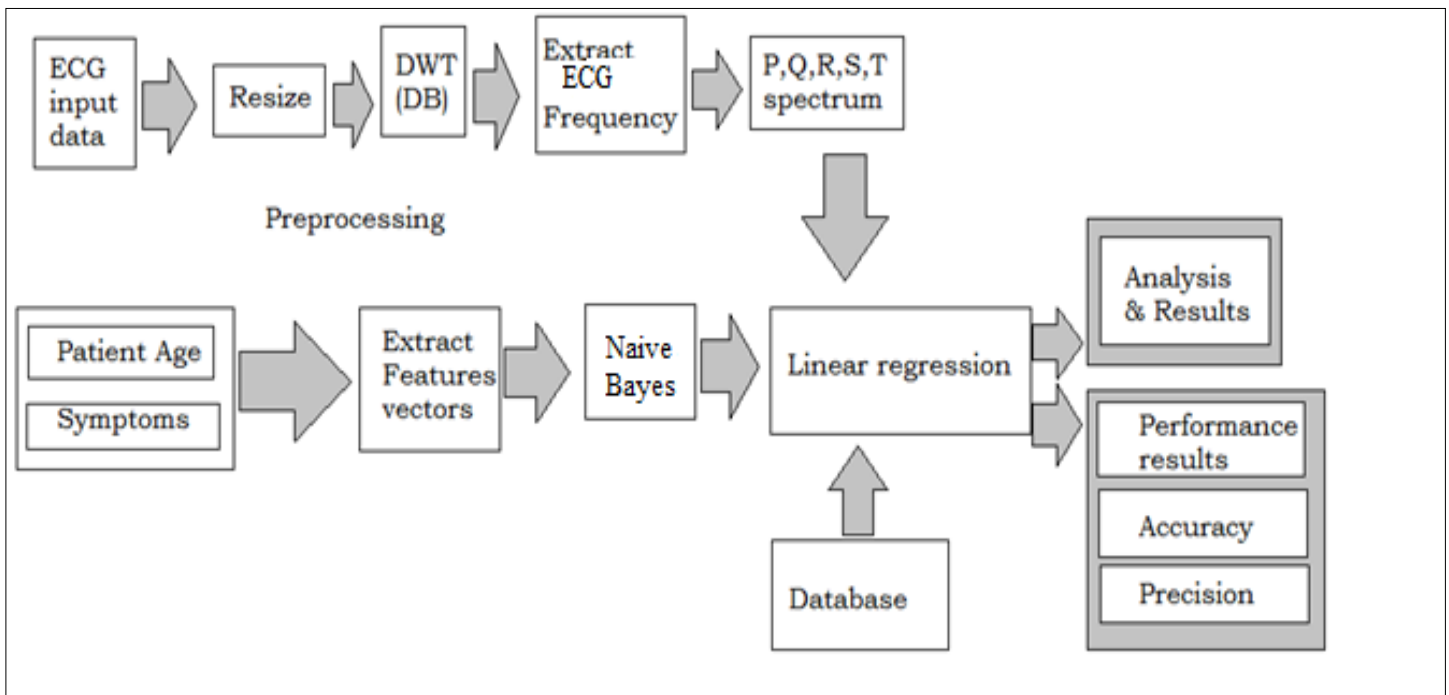
Fig – 3: Block diagram of Cardio vascular disease detection using Naïve Bayes technique

4. EVALUATION METRICS:

In this section we are measuring about performance results, accuracy and precision.

PERFORMANCE

For the performance evaluation, the test points are classified into two types, one is positive and the other one is negative. If positive, whether true and false and if negative whether true and false need to be evolved.

The above metrics are defined as follows:

TRUE POSITIVE (TP): The number of occurrence is correctly predicted as required.

FALSE POSITIVE (FP): The number of occurrence is incorrectly predicted as required.

TRUE NEGATIVE (TN): The number of occurrence is correctly predicted as not required.

FALSE NEGATIVE (FN): The number of occurrence is incorrectly predicted as not required.

The performance measurement shall be used for accuracy calculation, precision, recall, F1 score.

ACCURACY:

The accuracy rate is the ratio of true positive disease rate currently identified the total sample quantity.

$$\text{Accuracy} = \frac{TP}{(TP+TN+FP+FN)}$$

PRECISION:

Quality or state of being precisely exact.

$$\text{Precision} = \frac{TP}{(TP+FP)}$$

RECALL:

$$\text{Recall} = \frac{TP}{(TP+FN)}$$

F1 SCORE:

$$\text{F1 score} = \frac{2 \ X \ PR}{P+R}$$

5. CONCLUSION:

In this paper, we have studied about Naïve Bayes classifier and Linear Regression Techniques which is used in deducting various cardiovascular diseases from the database. Cardiovascular disease is very threatening; it exceeds the control line and reaches to maximum death rate in mankind. Thus it is concluded that with 85% to 90% accuracy one can assure various cardiovascular disease present in patients. From this project various diseases present in heart is suggested and it is not a conclusion.

6. FUTURE SCOPE:

Now a day one can find that all the technologies are computerized and well equipped and used by everyone in day today life style. This project is done as backend process but when frontend process is developed with

various software languages it can be developed as an App. So this App can be widely used in android mobile, clinics and in hospitals.

7. REFERENCES:

[1] Avinash Golande, Pavan Kumar T [2019] Heart Disease Prediction Using Effective Machine Learning Technique.

[2]Ravindhar NV, Anand, Hariharan Shanmugasundaram, Ragavendran, Godfrey Winster [2019]Intelligent Diagnosis of Cardiac Disease Prediction using Machine Learning

[3]Ning Meng, Peng Zhang, Junfeng Li, Jun He and Jin Zhu1[2016]Prediction of Coronary Heart Disease Using Routine Blood Tests

[4]Panel C Beulah ChristalinLathaS. CarolinJeeva[2019]Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques

[5] Resul Das , Ibrahim Turkoglu , Abdulkadir Sengur [2009]Effective diagnosis of heart disease through neural networks ensembles

[6] Chunyan guo, Jiabing Zhang, Yang liu, Yaying Xie, Zhiqiang Han, and JIanshe Yu Recursion Enhanced Random[2020]Forest With an Improved Linear Model (RERF-ILM) for Heart Disease Detection on the Internet of Medical Things Platform.

[7] Min Chen , Yixue Hao , Kai Hwang , Lu Wang and Lin Wang[2017] Disease Prediction by Machine Learning Over Big Data From Healthcare Communities.

[8] Hyeongsoo Kim, Musa Ibrahim M. Ishag , Minghao Piao ,Taeil Kwon and Keun Ho Ryu[2016]A Data Mining Approach for Cardiovascular Disease Diagnosis Using Heart Rate Variability and Images of Carotid Arteries.