ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Heart Disease Prediction and Treatment Suggestion.

L Mary Gladence[1], and T Ravi[2].

[1]Research Scholar, Sathyabama University, Jeppiaar Nagar, Chennai-119. India.
[2]Madanapalle Institute of Technology&Science,P.B.No:14,Angallu,Madanapalle-517325,India.

ABSTRACT

Data Mining tools are efficiently being used in medical fields, such as diagnosing of disease. Heart disease is causing death all over the world. Various Data Mining tools are available, whereas using any one of the tool has showed acceptable level of accuracy. This project determines the holes in the research about coronary illness analysis furthermore medicine furthermore construct a model should efficiently end the individuals holes and on discover, though applying information mining systems will coronary illness medication information might give acceptable likewise dependable execution likewise that attained over diagnosing coronary illness.

Keywords: Data Mining, Heart Disease, Naïve Bayes, Decision Tree

Corresponding author

ISSN: 0975-8585

INTRODUCTION

Data mining technique is a method of extracting useful information from large sets of data. Data mining techniques can also be used in medical field[1] to [13] in order to predict diseases. The healthcare industry collects huge amounts of healthcare data which, unfortunately, are not "mined" to discover hidden information for effective decision making [7]. Hidden patterns and relationship between them often goes unexploited. Several advanced data mining techniques can help to remedy this situation. This project uses several data mining techniques such as Decision Trees, Naïve Bayes, Neural Network, Association Rule, and Linear Regression. Results show that each technique has its unique strength in realizing the objectives of the defined mining goals. Applying these techniques to several data sets of heart disease can help us to predict the presence of the disease.

Related Works

Data mining techniques can be used for predicting various diseases such as, Cancer, Diabetes, and Heart Disease. Heart disease is the most important reason of casualty in the UK, USA, Canada, and England [1]. Jyoti Soni et. al [2] proposed three different supervised machine learning algorithms such as Naïve Bayes, k-NN, and Decision List algorithm and they are used for analyzing the heart disease dataset [3]. Data mining tool is used for classifying these data. These classified data is evaluated using 10 fold cross validation and the results are compared. Decision tree is one of the popular and important classifier which is easy and simple to implement. It doesn't have domain knowledge or parameter setting. It is more suitable for exploratory knowledge discovery and can handle large amount of data. The results obtained from Decision Tree are easier to interpret and read [4].

Hereditary calculation have been utilized as a part of [5], to decrease the genuine information size to get the ideal subset of credited sufficient for coronary illness forecast. Grouping is one of the regulated learning strategies to concentrate models portraying vital classes of information. Three classifiers e.g. Choice Tree, Naïve Bayes and Classification through bunching have been utilized to diagnose the Presence of coronary illness in patients. Affiliation Rule has been generally utilized for finding standards as a part of restorative applications [6].

PROPOSED WORK

Diagnosis of heart disease is a significant and a tedious work to perform. Disease diagnosis has been done in various ways. In the proposed system heart disease prediction is done by extracting the data from the UCI repository and mining it. Baseline accuracy is obtained after mining the datasets. Mining of the dataset is done using the algorithms such as Naïve Bayes, Decision tree, linear regression, Association rule and neural networks. The results obtained from the each mining model are combined together to obtain the optimal result.

A. Analyzing the Dataset.

Clinical Datasets contain large amount of information about the patients and as well as their medical report.

1) Dataset for Heart Disease: The term Heart disease encompasses of diverse diseases that affect the heart. Heart disease is considered as the major cause of several health issues around world. Record sets with medical attribute are collected and with the help of it the patterns significant to the heart attack prediction are extracted.

The attribute "Diagnosis" is identified as the predictable attribute with value "1" for patients with heart disease and value "0" for patients with no heart disease. In this module we analyze the records which are stored in the database. Originally thirteen attributes were involved in predicting the heart disease. Patients Identification number is taken as the key attribute to differentiate between various patients. Few of them which play a major role in predicting the heart disease are described below.

a) Angina: Also Known as "angina pectoris" is a sense of pain occurring in the chest. It occurs if the heart

muscle does not get enough amount of Oxygen-rich blood. Angina occurs in various numbers of patterns. Few among them are:

i. Chronic Angina: angina is most common type of angina, it is characterized by a long-term pattern of exercise-induced angina (Exang) occurring very predictably over months to years. The pattern of stable angina should be keenly observed in order to determine when the pain will usually occur. Stable angina isn't a heart attack whereas; it indicates the probability of occurrence of heart attack
ii. Unstable Angina: occurs at rest and does not follow a pattern. It occurs due to sudden worsening of symptoms in a person who had been experiencing stable angina. It is very dangerous and the patient must be provided with the treatment immediately. It indicates that the probability of heart attack is higher.
iii. Variant Angina: is also known as Prinzmetal's Angina. This type of angina is very rare and occurs usually at rest. The pain caused due to this is very severe and it can be treated by the course of medications.

b) Cholesterol: helps in building new cells in our body, insulate nerves and also to produce hormones. High cholesterol usually leads to heart attack. High cholesterol builds up in the walls of the arteries leading to a process called atherosclerosis. Atherosclerosis is one of the forms of heart disease. Cholesterol can be further classified as Low-density lipoprotein (LDL or "Bad" Cholesterol) and high-density lipoprotein (HDL or "good" Cholesterol). These define the form in which cholesterol travels in the blood. The cholesterol level is considered to be high if it is obtained 240 and above, borderline high if it's measured between 200-239 and desirable if it is less than 200.

c) Blood Pressure: It is the measure of the force at which the blood flows against the walls of your blood vessels. The rate of blood pressure can be measured in two ways known as systolic and diastolic. The measure of pressure force when your heart contracts and pushes out the blood is known as systolic. Diastolic number is the measure when a heart relaxes between the beats. The categorization of blood pressure based on systolic and diastolic can be stated as low risk if the value is 120/80, Medium risk if the value is 121-139/80-89, high risk if it is 140+/90.

d) Electrocardiogram: In order to diagnose heart block in a patient, doctors make use of a test known as electrocardiogram (ECG). This is used for detecting and recording the electrical activity of a heart when it undergoes depolarization (excitation) and polarization (recovery) in order to initiate each beat of the heart. ECG maps the hearts electrical activity into line tracings on a paper known as graph. Waves are referred to the spikes and dips in the graph. Doctor makes use of spikes and dips in order to determine the electrical activity of your heart. The grouping up of spikes and dips into different sections determine how your heart functions. The ECG results are considered to be normal if the heart beats in regular rhythm and it usually lies between 60 and 100 beats per minute. It's considered as abnormal if it beats too slowly that is less than 60 beats per minute or too fast such as more than 100 beats per minute.

B. Algorithm Implementation

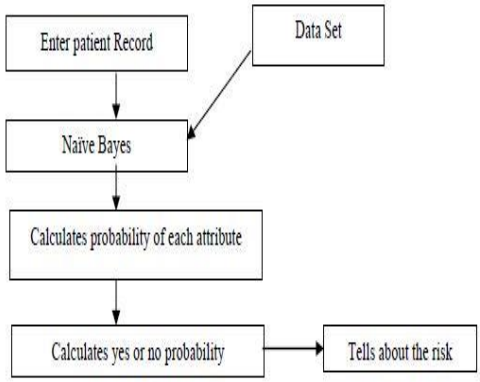

Figure1. Implementation of Naïve Bayes Algorithm of Patient Data

ISSN: 0975-8585

In this module, the data mining algorithm are applied to the datasets. Algorithm such as Decision Trees, Naïve Bayes, Neural Network, Association Rule, and Linear Regression are used. For example consider a person's medical report to be Age=<30 and Overweight=no and Alcohol Intake=never then Heart attack level is considered to be Low (Or) If Age=>70 and Blood pressure=High and Smoking=current then Heart attack level is considered high. Implementation of Naïve Bayes algorithm on the patient's data can be represented in(Fig.1)

C. Pattern Formation

This module results to the formation of a pattern based on the implementation of the algorithm. Each algorithm will create a pattern based upon the different data sets. Pattern formation is the process of obtaining certain binary values based upon the implementation of the algorithm and it has been shown in(Fig.2).

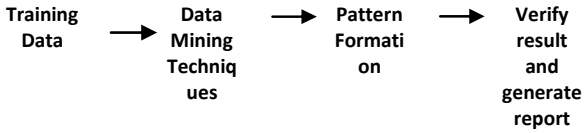

Figure 2: Process of Pattern Formation

D. Mining the Datasets

In this module, the input attributes provided by the patients are converted into some pattern by using data mining analysis algorithms. The pattern obtained is matched to the pattern which is generated by the algorithm implementation in dataset by using the different mining process. After mining process it will produce some binary values for both dataset and input attributes. The binary values are used for predicting the presence of the disease. Here we are using the analysis service of Microsoft sql server in order to obtain the baseline accuracy of the datasets. Based on the datasets, the model view of decision tree is shown in (Fig.3).

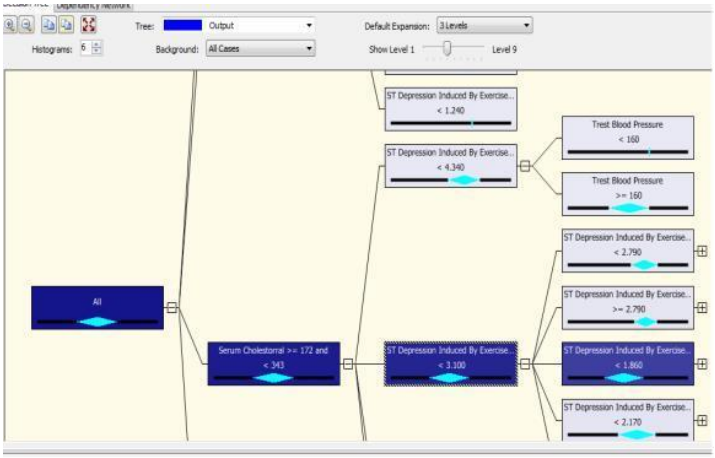

Figure 3: Model View of Decision Tree

E. Diagnosing the Disease:

In this module the binary values obtained are used for determining the disease. Implementation of each algorithm produces binary values. Depending upon the majority of the binary values the disease is predicted, that is, if the output of the algorithm implementation consists of majority of ones, the patient is considered to have the disease. If the output consists of majority of zeros, the patient is considered to be normal (No disease). The probability of disease is measured using which the treatment suggestion is provided to the patient. The probability of disease may differ for different patients and hence the treatment suggestion is given. Each algorithm produce its own output and based on the result of these algorithms the prediction of the disease was done and it is displayed in (Fig.4).

ISSN: 0975-8585

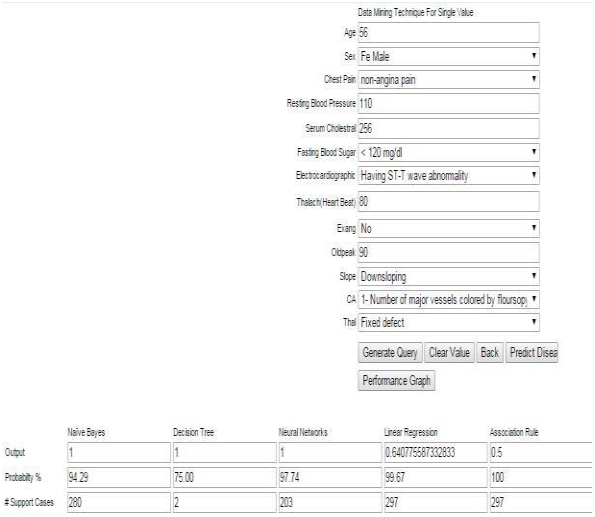

Figure 4: Output based on the results of algorithm

A Graphical representation of the output was displayed using the performance graph. The graph illustrates the output, probability and support cases for each algorithm. “Fig. 5” illustrates the performance graph of the given output.

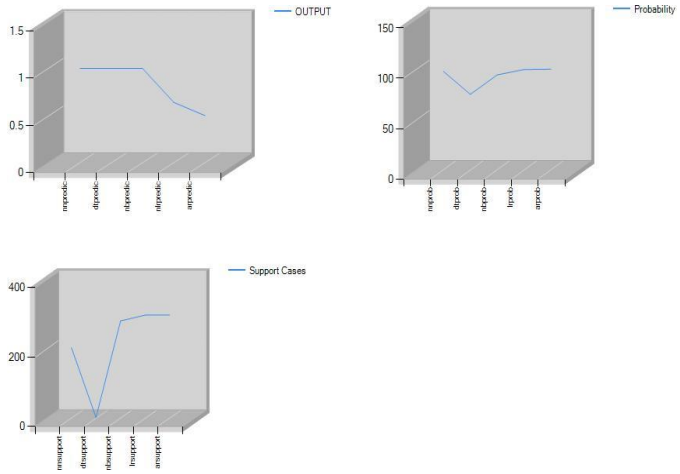

Figure.5 Performance Graph

The x-axis of the graph contains the various algorithm defined as in neural network, decision tree, naïve bayes, linear regression and association rule. The y-axis contains the different range values for each case such as output, probability and support cases.

CONCLUSION AND FUTURE ENHANCEMENT

Heart disease prediction is a major challenge in the health care industr8ffy. Predicting Heart Disease with less number of attributes is a difficult task. Decision Support in Heart Disease Prediction System is developed using all the five data mining techniques. The system extracts hidden knowledge from the heart disease database. This is an effective model to predict patients with heart disease. This model could solve complex queries and provide detailed information with accuracy. DSHDPS can be further enhanced and expanded by incorporating other data mining techniques such as clustering, Time series.

ACKNOWLEDGEMENTS

The authors are thankful to the management of Sathyabama University-India for supporting us to publish this work.

REFERENCES

[1] Shadab Adam Pattekari and Asma Parveen,"Prediction system for heart disease using Naïve Bayes", International Journal of Advanced Computer and Mathematical Sciences, vol.3,pp 290-294,2012

[2] Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", International Journal of Computer science and Engineering, Vol. 3, No. 6, June 2011.

[3] Han. J., Kamber, M, "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2006

[4] R.Rathna and A.Sivasubramanian. (2011), "TDMA Based Low Energy Consuming MAC Protocol for Wireless Sensor Networks in Environmental Monitoring Applications" in the proceedings of the International Conference on Parallel, Distributed Computing Technologies and Applications (PDCTA-2011) , Advances In Parallel Distributed Computing , Springer CCIS, ISSN 1865-0929, ISBN 978-3-642-24036-2,sep 2011, Vol. 203, Part 1, pp. 420-427.

[5] M. Anbarasi, E. Anupriya, N.CH.S.N.Iyengar, "Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm", International Journal of Engineering Science and Technology, Vol. 2(10), 2010, pp. 5370- 5376.

[6] C.A.Sathiyamoorthy, Dr.G. Kulanthaivel," Design And Develop A Computer Aided Design For Automatic Exudates Detection For Diabetic Retinopathy Screening", International Journal on Intelligent Electronic Systems,Vol.9,issue1 2015.

[7] Asha Rajkumar, G.Sophia Reena, "Diagnosis Of Heart Disease Using Datamining Algorithm", Global Journal of Computer Science and Technology, Vol. 10, Issue. 10, September 2010.

[8] R.Subhashini and V. Jawahar Senthil Kumar, "A Framework for Efficient Information Retrieval using NLP Techniques ", Proceedings of the International Conferences on Advances in Communication Network and Computing, CNC 2011, CCIS 142, pp. 391–393, 2011, Springer-Verlag Berlin Heidelberg 2011

[9] Hnin Wint Khaing, Data Mining based Fragmentation and Prediction of Medical Data, ISBN: 978-1-61284-840-2, 2011.

[10] Shruti Ratnakar, K. Rajeswari, Rose Jacob., Prediction Of Heart Disease Using Genetic Algorithm For Selection Of Optimal Reduced Set Of Attributes, International Journal of Advanced Computational Engineering and Networking, ISSN (PRINT): 2320-2106, Volume – 1, Issue – 2, 2013.

[11] Sathiyavathi R(2015) "A Survey: Big Data Analytics on Healthcare System", Contemporary Engineering Sciences, Vol. 8, 2015, no. 3, 121 – 125.

[12] Gowri, S. and Anandha Mala, G. S. (2015), "Classification of Breast Cancer Cells using Novel DPSC Algorithm", Journal of Pure and Applied Microbiology, Vol. 9, No. 2, pp. 1395-1400, ISSN: 0973-7510.

[13] Jabez J and Muthu Kumar B (2015), "Anomaly-Detection In Diabetes using SVM", International Journal of Applied Engineering Research, Vol.10, No.5, pp 12627-12636.