*International Journal of Latest Trends in Engineering and Technology*
*Special Issue SACAIM 2017, pp. 443-445*
*e-ISSN:2278-621X*

# A STUDY ONEFFECTIVE DATA MINING TECHNIQUES FOR PREDICTION OF HEART DISEASES

Aashritha Sequeira[1] & Nishel Roylin D'Souza[2]

**Abstract:In this modern world heart disease is a major health problem. Cardiovascular Disease(CVD) is one such threat.Unlesstreated at an early stage it will lead to illness and causes death.There are no adequate resources and effective analysis tools to discover relationships and trends in data especially in the medical sector.Health care industry today generates large amounts of complex clinical data about patients and other hospital resources. Data mining techniques are used to analyse this collection of data from different perspectives and derive useful information.The diagnosis of this disease using different features or symptoms is a complex activity.The purpose of this paper is to make a study on different data mining techniques such as decision Tree, neural network, Naive Bayes and predict the disease.**

**Keywords: Cardiovascular disease,Decisiontree,Neuralnetwork,Naïve Bayes.**

## 1. INTRODUCTION

Coronary illness is the main source of death on the planet in the course of recent years. Several distinct indications are related with coronary illness, which makes it hard to analyze. It seems sensible to have a go at using the learning and experience of a few pros gathered in databases towards helping the Diagnosis procedure. It also provides healthcare professionals an additional source of knowledge for deciding[8].

Most hospitals today utilize some kind of clinic data frameworks to deal with their medicinal services or patient information. These frameworks regularly create enormous measures of information which appear as numbers, content, diagrams and images[1].Data mining is the investigation of extensive datasets to extract hidden and already obscure examples, connections and learning that are hard to distinguish with customary factual strategies (Lee, Liao et al. 2000). Hence by executing a coronary illness expectation framework using Data Mining procedures and doing some kind of information mining on different coronary illness qualities, it can have the capacity to predict more probabilistically that the patients will be diaognised with heart disease[2].In this paper, we have made a study on the usage of different data mining techniques used in prediction of Cardiovascular Diseases(CVD).

The rest of the paper is organized as follows. The algorithms are explained in section II. Concluding remarks are given in section III.

## 2. ALGORITHMS

### *2.1.Naive Bayes*

It is a statistical classifier which assumes no dependency between the attributes.This classifier algorithm uses a conditional independence i.e, it assumes that an attribute value on a given class is independent of the values of other attributes. The advantage of using Naive Bayes is that one can work with the Naïve Bayes model without using any Bayesian methods[4]

1. Each data sample is represented by an n dimensional feature vector.LetX = (x1, x2….. xn), where n is the measurements made on the sample from n attributes, respectively A1, A2, An.

2. If there are m classes, C1, C2……Cm. Given an unknown data sample, X, the classifier will predict that X belongs to the class having the highest posterior probability, conditioned on X. That is, the naive probability assigns an unknown sample X to the class Ci if and only if:

P(Ci/X)>P(Cj/X) for all 1< = j< = m and j!= i

Therefore we maximize P(Ci|X). The class Ci for which P(Ci|X) is maximized is called the maximum posteriori hypothesis.

By Bayes theorem,

P(Ci/X)= (P(X/Ci)P(Ci))/P(X)

3. As P(X) is constant for all classes, only P(X|Ci)P(Ci) is maximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are equally likely, i.e. P(C1) = P(C2) = …..= P(Cm), and we would therefore maximize P(X|Ci). Otherwise, we maximize P(X|Ci)P(Ci). Note that the class prior probabilities may be estimated by P(Ci) = si/s ,

where Si=number of training samples of class Ci

s= the total number of training samples.

[1] Department of Software Technology, St.AloysiusCollege,AIMIT,Mangalore,Karnataka,India
[2] Department of Software Technology, St.AloysiusCollege,AIMIT,Mangalore,Karnataka,India

*2.2 Decision Tree Structure*

Thisis a classifier which is simple and easy to implement. There is no requirement of domain knowledge or parameter setting.High dimensional data can be handled easily. It produces results which are easier to read,understand and interpret. The drill through feature inorder toaccess detailed patients profiles is only available in Decision Trees[5].

If Age=<30 and Overweight=no and Alcohol
Intake=never
Then
Heart attack level is Low
(Or)
If Age=>70 and Blood pressure=High and
Smoking=current
Then
Heart attack level is high
Age

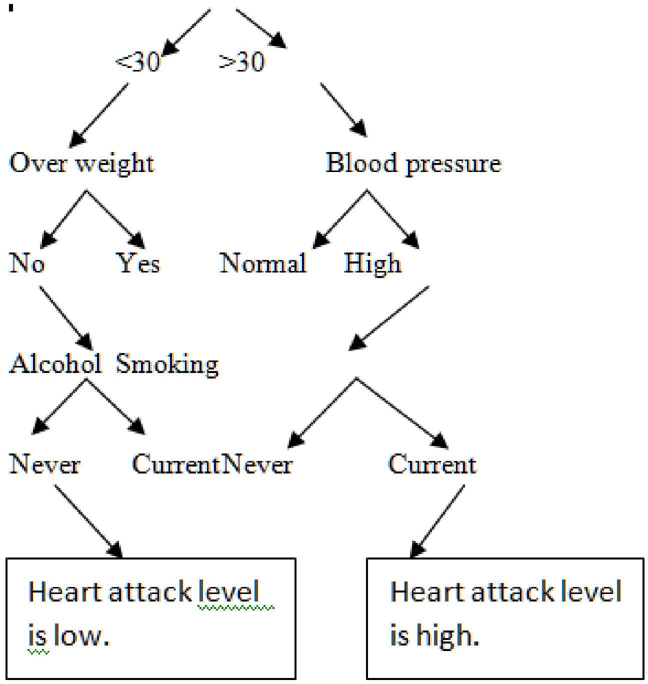

Figure 1**:** A decision tree for the concept heart attack level by information gain .

*2.3 Neural Network*

An artificial neural network (ANN), often called as "neural network" (NN), is a mathematical model or computational model based on biological neural network.In feed-forward neural networks the neurons of the first layer forward their output to the neurons of the second layer in a unidirectional fashion,which explains that the neurons are not received from the reverse direction. There is a connection between each layer & weights are assigned to each connection. The primary function of neurons of input layer is to divide input $x_i$ into neurons in hidden layer. Neuron of hidden layer adds input signal $x_i$along with weights $w_{ji}$ of respective connections from input layer.

The output$y_j$ function is$y_j = f (\Sigma w_{ji} x_i)$

where f is a simple threshold function such as sigmoid or hyperbolic tangent function[6].

Multi-Layer Perceptron Neural community(MLPNN):

The analysis unveils a continual utility of feed forward neural networks, from the diverse categories of connections for artificial neurons. A sort of feed ahead neural network mechanism is the Multi-layer Perceptron Neural Networks (MLPNN). The shape of MLPNN is proven in Fig.I

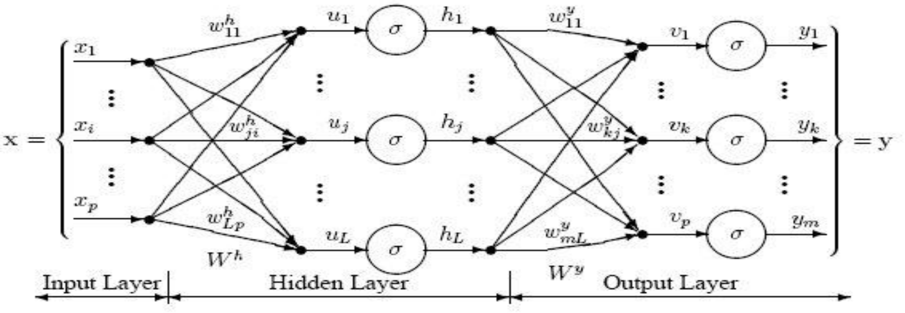

Fig.I

In MLPNN the primary challenge of the neurons within the input layer is the division of the input sign $x_i$ amongst neurons within the hidden layer. Every neuron j inside the hidden layer provides up its input indicators $x_i$ once it weights them with the strengths of the respective connections $w_{ji}$ from the enter layer and determines its output $y_i$ as a characteristic f of the sum, given as

$$y_i=f(\Sigma w_{ji} x_i) \qquad (1)$$

At this immediately it is feasible for f to be a simple threshold feature which includes a sigmoid, or a hyperbolic tangent function. The output of neurons in the output layer is decided in an equal style. The operating of Multi-Layer Perceptron Neural community is summarized in steps as stated under:
1) input records is provided to input layer for processing, which produces a expected output.
2) The anticipated output is subtracted from actual output and error price is calculated.
three) The community then uses a returned-Propagation algorithm which adjusts the weights.
four) For weights adjusting it begins from weights between output layer nodes and ultimate hidden layer nodes and works backwards via network.
5) When returned propagation is finished, the forwarding method begins again.
6) The process is repeated until the error between anticipated and actual output is minimized[6].

## 3. CONCLUSION

Health care industry today generates large amounts of complex clinical data about patients and other hospital resources. Data mining techniques are used to analyse this collection of data from different perspectives and derive useful information. Through this paper we have made a study on effective ways to predict heart diseases. In future, we would try to implement these concepts to predict the effectiveness and accuracy in heart disease prediction.

## 4. REFERENCES

[1] B.Venkatalakshmi, M.V Shivsankar.Heart Disease Diagnosis Using Predictive Data mining. International Journal of Innovative Research in Science, Engineering and Technology Volume 3, Special Issue 3, March 2014 2014 International Conference on Innovations in Engineering and Technology (ICIET'14).

[2] Jaymin Patel, Prof.TejalUpadhyay, Dr. Samir Patel.IJCSC 0973-7391,Volume 7,Number 1 sept 2015,March 2016 pp.129-137.Heart Disease Prediction Using machine learning and Data Mining Technique.

[3] Aditya Methaila, Prince Kansal,Himanshu Arya, Pankaj Kumar.Early Heart Disease Prediction Using Data Mining Techniques .

[4] N.AdityaSundar , P. PushpaLatha , M. Rama Chandra.PerformanceAnalysis Of Classification Data Mining Techniques Over Heart Disease Data Base . ADITYA SUNDAR* et al. ISSN: 2250–3676 [IJESAT] International Journal Of Engineering Science & Advanced Technology Volume-2, Issue-3, 470 – 478 IJESAT | May-Jun 2012 Available Online @ Http://Www.Ijesat.Org 470.

[5] M.A.NisharaBanu, B Gomathy,Disease Predicting System Using Data Mining Techniques.International Journal of Technical Research and Applications e-ISSN: 2320-8163, www.ijtra.com Volume 1, Issue 5 (Nov-Dec 2013), PP. 41-45.

[6] A. T. Sayad, P. Halkarnikar.88 Diagnosis Of Heart Disease Using Neural Network Approach .International Journal of Advances in Science Engineering and Technology, ISSN: 2321-9009 Volume- 2, Issue-3, July-2014.

[7] SellappanPalaniappan,RafiahAwang.Intelligent Heart Disease Prediction System Using Data Mining Techniques. IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.8, August 2008 343 Manuscript received August 5, 2008 Manuscript revised August 20, 2008.

[8] Abhishek Taneja. Oriental Journal Of Computer Science & Technology www.computerscijournal.org ISSN: 0974-6471 December 2013, Vol. 6, No. (4): Pgs. 457-466 An International Open Free Access, Peer Reviewed Research Journal Published By: Oriental Scientific Publishing Co., India.

[9] Fayyad, U.M., G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy., Advances in Knowledge Discovery and Data Mining, (AKDDM), AAAI/ MIT Press, Massachusetts (1996).