International Journal of Advanced Research in IT and Engineering

ISSN: 2278-6244
Impact Factor: 6.111

Comparative Study Based on Data Mining Techniques for Predicting Heart Disease

Ashwini Shetty A, PG Student (CSE), NMAM Institute of Technology, Karkala, India

Chandra Naik, Assistant Professor, Dept. of CSE, NMAM Institute of Technology, Karkala, India

Abstract— Data mining techniques are widely used in health care industry for determination of various diseases because of their ability to discover hidden pattern, effective decision making and for identification of relation. In this paper, we have studied data mining techniques like a Neural Network and Genetic Algorithm, which play an essential role in identifying the warning sign of heart diseases. Here first we depict the fundamental concept of the algorithms and study these algorithms how it is used in heart disease prediction.

Keywords— Data Mining, Genetic Algorithm, Heart Disease, Neural Network.

I. INTRODUCTION

"Data mining is technique of extracting useful information from large data source". Data mining is nothing but, analyze data and gathering useful information from it. Patterns that are hidden in data can be discovered by using different data mining methods. The discovered information used in healthcare industry requires lot of accuracy and uncertainty. Better performance with high accuracy can be achieved by combination of different machine learning algorithms.

Millions of people are dying due to cardiac disease every year in different countries due to mental stress, eating practice, lack of exercise and sleep, work load and consumption of alcohol etc. Since the diagnosis associated with this disease is complex, the major task to be efficiently and accurately executed. Heart disease has turn out to be a major cause of deaths globally; because this caused more deaths compared to any other diseases.

Nowadays data mining plays major role in medical field in predicting different diseases. Due to various reason people suffer from diseases of same sort so doctors cannot predict right disease which often lead to unwanted biases and unnecessary medical bill of treatments. Generally diagnosis is based on physician's experience. In this survey, different Data Mining concepts are discussed. This survey is based on neural network and genetic algorithms for predicting of heart disease.

II. LITERATURE SURVEY

Several studies have done that focus on prediction of heart disease and different data mining approaches applied for study and prediction of disease, and it is achieved by applying different techniques.

- Shatakumar B Patil et al.[1] develops system were data mining concepts such as Neural Network, k-means clustering MAFIA algorithms are used for obtaining and mining hidden patterns for classifying heart disease.
- K.Srinivas et al.[2] builds system were data mining concepts such as Neural Network, bagging, Decision tree and support vector machine are applied for prediction and classification of disease.
- Shulabha S. Apte et al.[3] proposed system to evaluate the prediction system for heart diseases by increasing input attribute and comparing performance of the data mining concepts like Naive Bayes, J48 algorithm and Neural Network.
- T John Peter et al.[4] proposed system using classification technique to detect relation and interaction between the variables in data source and to achieve this, classifiers like Naïve Bayes,Neural Network and KNN is used.
- M Akhil Jabbar et al.[5] developed a system to identify heart disease using techniques like Neural Network, feature subset selection and to reduce attributes PCA is used which decreases the irrelevant tests of patients.
- Kiyong Noh et al.[6] proposed a system where multiparametric feature extracted using classification technique for assessing heart rate variability from data pre-processing and ECG. For this experiment data sets are divided into two groups i.e. normal people and with heart disease.
- Sivagowry S et al.[7] develops a system where Neural Network and Particle Swarm Algorithm are applied before and after estimating efficiency by using confusion matrix and ROC curve.
- Bhuvaneswari Amma N.G[8] proposed system which classifies the chances of heart disease by using the Back-propagation algorithm which is used to train the data because of which the mean square error is decreased.
- A H Chen et al. [9] proposed system by using data mining concepts for attribute selection and based on attributes are neural network to identify the heart disease.

III. BASIC CONCEPTS

A. NEURAL NETWORK

An Artificial Neural Network is also called as "Neural Network". Behavior of Artificial Neural Network is same as biological neural network i.e. dense collection. It consist three layers i.e. input layer, hidden layer and output layer. There are connections between each layer and for every connection weight (w) is assigned. This network computes a series of transformation between their input (x) and output (y). So each layer it get new representation of the input in which things work similar in the previous layer might become less similar or things which are dissimilar in the previous layer might become more similar. In order to achieve we need the activities of neuron in each layer to be non-linear function of the activities in layer below. Structure of Neural Network is shown in Figure 1 below.

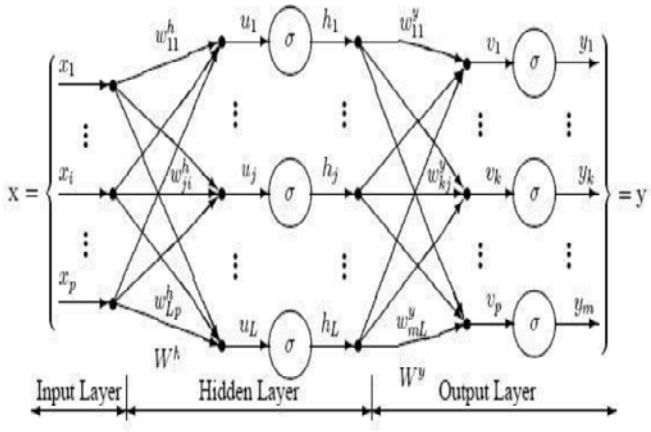

Figure 1: Structure of Neural Network [3]

Levenberg-Marquardt back propagation algorithm used for training and testing.trainlm is fastest back propagation algorithm.trainlm is a network training function. This function will be used to update weight and bias value. trainlm is rewarded as the fastest back propagation algorithm in the toolbox. It is one of the most proposed algorithms as first choice supervised algorithm, even though it requires less space than other algorithms. The neural network trained by using methods like a) Supervised training: In Supervised learning, gain information from labeled response. Output dataset is provided, which is used to train the machine and get desire output. b) Unsupervised training: In Unsupervised learning task, predict output from input dataset with unlabeled data. Here no dataset is provided, instead data is clustered.

B. GENETIC ALGORITHM

Genetic algorithm is connected to natural evolution process. It follows the "survival of the fittest" concept. It works on the human genetics and reproduction process. The algorithm will include population which is the set of solutions. These solutions are taken from one population and used in the other population. Solutions are taken from new off spring through the selection based on fitness criteria. That is, the more fit they are the more chances of reproduction. Individuals which are generated randomly are evaluated or inspected at first. Fitness is calculated for each individual and every generation. The individual is selected based on their fitness values and is modified to form a new population. Genetic Algorithm will be terminated when maximum generations are produced or fitness is reached to satisfactory level. Flow of genetic algorithm is shown in Figure 2 below.

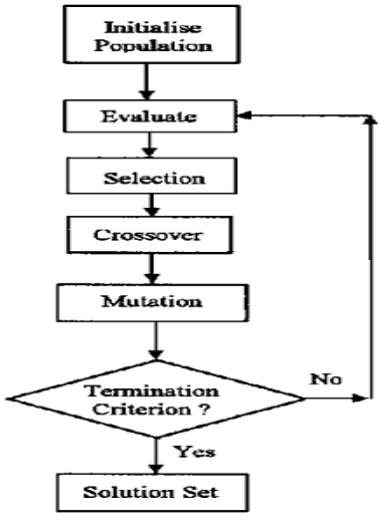

Figure 2 : Flow of Genetic Algorithm

IV. ARCHITECTURE OF PREDICTION SYSTEM

Benchmarked data sets are used [10]. 13 risk factors including the age of the person, gender of the person, blood sugar level of the person, blood cholesterol level, current blood pressure etc. The TABLE 1 shows the risk factor and its values which are used as inputs.

TABLE 1: Description of 13 Attributes

	Attributes	Values
1	Gender	Male=1,Female=0
2	Age	>79=2,61-78=1,51-60=0,35-50=-1, 20-34=-2
3	Resting heart rate	In mm-hg
4	blood pressure	In mm-hg[low=-1,normal=0,high=1]
5	cholesterol	In mg/dl[low=-1,normal=0,high=1]
6	blood sugar	Yes=1,No=0
7	exercise induced angina	Yes=1,No=0
8	resting electrocardiographic result	Normal=0,ST-T wave abnormality=1 left ventricle hypertrophy=2
9	Chest pain type	Typical angina=4,Atypical angina=3, Non angina=2,Asymptomatic=1
10	Slope	Upsloping =1,Flat=2,Downsloping=3
11	number of color vessels by fluoroscopy	Values 0-3
12	thal	Normal=3,Fixed defect=6,Reversible defect=7
13	ST depression	Yes=1,No=0

Analysis of data is considered to be critical step, since excessive noise will be present in the database. Data cleaning and data integration need to be performed, finds

the missing values and remove excessive data. These missing fields will lead to incorrect outputs. The best attribute is found by applying various techniques and algorithms. Some of the algorithms used are the hybrid algorithms taken from genetic algorithm or neural network. Input data will be trained by various techniques of data mining. Useful information could be derived from this. It will then predict if the patient could be affected by heart disease in the future or not.

V. CONCLUSION

In this paper, different data mining approaches have been discussed. The main perspective is to study different algorithms and evaluate them to find out the best method of heart disease prediction. The main techniques namely Neural Networks and genetic algorithm are focused and the objective of each technique is to achieve high accuracy and better performance in prediction of Heart Disease. Different data mining techniques such as clustering, association etc can also incorporated for future work.

REFERENCES

[1] Shatakumar B Patil, Y S Kumaraswamy, "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network", European Journal of Science Research, Volume 31, No.4, 2009, pp.642-656.

[2] K. Srinivas, B. K. Rani, and D. A. Govrdhan, "Application of data mining techniques in healthcare and prediction of heart attacks", International Journal on Computer Science and Engineering, vol. 2, no. 2, 2011, pp. 250–255.

[3] Shulabha S.Apte, Chaitrali S.Dangare, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", International Journal of Computer Applications (0975 – 888) ,Volume 47,No.10, June 2012,pp.44-48.

[4] T John Peter,K Somasundaran "An Empircal Study on Prediction of Heart Disease using Classification Data Mining Techniues", IEEE-International Conference on Advances in Engineering, Science And Management,2012.

[5] M Akhil Jabbar, B LDeekshatulu ,Priti Chandra, "Classification of Heart Disease using Artificial Neural Network and Feature Subset Selection", Global Journal of Computer Science and Technology Neural & artificial Intelligence, Volume 13, Issue 3, Version 1.0,2013

[6] Kiyong Noh, Heon Gyu Lee, Ho-Sun Shon, Bum Ju Lee, and Keun Ho Ryu, "Associative Classification Approach for Diagnosing Cardiovascular Disease", Springer, Vol:345, 2006, pp: 721- 727.

[7] Sivagowry S, Dr. Durairaj M, "PSO –an Intellectual Technique for Feature Reduction on Heart Malady Anticipation Data",International Journal of Advanced research in Computer Science and Software Engineering,Volume 4, Issue 9, 2014.

[8] Bhuvaneswari Amma N.G, "Cardiovascular Disease Prediction System using Neural Network",IEEE.

[9] AH Chen,SY Huang,PS Hong,CH Cheng,EJ Lin, "HDPS:Heart Disease Prediction System",10 Department of Medical Informatics,Tzu Chi University,Hualien City,Taiwan.

[10] Statlog-http://archive.ics.uci.edu/ml/machine-learning- databases/statlog/heart/