

# Hybridization of Data Mining Techniques for Prediction of Heart Disease

Shruti A. Darekar<sup>1</sup> Adina A. Dingankar<sup>2</sup> Praharsh Palanki<sup>3</sup> Pallavi Chitte<sup>4</sup>

<sup>1,2,3,4</sup>Department of Computer Engineering

<sup>1,2,3,4</sup>Ramrao Adik Institute of Technology Nerul, Navi Mumbai, India

**Abstract**— Medical Diagnosis Systems play a vital role in medical practice and are used by medical practitioners for diagnosis and treatment. In this system, a medical diagnosis system is presented for predicting the risk of cardiovascular disease. This system is built by combining the relative advantages of neural network and genetic algorithm. Artificial neural networks are particularly suited for complex classification problems. The weights of the neural network are determined using genetic algorithm because it finds acceptably good set of weights in less number of iterations. The dataset is used for training and testing. First, the dataset is pre-processed in order to make them suitable for training. Then, genetic based neural network is used for training the dataset. The final weights of the neural network are stored in the weight base and are used for predicting the risk of cardiovascular disease. Later, the system generates the charts which helps the people to know what further precautions are to be taken or followed after the prediction of the disease. The chart is generated on the basis of the results which is predicted after testing of the dataset.

**Keywords:** Artificial Neural Network, Back Propagation Algorithm, Genetic Algorithm, Data Mining, Heart Disease Prediction

## I. INTRODUCTION

The one of the most common reasons of death in India or other Asian countries is Heart Disease. In 2003, approx. 17.3 million people died all over the globe and out of that 10 million were due to coronary heart disease. The factors which are responsible for the risk of having a heart problem are smoking, alcohol, obesity, high blood pressure etc. However the introduction of artificial intelligence and medical science has helped to prevent all kind of diseases. For making a good decision, machine learning is used in extraction of relevant data from large databases which are available in hospitals. There are many kind of classification techniques such as decision trees, k-means, J48 etc. and all these weak classifiers need to be enhanced to improve their performances. In this paper, various kinds of techniques has been applied to predict the heart diseases and a proposed methodology using genetic algorithm and artificial neural network can be implemented in future so as to get efficient accuracy in less iterations and with least error. Thus proposed system will be implemented in java.

## II. LITERATURE SURVEY

Ankita Dewan and Meghna Sharma, "Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification", 2015.

The main objective of this paper is to set a standard which can not only determine and extract the unknown knowledge i.e. patterns and relations but also it retrieves the data related to heart disease from a past heart disease database record. It also helps in solving the complicated queries required for detecting the heart disease and hence it helps the

medical practitioners for making smart clinical decisions more accurately and precisely which was perhaps not possible with the traditional systems. Also with the help of efficient treatments it helps to reduce its cost [1].

Dr.K.Usha Rani , "Analysis of Heart Diseases Dataset using Neural Network Approach", International Journal of Data Mining & Knowledge Management Process (IJDMP) Vol.1, No.5, September 2011.

This paper emphasizes on the study of neural networks and its various techniques. The neural network consists of processing elements, also called as neurons, which when assembled in a closed interconnected network exhibits few features of the biological neural network. The main feature of neural network is iterative learning wherein all the data cases are presented to the network turn by turn and also the weights associated with the input values are adjusted each time. Neural networks is also well known for producing accurate results in the field of medical science & real world applications. Thus this Classifier is also combined with other data mining techniques to improve the performance [2].

Nilakshi P. Waghulde and Nilima P. Patil, "Genetic Neural Approach for Heart Disease Prediction", International Journal of Advanced Computer Research Vol.4, No.3, Issue-16, September-2014

In this paper, a Heart Disease Prediction System is developed using Neural Network and Genetic Algorithm. This system computes the number of hidden nodes for neural network which train the network with proper selection of neural network architecture and uses the global optimization of genetic algorithm for initialization of neural network. Then the genetic algorithm fitness function is used to predict the heart disease. For prediction, this system uses 12 parameters such as sex, age, blood cholesterol etc. From the result, it is found that genetic neural approach predicts the heart disease up to 98% accuracy [3].

Harshal Yeole, Sayali Ukirde, Sushma Khadse, and Priyanka Pedneka, "Prediction of Heart Disease Using BP-Neural Network & Genetic Algorithm", International Journal of Research in Advent Technology Special Issue National Conference "NCPCI-2016", 19 March 2016

In this paper, two algorithms are used namely Back Propagation Neural Network (BPNN) and Genetic Algorithm (GA). By using this algorithm on the dataset, the Heart Disease can be predicted with the reducing optimal attributes on Dataset with attributes. This scheme plans to sketch and generate discovery and expectation framework for heart illnesses given perceptive mining. A number of judgments have been headed to look at the implementation of different information quarrying methods including neural network and Genetic algorithm. This system uses a 13 attribute ordered medical database which is used as a foundation information [4].

Richa garg ,saurabh mittal, " Optimization ny genetic algorithm",International Journal of Advanced

Research in Computer Science and Software Engineering", Volume 4, Issue 4, April 2014.

This paper emphasizes on the complete working of genetic algorithm. Genetic algorithm is also one of the search and the optimization algorithms which works on the principles of natural evolution. Genetic algorithm also help in developing the various optimization strategies by distinguishing the existence of species through natural selection process. Genetic algorithm consists of two main processes out of which first process relates with the selection of the individual chromosome for the generation of new population whereby the second process relates with the manipulation of the selected chromosomes to form the best global population using the crossover and mutation techniques at every iteration respectively...Thus genetic algorithm works properly when used with any of the classifiers such naive Bayes, neural network etc....in order to give the accurate results with greater efficiency[5].

Preeti Gupta and Bikrampal Kaur, "Accuracy Enhancement of Artificial Neural Network using Genetic Algorithm", International Journal of Computer Applications, Vol.103, No 13, October 2014

This paper proposes the enhancement of the accuracy of the results by using Artificial Neural Network optimized with Genetic Algorithm in prediction of heart disease diagnosis. Here, neural network is optimized with Genetic Algorithm and proved experimentally. The proposed method achieved an accuracy of 97.83%. With this higher achieved accuracy, the heart disease can be diagnosed more accurately and much proper treatments can be suggested [6].

### III. PROPOSED METHODOLOGY

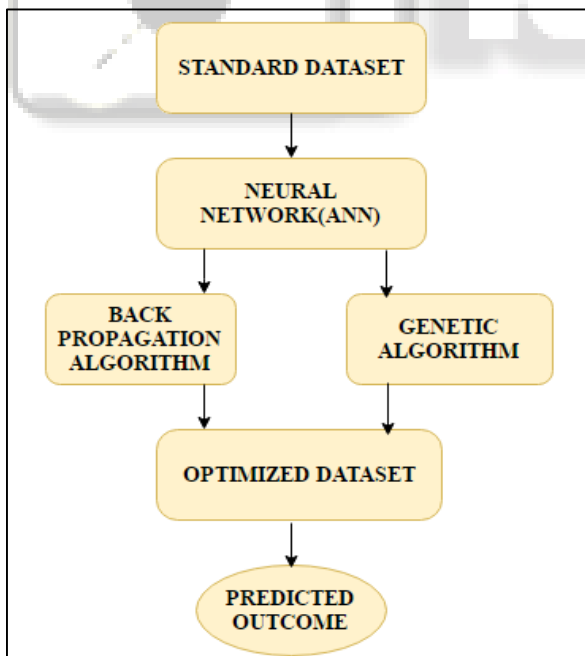


Fig. 1: Architecture Framework

Initially the data set is selected on which the prediction is meant to be performed and this data set is perhaps the standard dataset taken from the UCI repository for medical science. The second stage is data cleaning or pre-processing, in this stage the entire data set is scanned for the missing values and then the data set is processed to remove the

missing values. The third stage is Min-Max Normalization, here the missing value is replaced by the most probable value in the format of (#####). The fourth stage is Training and Prediction, this stage is divided into two phases weight optimization and prediction result which works simultaneously. In weight optimization a random population is selected, then the weights for the individual population is calculated and stored which is termed as best global population. This is where the genetic algorithm works hand in hand with the artificial neural network by back propagating at every new iteration of the calculating random population. Later, for prediction of results the weights of the best global population is then selected for testing as mentioned above in the Data Mining Techniques. Thus after the prediction performed on the dataset used earlier we get the accuracy in percentage given by the following formula:

$$\text{accur\%} = (\text{numCorrect} * 1.0) / (\text{numCorrect} + \text{numWrong})$$

Thus the result is then displayed in the graphical format either using pie chart or bar chart followed by the preliminary food links to cure the measure of heart disease or to take the precautions.

### IV. DATA MINING TECHNIQUES

There are various data mining techniques which are very helpful in analysing the data and data extraction of complex medical data. Many other practitioners of medical science uses these data mining techniques in different fields such as to detect cancer and stroke. To generate more optimized and efficient output the researchers have tried applying different techniques of machine learning like Artificial Neural Network, Back Propagation Algorithm and Genetic Algorithm.

#### A. Various Data Mining techniques are:

##### 1) Neural Network

Neural Networks are very useful in identifying and training very complex data in proper manner. One of the important characteristic of neural networks is that it tends to have a very high capacity of generalizing the input data and leads in generating the correct and accurate outputs for unfamiliar data. Therefore the problems of solving complex classification is effectively managed.

Artificial Neural Network (ANN) is one of the effective technique of Neural Network used for data mining in medical field. Generally, where the creation of a firm mathematical model is impossible or difficult but if there are enough representative set of samples then the usage of ANN's are well suited. This consists of three layers i.e. input layer, output layer and hidden layer where these layers are interconnected with respect to some weights. Here weights are calculated by the weight calculating function.

$$w = (iL * hL) + (hL * oL) + hL + oL$$

where, w is Weight

iL is input layer value

oL is output layer value

hL is hidden layer value

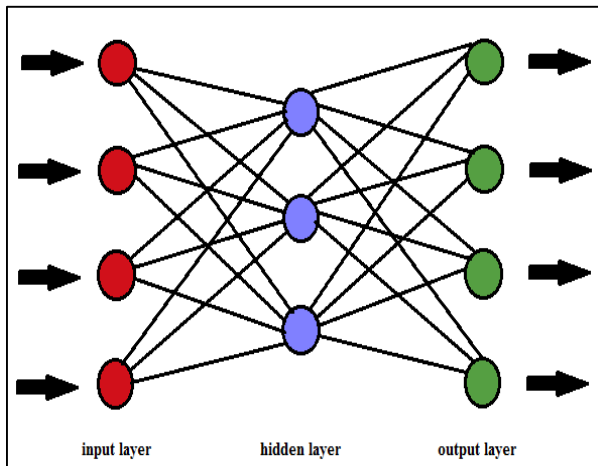


Fig. 2: Artificial Neural Network

Back Propagation Algorithm is a method of ANN. This algorithm is said to be the best predictive algorithm in neural network classification. In this algorithm the data and the targeted output has a non-linear relationship amongst them. The characteristic of this algorithm is that it is adaptive as well as tolerant to noisy data which is present in the dataset. Back Propagation algorithm consists of two steps ie firstly feed forwarding the values in network and secondly calculating errors and propagating it back. The following flowchart describes the steps included in this algorithm (fig 3):

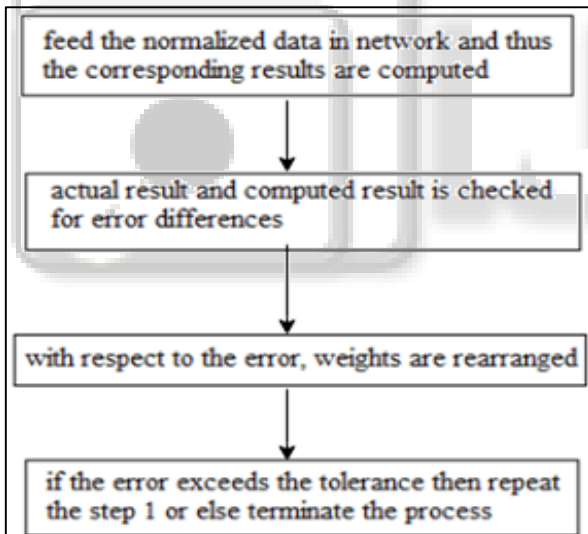


Fig. 3: Steps involved in Back Propagation algorithm

The following fig. 4 represents the attributes of Cleveland database.

Data Representation:

Number of attributes: 13 and one class attribute

Number of instances: 303

Category of class:

C0: normal person

C1: 1<sup>st</sup> stroke

C2: 2<sup>nd</sup> stroke

C3: death

Attribute	Description	Range
Age	Age in years	Continuous
Sex	(1=male; 0=female)	0,1
Cp	--Value 1:typical angina 1,2,3,4 --Value 2: atypical anginal --Value 3: non-anginal pain --Value 4: asymptotic	
trestbps	Resting blood pressure(in mm Hg)	Continuous
chol	Serum cholesterol in mg/dl	Continuous
fbs	(Fasting blood sugar .120mg/dl) (1=true; 0=false)	0,1
restecg	electrocardiography results --Value 0: normal --Value 1:having ST-T wave abnormality (T wave inversions and/or ST Elevation or depression of >0.05mV) --Value 2:showing probable or definite left ventricular Hypertrophy by Estes' criteria	0,1,2
Thalach	Maximum heart rate achieved	Continuous
Exang	Exercise induced angina(1=yes;0=no)	0,1
Oldpeak	ST depression induced by exercise relative to rest	Continuous
Slope	The slope of the peak exercise ST segment Value 1: up sloping Value 2: flat Value 3:down sloping	1,2,3
Ca	Number of major vessels (0-3) Colored by fluoroscopy	Continuous
Thal	Normal, fixed defect, reversible defect	3,6,7

Fig. 4: Attribute Description

## 2) Genetic Algorithm

Neural network is individually capable of predicting the data. The over training of data is one of the major disadvantage of neural network i.e. during the process of training the data there can be a possibility of occurring minute errors which may lead to improper prediction of data. So as to avoid this and to minimize the effect of the above issue we make use of one of the searching algorithms called as Genetic Algorithm which searches an optimal solution for the problem. Hence, when these two data mining techniques i.e. Artificial Neural Network and Genetic Algorithm when combined together leads to increase the accuracy of prediction. Both of these algorithms when combined together undergo various different phases which is well described in below flowchart (fig 5).

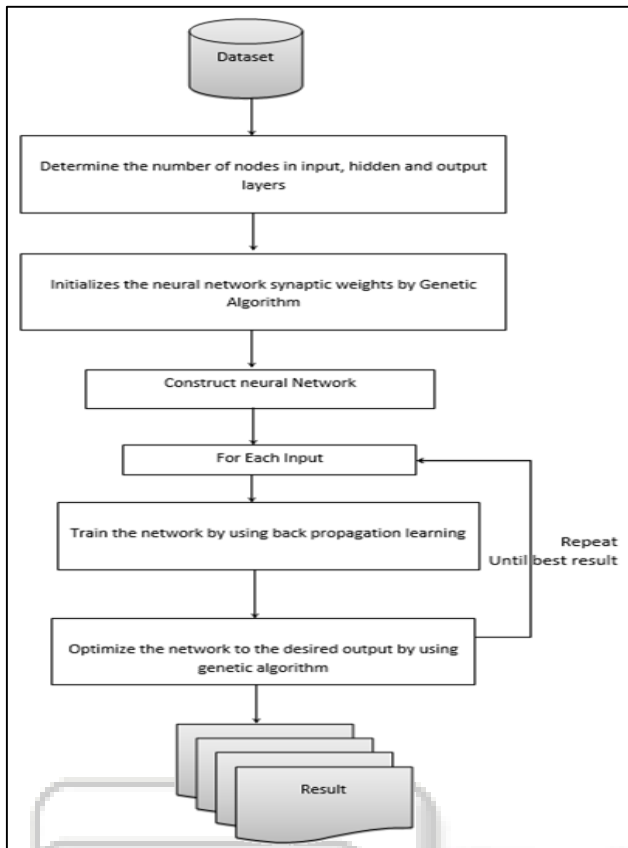


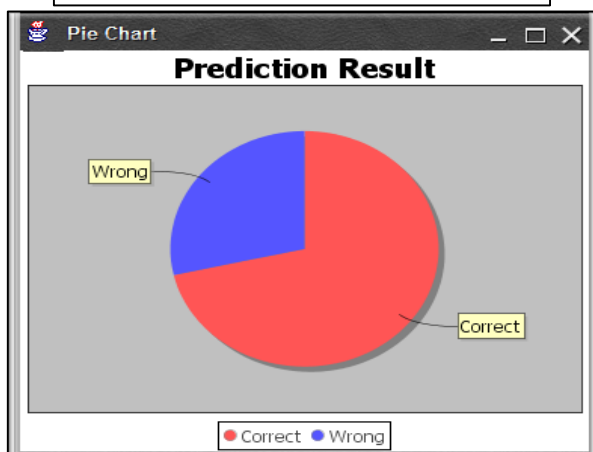
Fig. 5: Flowchart of the proposed system

## V. RESULT ANALYSIS

Original Outcome	Predicted Outcome	Result
2	2	c
0	0	c
0	0	c
0	0	c
3	0	w
0	3	w
0	0	c

Correct Results(c) count = 5  
Wrong Results(w) count = 2

Accuracy % =  $(5/7) = 71.429$



## VI. CONCLUSION

The Heart Disease Prediction system is meant to contribute a lot to the health and care of the society, thereby decreasing the rate of mortality even in youths and middle age people. Prediction of heart disease system can be an upper hand in the medical field by helping them to take the yearly medical statistics and comparing them with the previous one to show how the current prediction is accurate than the previous one. And our system will thereby also provide suggestions with the diet chart for the patients on the basis of the various parameters compared.

## REFERENCES

- [1] Ankita Dewan and Meghna Sharma, "Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification", 2015
- [2] Dr.K.Usha Rani , "Analysis of Heart Diseases Dataset using Neural Network Approach", International Journal of Data Mining & Knowledge Management Process (IJDMP) Vol.1, No.5, September 2011
- [3] Nilakshi P. Waghulde and Nilima P. Patil, "Genetic Neural Approach for Heart Disease Prediction", International Journal of Advanced Computer Research Vol.4, No.3, Issue-16, September-2014
- [4] Harshal Yeole, Sayali Ukirde, Sushma Khadse, and Priyanka Pedneka, "Prediction of Heart Disease Using BP-Neural Network & Genetic Algorithm", International Journal of Research in Advent Technology Special Issue National Conference "NCPCI-2016", 19 March 2016
- [5] Richa garg ,saurabh mittal, " Optimization ny genetic algorithm", International Journal of Advanced Research in Computer Science and Software Engineering", Volume 4, Issue 4, April 2014.
- [6] Preeti Gupta and Bikrampal Kaur, "Accuracy Enhancement of Artificial Neural Network using Genetic Algorithm", International Journal of Computer Applications, Vol.103, No 13, October 2014.