



Heart Disease Prediction System Using Data Mining

Shilpa M.Satre¹, Sunaina Manohar Bhagat², Shalini Manoj Thakur³
Professor¹, BE Student^{2,3}

Department of Information Technology
Bharati Vidyapeeth College of Engineering, India

Abstract:

There Abstract—Medical Diagnosis Systems play a vital role in medical practice and are used by medical practitioners for diagnosis and treatment. In this project, a medical diagnosis system is presented for predicting the risk of cardiovascular disease. This system is built by combining the relative advantages of genetic algorithm and neural network. Multilayered Back Propagation neural networks are particularly suited to 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. First, the preprocessing of data is done in order to make them suitable for training. Genetic- Neural network is used for training the system. The final weights are stored in the weight base and then, it is used for predicting the risk of cardiovascular disease. The classification accuracy obtained using this approach is 90.17%.

Keywords: Genetic-Neural Algorithm; Neural Network, Neural-Back propagation Algorithm; Cardiovascular Disease; Prediction Engine

I. INTRODUCTION

The one of the most common reasons of death in India or other Asian countries is Heart Disease. In 2003, approximately 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 the proposed system. 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: $\text{accuracy}\% = (\text{number Correct} * 1.0) / (\text{number Correct} + \text{number Wrong})$ 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. 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.

I.1 Motivation

The word itself implies that something of the proposed system must be beneficial to the society. As the prediction of heart diseases isn't a process known in its early stages just leading to mortality in big numbers. So this system proposes the idea of predicting the disease in its premature stage thus leading to saving the lives of people and thereby helping the society for their betterment. This system works by providing the various food links on the basis of the prediction made within the comparisons of the parameters of the trained set and the testing set with the help of various algorithms like Naive Bayes, classifier model and neural networks respectively. Thus this system is the proposal put by previous papers emphasizing on the incomplete work and working with few improvements.

I.2 Scope

As the healthcare domain is dynamic and it is a challenging issue to the data mining. In healthcare sector of data mining application, it is a forcing motivation. This vitality gives way to new heights and more data mining applications will be employed to discover new patterns and associations. Distributed data mining applications and text mining algorithms achieved a considerable weight. With the help of data mining algorithms, the classification performance increases. This can be further enhanced and expanded with more prediction algorithm for major life threatening diseases.

II. LITERATURE SURVEY

Intelligent Heart disease Prediction System Using Data Mining Techniques [1]. 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.

Classification and Prediction of heart Disease risk using data mining techniques of support vector machine and artificial neural network [2].This paper emphasizes on the study of neural networks and Support vector machine. 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 combined with SVM to improve the performance.

III. SYSTEM ARCHITECTURE

The training and test dataset are loaded in the system where its being pre-processed. The data gets cleaned after that and the unwanted data is removed. Then for testing the data is modeled using neural network and genetic algorithm. At last after running the algorithms on the data the prediction result is obtained.

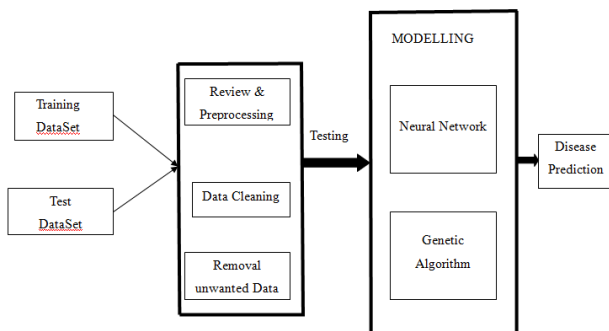


Figure.1. Architecture Diagram

IV. PROPOSED SYSTEM

This hybrid system uses back propagation algorithm for learning and training the neural network. The Multi-Layered Genetic-Neural Network is optimized to minimize the over fitting by calculating the number of nodes in hidden layer; which causes the overestimation of complexity of the resulting problem that leads to significant diversion in prediction process. As the initialization of the Genetic-Neural Network weights is a unseen process which makes it hard to find out globally optimized initial weights and the output of the network would run towards local optima hence the overall tendency of the network to find out a global solution is greatly affected. So the problem of local optimum solution is solved by optimizing the initial weights of neural network. For this a genetic algorithm is used which is specialized for global searching. Thus the system uses the back propagation algorithm to train the network by using the weights optimized by Genetic Algorithm. Error is calculated using equation to measure the differences between desired output and actual output that has been produced in feed forward phase. Error

then propagated backward through the network from output layer to input layer as represented below. The weights are modified to reduce the error as the error is propagated.

$$\text{Error} = 1/2(\text{Output desired} - \text{Output actual})^2$$

This process will be repeated iteratively until convergence is achieved (targeted learning error or 15 maximum number of iteration). The Genetic-Neural Network Approach for Heart Disease Prediction is being used for testing data to the desired value and predict whether the patient have a cardiovascular disease or not.

V. METHODOLOGY

IV.1 Neural Network

Neural Networks are very useful in classifying and training very intricate 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

IV.2 Back-Propagation Algorithm

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 i.e. 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.

V. CONCLUSION

The Heart 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 as well thereby making it efficient for the common people to get to know about their health status as the common people procrastinate to get their health checkup on a periodic basis .And our system will thereby also provide suggestions with the diet chart for the patients on the basis of the various parameters compared.

VI. REFERENCES

- [1]. Meghna Sharma and Ankita Dewan, "Prediction of Heart Disease Using a Hybrid Technique in Data Mining ", 2015.
- [2]. Dr.K.Usha Rani, "Analysis of Heart Diseases Dataset using Artificial Neural Network Approach", International

[3]. Nilima P. Patil and Nilakshi P. Waghulde, "Genetic-Neural Network System for Heart Disease Prediction", International Journal of Advanced Computer Research Vol.4, September-2014.

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

[5]. Preeti Gupta and Bikrampal Kaur, "Accuracy Enhancement of Neural Network using Genetic Algorithm", International Journal of Computer Applications, Vol.103, October 2014.

[6]. L. Hongjun, R. Setiono and H. Liu, "Effective Data Mining Using Artificial Neural Networks", IEEE Trans. On Knowledge and Data Mining, vol.8, Dec 1996.

[7]. P. Bajpai and D. M. Kumar, "Neural-Genetic Algorithm: An Approach to Solve Global Optimization Problem", vol. 1, pp 199-206, 2010.

[8]. A. Rangel Merine, L. L. Bonilla, R. Linares and Y. Miranda, "Optimization of DataSet based on Genetic Algorithm Approach", Oct 2005.