Prediction of Heart Diseases using Artificial Intelligence

Amreen Ahmed[1], Prof. Shilpa Verma[2]

Student, Computer Engineering, TSEC, Mumbai, India [1]

Asst. Professor, Computer Engineering, TSEC, Mumbai, India [2]

Abstract: Most nations face high and expanding rates of heart diseases or cardiovascular disease. There are number of factors which increase risk of heart diseases, like family history of heart disease, smoking, cholesterol, high blood pressure, obesity, lack of physical exercise etc. Heart disease is a major health problem in today's time. A major challenge facing healthcare organization is the provision of quality services at affordable costs. Thus, there arises a need to develop a decision support system for detecting heart disease of a patient. That is, to achieve a correct and cost effective computer-based treatment and support system that can be used to make good decision. Many hospitals use hospital information systems to manage their healthcare or patient data. These systems produce huge amounts of data in the form of images, text, charts and numbers. Sadly, this data is rarely used to support the medical decision making. There is a bulk of hidden information in this data that is not yet explored which give rise to an important query of how to make useful information out of the data. So there is necessity of creating a system which will help practitioners predict the heart disease before it occurs. By providing efficient predictions, it can help to reduce costs of treatment. A hybrid system of Back-Propagation algorithm for neural network and genetic algorithm is proposed. The characteristics of Back-Propagation algorithm are that it is adaptive and tolerant towards the noisy data or other outliers present in the medical data. Back propagation networks do not need the linear relationship between the data and the target output. But it has a disadvantage of getting stuck in the local minima and thus the data is first optimized using genetic algorithm and the concepts of crossover and mutation are applied and that is then again fed into the neural network for better results. In the case of heart diseases time is precious, proper diagnosis at the right time saves life of many patients. The system can be considered assisting the doctor to come to decision making.

Keywords: Neural Network, Genetic Algorithm, Crossover, Mutation, Back-Propogation.

I. INTRODUCTION

Heart disease refers to the class of diseases that involve the heart or blood vessels; it is usually used to refer to those related to atherosclerosis. Heart diseases include coronary heart disease, cardiovascular disease, raised blood pressure, peripheral artery disease, rheumatic heart disease, congenital heart disease and heart failure. In practice, heart disease is treated by cardiologists, thoracic surgeons, vascular surgeons and interventional radiologists, depending on the organ system that is being treated. The heart is the organ that pumps blood to all tissues of the body. If the pumping action of the heart becomes inefficient, vital organs like the brain and kidney suffer and if the heart stops working, death occurs within minutes. There arises a need to develop a decision support system for detecting heart disease of a patient. Every hospital has abundant database containing the details of patients suffering from heart diseases in which knowledge lies unexplored, this database can be effectively mined and utilized so that it can be helpful for the medical practitioners to predict the probability of a person suffering from heart disease saving a lot of precious time and money i.e. make an efficient computer based-decision support system. The automation of this system is very much needed to help the physicians to do better diagnosis and treatment. Heart disease patients have lot of these visible risk factors in which can be used very effectively for diagnosis. System based these risk factors would not only help medical professionals but it would give patients a warning about the probable presence of heart disease even before he visits a hospital or goes for costly medical checkups.

II. PROPOSED METHOD

Heart being one of the most important parts of the body, heart diseases turn to be a silent killer that is it is generally not detected till a person actually suffers from heart attack, causing a lot of deaths. Hospitals have huge amount of data but it is not utilized properly.
So, a novel technique is proposed that combines the various techniques of machine learning such as artificial neural network with back propagation algorithm and genetic algorithm and then develop a decision support system for detecting the heart conditions of a patient.

Data mining techniques are helpful in extracting & analyzing the complicated medical data using various kinds of techniques. Practitioners of medical science are also using these techniques in other fields such as detection of cancer and stroke. Researchers have been applying various techniques of machine learning such as Artificial Neural Network, Naïve Bayes, Decision Tree, BP (Back-Propagation algorithm),and genetic algorithm for the optimization purpose. The characteristics of BP algorithm are that it is adaptive and tolerant towards the noisy data or other outliers present in the medical data. The determination of neural network structure is a challenging task. Even though significant progress has been carried out in classification related areas of neural networks, there are number of issues that are not been solved successfully or completely applying to the neural network. A small network could not be able to provide good performance owing to its limited information processing power. A large network, on the other hand, may have some connections redundant Neural network trains the input data pattern through different layers with local convergences which does not provides optimal solution to the problem [3]. Thus the selection of proper network architecture plays key role in training the input data. Back propagation networks do not need the linear relationship between the data and the target output. The characteristics of BP algorithm are that it is adaptive and tolerant towards noisy data or other outliers present in the medical data. But it has a disadvantage of getting stuck in the local minima and thus the data is first optimized using genetic algorithm and the concepts of crossover and mutation are applied and that is then again fed into the neural network for better results.The dataset from the UCI repository is used for the same. It contains 14 attributes, 13 of which are the general attributes that can be used for the prediction and the last attribute being the detail of presence of heart disease in the patient and the intensity of the disease, as absent, low, medium, high or serious. The proposed prediction network will thus have 13 Input nodes, 7 hidden layer nodes and 1 output node. At first weights are assumed for every attribute, and when the system is trained the weights are adjusted at every level to form a weight base at the end of training. Here, genetic algorithm plays the role of optimizer for the weights, later the same neural network is used for prediction of the disease except that this time the weights won't be adjusted. Out of the 303 instances of data provided by the repository, 200 instances can be used for training the neural network and 103 can be used for testing.
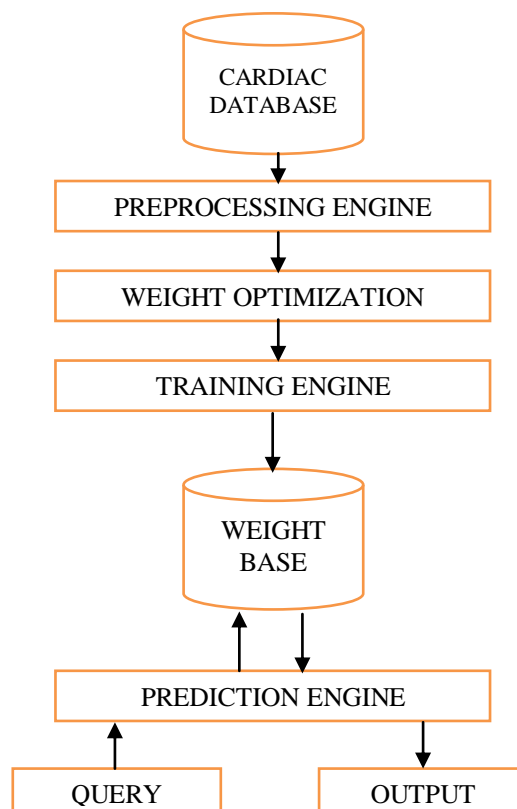

Fig. 1. System Architecture

III. IMPLEMENTATION DETAILS

All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified. The first step is build the cardiac dataset, the dataset is divided into a ratio, so that a part of it can be used for training and the remaining for testing. The dataset will not be having smooth data, there may be missing and wrong data present and

hence has to go through preprocessing engine, at this stage, data cleaning, data integration, data transformation and data reduction. The data will be different ranges of values and have to be made into a chromosome so that genetic algorithm can be applied and hence are also normalized at this stage. The weight optimization engine is the most critical part of the proposed system; here first a random population of weights is generated, using fitness function the best fit and worst fit individuals are found, and then the reproduction operators are applied in it till the best solution is found. In the training engine, back propagation algorithm is used for training the neural network and the mean square error between the actual and desired output is reduced to a predetermined level. The reason for using this algorithm is that it can find a good set of weights in a reasonable amount of time. The weight base will contain the final set of weights that can be used by the prediction engine. The prediction engine will predict the severity of the heart disease when the user submits the query to the prediction engine. It is the same neural network as the one used in the weight optimization engine except that it will not adjust the weights.
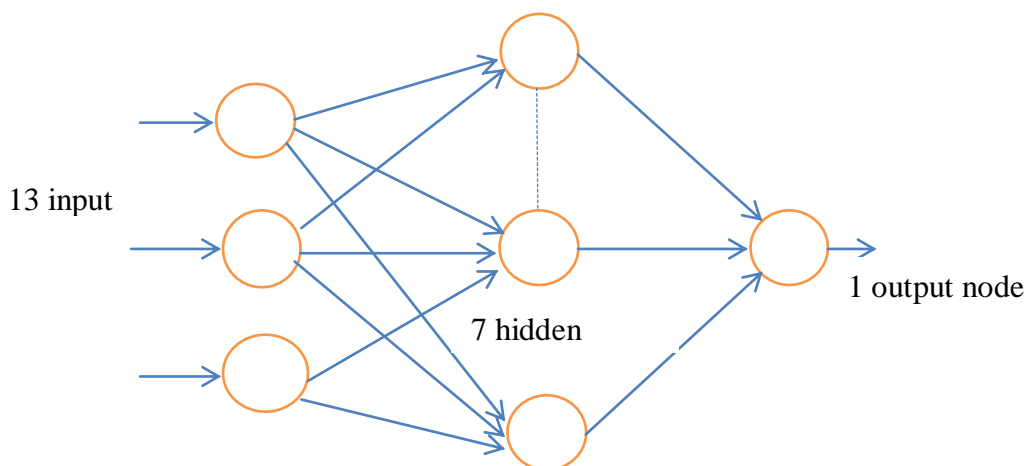

Fig. 2. Proposed Prediction Network

A.　　　　Cardiac Database
Dataset from the UCI repository [2] will be used for the prediction of heart diseases, it has 14 attributes. The detailed list of attributes is described below.

TABLE 1: SUMMARY OF THE HEART DISEASE DATASET

Attribute	Description	Domain of Value
Age	Age in years	29 to 77
Sex	Gender of the patient	Male(1) Female(2)
Cp	Chest pain type	Typical Angina(1) Atypical Angina(2) Non-Anginal(3) Asymptomatic(4)
Trestbps	Resting Blood Sugar	94 to 200 mg Hg
Chol	Serum Cholestrol	126 to 564 mg/dl
Fbs	Fasting Blood Sugar	>120 mg/dl True(1) False(0)
Restecg	Resting ECG result	Normal(0) ST-T wave abnormality(1) LV hypertrophy(2)
Thalach	Maximum heart rate achieved	71 to 202
Exang	Rest induced agina	Yes(1) No(0)
Oldpeak	ST depression induced by exercise relative to rest	0 to 6.2
Slope	Slope of peak exercise ST segment	Upsloping(1) Flat(2) Downsloping(3)
Ca	Number of major vessels colored by fluoroscopy	0-3
Thal	Defect Type	Normal(3) Fixed defect(6) Reversible defect(7)
Num	Heart Disease	0-4

B.   Pre-processing Engine

Preprocessing [4] is an important step in the knowledge discovery process, as real world data tend to be incomplete, noisy, and inconsistent. Data preprocessing includes data cleaning. Data cleaning routines attempt to remove columns with missing values, smooth out noise while identifying outliers and correct inconsistencies in the data. Input File, Target file and files for testing are prepared respectively. The data is normalized from the range of -1 to 1 using min max.

C.   Weight Optimization Engine

Weight optimization technique discussed in [4] is used for optimizing the weights of the neural network. Genetic algorithm uses a direct analogy of natural behavior, work with a population of individual strings, each representing a possible solution to the problem. Each individual string is assigned a fitness value which is an assessment of how good a solution is, to a problem. The best fit individuals participate in reproduction by cross breeding with other individuals in the population. A whole new population of possible solutions to the problem is generated by selecting the best fit individuals from the current generation. This new generation contains characteristics which are better than their ancestors.

D.   Training Engine

Back propagation algorithm discussed in [6] is used for training the neural network. The reason for choosing this algorithm is, it can find a good set of weights in a reasonable amount of time. Back propagation is a variation of gradient search. It uses a least-square optimality criterion. The key to back propagation is a method for calculating the gradient of the error with respect to the weights for a given input by propagating error backwards through the network. The neural network has the capability to quickly classify a dataset. It is trained on a set of training data until it reaches a predefined threshold level.

E.   Weight Base

Weight base contains the final weights of the trained neural network. These weights are used by the prediction engine to predict the severity of cardiovascular disease.

F.   Prediction Engine

Prediction Engine predicts the severity of cardiovascular disease. When the user submits the query to the prediction engine, it gets the weights from the weight base and predicts the severity of the disease as absent, low, medium, high or serious. It will contain 13 nodes in the input layer, 7 nodes in the hidden layer and only one node in the output layer. It gets the weight from the weight base and works the same as that of the back-propagation network's initial iteration, but here no error will be calculated.

IV.CONCLUSION

Here, an approach for Decision Support System is made for the analysis of heart related medical data. The Heart Disease dataset analysed to predict how severe the disease is. A hybrid of genetic algorithm and neural networks is developed for the prediction system. The data in the dataset is processed before to make it appropriate for classification. The weights for the neural network are found and the genetic algorithm optimizes it. The USI data is distinguished into five classes based on the intensity of the ailment using Back-propagation Algorithm and the final weights of the neural network are stored in the weight base. These weights are later used for predicting the presence and severity of cardiovascular disease. All the attributes are taken into considered for predicting the risk of cardiovascular disease. The accuracy obtained is about 94.17%. There are many interesting aspects for future work. If used it can save many lives.

REFERENCES

[1] AnkitaDewan, Meghna Sharma. "Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification", 2nd International Conference on Computing for Sustainable Global Development (INDIACom), IEEE, 2015.
[2] https://archive.ics.uci.edu/ml/datasets/Heart+Disease [Last Visited: 02/12/2015].
[3] Syed Umar Amin, Kavita Agarwal and Dr. Rizwan Beg, "Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factor", Proceeding of IEEE Conference on Information and Communication Technologies(ICT), pp. 1227-1231, April 2013.
[4] S Florence, N G Bhuvaneswari Amma, G Annapoorani, K Malathi, "Predicting the Risk of Heart Attacks using Neural Network and Decision Tree", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 11, November 2014.
[5] S. Rajasekaran and G.A. Vijayalakshmi Pai, "Neural Networks, Fuzzy Logic and Genetic Algorithms synthesis and Applications", PrenticeHall of India, 2007.
[6] http://statweb.stanford.edu/~susan/courses/s60/split/node60.html [Last Visited: 15/12/2015].
[7] Jiawei Han and Michelibne Kamber, "Data Mining Concepts and Techniques", Morgan Kauffman Publishers, 2009.