Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 08, August-2019

Heart Disease Prediction using Evolutionary based Artificial Neural Network

Tejali Mhatre [1], Satishkumar Varma [2]
Department of Information Technology,
Pillai College of Engineering, New Panvel, India

***Abstract*— Large number of people are diagnosed with heart disease. With increasing life expectancy of population and availability of effective treatment options to prevent heart disease patients has increased manifold. Most of deaths occurred due to heart disease. For preventing heart disease accurate identification of heart disease at early stage is very important. Classification of heart disease can be beneficial in medical field to identify risk of heart disease. The proposed methodology is useful for predicting the risk of heart disease in human being. This methodology is produced by combining the respective superiority of evolutionary algorithm and artificial neural network. Backpropagation algorithm is particularly suited for complex problem. Evolutionary algorithm is always used to find optimal solution.**

In the proposed system, genetic algorithm is act as a weight optimization engine for backpropagation network. Standard heart disease dataset from UCI repository is used to process the system. Evolutionary based backpropagation network is used for training the dataset. Final weights are used to store in weight base and used to predict the risk of heart disease. The classification accuracy obtained using this approach is 78.763%.

Keywords- Genetic Algorithm; Neural Network, Backpropagation Algorithm; Cardiovascular Disease; Prediction Engine

I. INTRODUCTION

Heart disease is considered as the class of diseases which involve the heart or heart vessels. Disease that influence the cardiovascular system would consider as heart disease. Cardiovascular diseases consist of coronary heart disease, cerebrovascular disease, raised blood pressure, peripheral artery disease, rheumatic heart disease, congenital heart disease and heart failure. In practice, heart disease is treated by cardiologists, vascular specialist, thoracic surgeons, neurologists, and interventional radiologists. It depends 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. The World Health Organization (WHO) has estimated that 17.9 million deaths occur worldwide, every year due to heart disease [2]. Medical diagnosis is vital yet convoluted task. It needs to be done accurately and efficiently. The automation of this system is very much needed to help the physicians to do better diagnosis and treatment. The representation of medical knowledge, decision making, choice and adaptation of a suitable model are some issues that a medical system should take into consideration. Medical progress is always supported by data analysis which improves the skill of medical experts and establishes the treatment technique for diseases. The purpose of medical diagnosis system is to assist physicians in defining the risk level of an individual patient. The heart disease dataset found in University of California, Irvine Machine Learning Repository is used for training and testing the system [10]. The purpose of using this dataset is to provide a complex, real world data example where the relationships between the features are not easily discovered by casual inspection. In this proposed system, the advantages of genetic algorithm and neural network are merged to predict the risk of cardiovascular disease. Genetic algorithm is an optimization algorithm that mimics the principles of natural genetics. It finds acceptably good solutions to problems acceptably quickly. In many applications, knowledge that describes desired system behavior is contained in datasets. When datasets contain knowledge about the system to be designed, a neural network promises a solution because it can train itself from the datasets. Neural networks are adaptive models for data analysis particularly suitable for handling nonlinear functions. By combining the optimization technique of genetic algorithm with the learning power of neural network, a model with better predictive accuracy can be derived.

II. RELATED WORK

Forecasting cardiac disease using techniques like classification, clustering and association rule mining is very prominent research. In this section cardiac disease prediction as a method is analyzed and reviews to introduce method in detail. There are more literature available before this hybrid technique introduced.

M. Anbarasi, E. Anupriya and N.Ch.S.N.Iyengar in[1] proposed method which predict risk of heart disease with feature selection using Genetic algorithm. Genetic algorithm is used to scale down the number of attributes from dataset. It is used to reduce thirteen attributes to 6 attributes using genetic search. They also used Naïve Bayes and Decision tree algorithm for prediction of disease

In [2], Mrs. Radhimeenaxshi proposed heart disease prediction system by utilizing two methods namely Support Vector Machine and Artificial Neural Network. In their work they compare result between two techniques by measuring accuracy. Statistical analysis is used to measure the performance of the system. Accuracy of SVM and ANN

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 08, August-2019

is 85.4%,84% respectively. They concluded that SVM is better than ANN.

In [3], Sa delma Banu N.K and Suma Swamy conducted survey from 2004 to 2015. It gives the idea of distinct models available for prediction. According to survey data mining techniques used for prediction are Decision tree, Naïve Bayes, Neural Network, Artificial Intelligence. They have mentioned the calculated accuracy of each method.

Yuan-Keun Kwon and Byuno-Ro Moon in [4], introduced a Neuro-Genetic approach for stock market analysis. They have tested data from 36 companies in NYSE from year 1992 to 2004.They used hybrid approach to predict accurate trading value. They have concluded that system is used to improve buy and sold strategy.

S.Florence, N.G.Bhuvaneswari Amma, G.Annapoorani, K.Malathi proposed a framework to predict the heart disease using Neural network and Decision tree(ID3)[5].They have tested algorithms with Acath heart attack dataset from UCI repository. Statistical analysis presented to understand output of both techniques.

In [6], they proposed different classification techniques for prediction of Coronary Heart Disease(CHD). They mainly focused on Support vector machine, Neural network and Decision tree algorithm. The experiment carried on dataset having 1000 records. Accuracy obtained for SVM, ANN, DT is 92.1%, 91%, 89.6% respectively.

III. SYSTEM ARCHITECTURE

The architecture of the proposed system is illustrated in Figure 1. The major components of this system are Cardiac Database, Preprocessing Engine, Weight Optimization Engine, Training Engine, Weight Base, and Prediction Engine. Figure1. System Architecture.

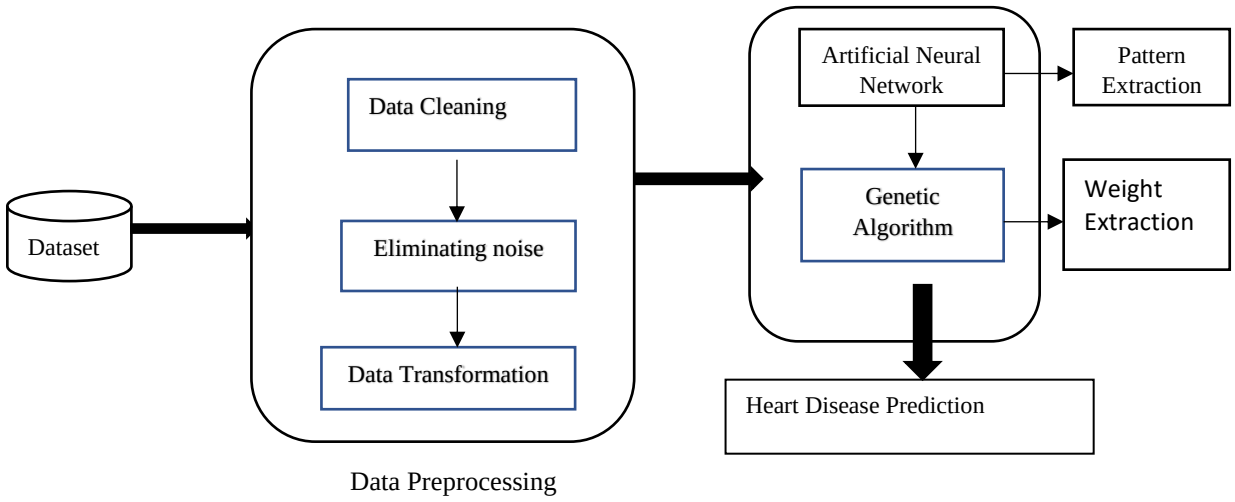

Figure 1. Block Diagram for Proposed system

A. Cardiac Database

The Cleveland heart disease data provided by the University of California, Irvine Machine Learning Repository [10] is used for analysis of this work. The dataset has 13 numeric input attributes namely age, sex, chest pain type, cholesterol, fasting blood sugar, resting ecg, maximum heart rate, exercise induced angina, old peak, slope, number of vessels colored and thal. It also has the predicted attribute i.e. the class label. The description of the dataset is tabulated in Table 1.

B. Preprocessing Engine

Preprocessing 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 integration, data transformation, and data reduction. Data cleaning routines attempt to fill in missing values, smooth out noise while identifying outliers and correct inconsistencies in the data. In this proposed work, the most probable value is used to fill in the missing values. Data transformation routines convert the data into appropriate forms for mining. Normalization is useful for classification purpose. By normalizing the input values for each attribute measured in the training tuples will speed up the learning process. In this work, the normalization technique used is min-max normalization.

The min-max normalization given in equation (1) discussed in [12] is defined as follows:

$$V = \frac{V1 - Vmin}{Vmax - Vmin} \dots\dots\dots\dots\dots [1]$$

In this work, the following attributes are normalized: age, trestbps, chol, thalach, and oldpeak.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 08, August-2019

Table 1. Summary of The Heart Disease Dataset

Attribute name	Meaning of the attribute	Range of the each attribute
Age	Age in years	Male(1) Female(2)
Cp	Chest pain type	Typical angina(1) Atypical Angina(2) Non-anginal(3) Asymptomatic(4)
Trestbps	Resting blood sugar	94 to200 mm Hg
Chol	Serum Cholesterol	126 to 564 mg/dl
FBS	Fasting blood sugar	>120 mg/dl True(0) False(1)
Restecg	Resting ECG signal	Normal(0) ST -T wave abnormality(1) LV hypertrophy(2)
Thalach	Maximum heart rate	71 to 201
Exang	Exercise induced	Yes(0) No(1)
Oldpeak	ST depression by exercise	0 to 6.2
Slope	Slope peak exercise	Upsloping(1) Flat(2) Downsloping(3)
Ca	Number of majr vessel	0-3
Thal	Defect type	Normal(3) Fixed Defect(6) Reversible defect(7)
Num	Heart disease	0-4

C. Weight Optimization Engine

Weight determination technique discussed in [11] is used for optimizing the weights of the neural network. Genetic algorithm uses a direct metaphor 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.

The following is the methodology for weight optimization used in this work as discussed in [11]:

Step 1: Get the input-output pairs (Ii,Ti), i=1, 2, ..,N where Ii = (I1i, I2i, …, Ini) and Ti=(T1i,T2i,…,Tni) of the neural network, Chromosome Ci, i=1,2,…,p belonging to the current population Pi whose size is p.

Step 2: Extract weights Wi from Ci .

Step 3: Keeping Wi as a fixed weight setting train the neural network for the N input instance. Calculate error Ei using equation (2):

$$Ei=\sum_j (Tij - Oji)^2 \ldots\ldots\ldots (2)$$

Where Oji is the output calculated by the NN

Step 4: Calculate the root mean square E of the errors using equation (4):

$$E= Ei/N \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

where Ei , i=1,2,…,N

Step 5: Calculate the fitness value Fi using equation (5) for each of the individual string of the population

$$Fi=1/\ Ei \ldots\ldots\ldots\ldots (4)$$

Step 6: Select parent using roulette wheel parent selection. Apply single point crossover and mutate child chromosome to the parent chromosome.

Step 7: Check for benefiting of child chromosome with the objective function. Replace the old generation by the new generation and name it the best chromosome.

Repeat steps 2 to 7 till the stopping criterion is met.

D. Training Engine

Back propagation algorithm discussed in [11] 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. Backpropagation is a variation of gradient search. It uses a least-square optimality criterion. The key to backpropagation 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. The Backpropagation algorithm can be outlined as follows:

Step 1: Get the number of input nodes, hidden nodes and output nodes.

Step 2: Get the weights from the weight optimization subsystem.

Step 3: For the training data, present the set of inputs and outputs. By using the linear activation function, the output of the input layer is evaluated as {O}I={I}I, where {I}I is the training data set.

Step 4: Compute the inputs to the hidden layer by multiplying the corresponding weights of synapses using equation (5):

$$\{I\}_H = [V]_T \{O\}_I \ldots\ldots\ldots\ldots (5)$$

where [V]T is the weight matrix for input to hidden layer obtained from weight optimization subsystem.

Step 5: Let the hidden layer units evaluate the output using the sigmoidal function as given in equation (6):

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 08, August-2019

$$\{O\}H= \frac{1}{1+e^{-I_{Hi}}} \text{.....(6)}$$

Step 6: Compute the inputs to the output layer by multiplying the corresponding weights of synapses as given in equation(7):

$$\{I\}o= [V]^T \{O\}_H \text{ .....(7)}$$

where [W]T is the weight matrix for hidden to output layer obtained from weight optimization subsystem.

Step 7: Let the hidden layer units evaluate the output using the sigmoidal function as given in equation (8):

$$\{O\}O= 1/ (1+e_j^{-Io}) \text{.....(8)}$$

Step 8: Calculate the error and the difference between the network output and the desired output using equation (9):

$$E_i=\sum_j (T_{ij} - O_{ji})^2 \text{ .....(9)}$$

Where Tji is the desired output and Oji is the output calculated by the neural network.

D. Weight Base

Final weights of the trained neural network. Are stored in weight extraction. These weights are used by the prediction engine to predict the risk of cardiovascular disease.

E. Prediction Engine

Prediction engine predicts the severity of heart disease. When the user start executing system, it gets the weights from the weight base and predicts the severity of the disease. The proposed prediction engine is shown in Figure 2. It consists of 13 nodes in the input layer, 7 nodes in the hidden layer and only one node in the output layer. It gets the weights from the weight base and works same as that of backpropagation network's initial iteration. But no error is calculated.

Figure 2. Proposed Prediction Network

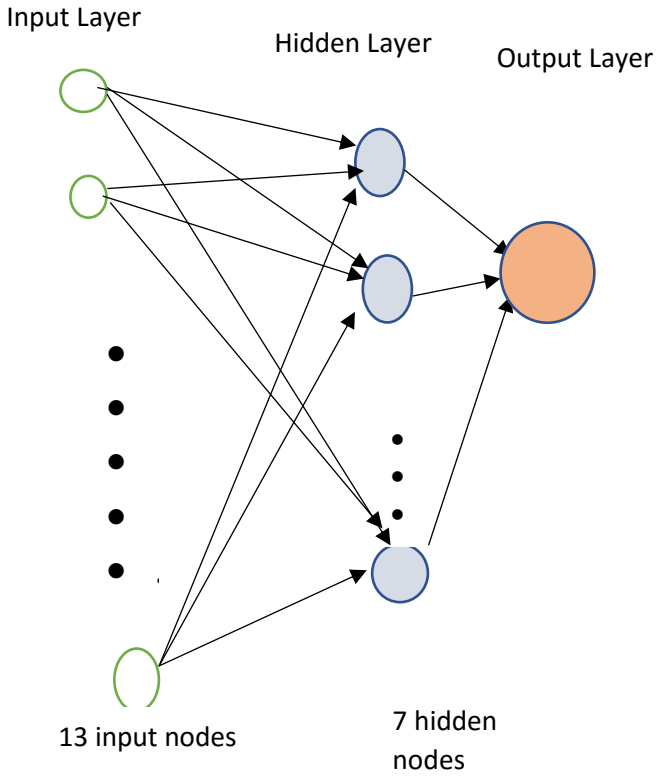

Figure 2 Proposed BPN network

IV. EXPERIMENTAL RESULTS

The Cleveland Heart Disease Dataset provided by the UCI Machine Learning Repository [8] is used for training and testing the medical diagnosis system. The distribution of dataset is given in Table 2. Among the 303 instances of data, 200 instances are used for training and 103 instances are used for testing.

TABLE 2. DISTRIBUTION OF DATA

Class label	Absent (0)	Low (1)	Medium (2)	High (3)	Serious (4)
Training	109	38	20	23	10
Testing	55	17	16	12	3

The cardiac dataset is preprocessed in order to make it suitable for further processing. The initial population of weights is randomly generated. The population size is determined by the number of nodes in the neural network. The number of nodes in the input layer, hidden layer, and output layer are 13, 7 and1 respectively. Therefore, the total number of weights needed for training is calculated using equation (11) which is discussed in [11]:

Weights = (Number of input nodes + Number of output nodes) x Number of hidden nodes ……(11)

Using the fitness function, the best fit and worst fit individuals are selected and then duplicate the best fit with the worst fit. Then, the reproduction operators are applied and the process is continued until the best solution is obtained. Backpropagation algorithm is used for training and the mean square error between the actual and desired output is reduced to a predetermined level. The final weights are stored in the weight base. This is used by the prediction engine to predict the risk of cardiovascular disease.

The training and testing dataset classification by genetic based neural network is given in Table 3. and Table 4. respectively. The classification accuracy of training set 75%. The accuracy of neural network and evolutionary based neural network is compared in Figure3. The classification accuracy of testing set is 78.756%.

Table 3. Classification of Training Data

Class label	Absent(0)	Low(1)	Medium(2)	High(3)	Serious(4)
Yes	108	37	20	23	10
No	1	1	0	0	0

Table 4. Classification of Testing Data

Class label	Absent(0)	Low(1)	Medium(2)	High(3)	Serious(4)
Yes	52	16	15	12	2
No	3	1	1	0	1

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 08, August-2019

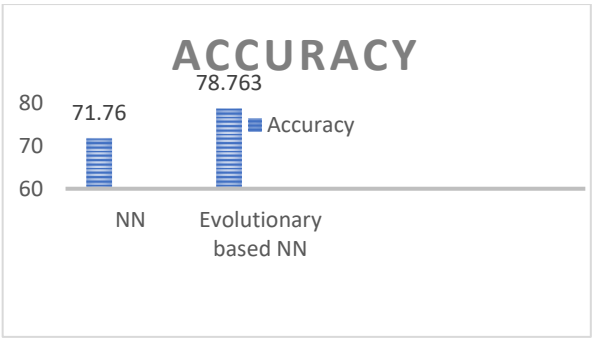

Figure 3. Accuracy comparison between neural Network and proposed method

V. CONCLUSION AND FUTURE WORK

In this paper, a framework for Decision Support System is developed for the analysis of medical data. The Heart Disease dataset is taken and analyzed to predict the severity of the disease. A genetic based neural network approach is used to predict the severity of the disease. The data in the dataset is preprocessed to make it suitable for classification. The weights for the neural network are determined using evolutionary algorithm. The preprocessed data is classified into five classes based on the severity of the disease using Backpropagation Algorithm and the final weights of the neural network are stored in the weight base. These weights are used for predicting the risk of cardiovascular disease. All the attributes are taken into consideration to predict the risk of cardiovascular disease. The accuracy obtained is 78.763%. There are many interesting aspects for future work. This system can be enhanced by using Evolutionary algorithms and Principal Component Analysis to reduce the dimension of the dataset and is used to predict the risk of cardiovascular diseases.

REFERENCES

[1] M.Anbarasi, E.Anupriya, N.CH.S.N.Iyengar, "Enhanced Prediction of Heart Disease with Feature Subset Selection using Evolutionary algorithm", International Journal of Engineering Science and Technology, Vol.2, No.10, pp.5370-5376, 2010.

[2] S.Goenka, D.Prabhakaran, V.S.Ajay, and K.S.Reddy, "Preventing Cardiovascular Disease in India – Translating Evidence to Action", Current Science, Vol.97, No.3, pp.367-377, 2009.

[3] Hai H.Dam, Hussain A.Abbass and Xin Yao, "Neural – Based Learning Classifier Systems", IEEE Transactions on Knowledge and Data Engineering, Vol.20, No.1, pp.26-39, 2008.

[4] Latha Parthiban and R.Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", International Journal of Biological and Life Sciences, Vol.3, No.3, pp.157-160, 2007.

[5] Niti Guru, Anil Dahiya and Navin Rajpal, " Decision Support System for Heart Disease Diagnosis using Neural Network", Delhi Business Review,Vol.8, No.1, pp.99-101,2007.

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

[7] Shantakumar B.Patil and Y.S.Kumaraswamy, "Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction", International Journal of Computer Science and Network Security ,Vol.9, No.2, pp.228-235, 2009.

[8] D.Shanthi, G.Sahoo and N.Saravanan, "Evolving Connection Weights of Artificial Neural Networks using Genetic Algorithm with Application to the Prediction of Stroke Disease", International Journal of Soft Computing, Vol.4, No.2, pp.95-102, 2009.

[9] Yung-Keun Kwon and Byungo-Ro Moon, "A Hybrid Neuro-Genetic Approach for Stock Forecasting", IEEE Transactions on Neural Networks,Vol.18, No.3, pp.851-864, 2007.

[10] http://www.ics.edu, UCI Repository of Machine Learning Data bases, Cleveland Heart Disease Dataset.

[11] S.Rajasekaran and G.A.Vijayalakshmi Pai, "Neural Networks, Fuzzy Logic, and Genetic Algorithms Synthesis and Applications", Prentice Hall of India, 2007.

[12] Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", Morgan Kaufman Publishers, 2009