Accuracy Enhancement of Artificial Neural Network using Genetic Algorithm

Preeti Gupta
M.Tech Research Scholar
Department of CSE
Chandigarh Engineering College, Mohali, Punjab

Bikrampal Kaur, PhD
Professor
Department of CSE
Chandigarh Engineering College, Mohali, Punjab

ABSTRACT

This research 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 with UCI dataset. In this study neural network is optimized with Genetic Algorithm and proved experimentally. The trained feed forward neural network and fitting neural network are optimized with genetic algorithm and is then compared with the scale conjugate gradient descent back-propagation algorithms trained feed forward neural network and fitting neural network respectively for the accuracy enhancement percentage. The proposed learning is much faster and accurate as compared to the other one. The proposed learning is designed and developed by using MATLAB GUI feature. 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.

General Terms

Learning enhancement, Optimization

Keywords

Data Mining, Diagnosis, Genetic Algorithm, Heart Disease, Neural Network.

1. INTRODUCTION

Data mining has great potential for examining the hidden patterns in the data sets of medical domain. These patterns can be used for clinical diagnosis. However the unprocessed medical data are widely distributed and heterogeneous in nature. These data require to be assembled. This collected data can then be integrated to form a hospital information system. Data mining technology provides a user-oriented approach to novel and hidden patterns in the data.

According to the World Health Organization (WHO) heart disease and stroke kills around 17 million people a year. By 2020, heart disease and stroke will become the leading cause of death all over the world. The term heart disease encloses the diverse diseases that affect the heart. Heart Disease was the major source of mortalities in the United States, England, Canada and Wales as in 2007 [1]. Heart disease kills one person every 34 seconds in the United States. Coronary heart disease and Cardio vascular disease are some categories of heart diseases. Hence more careful and efficient methods of cardiac diseases and periodic examination are of high importance. Data mining is the solution to this problem. Hence different data mining techniques can be applied to diagnose the disease more accurately. The main objective of this study is to develop a system by hybridizing ANN Technology and Genetic Algorithm for the diagnosis of heart disease with high accuracy.

1.1 Overview of Artificial Neural Network

The Artificial Neural Network (ANN) is a technique which is applied to solve data mining applications. Neural Network is a closely interconnected network, which offers rich structure providing some features of the biological neural network. Neural network provides an advantage to the user to execute parallel concept at each layer level. Neural Network structure consists of three layers: an input layers, one or more hidden layers and the output layer. Number of hidden layers and number of neurons in each layer strongly depends on the complexity of system studied. ANN learns from existing examples. Artificial neural network are trained based on two training methods:

1. Supervised Training- for supervised learning external teacher is available which provide the neural network input data and actual desired output. The network parameters are adjusted under the combined influence of training vector and error signal to get that desired output for a given specific input.

2. Unsupervised Training-Input data and computation function is given to the network and output is calculated. It does not require the desired output. Neural network is an iterative learning process in which input data cases are given to the network one at a time and weights associated with input interconnection are adjusted each time. Figure.1 shows how the inputs work in artificial neural networks.
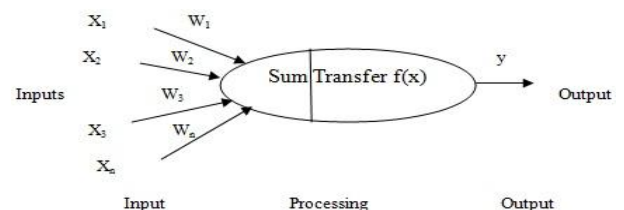

Figure.1 Artificial Neural Network

The advantages of Neural Networks for classification are:
- **Robustness** is more in Neural Network because of the weights.
- **Adaptive learning:** Performance of the Neural Network is improved by learning. Even after applying the training sets this may continues.
- **Real Time Operation**: For better performance the use of Neural Network can be parallelized as specified above.

- **Self-Organization**: Once the appropriate training has been performed, low error rate and high degree of accuracy has been achieved.
- **Fault Tolerance:** In noisy environment Neural Networks are more robust.

Types on ANN are:

1. Feed Forward Neural Network

Feed-forwrd ANNs is shown in figure.2 allow signals to travel from input to output i.e. only one way. There is no feedback (loops) i.e. the output of any layer does not influence that same layer. Feed-forward ANNs move to be straight forward networks that attach inputs with outputs. They are broadly used in pattern recognition. This type of organization is also referred to as top-down or bottom-up.
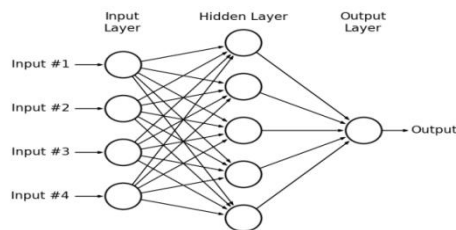

Figure.2 Feed-Forward Neural Networks

2. Fitting Neural Network

Fitting networks are feed forward neural networks used to fit an input-output relationship. Fitnet (hidden Sizes, trainFcn) takes these arguments and returns a fitting neural network. Figure.3 shows how the signal travels from the input nodes to the output node in the Fitting Neural Networks.
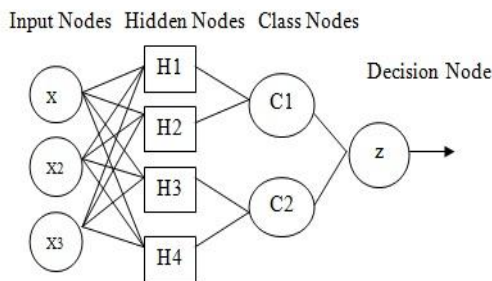

Figure.3 Fitting Neural Networks

1.2 Overview of Genetic Algorithm

Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimization problems. Basic ideas on Genetic algorithms were first developed by John Holland, and are mainly used as search and optimization methods. Given a large solution space, one would like to pick out the point which optimizes an object function while still fulfilling a set of constraints. In network planning, a solution point could be a specific link topology, a routing path structure, or a detailed capacity assignment with minimum costs [2]. The genetic algorithm is a random search algorithm that simulates natural selection and evolution. It searches the total solution space and can find the optimal solution globally over a domain. After an initial population is randomly generated, the algorithm evolves through three operators. Basic operations of Genetic Algorithm are as follows:

Choosing Operation/ Selection- The choosing operation means to choose the individuals from the old group and add them to the new group in a certain probability. The probability of the individual selected is relevant to the fitness value: the better fitness value of the individual, the greater probability of being chosen.

Crossover Operation- The crossover operation means to select two individuals from the existing individuals and generate a new better individual through the exchange and combination of two chromosomes. The crossover process is to choose two chromosomes from the groups and select randomly one point or more points to be exchanged. The crossover operation is shown below:

A :1100 01011111 A :1100 0101 0000

B :1111 0101 0000 B :1111 01011111

Mutation Operation- The mutation operation is to choose one individual from the group and then select a point of chromosomes for variation operation to generate better individuals. The mutation operation is shown below:
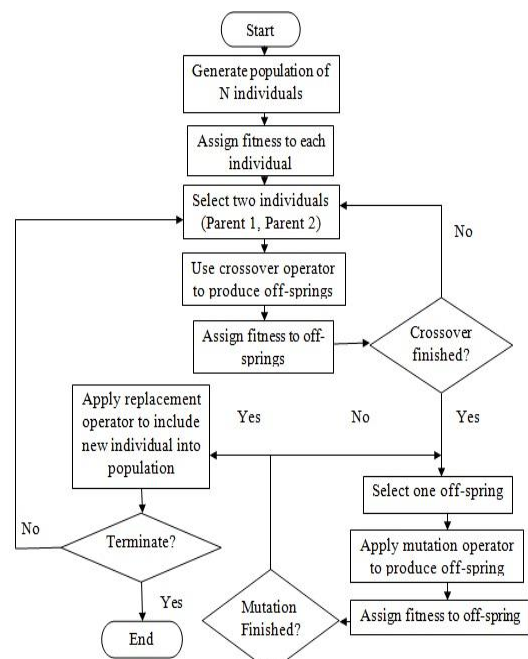
A :1100 0101 1111→ A :1100 0101 1101

Figure.4 Working of Genetic Algorithm

Figure.4 shows the basic working of Genetic Algorithm. First, an initial population of chromosomes is randomly selected. Then each of the chromosomes in the population is evaluated in terms of its fitness function. Next, a new population of chromosomes is selected from the given population by giving a greater change to select chromosomes with high fitness value. This is called natural selection. The new population may contain duplicates. If stopping criteria (e.g. no change in the old and new population, specified computing time, etc.) are not met then some specific genetic operations are performed on chromosomes of the new population. These operations produce new chromosomes, called offspring's. The same steps of this process, evaluation and natural selection are then applied to chromosomes of the resulting population. The whole process is repeated until given stopping criteria are met.

The solution is expressed by the best chromosome in the final population.

2. LITERATURE REVIEW

Amma, N.G.B, "Cardio Vascular Disease Prediction System using Genetic Algorithm and Neural Network", IEEE International Conference on Computing Communication and Applications, 2012.

In this paper, a diagnosis system is presented for estimating the risk of cardiovascular disease. This system combines the relative advantages of genetic algorithm and neural network. Multilayered feed forward neural networks are used to classify the complex problems. The weights of the neural network are driven using genetic algorithm because it finds decently best set of weights in less number of iterations. Genetic based Neural Network is used for training the system. The neural network final weights are stored in the weight base and are used for diagnosis the risk of cardio vascular disease. The accuracy of classification obtained using this approach is 94.17% [3].

Syed Umar Amin, Kavita Aggarwal, Dr. Rizwan Beg, "Data Mining in Clinical Decision Support and Treatment of Heart Disease", International Journal of Advanced Research in Computer Science and Technology, Vol.2 Issue.1, January 2013.

In this paper two data mining techniques are used Genetic Algorithm and Neural Network to predict the risk of heart disease with an accuracy of 89%. The hybrid system is implemented using the optimization advantage of genetic algorithm and has been proved better than back propagation in terms of stability and accuracy [4].

T. Manju, K. Priya, R. Chitra "Heart Disease Prediction System Using Weight Optimized neural Network", International Journal of Computer Science and Management Research, Vol 2, Issue 5 May 2013.

This paper presents the application of Multi Layer Feed Forward Neural Network that integrates Genetic Algorithm and Back Propagation network for heart prediction. GA is used to initialize and optimize the connection weights of MLFFNN. The optimized NN is trained and tested using 270 patient data [5].

Latha Parthiban, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", International Journal of Biological Biomedical and Medical Sciences, Vol.3 No.3, 2008.

This paper formulated an approach for the prediction of heart disease on the basis of coactive neuro- fuzzy inference system (CANFIS). The CANFIS model combined neural network capabilities and fuzzy logic approach which is then integrated with genetic algorithm to diagnose the presence of heart disease and the result shows that the proposed CANFIS model has great potential in prediction of heart disease [6].

Niti Guru, Anil Dhaiya, Navin Rajpal, "Decision Support System for Heart Disease using Neural Network", Delhi Business Review, Vol.8, No.1 (January-June 2007).

The prediction of Heart Disease, Blood Pressure and Sugar with the help of neural network was proposed by Niti Guru [7]. Tests were carried out on a specimen database of patient records. The Neural Network is tested and trained with 13 input variables such as Age, Blood Pressure, Angiography's report and etc. The supervised network has been focused for diagnosis of heart diseases. Training of the data is done with the help of back propagation algorithm. Whenever unrevealed data was fed by the doctor, the system identified the unrevealed data from comparisons with the trained data and generate list of probable diseases that the patients may prone to.

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

In this paper [8] they proposed an enhanced prediction of heart disease with feature selection using genetic algorithm. They predict better accurately the presence of heart disease with reduced number of attributes. They used Naïve Bayes, Clustering and Decision Tree methods to predict the diagnosis of patients with the same accuracy as accquired before the reduction of attributes. They concluded that the decision tree method outperforms the other two methods.

3. METHODOLOGY

The diagnosis process consists of two succeeding steps i.e. training and testing. When performing analysis with a set of existing data, split the data set into 70 % – 30 % for training and testing purpose respectively. The objective of this study is to measure the accuracy of this neural network and compare it with accuracy of the optimized neural network which is optimized using Genetic Algorithm. The Figure.4 shows the flow of work that will going to be carried out during our research.

The efficient neural network with improved accuracy can be obtained by following given steps.

- Firstly the dataset of biomedical site containing fields such as age, sex, blood group, blood pressure etc are defined.
- Select some (maximum part) among these entries for training of the neural networks and rest of the data is left for testing of the trained neural network.
- Therefore to the trained network, get the output after testing, then save the results.
- Get the weights values of the trained neural network and then update these weight values using Genetic Algorithm to train the network again with these new updated values.
- In final step test the system with these new neural network optimized by genetic algorithm, it will provide the efficient and accurate results.
- If results are optimized with respect to initial network without GA then again save the results and compare them with the initial trained neural network's output.
- If the results are not optimized with respect to the initial neural network without GA then again moves to update the weight values using GA.
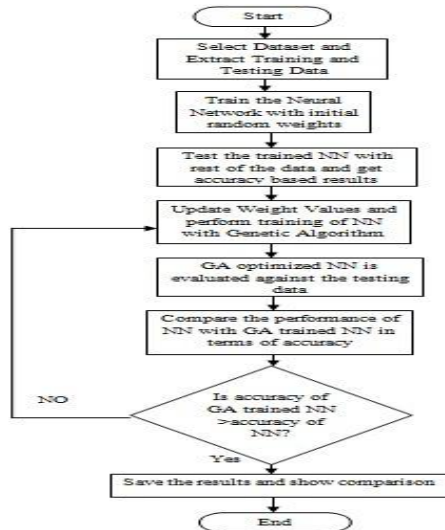- Show the results after the comparison is completed.

Figure.5 Basic flowchart of the Proposed System

3.1 DATA SET

In this study Cleveland heart disease data from UCI Machine Learning Repository [9] is used. This data was collected in Cleveland Clinic Foundation [10]. The purpose of this data is to separate the patients with coronary heart disease. The data comprises 303 instances, 13 attributes, and one decision attribute. The decision attribute has five values. The value ranges from 0 to 4. Zero indicates no presence of disease and 1-4 indicates the presence of disease. However, in this study only two class labels are considered. One label indicates no presence of the disease and the other indicates the presence of the disease. Thus, there are 164 cases with no presence of disease and 139 cases with presence of disease.

Table.1 Cleveland Heart Data Attributes

Attribute name	Description	Type
Age	Age of patient in years	Numeric
Sex	Sex (female = 0, male = 1)	Nominal
cp	Chest pain type (typical angina = 1, atypical angina = 2, non-angina pain = 3, asymptomatic = 4)	nominal
trestbps	Resting blood pressure (mm Hg) on admission to the hospital	numeric
chol	Serum cholesterol (mg/dl) numeric fbs If fasting blood sugar > 120mg/dl (yes = 1, no = 0)	nominal
fbs	If fasting blood sugar > 120mg/dl (yes = 1, no = 0)	nominal
restecg	Resting electrocardiographic result (normal = 0, having ST-T wave abnormality = 1, showing probable or definite left ventricular hypertrophy by Estes' criteria = 2)	nominal
thalach	Maximum heart rate achieved	numeric
exang	Exercise induced angina (no = 0, yes = 1)	nominal
oldpeak	ST depression induced by exercise relative to rest	numeric
slope	The slope of the peak exercise ST Segment	nominal
ca	Coronary artery disease	numeric
thal	Normal, Fixed defects, Reversible defects	nominal

4. RESULTS
4.1 Accuracy Comparison

Table.2 Comparison of Results obtained from ANN Model and GA-ANN Model

Accuracy	Prediction Model			
	Feed-Forward Neural Network	GA Trained Feed-Forward NN	Fitting Neural Network	GA Trained Fitting Neural Network
1st Execution	94.9	97.1	93.4	94.2
2nd Execution	95.6	97.1	95.6	96.4
3rd Execution	98.5	99.3	92.7	93.4
Mean Value	96.33	97.83	93.9	95.75

Table 2 shows the result after applying data sets on Feed Forward ANN model, Fitting ANN model and GA trained Feed Forward ANN model, GA trained Fitting ANN model. The performance comparison of the ANN and GA-ANN model is shown in Table 2.

Table.3 Improvement in the Prediction Accuracy of the ANN Model and GA-ANN Model

Prediction Model	Evaluation	
		Accuracy
Feed Forward ANN vs. GA-Feed Forward ANN	Feed Forward-ANN Model	96.33
	GA-FF-ANN Model	97.83
	Improvement %	1.50
Fitting ANN vs. GA Fitting ANN	Fitting ANN Model	93.90
	GA Fitting ANN	95.75
	Improvement %	1.85

Table 3 shows the accuracy improvement of Feed- Forward artificial neural network vs. Genetic Algorithm Feed Forward artificial neural network and Fitting artificial neural network vs. Genetic Algorithm Fitting artificial neural network.
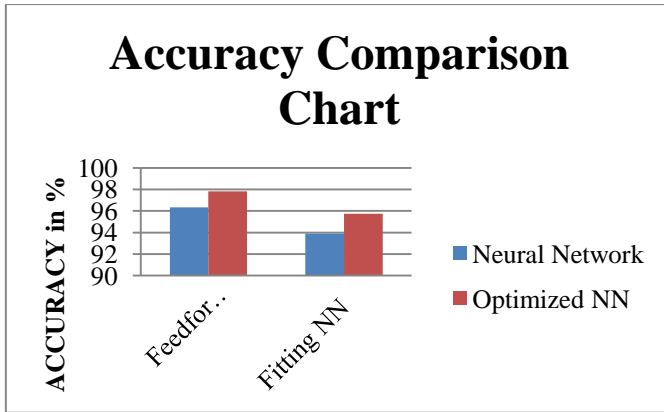
Accuracy Comparison Chart

Figure.6 Accuracy Comparison Chart

Figure.6 shows the accuracy comparison chart for both the neural networks i.e. Feed-forward normal neural network and optimized neural network, and Fitting normal neural network and optimized neural network. It shows the accuracy enhancement from normal neural network to optimized neural network.

4.2 Performance Comparison
Table.4 shows the Processing Time of Feed-forward NN and Fitting NN

Processing Time	Feed-Forward NN	Fitting NN
Iteration 1	10	25
Iteration 2	11	1
Iteration 3	98	1
Iteration 4	3.3	0.8
Iteration 5	18	29
Iteration 6	15	29

Table 4 shows the Processing Time taken by both the networks i.e. Feed-Forward Neural Network and Fitting Neural Network in order to calculate the accuracy.
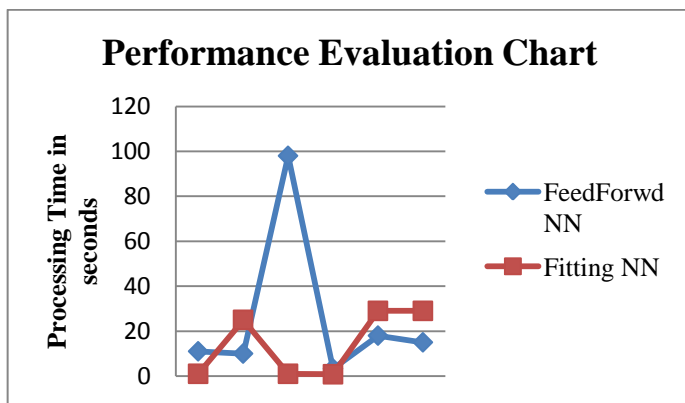
Performance Evaluation Chart

Figure.7 Performance Evaluation Chart

Figure.7 shows the Performance Evaluation Chart depicting the processing time taken by both the neural networks i.e. Feed-forward neural network and Fitting neural network. It's clearly depicted by the chart that how the processing time varies by both the networks i.e. the Feed-forward neural network and the Fitting neural network.

5. CONCLUSION
Accuracy enhancement for heart disease diagnosis has been proposed in this study. According to the results accuracy of neural networks were optimized, improved significantly with the use of genetic algorithm (GA). Feed-Forward neural networks and Fitting neural networks were used, their accuracy were calculated and optimized through genetic algorithm. Both normal and optimized accuracy were compared and time taken by both the neural networks was calculated. The GA trained ANN performed better than the normal ANN.

6. REFERENCES
[1] who.int/chp/chronic_disease_report/full_report.pdf.

[2] Midhuna.R, Shilpa Mehta, "Hybridization Of Neural Network Using Genetic Algorthim For Heart Disease Detection", International Journal of Application or Innovation in Engineering & Management, Volume 3, Issue 1, January 2014.

[3] Amma, N.G.B "Cardio Vascular Disease Prediction System using Genetic Algorithm and Neural Network", IEEE International Conference on Computing, Communication and Applications, 2012.

[4] Syed Umar Amin, Kavita Aggarwal, Dr. Rizwan Beg, "Data Mining in Clinical Decision Support and Treatment of Heart Disease", International Journal of Advanced Research in Computer Science and Technology, Vol.2 Issue.1, January 2013.

[5] T. Manju, K. Priya, R. Chitra "Heart Disease Prediction System Using Weight Optimized neural Network", International Journal Of Computer Science and Management Research, Vol 2, Issue 5 May 2013.

[6] Latha Parthiban and R. Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", International Journal of Biological Biomedical and Medical Sciences, Vol.3 No.3, 2008.

[7] Niti Guru, Anil Dhaiya, Navin Rajpal, "Decision Support System for Heart Disease using Neural Network", Delhi Business Review, Vol.8, No.1 (January-June 2007).

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

[9] A. Frank and A. Asuncion, "UCI Machine Learning Repository", University of California, Irvine, School of Information and Computer Sciences, 2010.

[10] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J.-J. Schmid, S. Sandhu, K. H. Guppy, S. Lee, and V. Froelicher, "International application of a new probability algorithm for the diagnosis of coronary artery disease", The American journal of cardiology, vol. 64, no. 5, pp. 304–310, 1989.