Volume 4, Issue 8, August 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Accuracy Enhancement of Heart Disease Diagnosis System Using Neural Network and Genetic Algorithm

Ms. Preeti Gupta
Research Scholar, Department of CSE
Chandigarh Engineering College Mohali, Punjab, India

Dr. Bikrampal Kaur
Professor, Department of CSE
Chandigarh Engineering College Mohali, Punjab, India

Abstract- Medical Diagnosis Systems plays an important role in medical field and are used by medical practitioners for diagnosis and treatment. In this paper, a medical diagnosis system is presented for predicting the risk of heart disease. In this paper the relative advantages of genetic algorithm and neural network are combined to achieve the desired accuracy. Feed-forward and fitting neural networks are used for the suited complex problems. ANN's are often used as a powerful discriminating classifier for tasks in medical diagnosis for early detection of diseases. The objective of this paper is to determine the weights of the neural network using genetic algorithm in less number of iterations. The dataset provided by University of Cleveland, is used for training and testing. In order to perform the training, the dataset is preprocessed to make it suitable. Genetic based neural network is used for training the system and accuracy comparison is done between the normal neural network and the GA based optimized neural network. The accuracy obtained using this approach is 97.75%.

Keywords- Diagnosis, Genetic Algorithm, Heart Disease, Medical System, Neural Network.

I. INTRODUCTION

Medical diagnosis is considered an art despite of all standardization efforts made, which is greatly due to the fact that medical diagnosis requires an expertise in coping with uncertainty which is not found in today's computing machinery. The researchers are inspired by the advancement in computer technology and machine learning techniques to develop software to assist doctors in making decision without requiring the direct consultation with the specialists. The medical diagnosis process can be interpreted as a decision making process, during which the physician activates the diagnosis of a new and unknown case from an available set of clinical data. This process can be computerized in order to present medical diagnostic procedures in a rational, obsolete, accurate and fast way.

Cardiovascular disease (CVD) grows along with the growth of economy and living standard [1]. In 2010, cardiovascular disease became the biggest cause of mortality in UK. The number of death caused by CVD is almost 180,000. As reported by British Heart Foundation, approximately 80,000 death was due to coronary heart disease [2]. Cardiovascular diseases, such as coronary heart disease and arrhythmia, are among diseases which endanger human life [1]. The presence of CVD can be detected by some symptoms, such as chest pain and fatigue. Nevertheless, it cannot be detected until an attack happened in 50% among reported cases [3]. Diagnosis of cardiovascular disease is an important stage before performing correct treatments. However, performing diagnosis is not an easy task. Therefore, highly skilled physician is required [3], [4].

The diagnosis of heart disease in most cases depends on a complex combination of clinical and pathological data. Because of this complexity, there exists a significant amount of interest among clinical professionals and researchers regarding the efficient and accurate prediction of heart disease. According to the statistic data from WHO, one third population worldwide died from heart disease; heart disease is found to be the leading cause of death in developing countries by 2010. It shows one third American adult have one or more types of heart diseases based on American Heart Association report. Computational biology is often applied in the process of translating biological knowledge into clinical practice, as well as in the understanding of biological phenomena from the clinical data. The discovery of biomarkers in heart disease is one of the key contributions using computational biology. This process involves the development of a predictive model and the integration of different types of data and knowledge for diagnostic purposes. Furthermore, this process requires the design and combination of different methodologies from statistical analysis and data mining [5].

II. ARTIFICIAL NEURAL NETWORK

A Neural Network consists of many Processing Elements, loosely called "neurons" and weighted interconnections among the Processing Elements. Each Processing Element performs a simple computation, such as calculating a weighted sum of its input connections, and computes an output signal that is sent to other Processing Elements. The training phase of a NN consists of adjusting the weights value of the interconnections, in order to produce the desired output. [6] The Artificial Neural Network is a technique that is commonly applied to solve data mining applications. A more sophisticated neurons the McCulloch and Pitts model. The difference from the previous model is that the inputs are "weighted"; the effect that each input has at its decision making is dependent on the weight of the selective input. The weight of an input is a number which when multiplied with the input gives the weighted input. These weighted inputs are then added together and if they excel a pre-set threshold value, the neuron fires. In any other case the neuron does not fire.

Types on ANN are:

1. Feed Forward Neural Network

Feed-forward ANNs (Figure 1) allow signals to travel one way only i.e. from input to output. There is no feedback (loops) i.e. the output of any layer does not affect that same layer and they are not dependent on each other. Feed-forward ANNs tend to be straight forward networks that accomplice inputs with outputs. They are largely used in pattern recognition. This type of organization is also referred to as bottom-up or top-down.

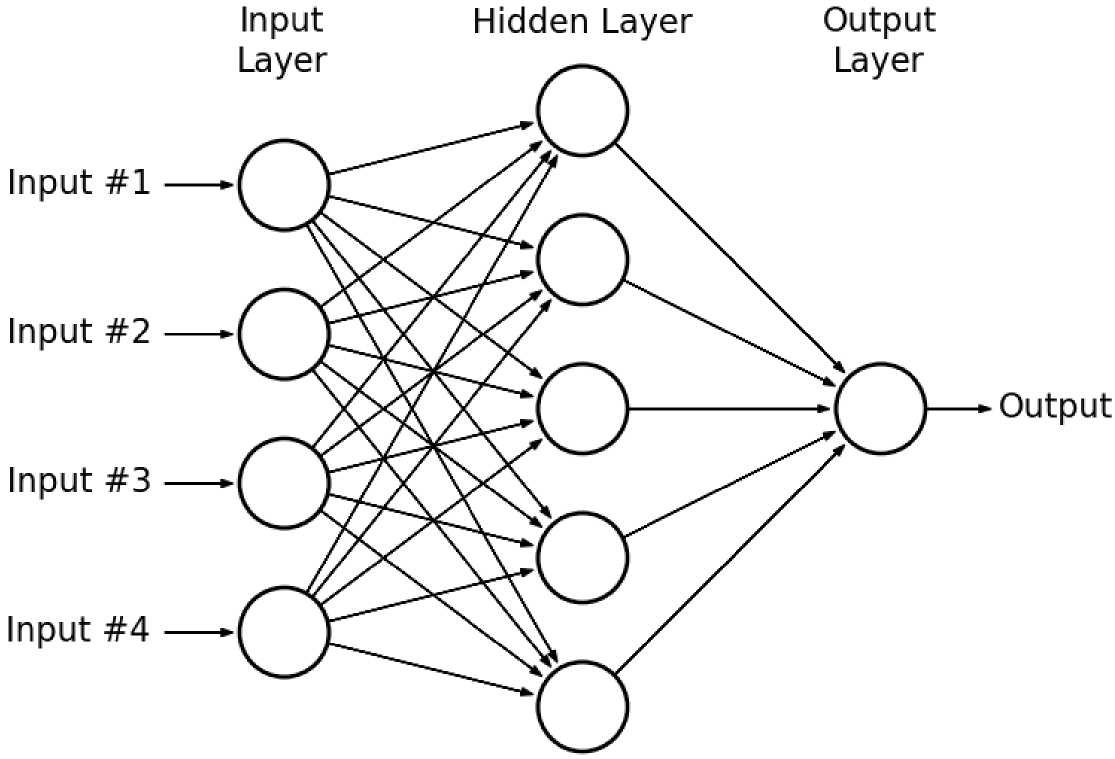

Figure.1 Feed-Forward Neural Networks

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.

III. GENETIC ALGORITHM

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 [8]. Genetic algorithms are based on the idea of natural selection. In nature, the properties of an organism are determined by its genes. Starting from a random first generation with all kinds of possible gene structures, natural selection suggests that over the time, individuals with "good" genes survive whereas "bad" ones are rejected. Genetic algorithms try to copy this principle by coding the possible solution alternatives of a problem as a genetic string. The genes can be bits, integers, or any other type from which a specific solution can be deduced. It is required that all solution points can be represented by at least one string. On the other hand, a specific gene string leads to exactly one solution. A set of a constant number of gene strings, each characterizing one individual, is called a generation. Since the different strings have to be evaluated and compared to each other, the notion of fitness is introduced. The fitness value correlates to the quality of a particular solution. Instead of working with the actual solution itself, genetic algorithms operate on the respective string representation. The following three basic operators are applied:

(i) Reproduction
(ii)Crossover
(iii) Mutation.

Pseudo code for Genetic Algorithm :

```
Algorithm: GA(n, χ, μ)

// Initialize generation 0:
k := 0;
Pk := a population of n randomly-generated individuals;
// Evaluate Pk:
Compute fitness(i) for each i ∈ Pk;
do
{ // Create generation k + 1:
// 1. Copy:
Select (1 − χ) × n members of Pk and insert into Pk+1;
// 2. Crossover:
Select χ × n members of Pk; pair them up; produce offspring; insert the offspring into Pk+1;
// 3. Mutate:
Select μ × n members of Pk+1; invert a randomly-selected bit in each;
// Evaluate Pk+1:
Compute fitness(i) for each i ∈ Pk;
while fitness of fittest individual in Pk is not high enough;
return the fittest individual from Pk;
```

Where n is the number of individuals in the population; χ is the fraction of the population to be replaced by crossover in each iteration; and μ is the mutation rate.

The reproduction process creates a new generation, starting from an existing generation; strings are reproduced with a probability respective to their fitness value. Strings which represent solutions with good properties have a higher chance to survive than strings depicting solution points with bad characteristics. This principle is also known as "survival of the fittest". The crossover operator exchanges genetic information between two strings. The strings of two randomly selected solutions are broken up at randomly chosen position, and parts of the strings are exchanged. One hopes that two solutions with good properties create an even better one. New genetic material is introduced by the mutation operator. The values of individual genes are changed and hence, new solutions are chosen. Mutation becomes important when after some generations the number of different strings decreases because strong individuals start dominating. In a situation of strong dominance of few strings, the crossover operator alone would not bring any changes and the search for an optimal solution would be ended. To partially shift the search to new locations in the solution space, the mutation operator randomly alters genes [7, 8].

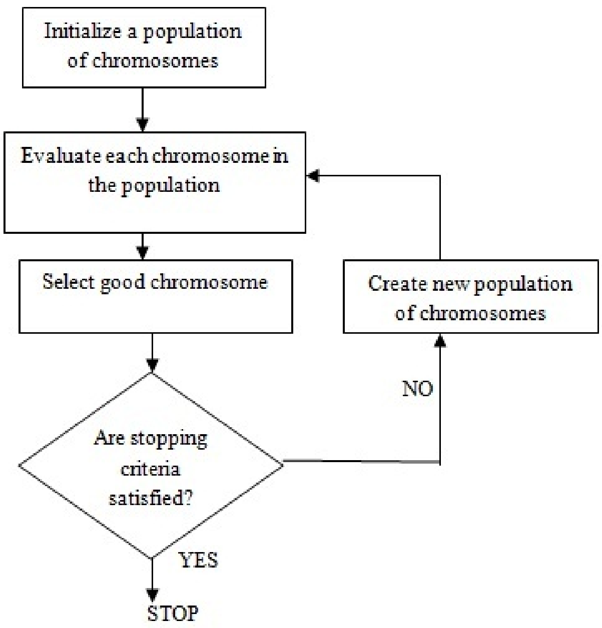

Figure.2 Flowchart of Genetic Algorithm

II. 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 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 feed forward 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**.** 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% [9].

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 [10].

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 [11].

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 [12].

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 [13] they proposed an enhanced prediction of heart disease with feature selection using genetic algorithm. They predict more 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 obtained before the reduction of attributes. They concluded that the decision tree method outperforms the other two methods.

III. METHODOLOGY

The Proposed Algorithm

Neural Networks is now a days the most promising area of interest. It is believed that for all the biomedical problems Neural Networks will prove to be the great solution in the coming years. Already it has been applied to various domains of medicine such as biochemical analysis, diagnostic system, drug development and image analysis. Neural Networks is a current research area at the moment. It will never replace human experts but they can help in screening and can be used by the experts to confirm their diagnosis. The advantages of such system are remarkable. People can be checked for heart diseases quickly and painlessly and thus detecting the disease at an early stage. In this thesis we have used hybrid data mining techniques in the diagnosis of heart disease patients. It investigates if integrating Neural Network with Genetic Algorithm can provide better performance than the normal neural network in the diagnosis of heart disease patients.

Steps showing the working of the algorithm:

- Firstly get the data set of any biomedical site or hospital which contains fields such as age, sex, blood group, blood pressure, smoking habits, obesity etc.
- Save the fetched data into MATLAB format so as to work upon it.
- Among these total entries, select maximum part say 70% for training of the networks and the rest 30% of the data is left for testing of the network.
- To train the network and to test the data we make use of Neural Networks. Neural Networks comprises of two facts i.e. weight values and the number of neurons.
- Now train the network using the training data set and rest 30% of data is executed for testing.
- Now calculate the accuracy of the trained network

  Accuracy= Actual Testing Data- Predicted Network Data.
- Save the results obtained after calculations.
- Now get the weight values of the trained neural network and then update these weight values using G.A. (Genetic Algorithm) and again train the network with these new updated values.
- Now the final step is to test the system with these new neural network values optimized by genetic algorithm and it will provide the accurate and efficient results.
- Again calculate the accuracy of the system using GA updated values.

Here comes a case:

1. If results get optimized with respect to initial network output without GA then again save the results and compare them with the initial accuracy of the trained neural network.
2. But if the results do not get optimized with respect to the initial neural network's output without GA then again move to update the weight values of ANN using GA.

- It will give value greater than the accuracy values of normal ANN and will stop the process when a greater value is achieved.
- Show the results in graphical representation for comparison.

Overview of Framework

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 swarm intelligence algorithm. The Figure 2 shows the flow of work that will going to be carried out during our research.

Steps which will be followed in our research are as follows:-

- To develop a system and to optimize it to get more accuracy after testing.
- To compare results of normal neural network and optimized neural network.
- Swarm intelligence algorithm i.e. Genetic Algorithm will be used for optimization.
- To develop better and more accurate proposed system.

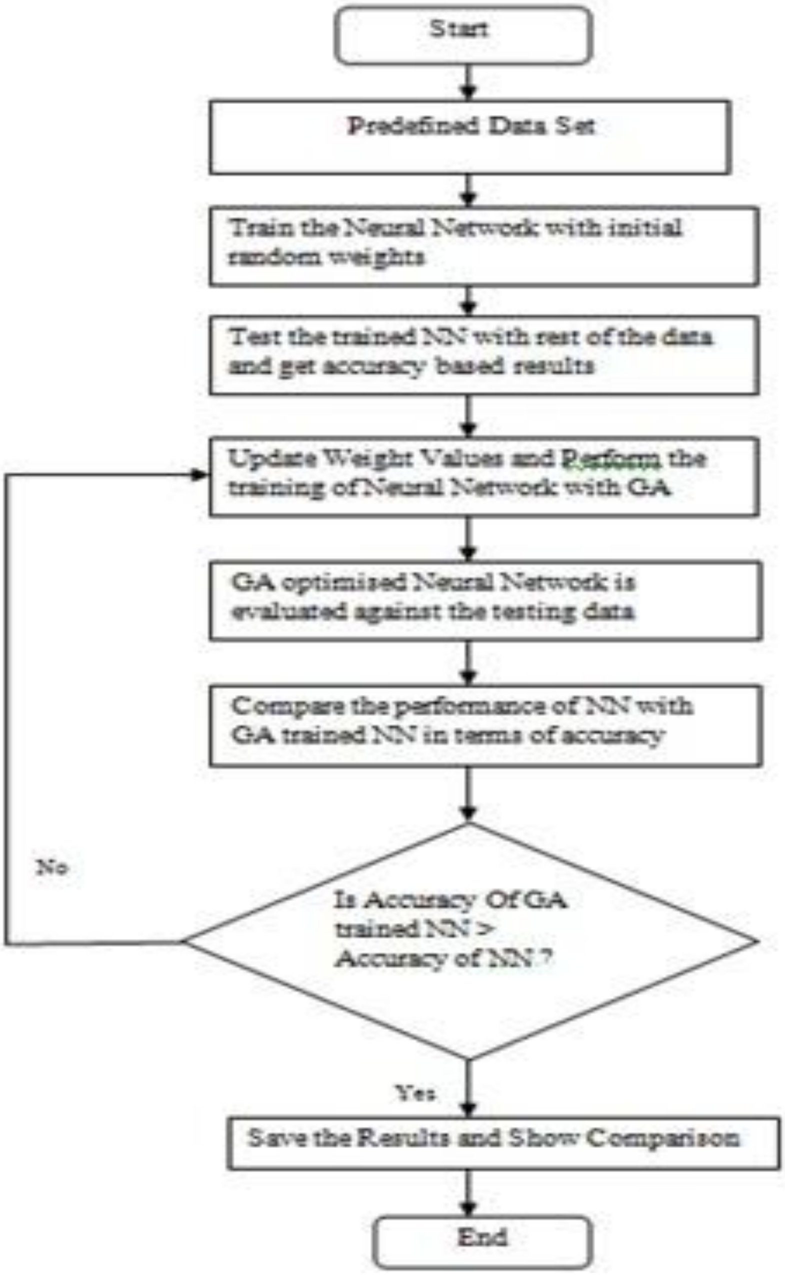

Figure.3 Basic flowchart depicting the training of neural network with genetic algorithm

IV. RESULTS

A. Accuracy Comparison

Table.1 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
4th Execution	94.9	95.6	95.6	99.3
5th Execution	95.6	97.8	96.4	97.1
6th Execution	99.3	99.6	95.6	97.1
Mean Value	96.46	97.75	94.88	96.25

Table 1 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 1.

Table.2 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	96.46
	GA-FF-ANN Model	97.75
	Improvement %	1.29
Fitting ANN vs. GA Fitting ANN	Fitting ANN Model	94.88
	GA Fitting ANN	96.25
	Improvement %	1.37

Table 2 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.

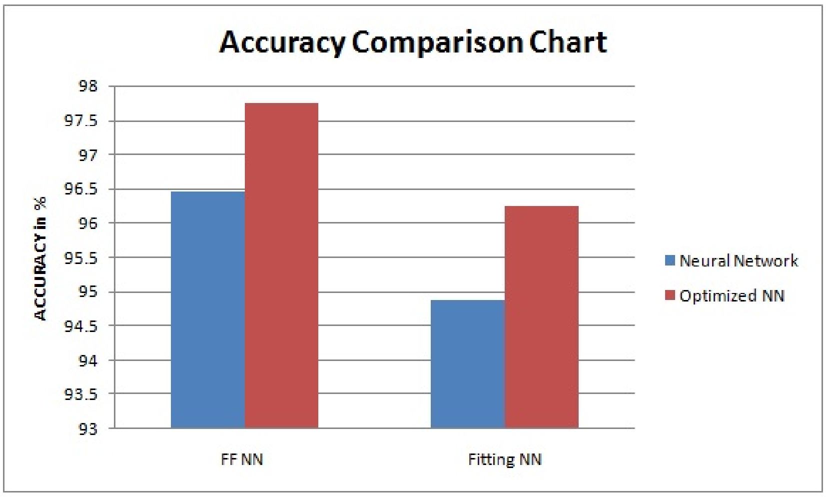

Figure.4 The above graph shows the accuracy enhancement of both the networks i.e. Feed-forward neural network and Fitting neural network

Figure.4 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.

V. CONCLUSION

In this paper, Accuracy enhancement for heart disease diagnosis has been proposed. The Heart Disease dataset is taken and analyzed to predict the accuracy of the disease. A genetic based neural network approach is used to predict the accuracy 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 genetic algorithm. The preprocessed data is classified using neural network optimized with genetic algorithm and the final weights of the neural network are stored in the weight base. These weights are used for predicting the risk of heart disease. Two types of learning algorithms are evaluated in this study, namely Feed forward neural network algorithm and Fitting algorithm. Improvement percentage of Feed-forward and Fitting neural network are 1.29% and 1.37% respectively. The accuracy obtained by the system is 97.75%.

REFERENCE

[1] F. Jin, J. Liu, and W. Hou, "The application of pattern recognition technology in the diagnosis and analysis on the heart disease: Current status and future", In 24th Chinese Control and Decision Conference (CCDC), pp. 1304–1307, 2012.

[2] Townsend N, Wickramasinghe K, Bhatnagar P, Smolina K, Nichols M, Leal J, Luengo-Fernandez R, and Rayner M, "Coronary Heart Disease Statistics 2012", British Heart Foundation, London, 2012.

[3] R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles", Expert Systems with Applications, vol. 36, no. 4, pp. 7675–7680, May 2009.

[4] M. C. Tu, D. Shin, and D. Shin, "Effective Diagnosis of Heart Disease through Bagging Approach", In 2nd International Conference on Biomedical Engineering and Informatics, 2009. BMEI'09., pp. 1–4, 2009.(above 1,2,3,4 ref are of 1st para cvd)

[5] A.H. Chen, S.Y. Huang, P.S. Hong, C.H. Cheng and E.J. Lin, "Hdps: Heart disease prediction system",IEEE Conference on Computing in Cardiology, pp. 557-560, December 2011.

[6] An introduction to neural computing. Aleksander, I. and Morton, H. 2nd edition.

[7] Jadhav, S.M. Dept. of Inf. Technol., Dr. Babasaheb Ambedkar Technol. Univ., Lonere, India Nalbalwar, S.L. ; Ghatol, A.A.''Data Fusion for Heart Diseases Classification Using Multi-Layer Feed Forward Neural Network ''Computer Engineering & Systems, 2008. ICCES 2008. International Conference on25-27 Nov. 2008 Page(s):67 -70.

[8] Valavanis, I.K.,School Of Electrical And Computer Engineering, National Technical University Of Athens, 9 heroon Polytechneiou Str,15780 Zographou, Geece, Mougiakakou, S.G; Grimaldi, Ka Nikita Ks,". Analysis Of Postprandial Lipemia As A Cardiovascular Disease Risk factor Using Genetic And Clinical Information: An Artificial Neural network Perspective", Engineering In Medicine And Biology Society 2008.Embs 2008.30th Annual International IEEE Embs Conference Vancouver, British Columbia, Canada, August 20-24, 2008 Pages 4609-4612.

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

[10] 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.

[11] 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.

[12] 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.

[13] 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.

