Heart Disease Prediction System using Multilayered Feed Forward Neural Network and Back Propagation Neural Network

Aditya A. Shinde
NBN Sinhgad School of
Engineering
Pune,Maharashtra

Rahul M. Samant
NBN Sinhgad School of
Engineering
Pune, Maharashtra

Atharva S. Naik
NBN Sinhgad School of
Engineering
Pune, Maharashtra

Sharad N. Kale
NBN Sinhgad School of
Engineering
Pune, Maharashtra

Shubham A. Ghorpade
NBN Sinhgad School of
Engineering
Pune, Maharashtra

ABSTRACT
In today's modern world heart disease is the most lethal one. Diagnosing patients appropriately on a timely basis is the most challenging task. The disease diagnosis is often made based on experience and knowledge of medical practitioners. Due to this, there are chances of unwanted biases, errors and it also takes longer time in the accurate diagnosis of the disease. Medical diagnosis systems play a fundamental role in medical practice and are used by medical practitioners for diagnosis and treatment of various diseases. The proposed system will use the multilayered feed forward neural network and back propagation neural network algorithms for the prediction of heart disease in four stages. The dataset provided by the University of California, Irvine [UCI] machine learning repository is used for training and testing. The dataset consists of 14 attributes of 303 patients including its class label [1]. The accuracy obtained using this approach is 92%.

General Terms
Multilayered Feed Forward, Back Propagation Neural Network, UCI machine learning repository.

Keywords
Heart Disease Prediction, Neural Network, Artificial Intelligence.

1. INTRODUCTION
Throughout history, humans have been affected by life-threatening diseases. Of the countless diseases, heart disease has received a great deal of attention from medical researchers. In India, statistics has shown that about 2 million people die of heart disease every year [2]. There are various factors that are responsible for the occurrence of heart disease. These factors are smoking, high cholesterol, high blood sugar level, inadequate physical exercises unhealthy diet and harmful use of alcohol.

Accurate identification and early treatment of such vital disease is important. The automation of the system is needed to help the medical practitioners for better diagnosis and treatment.

The purpose of the medical diagnosis system is to assist the physicians in defining the risk level of individual patients. Usefulness of knowledge detection from a medical data repository that could benefit medical diagnosis and prediction, patient health planning and progress, health care system monitoring and assessment, hospital and health services management and disease prevention.

In this proposed system, algorithms like multilayered feed forward neural network and back propagation neural network are combined to predict the risk of heart disease in four stages. The dataset used for training and testing of the system is taken from the University of California, Irvine [UCI]. This paper is motivated by these views and aforementioned issues and proposes a set of computational intelligence based on approaches for prediction of heart disease.

2. RELATED WORK
Mrs.G. Subbalakshmi, Mr. K. Ramesh, Mr. M. Chinna Rao proposed a system for heart disease prediction in 2011.The proposed system performs data mining using Naive Bayes classification algorithm. The accuracy achieved was 82.31%. The system is stable and simple to understand. Naïve Bayes classification can often outperform more complicated classification methods easily for large datasets. [3].

Dilip Roy Chowdhury, Mridula Chatterjee, R.K. Samanta proposed a model using artificial neural network for neonatal disease diagnosis in 2011. The model uses multilayer perceptron network. The system achieves 75% accuracy. One hidden layer is used with 11 input layers and 13 output layers. [4]

In 2012, Oleg Yu. Atkova, Svetlana G. Gorokhova, Alexandr G. Sboevc, Eduard V. Generozovd,Elena V. Muraseyeva, Svetlana Y. Moroshkinad, Nadezhda N. Cherniyc proposed a system for diagnosis of coronary heart disease. The system uses artificial neural networks including genetic polymorphisms and clinical parameters. A neural network model with multilayer perceptron with feed forward and statistical back propagation accuracy was improved with genetic algorithm. The accuracy achieved by the system is 64-90%. The model used dataset of 487 patients out of which 327 had CHD and 160 without CHD. [5]

Bhuvaneswari Amma N.G proposed a system for cardiovascular disease prediction using genetic algorithm and

neural network in 2012. The major components of this system are cardiac database, preprocessing engine, weight optimization engine, training engine, weight base, and prediction engine. The classification accuracy of training set is 99%. The classification accuracy of testing set is 94.17%. [6]

Jesmin Nahar, Tasadduq Imam, Kevin S. Tickle, Yi-Ping Phoebe Chen in 2013, proposed computational intelligence techniques for prediction of heart disease for prediction heart disease. Accuracy of 86.77% is achieved. Different computational algorithms on standard UCI dataset are applied. [7]

In 2014, C.Mahesh, K.Kiruthika, M.Dhilsathfathima proposed an expert system for diagnosing Hepatitis B using artificial neural network. The system uses general regression neural network. Accuracy of 83% is achieved. This is a fast learning model and has good convergence with a large number of training examples. [8]

In 2015, Ebenezer O. Olaniyi, Oyebade K. Oyedotun, Abdulkader Helwan proposed a system which used artificial neural network with 1 hidden layer for diagnosis of heart disease. The system has 85% accuracy. [9]

In 2016, Majid Ghonji Feshki and Omid Sujoodi Shijani proposed a system which uses PSO (Particle Swarm Optimization) and feed forward neural network. This system divides the instruction set into two groups of healthy and sick people in the first stage. In the second stage, 8192 subsets of the total features were extracted with the clear cost. In third stage PSO algorithm is applied for all the subsets to find the best subset with highest accessibility and accuracy and low time and cost. Subset includes 8 characteristics of age, sex, exercise tests (slope, old peak, and exang), trestbps (blood pressure), FBS and cholesterol by the PSO algorithm, accuracy have been improved by 2.38% and the accuracy has reached 9l.94%. [10]

3. METHODOLIES USED

There is a significant increase of interest in neural networks over the last decade and are being applied to various extraordinary problems in different domains like medicine, engineering, physics, geology, finance, etc. A neural network is a computational model or a system that provides the functionality to impersonate human brain. It is an adaptive, most often nonlinear system that learns to perform a function (an input/output map) from data.

Multilayer perceptron which makes use of multiple layers of the neural network is created by using the set of various parameters which are selected to adjust the models with the help of correlation between parameters and prediction of the disease. The input/output training data conveys the information required to discover the optimal operating points. A lot of flexibility to achieve any input-output is provided by the nonlinear nature of neural network processing elements. Therefore, the input/output training data is critical in neural network technology.

The system consists of a forward pass and backward pass. The forward pass calculates the output and compares it with the desired outputs. The error between desired and actual output is calculated. The backward pass then uses this error to alter or assign the weights to the network. To reduce the size of the error, Repetition of forward and backward pass is done until the error is low enough. Back propagation neural network is supervised learning and learns with the help of an example. It is widely used because it is easy to understand and can be easily applied. Back propagation neural network is a learning technique which adjusts its weight by transmitting its weight changes from sink to source. The goal of back propagation neural network is to have minimum error about output outputs which are the same as or closer to the target.

The data set is taken from data mining repository of University of California, Irvine (UCI). The study included 303 patients (males 241 —, females —62, mean age: 51.25 ± 9.74 years). The data set consists of 14 attributes which will be used in the system.

Table 1. Attributes Names and Description

Attribute	Description
Age	Age of the patient
Gender	Gender of the patient
Chest Pain	Abnormal flow of blood causing the pain in the chest
Trestbps	Resting blood pressure (in mm Hg on admission to the hospital)
Chol	Serum cholesterol in mg/dl
Fbs	(fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
Restecg	Resting electrocardiographic results --Value 0: normal -- Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) -- Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
Thalach	Maximum heart rate achieved
Exang	Exercise induced angina (1 = yes; 0 = no)
Oldpeak	ST depression induced by exercise relative to rest
Slope	The slope of the peak exercise ST segment --Value1:upsloping --Value2:flat -- Value 3: downsloping
Ca	Number of major vessels (0-3) colored by fluoroscopy
Thal	3 = normal; 6 = fixed defect; 7 = reversable defect
Num	Diagnosis of heart disease (angiographic disease status) -- Value 0: < 50% diameter narrowing -- Value 1: > 50% diameter narrowing (in any major vessel: attributes 59 through 68 are vessels)

3.1 Multilayered Feed Forward Neural Network

Multilayered feed forward neural network algorithm reads the history dataset imported from the University of California, Irvine [UCI] which consists of different parameters that are mentioned in the table above. This algorithm performs normalization/discretization operation for the above data values after reading the dataset. This algorithm then defines the total number of layers L and defines a total number of neurons Ni in each layer. After defining the total number of layers L and the total number of neurons operated in each layer, this algorithm will then generate the neural network.

The random weight values on the axioms of the network are assigned. Input values are assigned to the network on the input layer of the network. The values of each neuron are calculated in the hidden layer using the formula:

$$Val = \sum_{i=1}^{n}(Wij * Lij) \qquad (1)$$

The value of each neuron in the network is calculated with the help of equation (1) till the output layer is reached. Then limiter function is applied to the output value with the help of equation (2).
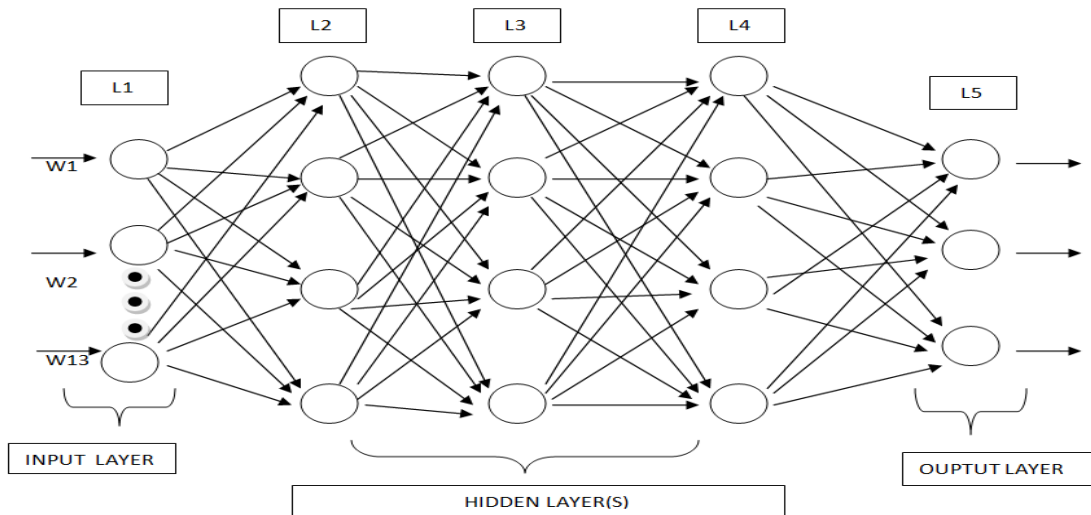
$$f(x) = 1/1 + e^{-x} \qquad (2)$$

Fig 1: Neural Network (13 input neurons, 3 hidden layers with 17 neurons each, 3 output neurons)

3.2 Back Propagation Neural Network

Back propagation algorithm looks for the minimum of error function using the method of gradient descent. The combination of weights which minimizes the error function is considered to be a solution of the learning problem. The output value of the multilayered feed forward neural network algorithm is then calculated using back propagation neural network algorithm to minimize the error function using the equation (3).

$$\Delta = (T - O) * (1 - O) * O \qquad (3)$$

While propagating backwards, the back propagation algorithm makes sure to change the weight values of the neurons according to error gradient descent function. The weight values are recalculated using the equation (4) till the input layer is reached.

$$\sum_{new}^{1} = W * \Delta * input \qquad (4)$$

4. EXPERIMENTATION AND RESULTS

The heart disease prediction system is designed by using NetBeans IDE with the help of Java development kit. This system creates 5 layers including one input layer, three hidden layers and one output layer. The network is trained using a supervised training algorithm. The system takes 14 health parameters as an input from the user to generate input layer. Each hidden layer consists of total 18 nodes. Output layer

contains 4 nodes which gives the output in the form of various stages which are 'Normal', 'Stage 1', 'Stage 2' and 'Stage 3' of the heart disease.

The dataset is imported in the system which is in CSV format. Serialization is applied on the CSV dataset to represent the data in tabular form which is suitable for the training purpose. The imported dataset is trained by each layer until more than 90% accuracy is achieved by the system. Classification is automatically processed for a new pattern to train the system. The developed heart disease prediction system can be used by a physician to automatically diagnose stage of heart disease by entering the basic information that is mentioned in the table 1. The heart disease prediction system minimizes the diagnosis time of a physician and increases the accuracy of diagnosis. The system is not only used for diagnosis but can also be used to store and view the results of the diagnosis for future reference.

Given figure shows the result of the following experiment. Based on the parameters shown in the figure 2, the results show that the person is normal and does not suffer from heart disease and figure 3 shows the person is having stage three heart diseases based on the parameters. From the present experiment, the system automatically identifies whether the person is 'Normal', or suffering from 'Stage 1', 'Stage 2' or 'Stage 3' heart disease. The system gives more precise results once the input parameters of a patient are clearly specified. The accuracy achieved through the proposed system is 92%.

Fig 2: Result of heart disease diagnosis of normal person

Fig 3: Result of heart disease diagnosis with stage three heart disease of a person

5. PERFORMANCE EVALUATION

Table 2 shows the performance of the neural network. The number of neurons used in the hidden layer was obtained by experimenting which number of neurons will represent the best patterns. Learning and momentum rates were introduced in the back propagation neural network. The learning rate determines the velocity at which the network learns the pattern, if the learning rate is minimal. It means that more time is required before the network can learn and on the other hand, if the network learning rate is too high, it makes the network to continuously hop over the optimum weights and the network fails to learn. The momentum rate blocks the network from settling down at the local minima. Therefore, the two parameters, learning rate and momentum rate are experimenting to obtain the best value. The learning rate and momentum rate used in the proposed system are 0.32 and 0.73 respectively, with all hidden neurons. The time for the network to learn was 20 sec at 2000 epoch. The table 2 below shows the result obtained from the multilayered feed forward neural network with its corresponding minimum error curve. Figure 4 shows the graph of Mean Square Error against the epoch.

Table 2. Performance table of the Neural Network

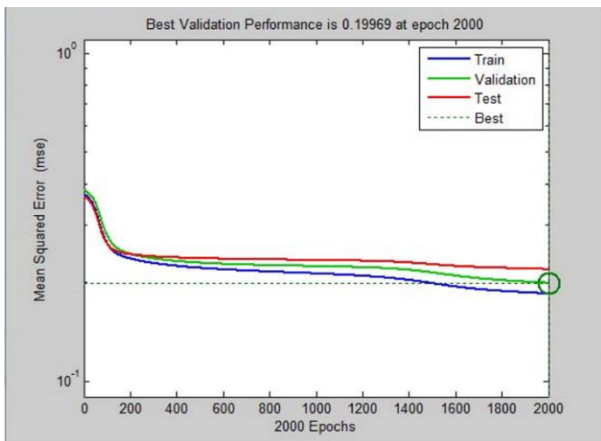| No. of Input Neurons	No. of Hidden Layers	No. of Output Neurons	Epoch
14	3	3	2000
Learning Rate	Momentum rate	Performance	Recognition rate
0.32	0.73	0.199	92%

Fig 4: The Graph of Mean Square Error versus Epoch

6. CONCLUSION

Due to increase in population over the years there is significant increase in the number of people suffering from heart disease. As a result, there is increase in the unavailability of physicians and also errors or inaccuracy in the diagnosis of the disease. This paper proposes/presents a system which can be used to speed up the process as well as to increase the accuracy of diagnosing the disease accurately within minimum time which results in better performance than the traditional diagnosis methods.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] http://archive.ics.uci.edu/ml/datasets/Heart+Disease for dataset. Checked on 2nd February, 2017.

[2] http://www.dnaindia.com/health/report-heart-disease-no1-killer-of-indians-1866673 for heart disease statistics. Checked on 28 February, 2017

[3] Mrs.G.Subbalakshmi, Mr. K. Ramesh, Mr. M. Chinna Rao. 2011. Decision Support in Heart Disease Prediction System using Naive Bayes. Vol: 2 No: 2 Apr-May 2011. ISSN: 0976-5166, 170-176.

[4] Dilip Roy Chowdhury, Mridula Chatterjee, R.K. Samanta. 2011. An Artificial Neural Network Model for Neonatal Disease Diagnosis. International Journal of Artificial Intelligence and Expert Systems (IJAE), Volume (2): Issue (3), 2011. 96-106.

[5] Oleg Yu. Atkova, Svetlana G. Gorokhova, Alexandr G. Sboevc, Eduard V. Generozovd, Elena V. Muraseyeva, Svetlana Y. Moroshkinad, Nadezhda N. Cherniy. 2012. Coronary heart disease diagnosis by artificial neural networks including genetic polymorphisms and Clinical parameters. 0914-5087, 190-194.

[6] Bhuvaneswari Amma N.G. 2012. Cardiovascular Disease Prediction System using Genetic Algorithm and Neural Network.

[7] Jesmin Nahar, Tasadduq Imam, Kevin S. Tickle, Yi-Ping Phoebe Chen. 2013. Computational Intelligence for Heart Disease Diagnosis. 0957-4174. 96-104.

[8] C.Mahesh, K.Kiruthika, M.Dhilsathfathima. 2014. DIAGNOSING HEPATITIS B USING ARTIFICIAL NEURAL NETWORK BASED EXPERT SYSTEM. 978-1-4799-3834-6/14/$31.00.

[9] Ebenezer O. Olaniyi, Oyebade K. Oyedotun, Abdulkader Helwan, Khashman Adnan. 2015. Neural Network Diagnosis of Heart Disease. 978-1-4673-6516-1/15/$31.00. 21-24.

[10] Majid Ghonji Feshki, Omid Sujoodi Shijani. 2016. Improving the Heart Disease Diagnosis by Evolutionary Algorithm of PSO and Feed Forward Neural Network. 978-1-5090-2169-7/16/$31.00. 48-53.