Neural Network Diagnosis of Heart Disease

[1]Ebenezer O. Olaniyi, [2]Oyebade K. Oyedotun, [3]Abdulkader Helwan
[1,2,3] Near East University, Lefkosa, via Mersin 10, Turkey,
[1,2]Department of Electrical/Electronic Engineering
[1,2]Member Center of Innovation for Artificial Intelligence
[1]obalolu117@yahoo.com,[2]oyebade.oyedotun@yahoo.com, [3]abdulkader.helwan@gmail.com

[4]Khashman Adnan
[4]Founder, (and Director) Center of Innovation for Artificial Intelligence
[4]British University of Nicosia, Via Mersin 10, Turkey
[4]adnan.khashman@bun.edu.tr

***Abstract*— Mortality rate increases all over the world on a daily basis. The reason for this could be largely adduced to the increase in the number of patients with cardiovascular diseases. To worsen the case, many physicians have been known for misdiagnosis of patients reporting heart related ailment. In this paper, an intelligent system has been design which will help in effective diagnosis of the patient to avoid misdiagnosis. The dataset of UCI statlog heart disease has been used in this experiment. The dataset is comprises thirteen features which are vital in diagnosis of heart diseases. The system is model on a multilayer neural network trained with backpropagation and simulated on feedforward neural network. The recognition of 85% was obtained from testing of the network.**

Keywords—Heart Disease; Diagnosis; Neural Network; Death rate.

I. INTRODUCTION

Heart can be described as the compound organ in the body that is made up of nerves and muscles. Heart pump 51% of the blood to different parts of the body providing human body renewed materials [1]. Any failure or defect in the heart may result into sudden death. In the united state of America, statistics has shown that about 800,000 people die of heart disease every year [2].This is one of the reasons why researcher has focus more in designing intelligent system that can be used to diagnose heart diseases with high accuracy, to avoid misdiagnosis. Besides, many people are living with heart disease without awareness. This kind of heart disease is referred to as silent killer. Also, heart diseases have resulted into death of several people that have the symptoms of the disease without taking it into consideration. There are features that increase the possibilities of the disease. These are smoking, inadequate physical exercises, high blood pressure, unhealthy diet, harmful use of alcohol, high cholesterol, and high blood sugar level [3],[4],[5]. Cardiovascular heart diseases incorporated with coronary heart, peripheral artery, rheumatic heart, inflammatory heart disease and other forms of heart disease [3]. The other form of heart diseases includes tumor of the heart, vascular tumor of the brain, and disorder of the lining of the heart.

The statistics of the death recorded in the year 2012 caused by cardiovascular diseases are shown in the table below:

Table 1: Statistics of Death Recorded in 2012 due to heart diseases [6].

Types of Cardiovascular Disease	The population of people die from cardiovascular disease
Coronary heart disease	7.4million
Hypertensive heart disease	0.9million
Rheumatics heart disease	0.3million
Inflammatory heart disease	0.4million
Stroke	6.7million
Other forms of heart disease	1.8million
Total Death recorded by CVD	17.5million

The results from the death population table shown above, presented coronary heart disease and stroke to be the leading causes of death out of all types of cardiovascular heart disease. The coronary heart disease can be described as the diseases that affect the blood vessels supplying the heart muscles while the stroke are caused by the disruption of the blood supply to the brain. This may be as a result of the blockage or rupturing of the blood vessel. These two types of cardiovascular heart disease occur mostly in people of age 70years to 79years [6]. The major risk factor of coronary heart disease and the stroke are similar. These are high blood pressure, smoking of tobacco, high blood cholesterol, unhealthy diet, diabetes, physical inactivities, atrial fibrillation and advancing in age.
Due to the high risk involves in the diagnosis of this diseases, there is a need to design an intelligent system that will help in diagnosis of the disease to prevent misdiagnosis. This system can be model by modeling human activities on a computer.
Artificial neural network is the modeling of human brain on the system. This is done in order to make a system to be able to perform the activities of the human brain. The artificial neural network has three parts which are the input layer, hidden layer and the output layer. The input layer is the layer where the input data is presented into the neural network, the input layer is non-processing layer of the neural network. The hidden layer is the layer in between the input layer and the

978-1-4673-6516-1/15/$31.00 ©2015 IEEE

output layer. Hidden layer consist of interconnection element called neurons where the complicated associations between patterns are computed [7]. It is a processing layer of the neural network, it made up of the summation function and activation function. The three layers ; the input, the hidden and output layer interact with each other with the aid of what is referred to as connection weight. The output layer is also a processing layer which comprises the summation and the transfer function. The output layer holds the results of the network.

In this research work, neural network has been used as a tool to diagnose heart disease. The neural network was trained with backpropagation algorithm and tested with feedforward neural network. The recognition rate of 85% was obtained compared with other results of heart disease diagnosis obtained from previous research work.

In the recent research work, Sundar et.al [8] proposed a prototype using data mining techniques namely Naïve Bayes and WAC (weighted associative classifier). This system answers the queries which the traditional likelihood fails to answer. The system detects the likelihood of patient getting a heart disease. The recognition rate of 84% and 78% were obtained from weighted associative classifier and Naïve Bayes. Niti, Anil and Navin (2007) in their work proposed a decision support for heart disease diagnosis using neural network. They trained their system with 78 patient records and errors made by human are avoided [9].

Vikas et al.,[10] provided a survey of current technique of knowledge discovery in databases using data mining technique which will be useful for medical practitioners to make effective decision. The authors used the heart disease database obtained from from UCI machine learning repository. The database contains 13 attributes. They selected 11 attributes for the research work. The authors modeled on Naïve Bayes and J48 decision tree with recognition rate of 82.31% and 84.35% respectively.

The remaining parts of the paper is arranged as follows: Section II material and method, section III is the design of the heart disease neural network, section IV is the evaluation of the performance and section V is the conclusion of the paper.

II. MATERIAL AND METHOD

This dataset was taken from the UCI machine learning repository. The heart disease dataset is made up of 75 raw features from which 13 features were published [11]. These features are very vital in the diagnosis of heart diseases. The features include fasting blood sugar test which must indicate $\prec 120mg/dl$ for a patient with absent test result and test result of $\succ 120mg/dl$ for a patient that has heart disease. Also, a patient that has serum cholesterol greater than 180mg/dl is also considered as heart disease present.

The 13 features considered in this research work are stated below [11]:

Feature 1: Age
Feature 2: Sex
Feature 3: Chest pain type 4values
Feature 4: Resting blood pressure
Feature 5: Serum cholesterol in mg/dl
Feature 6: Fasting blood sugar $\succ 120mg/dl$
Feature 7: Resting electrocardiographic result (value 0,1, 2)
Feature 8: Maximum heart rate achieved
Feature 9: Exercise induced angina
Feature 10: Old peak = ST depression induced by exercise relative to the rest
Feature 11: The slope of the peak exercise ST segment
Feature 12: Number of major vessels (0 - 3) coloured by flourosopy
Feature 13: Thal: 3= normal; 6 = fixed defect; 7 = reversible defect.

A. Data Pre-Processing

In order to increase the stability and performance of the network. There is a need for pre-processing the data. This pre-processing of the data is known as normalization. Normalization of data transforms the input data to the form that is suitable for the network to learn easily. The normalization process input data to optimal training effect the network which considered as input pre-processing function [12], [13]. Normalization can be done by dividing each sample of a feature by their corresponding highest sample value.

$$Normalization = \frac{Sample\ of\ a\ feature}{Highest\ sample\ value} \quad (1)$$

Table 2 shows the result obtained after normalization of the first five features.

Table 2: The Normalized values of the dataset

Features	Sample				
Age	0.9090	0.8701	0.7403	0.8312	0.9610
Sex	1	0	1	1	0
Chest pain type	1	0.7500	0.5000	1	0.5000
Resting blood pressure	0.6500	0.5750	0.6200	0.6400	0.6000
Serum Cholesterol	0.5709	1	0.4628	0.4663	0.4769

III. DESIGN OF THE NETWORK

In this research work, the neural network for diagnosis of heart diseases was design with the heart disease sample obtained from UCI machine learning repository. The dataset set is made up of thirteen (13) features which contain 270 samples. The dataset is divided into training set and testing set. The dataset was divided using ratio 60:40 i.e. 60% of the

dataset for training and 40% of the dataset for testing of the network. This is the standard ratio for dividing dataset in machine learning dataset. The advantage of this division is that it provides adequate data for training and testing the system in a way that it avoid underfitting that may occur if the training dataset is smaller than the testing dataset. Also, if the training dataset is far greater than the testing dataset, this can result into overfitting of the system.

Batch training was employed in which the entire patterns were introduced to the neural network at once. Therefore, at the input layer we have a matrix 13 * 162 for training the network with its corresponding target which forms a matrix 2*162 and matrix 13 * 108 for testing the network after training to determine the performance of the network.

The number of the input neurons used in the design is the number of features of the heart disease dataset. Therefore, the number of the input neurons for the design is thirteen and the number of the output neurons in the output layer are two neurons, which represent whether the heart disease present (1 0) or the heart disease absent (0 1). At the hidden layer and output layer, the sigmoid activation function was used because of its soft switching nature and its non-linearity characteristics. Figure 1 is the architectural representation of the neural network diagnosis of heart disease.

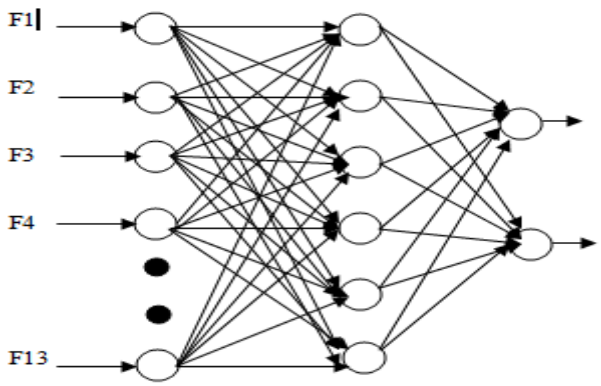

Figure 1: The design of the neural network for diagnosis of heart disease.

In the diagram above, the input of the network dataset is represented by F1, F2 until the last feature which is F13.

III. PERFORMANCE EVALUATION

The number of neurons used at the hidden layer of the research work was obtained by experimenting which number of neurons will best represent the patterns. Also, as the neurons in the hidden layer is varied, the learning rate which is the learning power of the network and the momentum rate which determine the speed of the network are also varied to allow the network to learn and prevent the network from settling at the local minima. The neurons in the hidden layer were experiment from 2 neurons until its reach 6 neurons which are the perfect hidden neurons which can represent the pattern accurately for the system to have a better performance.

The learning rate and momentum rate that was finally used for the network were 0.32 and 0.73 respectively.

Table 3 shows the performance of the neural network and Figure 2 shows the minimum square error against the number of iteration.

Table 3: The performance table of the neural network

No. of Input Layer	13	No. of Hidden Layer	6
No. of output layer	2	Momentum rate	0.72
Learning rate	0.32	Epoch	2000
Performance	0.199	Recognition rate	85%

The recognition rate of this research work was compared with previous work on the diagnosis of heart disease using neural network. Table 4 shows the comparison of the previous work with this current research work.

From the table 4, BPNN has the highest recognition rate compared to other algorithms used by other author considering the same UCI machine learning heart disease database. This shows that BPNN is more effective and accurate in diagnosis of heart disease in medical field as compared with other algorithms previously applied.

Table 4: The comparison table of the research work

KNN	45.67%
Decision Tree	84.35%
Naïve Bayes	82.31%
WAC	84%
BPNN	85%

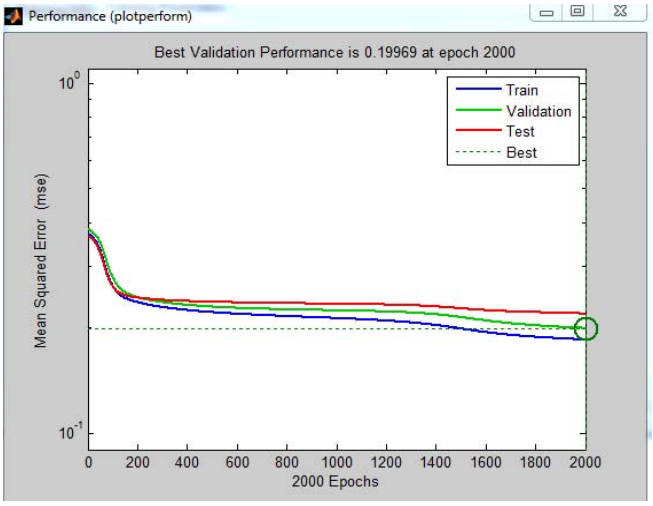

Figure 2: The graph of Minimum Square Error versus Epoch

IV. CONCLUSION

In this research work, discussion has been done concerning the death rate as a result of different types of cardiovascular heart disease. The coronary heart disease leads the population of death in the world followed by stroke. The major factor that caused death from heart disease is misdiagnosis. Therefore, an intelligent system has been designed which will prevent misdiagnosis of heart disease.

In this research work, recognition rate of 85% was obtained and this result was compared with other algorithm to ascertain the best algorithm with best result. Multilayer neural network with trained with backpropagation was discovered to have the best result that is suitable for diagnosis of heart disease without misdiagnosis.

REFERENCE

[1] S. Ghwanmeh, A. Mohammad, A. Ibrahim, "Innovative Artificial Neural Networks-Based Decision Support System for Heart Disease Diagnosis," Journal of Intelligent Learning Systems and Applications, Vol. 5, pp. 176-183, August, 2013.

[2] S. I.McFarlane et.al. "Hypertension in high cardiovascular-risk population", International Journal of Hypertension, Vol. 2011, pp. 1-3, doi 10.4061/2011/746369

[3] H. Masethe, M.A. Masethe, "Prediction of Heart Disease Using Classification Algorithms," Proceedings of the World Congress on Engineering and Computer Science, Vol. II, October 2014.

[4] V. Chaurasia,"Early Prediction of Heart Disease Using Data Mining", Carribbean Journal of Science and Technology, Vol. 1, pp 208-217, 2013.

[5] D. S Chaitrali and A.S Sulabha, "A Data Mining Approach for Prediction of Heart Disease Using Neural Networks", International Journal of Computer Engineering and Technology, Vol. 3, pp. 30-40, 2012.

[6] J. V. Goethe, "Cardiovascular Disease," World Health Organization, 2012

[7] A. T. C. Goh, "Back-propagation neural networks for modeling complex systems," Artificial Intelligence in Engineering, vol. 9, pp. 143-151, 1995.

[8] N. A Sundar, P.P Latha and M.R. Chandra "Performance Analysis of Classification Data Mining Techniques Over Heart Disease Data Base", International Journal of Engineering Science and Advance Technology, Vol. 2, pp. 470-478, June, 2012.

[9] N. Guru, A. Dahiya and N. Rajpal, " Decision Support System for Heart Disease Diagnosis using Neural Network", Delhi Business Review,Vol.8, pp.99-101, 2007.

[10] V. Chaurasia and S. pal, "Data Mining Approach to Detect Heart Disease," Vol. 2, Issue 4, pp. 56-66, 2013.

[11] fftp.ics.uci.edu/pub/machine-learning-databases (last accessed: February 10, 2015).

[12] F.S. Gharehchopogy, M. Molany and F. Dabaghchi, "Using Artificial Neural Network in Diagnosis of Thyroid Disease: a case Study", International Journal on Computational Science and Applications, Vol. 3 pp. 49-61, August 2013.

[13] J.Sola, J.Sevilla, "Importance of input data normalization for the application of neural networks to complex industrial problems", IEEE Transactions on Nuclear Science, Vol. 44, No. 3, pp. 1464 – 1468,1997.