International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 4 No. 4 Dec. 2013, pp. 679-689

http://www.issr-journals.org/ijias/

A New Generalized Regression Artificial Neural Networks Approach for Diagnosing Heart Disease

Behnam Zebardast[1], Ali Ghaffari[2], and Mohammad Masdari[1]

[1]Department of Computer Engineering,
Science and Research Branch, Islamic Azad University,
West Azerbaijan, Iran

[2]Department of Computer Engineering,
Tabriz Branch, Islamic Azad University Tabriz, Iran

ABSTRACT: Artificial Neural Networks (ANNs) play an important role in the field of medical science in solving health problems and diagnosing diseases both in critical illnesses and in common diseases. Since it is important to diagnose accurately the people' disease condition, therefore for the precisely diagnosing those condition, we must use appropriate methods that to minimize the errors in diagnosis. So, using an appropriate method to diagnose heart disease and to prevent complications of the disease is an important step toward patients' improvement. Therefore, in this paper the presence or the absence of heart disease of the four datasets using Generalized Regression Neural Networks (GRNN) will be discussed. Each of the four datasets contains of 14 features that they are used to diagnose heart disease with GRNN. In this paper, GRNN have been implemented in MATLAB environment. The aim is maximizing the precision of measurement in accurately diagnosing heart disease in the process of training and testing. By comparing the results of each dataset, we found the best accuracy in the training phase that is equal to 100% which belongs to Switzerland and Long Beach VA datasets, and the best accuracy in the testing phase belongs to the Cleveland dataset that is equal to 96.6667%.

KEYWORDS: Generalized Regression Neural Network (GRNN), heart disease, datasets, training accuracy, test accuracy.

1 INTRODUCTION

Nowadays, heart diseases are the most common in the industrialized and the developing countries followed by the high mortality and morbidity. The heart is a muscle which pumps blood from the lungs to the lower part of the body around the body. If blood circulation in the body be inefficient, another organism like brain suffers from this problem. If the heart stops working, death can occur in minutes. Life is totally dependent on the work of the heart, a number of factors that increase the risk of heart disease include family history, smoking, poor diet, high blood pressure, high blood cholesterol, obesity, physical inactivity, and ... [1]. Factors such as these are used for the analysis of heart disease, In many cases, the diagnosis is generally based on the results of the test, the patient and physician, the diagnosis is a complex task that requires high skills and experience, early diagnosis and intensive medical care of patients with this disease can largely prevents them from sudden death [1], [2]. Different methods have been proposed to solve this problem. One of these methods is ANNs. In recent years, machine learning techniques are widely used in prediction and diagnosis, especially in medical diagnosis, ANNs are 'hot' topics for medical research, particularly in the fields of radiology, urology, cardiology, cancer, etc. One of the major problems is a medical diagnosis because the disease is diagnosed by a physician who requires medical attention and previous experience [3]. Several research groups around the world working on the development of ANNs in medical field. An ANN is used for the purpose of enhancing the precision, accuracy and objectivity in medical diagnose. The purpose of diagnosing a disease by ANN is to achieve a high rate of accuracy and tangibility of medical diagnose in the training and testing of a dataset [3]. It is a technique that attempts to simulate the behavior of the human brain's neurons. This technique has been used for several recent years in many cases including the detection and estimation and prediction of the capabilities of ANNs [4]. The most important advantage of using ANNs is that they are used for solving the problems that are too complex for conventional technologies that there is no algorithmic solution or solutions are highly complex. This attributes is widely used in medical fields [5]. ANNs have been successfully used in various areas of medicine like diagnosing systems [6], the analysis of medicine [7], medical images' analysis [8]. Diagnosis of the illness in patients is the most important feature for medical treatment in which ANNs are suitable system for the operation. How to obtain information on individual patients is of crucial importance. The more patients are used in implementing an ANN, the better performance of the ANN will appear [3], [9]. Another advantage of ANNs regarding the diagnosis of diseases is that feature like fatigue, exhaustion, emotional state have no influence on them. Its reason is the implementation of ANNs on computer systems; these cases are vacuous for computer systems. ANNs are able to learn and an ANN can be trained before performing major operation, and use it after training and testing. Learning in ANNs can be done by adjusting the input weights and the act of learning can be done, if weight in the input changes in the case that the desired results happen [10], [11]. There are many relationships between stored data on the computer that at first the designed ANNs must be trained for learning patterns and the relationship among data and extracting hidden information and after those procedures apply those ANNs [12], [13]. ANNs are used in any situation where a relationship between some of the variables considered as input and some of them as output variables (diagnosis) [14]. In this paper, we tried to diagnose heart disease to a degree of accuracy using the GRNN and find maximum optimized answer. Having information (Data's) is one of the critical issues in the use of ANNs. In this paper, we used the 4 dataset named Cleveland, Hungarian, Switzerland and the Long Beach VA with specific examples that the GRNN are trained by these data. GRNN have been implemented in MATLAB environment.

In [15], Researchers diagnosed heart disease with GRNN and Radial Basis Function (RBF) of ANNs which were used from a data set containing 300 samples. In This method features such as high blood pressure, diabetes, hyperthyroidism, chest pain, shortness of breath, smoking, tobacco, alcohol and … are used for patients. In this way the data have been normalized in the range of 0 and 1. After training the data, the results showed that the performance of RBF has been much better than GRNN for the diagnosis of heart disease. We discuss in the second section of the paper about datasets that is used for heart disease diagnose. In the third section of the paper we will discuss GRNN and the proposed method and the results of GRNN training and testing data. As soon as we will deal with the conclusions and future works in fourths section.

2 DATASETS FOR THE DIAGNOSIS OF HEART DISEASE

According to statistics from the World Health Organization, heart disease is the most common cause of death among other diseases [16]. Early diagnosis and intensive medical care of patients with this disease can largely prevents them from sudden death. Healthcare management of heart failure patients is a very complicated and difficult task. It depends on many factors and how to approach life. For example, age, tobacco use, personal fitness, blood pressure, stress, etc. are some of these factors [17.18]. Without a doubt, one of the most important parts of the development of ANNs is selecting the data. Our collected data which are from four hospitals in Cleveland, Hungarian, Switzerland and the Long Beach VA are from 14 features for doing this research. The data can be used for training and testing the GRNN. Table 1 shows all 4 dataset and the number of each dataset has been written in Table 1.

Table 1. Dataset names and the number of each dataset

Dataset	Total
Cleveland	303
Hungarian	294
Switzerland	123
Long Beach VA	200

Each dataset contains of 14 features that they are used to diagnose heart disease with GRNN. Table 2 shows the 14 used features in this paper.

Table 2. The names of used features for diagnosing heart disease, along with the kind of their values

No. of Feature	Feature	Descriptions and Feature values
1	Age	Numerical values
2	Sex	Male=1 Female=0
3	Chest pain type	Typical angina=1 Atypical angina=2 Non-angina pain=3 Asymptomatic=4
4	Resting blood pressure	Numerical values in mm hg
5	Serum cholesterol	Numerical values in mm/dl
6	Fasting blood sugar	Fasting blood sugar > 120 mg/dl (True=1; False=0)
7	Resting electrographic results	Normal=0 Having ST-T wave abnormality=1 Left ventricular hypertrophy=2
8	Maximum heart rate achieved	Numerical values
9	Exercise induced angina	Yes=1 No=0
10	ST depression induced by exercise relative to rest	Numerical values
11	Slope of the peak exercise ST segment	Up sloping=1 Flat=2 Down sloping=3
12	Number of major vessels colored by fluoroscopy	Value = 0-3
13	Defect type	Normal=3 Fixed defect=6 Reversible defect=7
14	Diagnosis of heart disease	Yes=1 No=0

According to Table 2, the first 13 features are used as inputs for GRNN and the feature number 14 called heart disease' diagnosis used as output's feature. It is used as an attribute to clarify persons' healthiness or sickness. As shown in Table 2, the values of the feature number 14 are 0 and 1, 0 means as being healthy and 1 as being patient.

3 PROPOSED MODEL

Generally, the ANNs are parallel processing systems which are used for recognizing complex patterns in the between data. In fact, an ANN is an information processing system consisting of some common features with biological neural networks. Then each ANN consists of a set of neurons with special makeup. The chief part of an ANN composed of neurons and communication links between them. The neurons are interlinked processing elements which interacted coordinately to solve an issue [19]. ANNs are capable of learning. This capability through experience and generalizability in solving new problems is the advantage of this method over other methods [20].

The proposed model in this paper is to diagnose heart disease using GRNN. GRNN is often used in function approximation [21]. In fact, these types of ANNs are a kind of ANNs based on RBF. GRNN are trained and respond quickly. The main advantage of GRNN model is that it is much faster than trained sample back- propagation [21]. GRNN is exactly four layers including an input layer, a layer called the radial basis (a hidden layer), a layer of regression Neurons, and an output layer [22]. Neurons of radial basis layer (hidden layer) recognized as instructional centers for data. The hidden layer transfer function is non-linear function radbas (Gaussian). The number of neurons in this layer is equal to the number of training data. In GRNN, regression layer must be exactly one neuron more than the output layer. Regression layer composed of linear units (linear neurons). The output layer transfer function is purlin linear function. Formula (1) is a predicted output by GRNN.

$$y_i = \frac{\sum_{i=1}^{n} h_i w_{ij}}{\sum_{i=1}^{n} h_i} \tag{1}$$

There are two sections in the regression layer. One section is the sum of the denominator unit and the other is the sum of the fraction unit. According to formula (1), the upper part is the sum of the fraction unit and lower part is the sum of denominator in the regression layer. Output layer divides the obtained value of the sum of the fraction on the sum of the denominator unit. W_{ij} is the output corresponding aim (purpose) with training of the input vector X_i and the output j. h_i is the output of each neuron in hidden layer in the form of formula (2) that it is in the form of nonlinear transfer function in Gaussian.

$$h_i = \exp\left(\frac{-D_i^{\,2}}{2(spread)^2}\right) \tag{2}$$

According to formula (2), spread is a Constant value to increase the accuracy of training and testing and it is used for neurons in hidden layer. D_i^2 is squared distance between the input vector X and the training vectors u in formula (3).

$$D_i^{\,2} = u_i (x - u_i)^T (x - u_i) \tag{3}$$

According to formula (3), X is the input vector and u_i is the center of the hidden ith neuron.

Since the range of variation of the reviewed data is very high, and also the nature of transfer function using ANNs often need to have their input in limited scope, these functions are usually flat in small and large in quantities, and their sequences are less sensitive to changes in input values.Thus, their normalization value is often used instead of data values. Employing methods to put input data in a limited range is called normal values which are one of the pre-processing methods. Other advantage of putting inputs in a limited range is to inhibit the excessive growth of the weight. This limitation reduces the convergence time in ANN and minimizes achievable error. In other words, entering data in raw form reduces the speed and accuracy of the ANN, to prevent this, the data must be normalized before training [23], [24]. Different methods were tested to normalize the data, finally using formula (4) the data were normalized in the range [-1, 0] to increase the training performance.

$$In = \frac{In - Max_{In}}{Max_{In} - Min_{In}} \tag{4}$$

Input data in GRNN are classified into two training and testing categories. We have chosen randomly 90% of all the data as training data. After that GRNN was trained by the data, they find their final weights, so that the GRNN gives the least error for the training of data. The rest of the data that are 10%, after that the GRNN was trained by the training data to achieve the minimum error, this 10% of the data that have not a role in the training given as input to the GRNN, the GRNN response and the desired response is compared and in this way the trained GRNN is measured. If GRNN that are tested in a time period, answer correctly the training is over, otherwise; GRNN training begins again. When at the final step the program got out of the training and on all inputs produced the right outputs, in this case, it is clear that the weights of GRNN are adjusted properly, so from now on we will use these values to diagnose the action.

We considered 4 datasets in this paper. The number of data in GRNN for training is 90%, and for testing is 10% which they are classified as follows in Table 3.

Table 3. The number of training and testing of data in each datasets in GRNN

Datasets	Total	Number of training samples	Number of tested samples
Cleveland	303	273	30
Hungarian	294	265	29
Switzerland	123	111	12
Long Beach VA	200	180	20

According to the given description, (Fig. 1.) shows the used architecture on GRNN to diagnose heart disease.

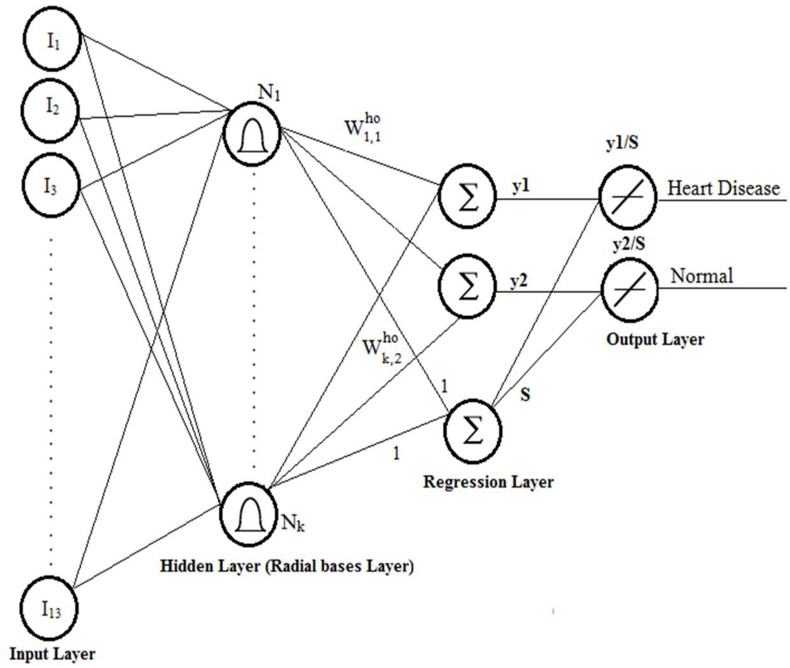

Fig. 1. Proposed scheme for GRNN to diagnose heart disease

According to (Fig. 1.), GRNN has an input layer, a hidden layer, a regression layer and an output layer. The Number of neurons in the input layer is equal to the first 13 features in Table 2 (I_1 to I_{13}), the number of hidden layer neurons is equal to N_k (N_1 to N_k) which is equal to the number of training data. The number of neurons in the regression layer is equal to 3. Because the number of neurons in this layer is a unit more than the neurons of the output layer, so the incoming weights of the third neuron in the regression layer is always equal to 1. y1 and y2, individually are the sum of the fractional of the regression layer based formula (1). Also S is the sum of the denominator of the regression layer based formula (1). And the number of output layer neurons is the number of defined classes containing 2 neurons that is the existence of disease (1) and the lack of it (0).

The construction of the implementation of newgrnn function is [25] that this function is used to construct GRNN. Also, the spread value has been set 0.3 in GRNN. Determination of such value is basically experimental that the amount of training and testing accuracy increases based on the value. We showed the accuracy rate of training, testing, and the points (actual data) that diagnosed properly by GRNN for the diagnosis of heart disease for all four dataset based on to the proposed GRNN scheme and used functions in GRNN.

Fig. 2 and Fig. 3, show the rate of training' accuracy, testing, the points (actual data) that diagnosed properly for heart disease diagnose by GRNN for Hungarian dataset.

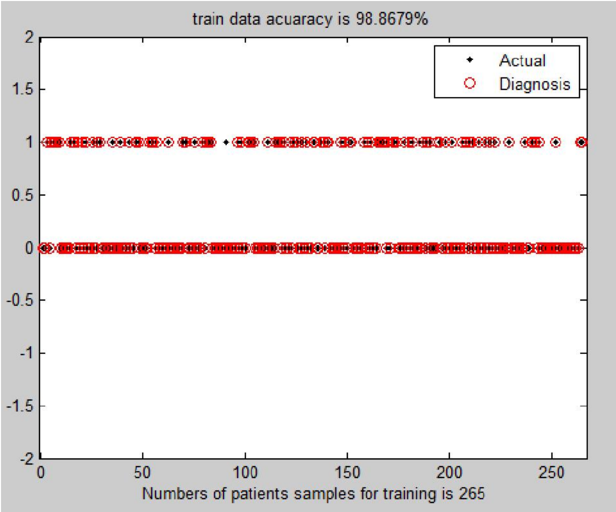

Fig. 2. The rate of accuracy of training in Hungarian dataset

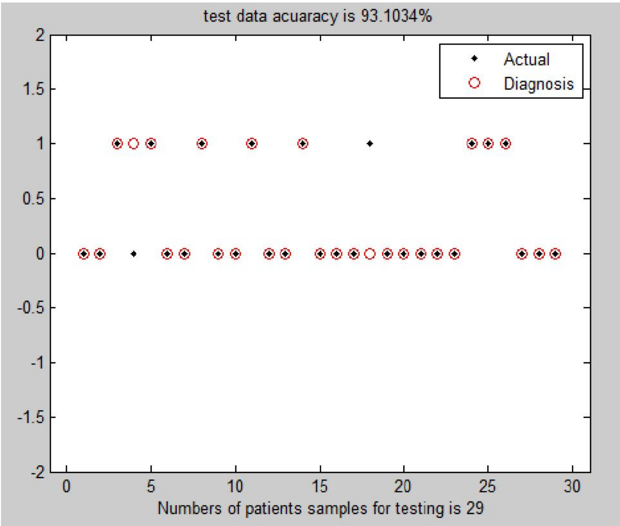

Fig. 3. The rate of accuracy of test in Hungarian dataset

According to Fig. 2 and Fig. 3 as we can see the training and testing rate of accuracy are 98.8679% and 93.1034%, respectively.

The horizontal axis (in Fig. 2, sets of training data and in Fig. 3 set of test data) and the vertical axis (0 and 1 points) are the black points (actual data) and red points (diagnostic data). The black points include the values of 0 (non-disease) and 1 (disease) which they are covered by red points. The red points are the output of GRNN in training and testing stage. This cover indicates the correct diagnose of the data (0 and 1 points) by the output of the GRNN in training phase in Fig. 2 and testing in Fig. 3. The procedures are the same for all datasets in testing and training.

Fig.s 4 and 5 show the rate of training' accuracy, testing, the points (actual data) that diagnosed properly for heart disease diagnose by GRNN for Cleveland dataset.

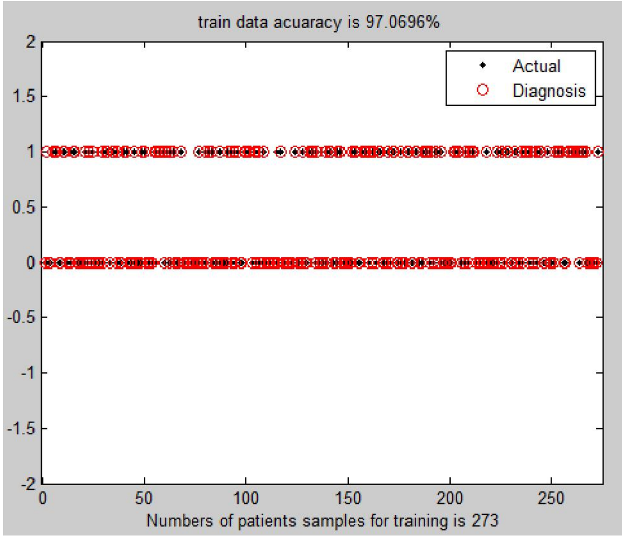

Fig. 4. The rate of accuracy of training in Cleveland dataset

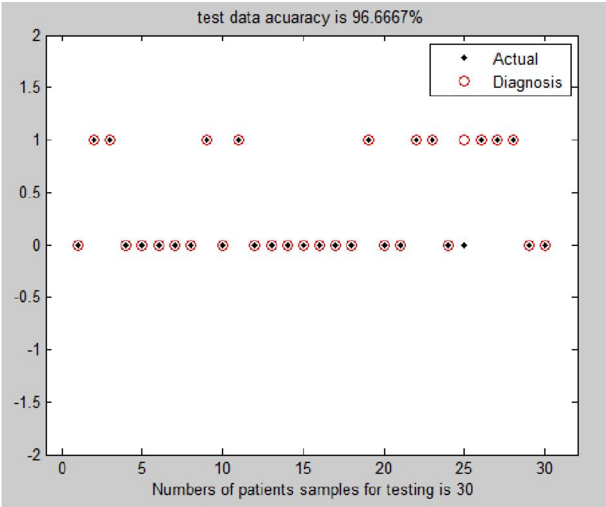

Fig. 5. The rate of accuracy of test in Cleveland dataset

According to Fig. 4 and fig.5 we can see that the training and testing accuracy rate are 97.0696% and 96.6667%, respectively.

Fig. 6 and Fig.7 show the rate of training' accuracy, testing, the points (actual data) that diagnosed properly for heart disease diagnose by GRNN for Switzerland dataset.

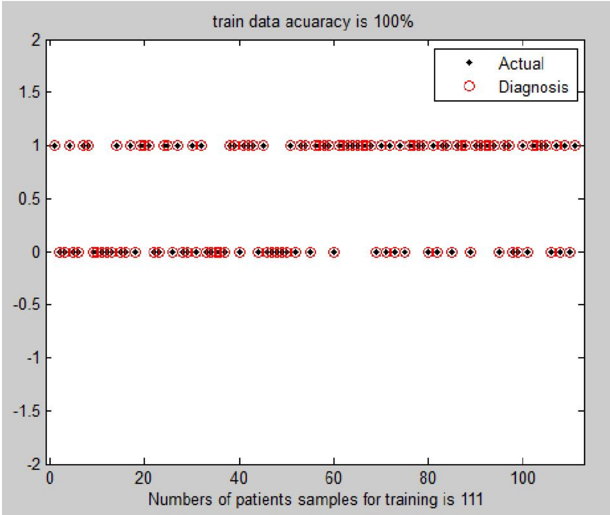

Fig. 6. The rate of accuracy of training in Switzerland dataset

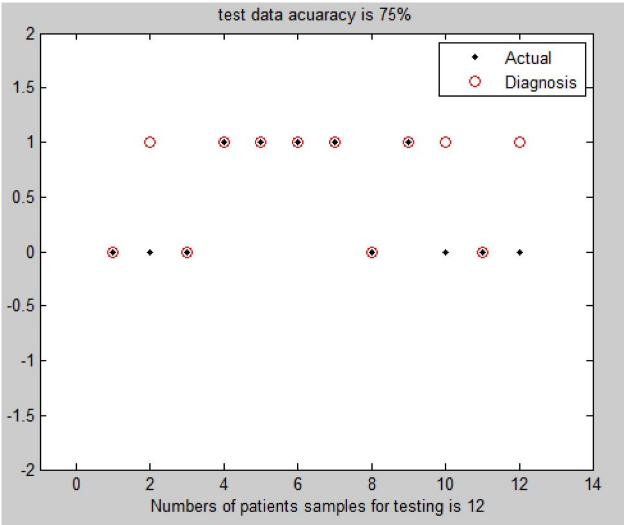

Fig. 7. The rate of accuracy of test in Switzerland dataset

According to Fig. 6 and Fig.7 we can see that the training and testing accuracy are 100% and 75%, respectively. Fig. 8 and Fig. 9, show the rate of training' accuracy, testing, the points (actual data) that diagnosed properly for heart disease diagnose by GRNN for Long Beach VA dataset.

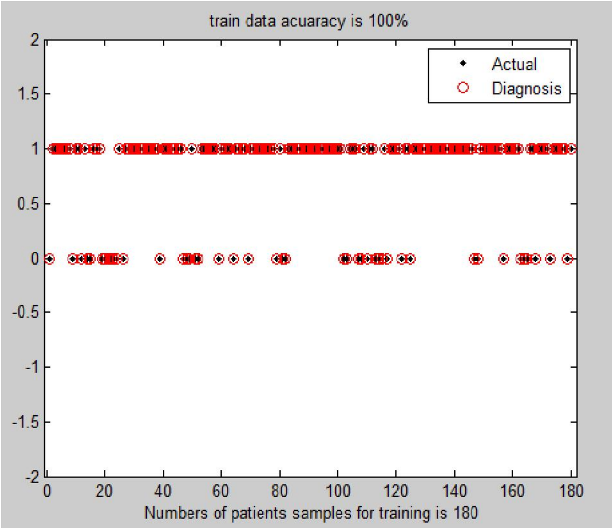

Fig. 8. The rate of accuracy of training in Long Beach VA dataset

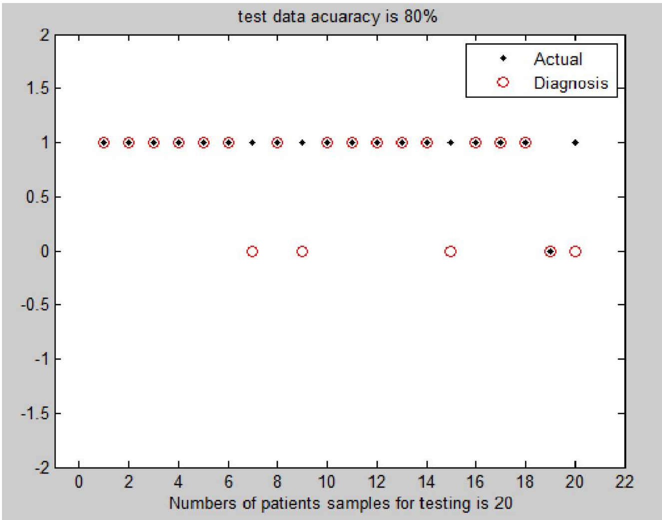

Fig. 9. The rate of accuracy of test in Long Beach VA dataset

According to Fig. 8 and Fig. 9 we can see that the training and testing accuracy are 100% and 80%, respectively. Comparing the accuracy of training and testing of any datasets, we can say that Switzerland and Long Beach VA datasets have the best diagnose accuracy on the training phase, which is equal to 100%, even including the smallest samples of patients. In other words, they have been able to diagnose all the training data. But the best accuracy in the testing phase belongs to Cleveland dataset which is equal to 96.6667%.

4 Conclusion and Future Works

In this paper, we tried to use the GRNN to provide a method for heart disease diagnose. So that we can reduce human error in diagnosing of heart disease using GRNN to solve significantly one of the problems i.e. the heart disease diagnose as the first step. Then, as a result of using this way the patient can prevent the so-called disease and maintain his/her health. We used 4 datasets with the same features and different samples for the work using MATLAB environment to implement GRNN. The GRNN consists of four layers, an input layer, a layer called radial basis (a hidden layer), a layer of regression Neurons, and an output layer. We used non-linear transfer function of radbas (Gaussian) for the hidden layer and a linear function purelin for regression layer and output layer. After the normalization of the data, we divided the data into training and testing data that 90% of total data were for training and 10% for testing. As a result, the training and testing of each dataset was determined that the best accuracy in the training phase belongs to Switzerland and Long Beach VA datasets that is equal to 100%, and the best accuracy in the testing phase belongs to Cleveland dataset that is equal to 96.6667%. GRNN can take advantage of this new model to replace the traditional methods for detecting heart disease associated with human cited errors. It reduces the human cost that is a major contribution in the field of medical science. What we can do in the future is that other models ANNs such as multilayer perceptron of ANNs, probabilistic neural networks and RBF of ANNs can be also used in training and testing phase to achieve better diagnose accuracy, then each of these models can be compared with each other, especially with GRNN to see which one has better diagnose accuracy.

REFERENCES

[1] C.S. Dangare, S.S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", *International Journal of Computer Applications*, vol. 47– No. 10, pp. 44-48, June 2012.

[2] X.Yanwei, J. Wang ,Z. Zhao, Y.Gao, "Combination data mining models with new medical data to predict outcome of coronary heart disease", *IEEE, Proceedings International Conference on Convergence Information Technology*, Gyeongju, pp. 868–872, 2007.

[3] N. Ganesan, K. Venkatesh, M.A. Rama, A. Malathi Palani, "Application of Neural Networks in Diagnosing Cancer Disease Using Demographic Data", *International Journal of Computer Applications*, vol. 1, no. 26, pp. 76-85, 2010.

[4] F.S. Gharehchopogh, "Neural Network Application in Software Cost Estimation: A Case Study", *International Symposium on Innovations in Intelligent Systems and Applications (INISTA 2011),* pp. 69-73, IEEE, Istanbul, Turkey, 15-18 June 2011.

[5] W. Sibanda, P. Pretorius, "Novel Application of Multi-Layer Perceptrons (MLP) Neural Networks to Model HIV in South Africa using Seroprevalence Data from Antenatal Clinics", *International Journal of Computer Applications,* vol. 35, no. 5, pp. 26-31, December 2011.

[6] Q. K. Al-Shayea, "Artificial Neural Networks in Medical Diagnosis", *International Journal of Computer Science Issues*, Vol. 8, No. 2, pp. 150-154, March 2011.

[7] J.L. Patel, R.K. Goyal, "Applications of artificial neural networks in medical science," *Curr. Clin. Pharmacol*., 2(3), pp. 217-226, 2007.

[8] J. Jiang, P. Trundle, J. Ren, "Medical image analysis with artificial neural networks", *Computerized Medical Imaging and Graphics,* pp. 617-631, 2010.

[9] S.M. Patil, R.R. Mudholkar, "An Osteoarthritis Classifier Using Back Propagation Neural Network", *International Journal of Advances in Engineering & Technology (IJAET),* Vol. 4, pp. 292-301, Sept 2012.

[10] F.S. Gharehchopogh, E. Ahmadzadeh, "Artificial Neural Network Application in Letters Recognition for Farsi/Arabic Manuscripts", *International Journal of Scientific & Technology Research*, vol. 1, No. 8, pp. 90-94, September 2012.

[11] S. Mozaffari, K. Faez, V. Margner, H. El-Abed, "Lexicon reduction using dots for off-line Farsi/Arabic hand written word recognition," *Pattern Recognition Letters*, vol. 29, No. 6, pp. 724-734, April 2008.

[12] S. Sathasivam, "Application of neural networks in predictive data mining", *2nd international conference on business and economic research, Langkawi Kedah, Malaysia,* pp. 371-376, 14-16 March 2011.

[13] W. Gao, "New Evolutionary Neural Networks", *International Conference on Neural Interface and Control, Wuhan Polytech. Univ., China,* pp. 167-171, 26-28 May 2005.

[14] P.A. Maiellaro, R. Cozzolongo, P. Marino, "Artificial Neural Networks for the Prediction of Response to Interferon Plus Ribavirin Treatment in Patients with Chronic Hepatitis C", *Current Pharmaceutical Design*, Vol. 10, No. 17, pp. 2101-2110, 2004.

[15] S. A. Hannan, R. R. Manza, R. J. Ramteke, "Generalized Regression Neural Network and Radial Basis Function for Heart Disease Diagnosis", *International Journal of Computer Applications*, vol. 7, No. 13, October 2010.

[16] N. Elfadil, A. Hossen, "Identification of Patients With Congestive Heart Failure Using Different Neural Networks Approaches", *Journal Technology and Health Core*, vol. 17, No. 4, December 2009.

[17] F.S. Gharehchopogh, Z.A. Khalifelu, "Neural Network Application in Diagnosis of Patient: A Case Study", *IEEE, International Conference on Computer Networks and Information Technology (ICCNIT 2011),* Abbottabad, Pakistan, pp. 245-249, 11-13 July 2011.

[18] V.G. Koutkias, I. Chouvarda and N. Maglaveras, "Multi-Agent System Architecture for Heart Failure Management in a Home Care Environment", *IEEE, Computers' in Cardiology*, vol. 30, pp. 383-386, 2003.

[19] S. M. Jadhav, S. L. Nalbalwar, A.A. Ghatol, "Artificial Neural Network Models based Cardiac Arrhythmia Disease Diagnosis from ECG Signal Data", *International Journal of Computer Applications*, vol. 44, No. 15, pp. 8-13, April 2012.

[20] G.T. Tsenov, V.M. Mladenov, "Speech recognition using neural networks", *Neural Network Applications in Electrical Engineering (NEUREL), 10th Symposium on, Belgrade,* pp. 181-186, 23-25 Sep. 2010.

[21] J. Parojcić, S. Ibrić , Z. Djurić, M. Jovanović , O.I. Corrigan, "An investigation into the usefulness of generalized regression neural network analysis in the development of level A in vitro-in vivo correlation", *European journal of pharmaceutical sciences,* vol. 30, pp. 264-272, 2007.

[22] O. Er, N. Yumusak, F. Temurtas, "Chest diseases diagnosis using artificial neural networks", *Expert Systems with Applications,* Vol. 37, pp. 7648–7655, 2010.
[23] S. Chittineni, R. B. Bhogapathi, "A Study on the Behavior of a Neural Network for Grouping the data", *International Journal of Computer Science,* vol. 9, No. 1, pp. 228-234, January 2012.
[24] A. Roy, D. Dutta, K. choudhury, "Training Artificial Neural Network using Particle Swarm Optimization Algorithm", *International Journal of Advanced Research in Computer Science and Software Engineering,* Vol. 3, No. 3, pp. 430-434, March 2013.
[25] Demuth, Beale, *Neural Network Toolbox for Use with MATLAB, User's Guide,* Version 4, The MathWorks, Inc. 3 Apple Hill Drive Natick, MA 01760-2098, 840 pages, 2002.