IEEE International Symposium on Industrial Electronics (ISIE 2009)
Seoul Olympic Parktel, Seoul, Korea July 5-8, 2009

Multilayer Perceptron Neural Network with Supervised Training Method for Diagnosis and Predicting Blood Disorder and Cancer

Vahid Aeinfar[1, 2]*, Hoorieh Mazdarani[1, 2], Fatemeh Deregeh[1, 2], Mohsen Hayati[1, 2], Mehrdad Payandeh[3]
[1] Electronic Department, Faculty of Technology, Razi University, Kermanshah, Iran.
[2] Computational Intelligence Research Center, Razi University, Kermanshah, Iran
[3]Hematology-Oncology and stem cell transplantation research center,
Shariati Hospital, University of Medical Sciences, Tehran, Iran
* E-mail: V.Aeinfar@Modares.ac.ir

Abstract- This paper represents a novel use of artificial neural networks in medical science. The proposed technique involves training an MLP with BP learning algorithm to recognize the pattern of diagnosing and predicting five blood disorders, through the results of blood tests. The blood test parameters and diagnosis of physician about the diseases for 450 cases of patients from Taleghani hospital in Kermanshah, Iran, are used in the supervised training method to update the network parameters. This method was implemented to diagnose these disorder and cancer: Megaloblastic Anaemia, Thalassemia, Idiopathic thrombocytopenic pupura (ITP), chronic myelogenous leukemia and Lymphoproliferative.

I. INTRODUCTION

One of the major problems in medical life is setting the diagnosis. A lot of applications tried to help human experts, offering a solution. This paper describes how Artificial Neural Networks can improve this domain [1].

On average, the human body contains five liters of blood, and your red blood cells are replaced every 120 days. Blood diseases can range from anemia, which is common, to rare disorders that affect only a few. Many different diseases affect blood. Many people have some form of blood disease, either detected or not. In the United States alone approximately 72,000 people have sickle cell anemia with about 2,000,000 carrying the trait. There are 20,000 hemophilia patients in the U.S. Each year, nearly 27,000 adults and more than 2,000 children in the United States learn that they have leukemia (statistics are from NIH and Cancernet web sites)

The health of population, which is based primarily on the result of medical research, has a strong impact upon all human activities. The hematologic disorders are one of the important branches of internal medicine, that with the wide progression that implemented in recent years in this field, we think that a special attention for improving conventional methods is needed. At to this background, we presented a new idea in diagnosis of wide spread of these disorders.

We know that in medical sciences, the good interpretation of data and setting the correct and early diagnosis are very important which can be the base of a good and effective treatment, especially in hematology. (Like other fields of internal medicine disorders) Medical decision making becomes a very hard activity because the expert human, who have to make decisions, can hardly process the huge amounts of data and usually suffering from absence of good and accurate analysis of these laboratory documents; so they need a tool that helps them to make a good decision. They could use some expert systems or Artificial Neural Networks, which are part of Artificial Intelligence [1].

According to this idea and for the best doing of referral system in decision-making for the patients, we presented a new method with a high quality and quantity analytic potency that resolves these weaknesses.

Because of that the number of hematologists (Physicians specialized in hematology, Hematology is the branch of biology, pathology, clinical laboratory, internal medicine, and pediatrics that is concerned with the study of blood, the blood-forming organs, and blood diseases. Hematology includes the study of etiology, diagnosis, treatment, prognosis, and prevention of blood diseases. The lab work that goes into the study of blood is performed by a Medical Technologist.) is limited and in most small towns and clinics there is not any physician, this method can be very applicable and can be used in any general hospital, clinic and even in laboratories for primary diagnosing and sending to hematologists.

At first we searched for disorder diagnosis and predicting with Artificial Neural Network, and we find that maximum research and solutions in medicine branch are done in very few disorder.

For example most of Artificial Neural Network was used to detecting these disorder: breast cancer (that intends to an integrated view of implementing automated diagnostic systems for breast cancer detection[2,3,4,5].), prostate cancer (using cohorts examined by extended biopsy, they developed and validated multivariate models predicting prostate cancer on initial biopsy and examined whether these extended biopsy-based models [6,7] and neural network model which is compared with Logistic Regression to predict this cancer [8,9,10]), heart disease like coronary artery disease [11], or classify heart disease [12] , Alzheimer's disease [13,14], brain disease such as A biomedical system based on fuzzy discrete hidden Markov model for the diagnosis of the brain diseases [15], cervical cancer [16], dermatologic diagnosis in order to improve diagnosing [17], diabetes such as Prediction of Progression of Diabetic Nephropathy [18], optic nerve disease

978-1-4244-4349-9/09/$25.00 ©2009 IEEE

[19],Ovarian cancer [20] , pancreatitis and pancreatic cancer [21].

The outline of this paper is as follows: in Section 2 details of ANN method were derived and we explained our using architecture. Section 3 presents our numerical result for the various cases of Artificial Neural Networks that we designed. Section 4 presents conventional method in order to compare the designed ANN performance in predicting mentioned disorder, with the classical methods such as statistic solutions, multivariable nonlinear regression (with SPSS V.13 software) is used. Results of this method completely presented in this section. The last section explained comparison of these two method results to illustrate the efficiency of neural networks, fast convergence and low use of memory for any kind of diagnose and predicting disorder.

II. METHOD

A. *Neural Networks*

Neural Networks were composed of simple interconnected processing elements, called *neurons*, which is operating in parallel. These elements were inspired by biological nervous systems. As in nature, the network function is determined largely by the connections between elements. We can train a Neural Network to perform a particular function by adjusting the values of the connections (weights) between elements.

An artificial neuron is a simplistic representation that emulates the signal integration and threshold firing behavior of biological neurons by means of mathematical equations. Like their biological counterpart, artificial neurons were bounded together by connections that determine the flow of information between peer neurons. Stimuli were transmitted from one processing element to another via *synapses* or interconnections, which can be excitatory or inhibitory. If the input to a neuron is excitatory, it is more likely that this neuron will transmit an excitatory signal to the other neurons connected to it. Whereas an inhibitory input will most likely be propagated as inhibitory. A neuron with a single scalar input is shown in Fig.1.

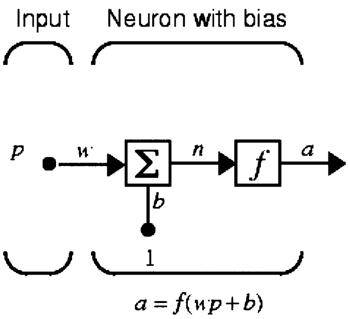

Fig. 1 A neuron with a single scalar input

Neural Networks were usually adjusted, or trained, so that a particular input leads to a specific target output. There, the network was adjusted, based on a comparison of the output and the target, until the network output matches the target. Typically many such input/target pairs are needed to train a network. A simple block diagram of a neural network is defined by Fig.2.

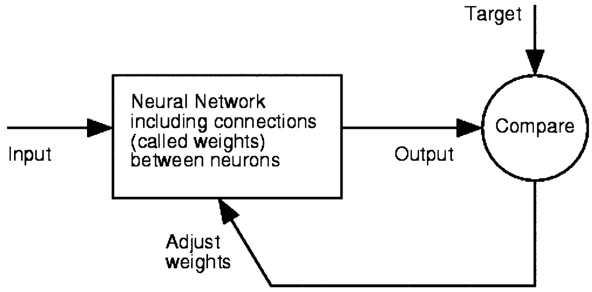

Fig. 2 A simple block diagram of a neural network

Neural Networks have been trained to perform complex functions in various fields, including pattern recognition, identification, classification, speech, vision, and control systems. Today neural networks can be trained to solve problems that are difficult for conventional computers or human beings. Neural networks are good at fitting functions and recognizing patterns. In fact, there is a proof that a fairly simple neural network can fit any practical function.

B. *MLP networks*

There are many types of Neural Networks for various applications available in the literature. Multi-Layer Perceptrons (MLPs) are the simplest and therefore most commonly used Neural Network architectures due to their structural flexibility, good representational capabilities and availability of a large number of training algorithms. MLPs are feedforward Neural Networks and universal Approximators, trained with the standard backpropagation algorithm. They are supervised networks so they require a desired response to be trained. They learn how to transform input data into a desired response, so they are widely used for pattern classification. With one or two hidden layers, they can approximate virtually any input-output map.

An MLP consists of three layers: an input layer, an output layer, and an intermediate or hidden layer. In this network every neuron is connected to all neurons of the next layer, in other word MLP is a fully connected network. Fig.3 shows the structure of MLP.

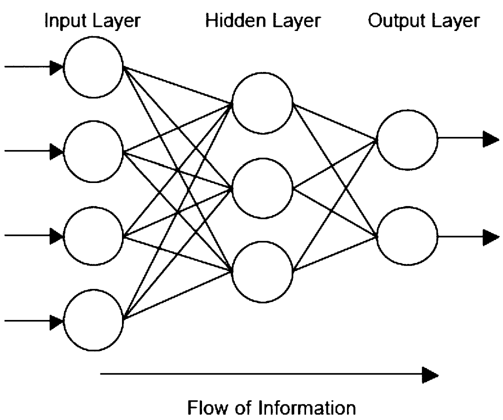

Fig.3 MLP network structure

Processing elements (PE) or neurons in the input layer only act as buffers for distributing the input signals x_i (i shows the i-th input PE) to PEs in the hidden layer. Each PE $_j$ (j shows the j-th PE in the hidden layer and output layer) in the hidden layer sums up its input signals x_i after weighting with the values of the respective connections w_{ji} from the input layer and computes its output y_j as a function f of the sum, viz.,

$$Y_j = f(\sum w_{ji} x_i) \quad (1)$$

There are some choices for the transfer function f which can be globally supported. The only limitation about this function is that it must be differentiable.(Fig.4)

$$\textit{Linear:} \quad f(x) = x \quad (2)$$

$$\textit{Log-Sigmoid:} \quad f(x) = \frac{1}{1+e^{-x}} \quad (3)$$

$$\textit{Tan-Sigmoid:} \quad f(x) = \frac{2}{1+e^{-2x}} - 1 \quad (4)$$

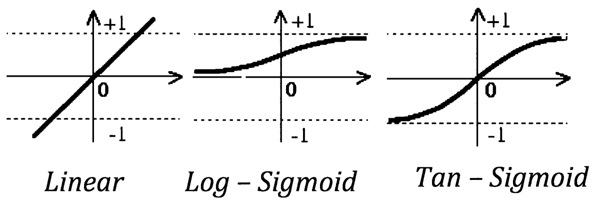

Fig. 4 some useful transfer functions

The output of processing elements in the output layer is computed similarly. You can see the architecture of the multiple layers of neurons in Fig.5.

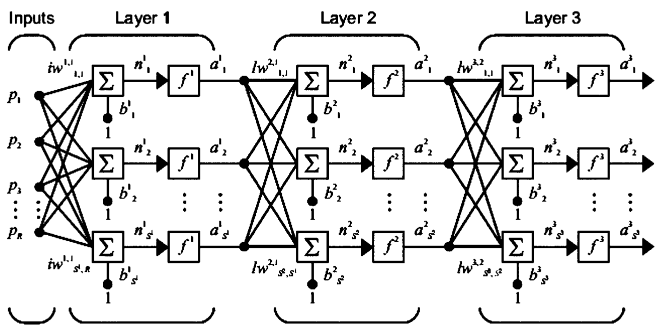

Fig. 5 Architecture of the multiple layers of neurons

C. *Training the network*

Before training a feedforward network, you must initialize the weights and biases. Once the network weights and biases are initialized, the network is ready for training. Training a network consists of adjusting its weights using a training algorithm. Training process requires a set of examples of proper network behavior, network inputs and target outputs. During training the weights and biases of the network are iteratively adjusted to minimize the network performance function.

The MLP used in this work was trained with the backpropagation (BP) learning algorithm. This training algorithm optimizes the weights by attempting to minimize the sum of squared differences between the desired and actual values of the output neurons.

$$E = \frac{1}{2} \sum_j (Y_{dj} - Y_j)^2 \quad (5)$$

Where y_{dj} is the desired value of output neuron j and y_j is the actual output of that neuron. Each weight w_{ji} was adjusted by adding an increment Δw_{ji} to it. Δw_{ji} was selected to reduce E as fast as possible. The adjustment was carried out over several training iterations until a satisfactory small value of E was obtained or a given number of iterations had been reached. How Δw_{ji} is computed depends on the training algorithm adopted.

The backpropagation computation was derived using the chain rule of calculus. The basic backpropagation training algorithm, moves weights in the direction of the negative gradient.

The available ANN softwares today provide many Neural-Network architectures and learning algorithms, and also help users to apply ANN to their specific problems easily. MATLAB Neural-Network Toolbox might be a good example of these softwares. One can also write ANN software using available compilers. In this paper we used MATLAB Neural-Network Toolbox for developing the network.

III. ANN APPLICATIONS TO DIAGNOSING AND PREDICTING BLOOD DISORDER AND CANCERS

In this paper by using Multilayer Perceptron Neural Networks, a novel applicable method for diagnosing and predicting five blood diseases and cancers through the results of blood tests, is presented.

In order to train and test the network, 450 data series which was obtained from blood tests of 450 patients that were examined by haematologists, were used. Eleven parameters playing role in diagnose these disorders, in these blood tests. So these parameters have been used as the inputs of network. The developed system is such as Fig.6. that its input parameters are White Blood Cell (WBC), Red Blood Cell (RBC), Hemoglobin (HGB), Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin (MCH), Mean Corpuscular Hemoglobin Concentration (MCHC), Red Cell Distribution Width (RDW), Platelet (PLT), Neutrophil (NEUT), Lymphocyte,Lecocyte(LUC) (the result parameters of blood tests) and its outputs are five blood diseases: Megaloblastic Anaemia, Thalassemia, Idiopathic thrombocytopenic pupura (ITP), Chronic myelogenous leukemia and Lymphoproliferative.

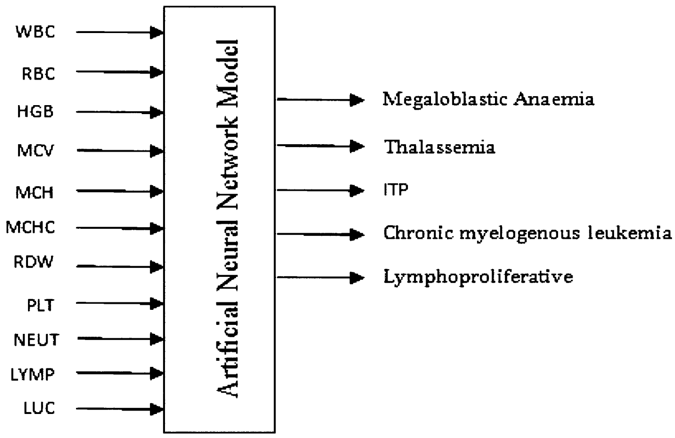

Fig. 6: Input-output schematic of system

So in order to present the problem in mathematical form, a multilayer perceptron network with eleven inputs and five outputs has been created. This network involves three layers (an input layer, a hidden layer, and an output layer). Number of neurons in input layer is equal to the number of input variables (eleven), and the number of neurons in output layer is equal to the number of outputs (five). This network has been simulated with MATLAB Neural Network Toolbox with 360 series of 450 available data series which was used for training the network and 90 series for testing it.

Because of the outputs of network always have two values, 0 or 1, the log-sigmoid transfer function for the output-layer was selected. To achieving the best transfer functions for input and hidden layer several trials were made, finally the best result was obtained with a network with tan-sigmoid transfer function in both input and hidden layers.

There are many choices for number of neurons in hidden layer. So in order to achieving the best network with optimum parameters, a comparison between networks by different neurons in the hidden layer has been done. Results of testing networks with 6, 7, 8... 21 neurons in hidden layer are shown in Table.1, Table.2, Fig.7 and Fig.8.

TABLE.1
Comparison of training error in networks having different number of neurons in hidden layer

Network	Number of neurons in hidden layer	Test error	Number of epochs
ANN.1	6	0.004444	113
ANN.2	7	0.008889	52
ANN.3	8	0.008889	171
ANN.4	9	0.028889	82
ANN.5	10	0.006667	131
ANN.6	11	0.008889	38
ANN.7	12	0.017778	65
ANN.8	13	0.017778	40
ANN.9	14	0.008889	91
ANN.10	15	0.004444	67
ANN.11	16	0.006667	53
ANN.12	17	0.013333	98
ANN.13	18	0.026667	57
ANN.14	19	0.024444	23
ANN.15	20	0.013333	46
ANN.16	21	0.013333	100

TABLE.2
Comparison of training error with different training functions

Network	Number of neurons in hidden layer	Train function	Percent of test error	Number of epochs
ANN.10.1	15	Trainlm	0.004444	67
ANN.10.2	15	Trainscg	0.008889	976
ANN.10.3	15	Traingdm	0.006667	5000
ANN.10.4	15	Trainrp	0.006667	481

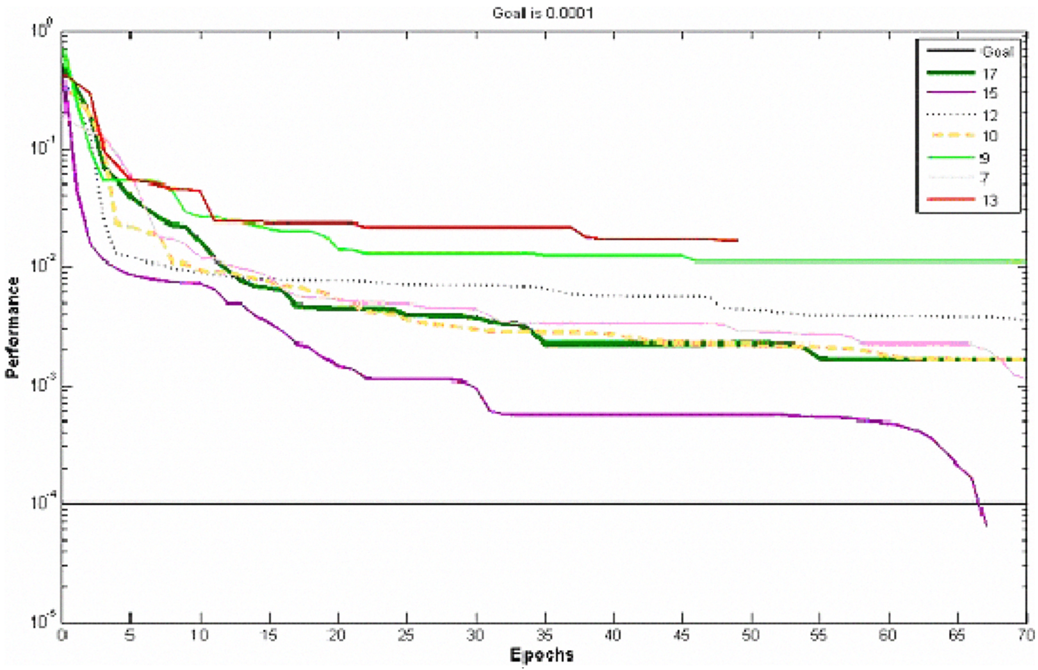

Fig. 7 Ttraining networks having different number of neurons in hidden layer

Comparison between the range of test error and the number of epochs that was required to reach the minimum performance of 0.0001, shows that a network with fifteen neurons in hidden layer and trainlm function as its training function, has the best performance for this problem. (Don't forget that increasing in number of neurons in hidden layers may cause overfitting problem in the network, so for preventing this problem, the number of neurons must be less than the cases in the training algorithm). It must be say that in this problem the value of test error is more important than the number of epochs, so the error value has the main rule in our decision making. Table.3 and Fig.8 Shows the final network architecture and the results of its performance.

TABLE.3
Optimal structure of developed MLP Neural Network for obtaining minimum prediction error

Number of hidden layers	1
Number of neurons in hidden layer	15
Hidden layer Transfer Function	Tan-sigmoid
Output layer Transfer Function	Log-sigmoid
Training function	Trainlm
Test error	0.004444
Number of epochs	67

The network memorizes the training set and does not generalize well when the network is trained too much (over-fitting). The training holds the key to an accurate solution, so the criterion to stop training must be very well described. Cross-validation is a highly recommended criterion for stopping the training of a network. When the error in the cross validation increases the training should be stopped. A practical way to find a point of better generalization is to use a small percentage (around 10%) of the training set for cross-validation [22].

For obtaining a better generalization of the network presented in this work, 90 series of training data (which were selected randomly) were used as cross validation set. The values of error for 2 different training algorithms with cross-validation and without it are given in Table.4 and Fig.9. As it is seen from Table.4, training of the MLP with cross-validation has less error value.

TABLE.4
Comparison networks training with Cross-Validation and without it

Network	Number of Training data series	Number of Validation data series	Number of Testing data series	error	epochs
Training without cross-validation	360	-	90	0.004444	67
Training with cross-validation	270	90	90	0.003333	14

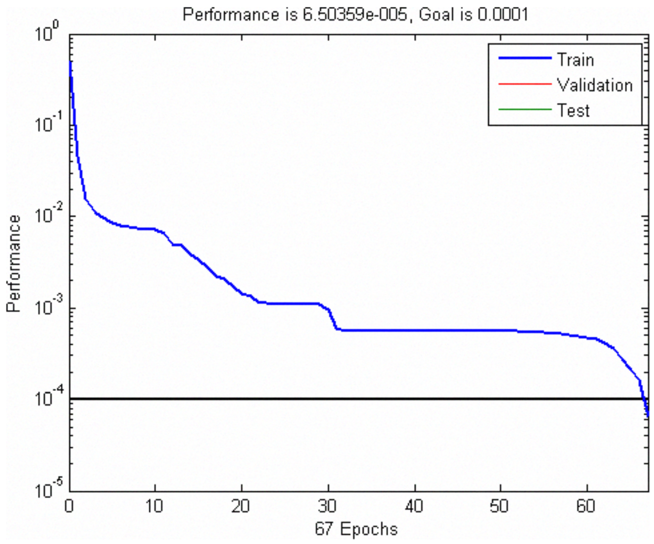

Fig. 8 Performance changes in the course of network training

IV. CONVENTIONAL METHOD

In order to show the ability, accuracy and capability of ANN application in predicting diseases, we compared the ANN method with one of the conventional methods which can do this task similarly. One of these conventional methods is using statistic solutions. The statistic method that we used is multivariable nonlinear regression for finding relation between each diseases and inputs for analyze cases, and predicting diseases for new cases, based on the founded relation. In this method all 450 cases were randomly divided to two sets, selected cases for analyzing and unselected cases for testing the method accuracy.(These analyzes are done with SPSS 14th version software.)

The final results of regression analyze for all five outputs and average of these five diseases is shown in Table.5. For comparing accuracy of this method with ANN method we should have average error from all five outputs, because the ANN method simulates and calculates all five outputs errors simultaneously. Now from these results we can compare these two methods.

TABLE.5
Final results of Regression Analyze

Disorder	Error percentage	
	Selected cases (train)	Unselected cases (test)
Megaloblastic Anemia	0	0.8
Thalassemia	0.3	0.8
ITP	1.6	0
Chronic Myelogenous Leukemia	1.9	0.8
Lymphoproliferative	0.6	1.5
Average	0.88	0.78

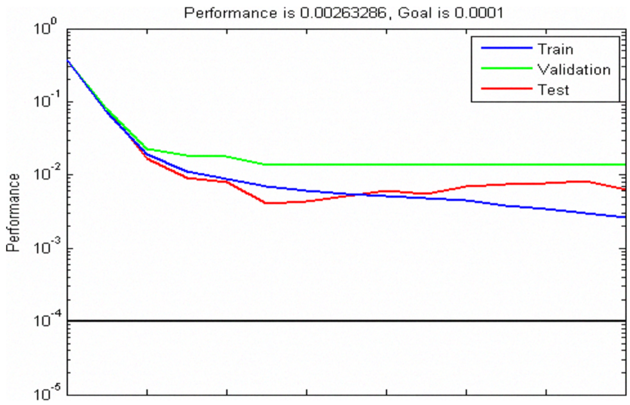

Fig.9 Performance changes in the course of network training

V. CONCLUSION

The first topic that needs to be discussed is the data that was sampled to train and test the MLP network. The data were collected from blood test results of 450 patients that were examined by hematologists, in Taleghani Hospital in Kermanshah, Iran. The data samples were used for training, validating and testing the Neural Network (60-20-20 ratio for training with cross-validation and 80-20 ratio for training without it) and also for multivariable nonlinear regression analysis (70-30 ratio).

The number of hidden layers and neurons in each layer were determined through trial and error to be optimal including with different transfer functions as tangent-sigmoid and log-sigmoid. After several trials, the best result was obtained from a three-layered network. In this network the tangent-sigmoid function was used in the input and hidden layer, and log-sigmoid function in the output layer. And the most suitable network configuration found was 11 x 15 x 5. It means that the number of neurons was 15 for the hidden layer.

For training network with backpropagation learning algorithm, The MSE performance function with 0.0001 goal value, and trainlm function were used and cross-validation method was used to stop training in order to prevent over-fitting problem.

From the results achieved in two last chapters, the designed Neural Network has 0.3333 error percentage but the SPSS analysis has 0.78 error percentage. By these results we can conclude that the Artificial Neural Network is more applicable by reason of high accuracy, fast convergence and low use of memory for diagnosis and predicting disorder.

Another eminent property of this network is its ability to diagnose all five mentioned disorder simultaneously unlike multi variable nonlinear regression that can do analysis only for one disorder in each time.

V. REFERENCES

[1] 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, 2004, 10, 2101-2109

[2] Murat Karabatak and M. Cevdet Ince, An expert system for detection of breast cancer based on association rules and neural network, 2008.

[3] TOBB Ekonomi ve Teknoloji Üniversitesi,Implementing automated diagnostic systems for breast cancer detection , Department of Electrical and Electronics Engineering, Faculty of Engineering, 06530 Söğütözü, Ankara, Turkey September 2006.

[4] T.Z. Tan, C. Quek, G.S. Ng and E.Y.K. Ng, A novel cognitive interpretation of breast cancer thermography with complementary learning fuzzy neural memory structure, 2006.

[5] Paulo J.G. Lisboa, Terence A. Etchells, Ian H. Jarman, M.S. Hane Aung, Sylvie Chabaud, Thomas Bachelot, David Perol, Thérèse Gargi, Valérie Bourdès, Stéphane Bonnevay and Sylvie Négrier, Time-to-event analysis with artificial neural networks: An integrated analytical and rule-based study for breast cancer, 2007

[6] Carsten Stephan, Henning Cammann, Kurt Miller, Serdar Deger, Michael Lein and Klaus Jung,THE FIRST ARTIFICIAL NEURAL NETWORK FOR FIVE DIFFERENT PSA-ASSAYS IN PROSTATE CANCER DIAGNOSTICS , 2008.

[7] Carsten Stephan, Chuanliang Xu, Patrik Finne, Henning Cammann, Hellmuth-Alexander Meyer, Michael Lein, Klaus Jung and Ulf-Hakan Stenman, Comparison of Two Different Artificial Neural Networks for Prostate Biopsy Indication in Two Different Patient Populations, 2007.

[8] Michael W. Kattan, Editorial Comment on: Development, Validation, and Head-to-Head Comparison of Logistic Regression-Based Nomograms and Artificial Neural Network Models Predicting Prostate Cancer on Initial Extended Biopsy, 2008.

[9] Satoru Kawakami, Noboru Numao, Yuhei Okubo, Fumitaka Koga, Shinya Yamamoto, Kazutaka Saito, Yasuhisa Fujii, Junji Yonese, Hitoshi Masuda, Kazunori Kihara and Iwao Fukui, Development, Validation, and Head-to-Head Comparison of Logistic Regression-Based Nomograms and Artificial Neural Network Models Predicting Prostate Cancer on Initial Extended Biopsy, 2008.

[10] Felix K.-H. Chun, Markus Graefen, Alberto Briganti, Andrea Gallina, Julia Hopp, Michael W. Kattan, Hartwig Huland, and Pierre I. Karakiewicz, Initial Biopsy Outcome Prediction—Head-to-Head Comparison of a Logistic Regression-Based Nomogram versus Artificial Neural Network**,** 2006.

[11] Imran Kurt, Mevlut Ture and A. Turhan Kurum,Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease , 2006**.**

[12] Samit Ari and Goutam Saha In search of an optimization technique for Artificial Neural Network to classify abnormal heart sounds, 2008

[13] Lennart Mucke, Strategies to prevent neural network dysfunction in Alzheimer's disease, Gladstone Institute of Neurological Disease and UCSF, San Francisco, CA, USA, 2008.

[14] Monica Di Luca1, Enzo Grossi, Barbara Borroni, Martina Zimmermann, Elena Marcello, Francesca Colciaghi, Fabrizio Gardoni, Marco Intraligi, Alessandro Padovani and Massimo Buscema, Artificial neural networks allow the use of simultaneous measurements of Alzheimer Disease markers for early detection of the disease, Journal of Translational Medicine 2005, 3:30doi:10.1186/1479-5876-3-30

[15] Harun Uğuz, Ali Öztürk, Rıdvan Saraçoğlu and Ahmet Arslan, A biomedical system based on fuzzy discrete hidden Markov model for the diagnosis of the brain diseases, Turkey, 2007.

[16] Xiaoping Qiua, Ning Taob, Yun Tana and Xinxing Wua, Constructing of the risk classification model of cervical cancer by artificial neural network, China, 2006.

[17] Chun-Lang Chang, Chih-Hao Chena, Applying decision tree and neural network to increase quality of dermatologic diagnosis, 2008.

[18] Mohammad Mataria, Michae G. Janech, Jonas Almeida, Evelyn Lewis, Alison Bland and John Arthur, Prediction of Progression of Diabetic Nephropathy in a Small Set of Patients by Artificial Neural Networks and Proteomic Analysis, 2008.

[19] Kemal Polata, Sadık Karab, Ayşegül Güvenc, Salih Güneşa, Utilization of Discretization method on the diagnosis of optic nerve disease, 2008.

[20] Tuan Zea Tan, Chai Quek, Geok See Ng and Khalil Razvi, Ovarian cancer diagnosis with complementary learning fuzzy neural network,2008

[21] Adrian Săftoiu, Peter Vilmann, Florin Gorunescu, Dan Ionuţ Gheonea, Marina Gorunescu, Tudorel Ciurea, Gabriel Lucian Popescu, Alexandru Iordache, Hazem Hassan and Sevastiţa Iordache, Neural network analysis of dynamic sequences of EUS elastography used for the differential diagnosis of chronic pancreatitis and pancreatic cancer,2008.

[22] E.D. Ubeyli, I. Guler, Neurocomputing 62 (2004) 349.