Vol. 5(17) Oct. 2015, PP. 2455-2467

Contents list available at IJMEC

International Journal of Mechatronics, Electrical and Computer Technology (IJMEC)

Journal Homepage: www.aeuso.org

PISSN: 2411-6173 - EISSN: 2305-0543

A New Method Based on Genetic Algorithm to Achieve the Best Artificial Neural Network Model for the Diagnosis of the Breast Cancer

Mahyaar Daaraaee[1*], Javad Vahidi[2], Abbas Alipour[3] and Masoud Darayee[4]

[1]*Department of Computer Engineering, Ayatollah Amoli Branch, Islamic Azad University, Amol, Iran.*

[2]*Iran University of Science & Technology, Tehran, Iran.*

[3]*Department of Social Medicine, Faculty of Medicine, Mazandaran University of Medical Sciences, Sari, Iran.*

[4]*Department of General Surgery, Faculty of Medicine, Mazandaran University of Medical Sciences, Sari, Iran.*

***Corresponding Author's E-mail:** : mahyar354@yahoo.com*

Abstract

Methods such as biopsy and surgery for breast cancer diagnosis, are expensive, invasive and with risks. The purpose of this study is to propose a new method for achieving the best artificial neural network (ANN) model using cytology results such as clump thickness,… so that can be applied for the diagnosis of benign or malignancy of breast tumors with minimum error and maximum reliability instead of invasive methods. The Wisconsin breast cancer database was used. We applied the genetic algorithm (GA) for determination of the best structure and training of multi-layer NN i.e. optimization of the values of the weights and bias. The implementation was done by MATLAB showed that GA is able to determine the best structure for a multi-layer NN and train it properly too. Then the error back propagation algorithm (EBPA) was used to train the models proposed by GA and we found that EBPA trains the neural networks (NNs) equal to GA, even better in some cases. Accordingly, the EBPA was determined as a basic algorithm for training NNs which their structure was proposed by GA. 5-Fold Cross-Validation was used for a closer evaluation the performance of these NNs. Based on the results obtained in different performances, we achieved the best neural network model which its structure was (9-6-4-1) with an average accuracy, sensitivity and specificity 0.974, 0.979 and 0.971 respectively, and accordingly a new method to obtain an optimized model based on the available data, is proposed in which the structure of the NN is determined by the GA and Training is done by EBPA. The proposed NN model performance was compared with the logistic regression (LR) that ANN had better performance rather than LR in the diagnosis of benign or malignant tumors. Thus, the ANN model which is obtained by the method described, can replace invasive medical methods, and determine patients who do not need a biopsy and surgery, with high accuracy and sensitivity.

Keywords: *Invasive methods, Cutoff prediction, Accuracy, Sensitivity, Specificity, Artificial neural network, False positives, False negatives, 5-Fold Cross Validation, ROC curve*.

1. Introduction

One of the most widely studied fields in statistics, decision science and computer science has been a data classification manner applying knowledge achieved from known historical data. It has been used in issues of medicine, social science, management and engineering [1,33]. Different issues such as disease diagnosis, image segmentation, face recognition and credit management are using classification techniques [2,34]. Before, linear programming methods were efficient and effective approaches in medical and other fields [3-6]. Recently, machine-learning technology has been

Article History:

Received Date: Jun. 02, 2015

Accepted Date: Sep. 14, 2015

Available Online: Sep. 19, 2015

intensively used for classification tasks [7]. These researches have included a range of techniques including support vector machines (SVM) [8], evolutionary processes [9,10], K-nearest neighbors [11,12], decision trees [13-15], artificial neural network (ANN) [16-18] and clustering [19].

Data mining and knowledge discovery in database are methods to detect relationships between entities in data [20]. Very large data sets can be gathered in medicine, by using the new facilities. These data sets require special methods for analyzing, processing and efficient use of them [1]. Intelligent diagnostic systems are one of the application domains of analyzing database and data classification. The goals of these studies are aiding to physicians in making diagnostic decision.

These days, breast cancer is rising in all societies as a social problem and major cause of death in women aged 40 to 44. When normal cells in the breast begin to change and grow out of control, lead to a mass called tumor. The tumor can be benign (non-cancerous) or malignant (cancerous, meaning it can spread to other parts of the body). Accurate diagnosis of breast cancer is one of the main issues in the medical world. Methods such as biopsy and surgery are the best method for the evaluation of the breast cancer, but are expensive, invasive and with risks. Therefore, to identify and assess the extent of the disease before the mentioned methods, non-invasive methods (such as Radiography, Mammography, etc.) are done that due to the low accuracy of these methods may be in the results, false positive or false negative which is dangerous for patients. Thus, the decision support system alongside other methods to minimize false results seems to be necessary and if that can be achieved with a model that can accurately predict breast disease status, in addition to preventing invasive methods such as surgery and biopsy which is associated with the side effects such as anesthesia, stress,… for patients, will considerably help physicians in the early detection, control and treatment of the disease. Since the parameters of the result Cytology, such as the Uniformity of Cell Size,… are considered as indicators for the assessment of the disease, can be used as input parameters of the proposed model in the diagnosis of the disease statue (Benign or Malignant) the mass of the breast and the impact of these variables are also among the subjects which can be studied.

1.1. Background

Artificial neural networks (ANNs) inspired by biological neural networks, count a mathematical model of the human diagnostic system and have been widely used in various fields, particularly in medicine; few of them related to the prediction and diagnosis of breast cancer, will come in the following.

Marchevsky et al. [22] applied neural networks (NNs) based on genetic algorithm (GA) and logistic regression (LR) classifiers to evaluate 19 prognostic features of 279 patients. Lancashire et al. [23] identified gene signatures corresponding with estrogen receptor and axillary lymph nodes (ALN) status in breast cancer through applying the multilayer perceptron (MLP) to a gene microarray dataset that consisted of 49 samples. In these studies, the accuracy values were acquired 0.89 and 1, respectively.

Karakis et al. [24] applied pattern recognition methods such as MLP, Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) and k-Nearest Neighbor (k-NN) to categorize ALN status of 270 breast cancer patients. MLP network that applied Levenberg–Marquardt algorithm in back propagation training was consisted of a middle layer with 4 nodes. MLP classifier obtained the correlation coefficient and the correct rate of testing dataset 0.872 and 0.944, respectively.

In a study by Salehi et al. [25] for Comparison of 3-layer perceptron neural network and cox regression in survival prediction of breast cancer patients, the result of this comparison showed the relative superiority of ANN.

Karabatak and Cevdet [1], presented an automatic diagnosis system for detecting breast cancer based on association rules (AR) and ANN. In this study, AR was used for reducing the dimension of breast cancer database, and ANN was used for intelligent classification. The proposed AR + NN

system performance was compared with ANN model. On test stage, the predictive accuracy value of the proposed system applied to the Wisconsin breast cancer database (WBCD), in the classification of benign and malignant tumors, was 0.956.

Cai and Jiang [26] developed a novel ANN method based on Matrix Pseudo-Inversion (MPI) for use in diagnosis of breast tumor status. The MPI-ANN was constructed as a 3-layer (i.e., input, hidden, and output layers) feed-forward NN, and the weights connecting the middle and output layers were directly determined based on MPI without a lengthy learning iteration. On test stage, the accuracy, sensitivity and specificity of this method applied to the WBCD was 0.897, 0.93 and 0.85, with respectively.

In the last three decades, some attempts made to combine evolutionary algorithms and NNs [27,35]. Karakis et al. [28] predicted ALN status in breast cancer patients using LR and Genetic Algorithm (GA) based MLP models. GA Provided to select the best features as MLP inputs and optimized the weights of MLP. The values of regression and accuracy of the GA based MLP were obtained 0.96 and 0.98, respectively.

The present study by applying GA attempts to propose a new method for achieving the best ANN model structure using cytology results such as clump thickness,… as input features, so that can be applied for the diagnosis of benign or malignancy of breast tumors with minimum error and maximum reliability.

2. Tools & Methods

In this section, at first, we will talk about ANN as the most important tool used in this study while the data set employed to evaluate our ANN Model, is described. Afterward, we review GA and then the method used in this research will be expressed. It should be mentioned; we used Excel and SPSS for statistical works and regression, and MATLAB for ANN implementation and its combination with the GA.

2.1. Artificial Neural Networks

The main idea (partly) has been inspired by the biological nervous system function for processing data and information in order to learn and knowledge creation. A key element of this idea, creating new structures for information processing systems. This system consists of a large number of super interconnected processing elements called neurons or nodes or cell that act in concert together to solve a problem.

2.1.1. Perceptron: A type of ANN is made based on a calculation unit named perceptron. A perceptron gets a vector of inputs with real values and calculates a linear combination of the inputs. If the result is more than a threshold, perceptron output will be equal to 1, otherwise it equal to 0. Added bias causes the use of perceptron a simpler. Thus, the network output is obtained from the following equation:

$$\hat{y} = b + \sum_{i=1}^{n} x_i w_i$$

2.1.2. Multilayer Neural Network: The single-layer neural networks can be used for pattern classification. These networks can only separate models which are linearly independent in input space. In this study, we actually need a mapping to be approximated between inputs and outputs. In other words, it is necessary to be learned an appropriate mapping which relates system inputs to system output. Unlike Perceptron, multi-layer networks can do this.

An ANN consists of layers and weights. Layers are based on multiple of interconnected neurons, which connect inputs to outputs. Both layers of a network are connected with weights which indeed are connectors. In general, NNs have three types of layers:

Input Layer: It includes incoming data that is fed into the network, i.e network inputs are characteristics of the samples.

Hidden layer(s) (middle layer(s)): Function of these layers is determined by inputs and weight of connection between them and hidden layer. Weights between inputs and hidden layer neurons determine when a hidden layer neuron needs to be enabled.

Output Layer: Function of this layer depends on activity of hidden layer neurons and weight of connection between them and output layer neurons.

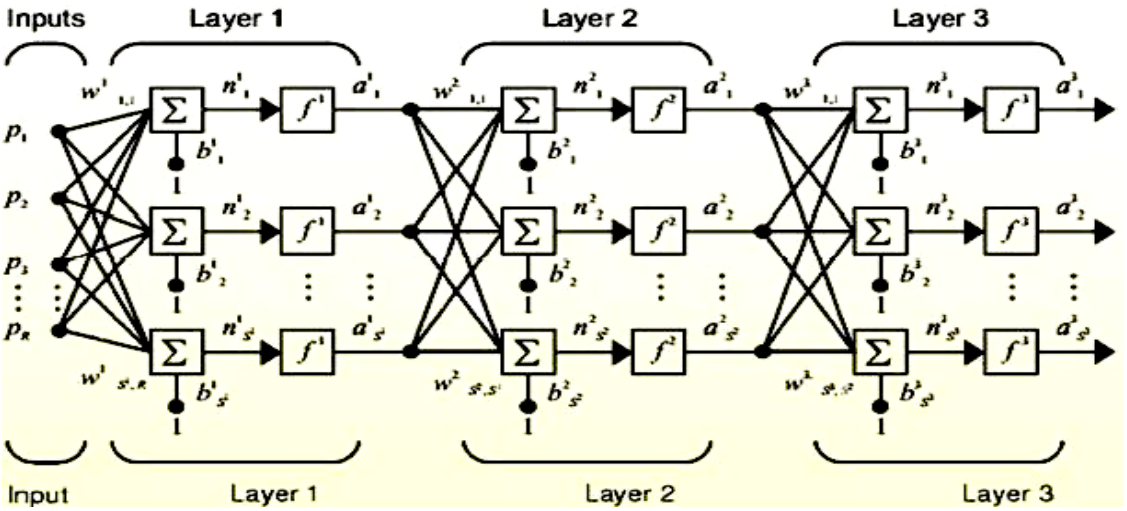

Figure 1: Multilayer ANN Structure

If we assume a network equivalent to a graph, network training process determines optimal values for the weights of the edges and bias.

2.2. Data set

Wisconsin Breast Cancer database (WBCD) was employed for validation of the proposed method. This set contains 699 data samples that have been obtained through the cytology of the breast tumors. Each data sample has 9 features with a range of values [1, 10]. The data are classified in two classes or groups, benign and malignant. 16 samples have one field without value; So we excluded them from the data set, therefore of 683 samples were left.

2.2.1. Preliminary assessment of attributes: For this purpose, we applied T-Test in SPSS.

Table 1: Results of T-Test upon the data features

Attribute Number	Features	Class	Number	Mean	Standard Deviation	P-value*
1	Clump Thickness	Benign (1)	444	2.96	1.673	0.000
		Malignant (2)	239	7.19	2.438	
2	Uniformity of Cell Size	Benign (1)	444	1.31	0.856	0.000
		Malignant (2)	239	6.58	2.724	
3	Uniformity of Cell Shape	Benign (1)	444	1.41	0.957	0.000
		Malignant (2)	239	6.56	2.569	
4	Marginal Adhesion	Benign (1)	444	1.35	0.917	0.000
		Malignant (2)	239	5.59	3.197	
5	Single Epithelial Cell Size	Benign (1)	444	2.11	0.877	0.000
		Malignant (2)	239	5.33	2.443	

6	Bare Nuclei	Benign (1)	444	1.35	1.178	0.000
		Malignant (2)	239	7.63	3.117	
7	Bland Chromatin	Benign (1)	444	2.08	1.062	0.000
		Malignant (2)	239	5.97	2.282	
8	Normal Nucleoli	Benign (1)	444	1.26	0.955	0.000
		Malignant (2)	239	5.86	3.349	
9	Mitoses	Benign (1)	444	1.07	0.510	0.000
		Malignant (2)	239	2.60	2.564	

*P-value (significance level) has been calculated in confidence interval 95%.

As can be seen in Table 1, the mean and standard deviation of all of the features in both healthy (Benign) and patient (Malignant) groups have significant differences. This indicates the values of all of these features relate to benignity or malignancy of breast tumors. Accordingly, we used all of the 9 features in the NN model as input variables.

2.2.2. Normalization: Data must be normalized for use in ANN. For this purpose, we divide all data in each column (attribute) by the maximum value of each column that is 10. Thus, all data became within [0,1].

2.2.3. 5-Fold Cross Validation: The 5-fold cross validation method [30] was used for the data set to get the correct rates of training and testing. The dataset was divided into five parts based on the proportion between benign and malignant of the original data set, i.e., the proportion in each of the five parts were approximately the same as the one of the original data set (i.e., 444:239 for WBC real data). For each fold of cross validation, one part of the five parts of the data set was selected as the testing data while the remaining four parts were considered as the training data. This was repeated for all five parts and the average value of the correct rate was generated for both training and testing.

2.3. Using Genetic Algorithm(GA)

In this study, the goal is to determine the optimal weights and the optimal structure for a multi-layer NN. To achieve this, we used GA. Mainly, the field of evolutionary algorithms (GA in this study) in NN(s) is where the neural network training, in other words weighting the edges of a NN leads to an optimization problem that in this research, we want, additionally, to use of GA to determine the structure of the NN, too.

2.3.1. A review of the GA: GA is one of the types of evolutionary algorithms that its learning method is based on biological evolution. This method was introduced in 1970 by John Holland. For solving a problem the algorithm produces a large set of possible solutions (hypothesis). Each of these solutions is evaluated by using a "fitness function". Then some of the best solutions produce new solutions (generations). Thus, the solutions will evolve.

2.3.1.1 Genetic Algorithm:

1. Initialize: Initialize population P randomly with number of p hypotheses.
2. Evaluate: ∀ hypothesis $h \in P$; Calculate Fitness (h).
3. While (Fitness $(h)_{Best}$< Threshold)

{Elite selection: Select $(1\text{-}r)\times p$ hypotheses from P & append them to P_s .
Crossover: Using the probability obtained by the equation

$$P(h_i) = \text{Fitness}(h_i) / \Sigma_j \text{Fitness}(h_j)$$

select $(r \times p)/2$ pairs of hypotheses through P & create 2 children from them using crossover operator.
Append children to P_s .
Mutation: Select m% of P_s with equal probability & reverse a bit of each of them randomly.
Update: $P_s \rightarrow P$.
Evaluate: ∀ hypothesis h ∈ P; Calculate Fitness(h).}
4. Return h | Fitness(h) = Fitness$(h)_{Best}$.

2.3.2. Initialize of population: At first, we produce a set of solutions (answers, hypotheses, chromosomes) named primary population. The answers include various parameters that we want based on them using GA to determine the best possible structure for the proposed MLP model, while training it. Search space includes the structure of NN(9-1-1) to NN(9-28-2) and NN(9-14-14-2) for networks with one middle layer and two middle layers, respectively that activation functions of each node may be linear, tangent sigmoid or logarithmic sigmoid.

Common criteria usually used for determining the number of the middle layer nodes is:

$$Number\ of\ middle\ layer\ nodes = 2\times\ (Number\ of\ features) + 1$$

We considered search space for the number of the middle layer nodes within:
[1, 3× $(Number\ of\ features)$] approximately, that is [1,28].
Each of these solutions has the attributes (genes) as follow:

Network Bias and weights: Include random values within (-1,1) and with regard to the network which has the most nodes, NN (9-27-14-2), we totally assign the first 692 attributes (genes) of each chromosome for weights and bias.

Number of output layer nodes: 693th gene of each chromosome is assigned to it that randomly gets a value of 1 or 2.

Number of middle layers: 694th gene of each chromosome is assigned to it that randomly gets a value of 1 or 2.

Number of middle layers nodes: In the Case of the NN has one middle layer, 695th gene indicates the number of middle layer nodes that randomly gets an integer value within [1,27] and in the case of the NN has two middle layers, this gene indicates the number of the nodes in the first middle layer that randomly gets an integer value within [1,27] and 696th gene indicates the number of the nodes in the second middle layer that randomly gets an integer value within [1,14].

Type of the activation functions of the nodes of the middle layers and output layer: 697th, 698th and 699th genes of each chromosome are assigned to them.

2.3.3. Cost function: This function is defined to compare the performance of different solutions and select the best hypotheses of the population so that after the full running of GA, the hypothesis that its cost function is less, is chosen as the optimum solution. We defined the error of models in predicting of response variables based on the input data (1- accuracy) on the first priority, and the error of models in diagnosis of malignant tumors (1- sensitivity) on the second priority, as the cost function. In the Case of the hypothesis has only one node in the output layer, the proposed approach is, finding the best Cutoff Prediction to classify neural network outputs in which the values of the cost function, is minimal.

2.3.4. Chromosomes selection for crossover: Roulette wheel selection method was used for this purpose. In this way, the probability of the selection of a chromosome for next population depends on the ratio of its fitness to summation of the remaining chromosome's fitness, that is: $P(h_i)$ = Fitness (h_i) / Σ_j Fitness (h_j)

This method increases the probability of the selection of chromosomes with less cost function for crossover while each of the chromosomes has a chance to crossover.

2.3.5. Structure determination, training and testing of NN: In each run, the initial population of more than 2,500 chromosomes that any of them involving various characteristics of a NN structure was formed and accordingly made a multilayer NN for each chromosome. Then performance of each structure was calculated using the cost function based on the training data set. Afterwards, the chromosomes were sorted in order of the performance; 5% of them which had the best performance, were transferred unchanged to the next generation, 60% were combined with each other using crossover operator and formed another part of the population of the next generation and the remaining 35% were mutated and transferred to the next generation. This process continued until the 100 consecutive generations in which the best answer (chromosome) was not changed. Thus, the structure and weights of a trained NN which had the least error based on training data was obtained. In the last step, we simulate the obtained NN by test data, and got its performance. These steps were repeated using the 5-fold cross validation.

3. Results

Due to the use of 5-fold cross validation technique, following steps was performed for each different combination of training and test data (total 5 combinations).

3.1. Step 1

Structure determination and training of the neural network were performed using GA that the results are in the table below:

Table 2: Results of structure determination and training of network by GA

ID	Dtrain	Gen	N.H	N.H.1	N.H.2	N.O	Func1	Func2	FuncO	Dtest	Cut	Acc	Sen	Sp
1	All-{F1}	122	2	6	5	1	tansig	tansig	purelin	F1	0.63	0.971	1.000	0.955
2	All-{F2}	107	2	8	6	1	tansig	purelin	logsig	F2	0.61	0.956	0.958	0.955
3	All-{F3}	103	2	6	4	1	tansig	tansig	tansig	F3	0.26	0.942	0.938	0.944
4	All-{F4}	222	2	1	8	1	tansig	purelin	logsig	F4	0.81	0.964	0.958	0.966
5	All-{F5}	124	2	12	5	1	logsig	tansig	tansig	F5	0.17	0.985	1.000	0.977
										Average =		**0.963**	**0. 971**	0.959

ID: Run number. **Dtrain** and **Dtest:** Train and test data sets respectively. **Gen:** Number of generation in which after 100 consecutive generations the results of GA haven't changed. **NH:** Number of hidden layers proposed by GA. **N.H1**, **N.H2** and **N.O:** Respectively, number of nodes in the 1st hidden layer, 2nd hidden layer and output layer proposed by GA. **Func1**, **Func2** and **Funco:** Respectively, mean type of activation function of the nodes of the 1th hidden layer, 2nd hidden layer and output layer proposed by GA while **purelin**, **tangsig** and **logsig** mean linear, hyperbolic tangent sigmoid and log-sigmoid functions, respectively. **F1 ... F5:** Indicate 5 parts of data set selected by 5-Fold Cross validation. **Cut:** Cutoff classification of output vlues for achieve the highest model accuracy and sensitivity obtained based on train data. **Acc** ,**Sen**, and **Sp:** Respectively indicate the rate of the accuracy (precision) of the prediction, sensitivity and specificity of model based on test data.

As can be seen, after 5 runs, all the structures proposed by GA have two hidden layers and one node in the output layer. To show you how GA closes to the optimal solution, error reduction (in %) process of GA during successive generations, in prediction of benign and malignant masses in runs of 4 and 5 is given in the following. The horizontal and vertical axes show number of generations and error (in %) respectively.

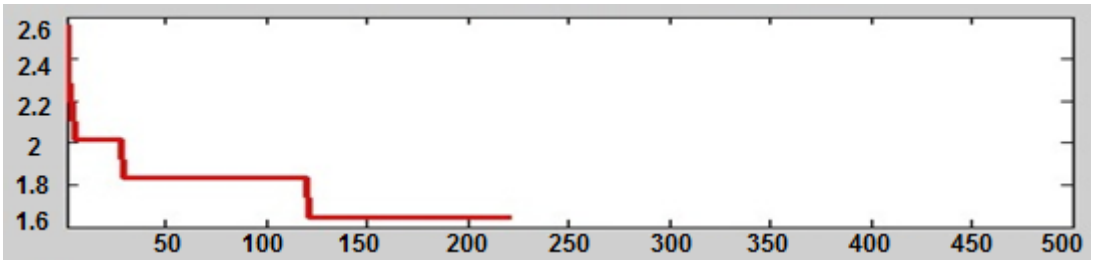

Figure 2: Error reduction (in %) process of GA in the run of 4

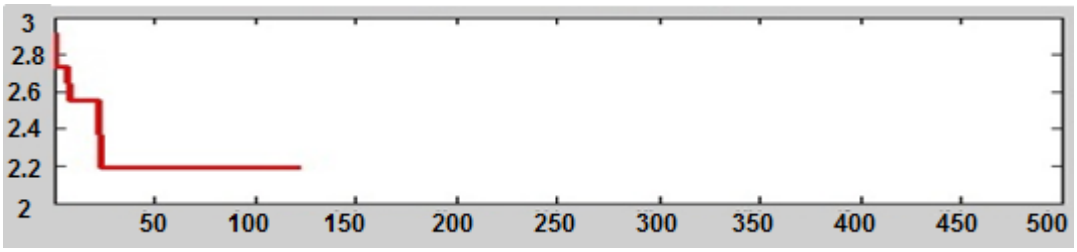

Figure 3: Error reduction (in %) process of GA in the run of 5

3.2. Step 2

Training of networks trained while their structures were determined by the GA, continued by error back propagation algorithm (EBPA) that the performance indices (accuracy, sensitivity and specificity) obtained in the test phase, are in table 3.

Table 3: Results of structure determination by GA, training by both GA and EBPA

ID	Dtrain	N.H	N.H.1	N.H.2	N.O	Func1	Func2	FuncO	Dtest	Cut	Acc	Sen	Sp
1	All-{F1}	2	6	5	1	tansig	tansig	purelin	F1	0.01	0.964	0.958	0.966
2	All-{F2}	2	8	6	1	tansig	purelin	logsig	F2	0.01	0.949	0.917	0.966
3	All-{F3}	2	6	4	1	tansig	tansig	tansig	F3	0.01	0.949	0.917	0.966
4	All-{F4}	2	1	8	1	tansig	purelin	logsig	F4	0.01	0.956	0.979	0.944
5	All-{F5}	2	12	5	1	logsig	tansig	tansig	F5	0.01	0.963	0.936	0.977
									Average =		**0.956**	**0.941**	0.964

As can be seen from the table above and according to the definition of the cost function mentioned in the section 2.3.3, prediction results in most runs and on average, not only didn't improve, but also declined.

3.3. Step 3

At this stage, only the structures of NNs were determined by the GA, and NNs with the weights that were initialized randomly, were trained by EBPA while using gradient descent algorithm, that the performance indices obtained in the testing phase, are in table 4.

Table 4: Performances of NNs whose structures were determined by GA and were trained by EBPA

ID	Dtrain	N.H	N.H.1	N.H.2	N.O	Func1	Func2	FuncO	Dtest	Cut	Acc	Sen	Sp
1	All-{F1}	2	6	5	1	tansig	tansig	purelin	F1	0.01	0.971	1.000	0.955
2	All-{F2}	2	8	6	1	tansig	purelin	logsig	F2	0.01	0.971	0.979	0.966
3	All-{F3}	2	6	4	1	tansig	tansig	tansig	F3	0.01	0.971	0.979	0.966
4	All-{F4}	2	1	8	1	tansig	purelin	logsig	F4	0.01	0.964	0.979	0.955
5	All-{F5}	2	12	5	1	logsig	tansig	tansig	F5	0.01	0.985	0.979	0.989
									Average =		**0.972**	**0.983**	0.966

Based on Table 2 and table 4, the following charts compare the performance of different structures of the NNs trained with GA and EBPA in the testing phase.

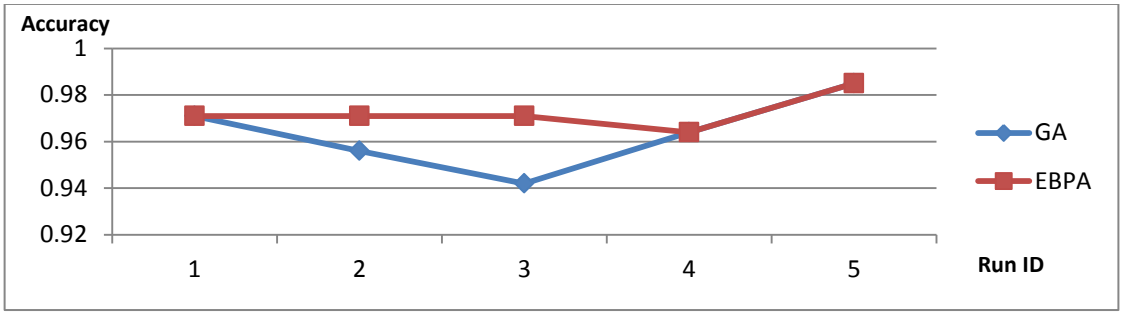

Figure 4: The Comparison of the accuracy of different structures of NNs trained with GA and EBPA

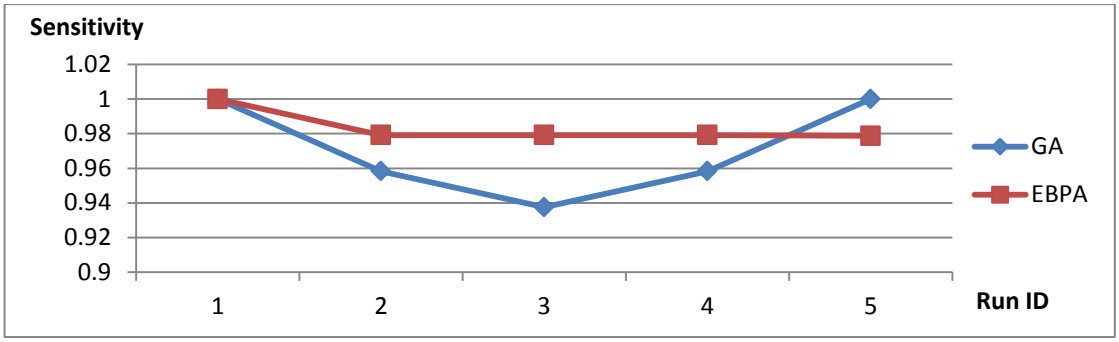

Figure 5: The Comparison of the sensitivity of different structures of NNs trained with GA and EBPA

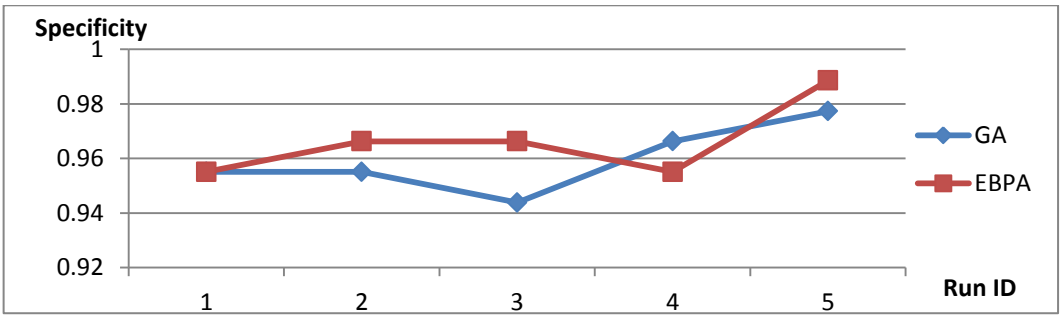

Figure 6: The Comparison of the specificity of different structures of NNs trained with GA and EBPA

Looking at above charts, table 2 and table 4, we see the performance of the structures trained by EBPA was better than the same structures trained by GA when facing the test data sets in the most runs and on average too. So, we trained each of five structures recommended by GA, with other data sets based on 5-fold cross validation technique, by EBPA, i.e. each structure was trained five times by EBPA. Thus, training phases 25 times were done totally, and after each phase, we evaluate the performance of NNs with the test data sets. Based on the results obtained in testing phases, the model which its structure was (9-6-4-1), i.e. with 6 nodes in the first hidden layer, 4 nodes in the second hidden layer and 1 node in the output layer, and activation function of nodes in each three layer was hyperbolic tangent sigmoid, with an average accuracy, sensitivity and specificity 0.974, 0.979 and 0.971 respectively, had better performance. (We passed up the detailed results of 25 runs for brevity.) Thus, we select the described structure, NN(9-6-4-1), as a typical NN model to predict benignity or malignancy of breast tumors using 9 features (independent variables) of this study data set, WBCD.

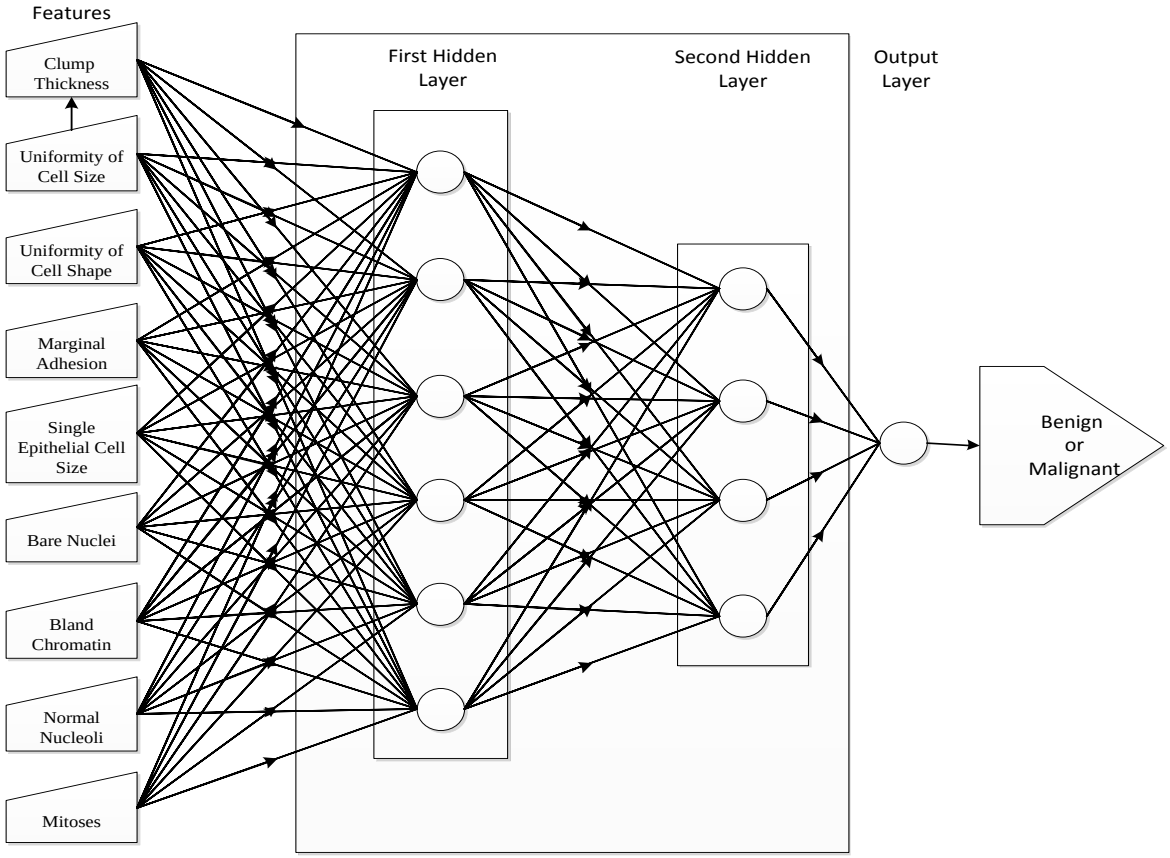

Figure 7: The topology of the typical ANN model

3.4. Comparison with logistic regression (LR)

In each run of 5 different runs, we obtained the LR model based on one of the combinations of the training data and determined the best cutoff prediction for separation of data, class 1 from class 2, then tested the obtained LR model with the rest of the data. The results come on the following table:

Table 5: Performance of the LR on training and testing data sets

ID	Training Data	Accuracy	Sensitivity	Specificity	Cutoff Prediction	Testing Data	Accuracy	Sensitivity	Specificity
1	All-{F1}	0.978	0.995	0.969	0.16	F1	0.971	1.000	0.955
2	All-{F2}	0.978	0.995	0.969	0.17	F2	0.964	0.958	0.966
3	All-{F3}	0.978	0.990	0.972	0.23	F3	0.956	0.958	0.955
4	All-{F4}	0.978	0.995	0.969	0.17	F4	0.964	0.979	0.955
5	All-{F5}	0.976	0.990	0.975	0.25	F5	0.985	0.979	0.989
	Average	**0.978**	**0.993**	**0.971**		**Average**	**0.968**	**0.975**	**0.964**

According to the above table it can be seen that the average accuracy, sensitivity and specificity of this model in the face of the testing data were 0.968, 0.975 and 0.964, respectively. By comparison with the average corresponding indices of the NN (9-6-4-1), we find that the ANN model has been more accurate, sensitive and specific than LR.

4. Discussion and conclusion

In this section, we will first discuss the findings and will then describe some general conclusions.

4.1. Relatively conservative approach

The approach of the proposed method, find the answers in order of priority, firstly, have the maximum accuracy in the diagnosis of benign and malignant tumors (to achieve higher accuracy), secondly, have the highest accuracy in the detection of malignant tumors. This means that the

answers have the same accuracy but different sensitivity and specificity, the answer is selected which be more sensitive; That is, this approach prefers false positives [32] than false negatives [32]; Because false negative (model diagnosis based on non-malignant tumor while the patient has a malignant tumor) could be dangerous for the patient, since it could exempt him/her from medical treatment while positive result (the model diagnoses that tumors are malignant) can be more precisely determined by other medical tests.

4.2. Structure Determination by GA, training by EBPA

As stated in section 3, determination of the optimal structure and weights to achieve the best NN model has totally done in three steps:

In the first step both optimized structures and weights were determined by the GA.

In the second step, for error minimization, the networks trained by the GA, were trained again by the EBPA, i.e. before the start of training by the EBPA, the primary weights of the networks were weights that had been optimized (trained) by the GA, before. At this step, in the testing phase the prediction results not only didn't improve, but also declined because of over fitting phenomenon. This means that the increasing of the repetition number of the network training to adjust the weights to consider the rare examples that may not match the overall distribution of data, reduces the network performance in the face of examples have never seen (the test data). Therefore, using this method is not recommended. In the third step, weights and bias of the networks whose structure was obtained by the GA, initialized randomly, and then the networks trained by the EBPA. In this case and looking carefully at figure 4,5 and 6, we see the structures trained by EBPA (the common method of the NN training) have been more accurate, sensitive and specific when facing the test data sets in the most runs than the same structures trained by the GA.

4.3. The presentation of a new method to achieve the best ANN model

Thus, we could present a new method to achieve the best ANN model with the help of the GA and EBPA that diagnoses the benignity or malignancy of the breast tumors, accurately. In fact, the GA population included a large number of ANN in which their structural features and the weights of the edges were the attributes of each member of the population (chromosomes) and chromosomes with better performance enhanced over many generations and produced better results. Then the bias and weights optimized by EBPA. Thus we achieved the way to the best ANN model based on existing data set. In the next figure, this method has been drawn in the form a flow chart, briefly.

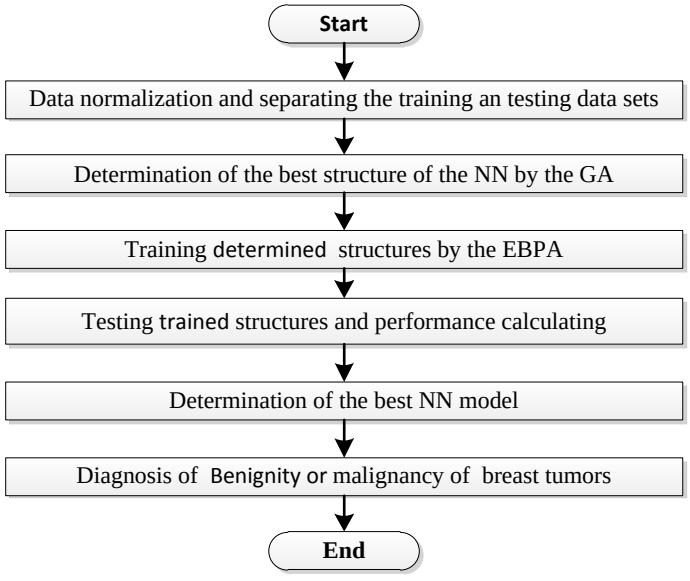

Figure 8: The flow chart of the presented method

4.4. Comparison of the results with the findings of other researchers

Cai and Jiang [26] using ANN based on matrix pseudo-inversion, predicted the benignity and malignancy of the breast tumors and the dataset used (WBCD), was the same used in our study. The accuracy, sensitivity and specificity of the proposed model on the test data, were 0.897, 0.928 and 0.841, respectively, which are lower than the corresponding values in our study. Also the accuracy of NN (9-11-1) that Karabatak and Cevdet [1] have achieved by applying on WBCD while using all of 9 features was 0.952 that is still less than the accuracy of our proposed model, NN(9-6-4-1).

4.5. General conclusions

Thus, the ANN model which is obtained by the method described, can replace invasive medical methods, and determine patients who do not need a biopsy and surgery, with high accuracy and sensitivity. This result is very important because of the possible complications and injuries of the biopsy and surgery in patients who do not need them can be avoided. On the other hand, patients who truly need diagnostic and therapeutic methods can be detected at the fastest possible time and with the highest accuracy.

5. Future Works

The approach considered in this study, i.e. the use of GA to determine the structure and EBPA for the training of the ANN, can more widely be the field of the research in the future for other researchers, including:

- Use of other evolutionary algorithms instead of GA to determine the structure of ANNs.
- Generalization of the proposed method to apply other breast cancer risk factors as input variables instead of features used in this data set.
- Use of this research approach to achieve efficient models of ANNs for diagnosis of other diseases.

ACKNOWLEDGEMENTS

The breast cancer data set used in this study was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg and has been downloaded from the UCI Machine Learning Repository web site.

References

[1]. M. Karabatak and M. Cevdet Ince, "An expert system for detection of breast cancer based on association rules and neural network", Expert Systems with Applications, Vol. 36, 2 (2009), pp. 3465-3469.

[2]. D. Michie, D. J. Spiegelhalter and C. C. Taylor, "Machine learning, neural and statistical classification", Ellis Horwood, London, (1994).

[3]. K.P. Bennett and O. L. Mangasarian, "Robust linear programming discrimination of two linearly inseparable sets", Optimization methods and software, Vol. 1, 1 (1992), pp. 23-34.

[4]. E. Freed and F. Glover F, "A linear programming approach to the discriminant problem", Decision Sciences, Vol. 12, 1 (1981), pp. 68-74.

[5]. R.C. Grinold, "Mathematical programming methods of pattern classification. Management Science", Vol. 19, 3 (1972), pp. 272–289.

[6]. F. W. Smith, "Pattern classifier design by linear programming", Computers, IEEE Transactions on Computers, Vol. 17, 4 (1968), pp. 367-372.

[7]. Y. U. Ryu, R. Chandrasekaran and V. S. Jacob, "Breast cancer prediction using the isotonic separation technique", European Journal of Operational Research, Vol. 181, 2 (2007), pp. 842-854.

[8]. S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C. H. Yeang, M. Angelo, C. Ladd et al. "Multiclass cancer diagnosis using tumor gene expression signatures", Proceedings of the National Academy of Sciences, Vol. 98, 26 (2001), pp. 15149-15154.

[9]. B. E. G. Rothberg, A. J. Berger, A. M. Molinaro, A. Subtil, M. O. Krauthammer, R. L. Camp, W. R. Bradley, S. Ariyan, H. M. Kluger and D. L. Rimm, "Melanoma prognostic model using tissue microarrays and genetic algorithms", Journal of Clinical Oncology, Vol. 27, 34 (2009), pp. 5772-5780.

[10]. K. Stadlthanner, F. J. Theis, E. W. Lang, A. M. Tomé, C. G. Puntonet and J. M. Górriz, "Hybridizing sparse component analysis with genetic algorithms for microarray analysis", Neurocomputing, Vol. 71, 10 (2008), pp. 2356-2376.

[11]. C. P. Lee, W. S. Lin, Y. M. Chen and B. J. Kuo, "Gene selection and sample classification on microarray data based on adaptive genetic algorithm/k-nearest neighbor method", Expert Systems with Applications, Vol. 38, 5 (2011), pp. 4661-4667.
[12]. R. M. Parry, W. Jones, T. H. Stokes, J. H. Phan, R. A. Moffitt, H. Fang, L. Shi et al, "k-Nearest neighbor models for microarray gene expression analysis and clinical outcome prediction", The pharmacogenomics journal, Vol. 10, 4 (2010), pp. 292-309.
[13]. H. Y. Chen, S. L. Yu, C. H. Chen, G. C. Chang, C. Y. Chen, A. Yuan, C. L. Cheng et al, "A five-gene signature and clinical outcome in non–small-cell lung cancer", New England Journal of Medicine, Vol. 356, 1 (2007), pp. 11-20.
[14]. Y. C. Hsu, Y. Chiung, S. Yuan, H. Y. Chen, S. L. Yu, C. H. Liu, P. Y. Hsu, G. Wu et al. "A four-gene signature from NCI-60 cell line for survival prediction in non–small cell lung cancer", Clinical Cancer Research, Vol. 15, 23 (2009), pp. 7309-7315.
[15]. G. Stiglic, M. Bajgot and P. Kokol, "Gene set enrichment meta-learning analysis: next-generation sequencing versus microarrays", BMC bioinformatics, Vol. 11, 1 (2010), 176.
[16]. H. B. Burke, P. H. Goodman, D. B. Rosen, D. E. Henson, J. N. Weinstein, F. E. Harrell, J. R. Marks, D. P. Winchester and D. G. Bostwick, "Artificial neural networks improve the accuracy of cancer survival prediction", Cancer. Vol. 79, 4 (1997), pp. 857-862.
[17]. G. Salomoni, M. Grassi, P. Mosini, P. Riva, P. Cavedini and L. Bellodi, "Artificial neural network model for the prediction of obsessive-compulsive disorder treatment response", Journal of clinical psychopharmacology, Vol. 29, 4 (2009), pp. 343-349.
[18]. K. Katoh and H. Toh, "Parallelization of the MAFFT multiple sequence alignment program", Bioinformatics, Vol. 26, 15 (2010), pp. 1899-1900.
[19]. K. Shedden, J. M. G. Taylor, S. A. Enkemann, M. -S. Tsao, T. J. Yeatman, W. L. Gerald, S. Eschrich et al, "Gene expression–based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study", Nature medicine, Vol. 14, 8 (2008), pp. 822-827.
[20]. S. M. Chou, T. S. Lee, Y. E. Shao and I. F. Chen, "Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines", Expert Systems with Applications, Vol. 27, 1 (2004), pp. 133-142.
[21]. F. Brunicardi and D. Andersen, "Schwartz's Principles of Surgery", 10th edn, McGraw-Hill Professional, (2014), 2069 p.
[22]. A. M. Marchevsky, S. Shah and S. Patel, "Reasoning with uncertainty in pathology: artificial neural networks and logistic regression as tools for prediction of lymph node status in breast cancer patients", Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc 12, 5 (1999), pp. 505-513.
[23]. L. J. Lancashire, R. C. Rees and G. R. Ball, "Identification of gene transcript signatures predictive for estrogen receptor and lymph node status using a stepwise forward selection artificial neural network modelling approach", Artificial intelligence in medicine, Vol. 43, 2 (2008), pp. 99-111.
[24]. R. Karakis, M. Tez and N. F. Guler, "Classification the axillary lymph node status of breast cancer patients with the analysis of pattern recognition", In Signal Processing and Communications Applications (SIU), (2011) IEEE 19th Conference on, pp. 988-991.
[25]. M. Salehi, M. R. Gohari, N. Vahabi, F. Zayeri, S.H. Yahyazadeh and MR Kafashian, "Comparison of Artificial Neural Network and Cox Regression Models in Survival Prediction of Breast Cancer Patients", Scientific Journal of Ilam University of Medical Sciences, Vol. 21, 2 (2013), pp. 120-127.
[26]. B. Cai and X. Jiang, "A novel artificial neural network method for biomedical prediction based on matrix pseudo-inversion", Journal of biomedical informatics, 48 (2014), pp. 114-121.
[27]. S. Belciug and F. Gorunescu, "A hybrid neural network/genetic algorithm applied to breast cancer detection and recurrence", Expert Systems, Vol. 30, 3 (2013), pp. 243-254.
[28]. R. Karakış, M. Tez, Y. A. Kılıç, Y. Kuru and İ. Güler. "A genetic algorithm model based on artificial neural network for prediction of the axillary lymph node status in breastcancer", Engineering Applications of Artificial Intelligence, Vol. 26, 3 (2013), pp. 945-950.
[29]. M. B. Menhaj, "Fundamentals of Neural networks", 8th edn, Amir Kabir University Publishing, Tehran, Iran, (2013) , 718 p.
[30]. R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection", In Proceeding of the international joint conference on artificial intelligence, Vol. 14, 2 (1995), pp. 1137-1145.
[31]. M. Mitchell, "An Introduction to genetic Algorithms, fifth printing", London, (1999), 158 p.
[32]. B. A. Mobley, E. Schechter, W. E. Moore, P. A. McKee and J. E. Eichner, "Predictions of coronary artery stenosis by artificial neural network", Artificial intelligence in medicine, Vol. 18, 3 (2000), pp. 187-203.
[33]. M. Louzazni and E. Aroudam, "An intelligent Fault Diagnosis Method Based on Neural Networks for Photovoltaic System", International Journal of Mechatronics, Electrical and Computer Technology, Vol. 4(11) (2014), pp. 602-609.
[34]. S. M. R. Hashemi, M. Kalantari and M. Zangian, "Giving a New Method for Face Recognition Using Neural Networks", International Journal of Mechatronics, Electrical and Computer Technology, Vol. 4(11) (2014), pp. 744-761.
[35]. D. Rahbari, " A Novel approach in Classification by Evolutionary Neural Networks", International Journal of Mechatronics, Electrical and Computer Technology, Vol. 4(10) (2014), pp. 33-52.