UGC Approved Journal

IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified

Vol. 5, Issue 10, October 2017

Diagnosis of Breast Cancer by Combining the Perceptron Neural Network Data Mining Techniques and Artificial Immune System

Esmat Banihashem[1], Touraj Banirostam[2]

Department of Computer Engineering, Electronic Branch, Islamic Azad University, Tehran, Iran[1]

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

Abstract: Breast cancer is one of the most common cancers with a high mortality rate among women, which is the most common cause of death in the female population. Early diagnosis of breast cancer increases patient's chance of survival from 56% to over 86%. In this paper we used from new method NNCAIS[i] that combination of two perceptron neural network algorithm and artificial immune system algorithm for more accurate diagnosis of tumors in breast cancer. In the combined model artificial neural network weights are trained by artificial immune algorithm and the best antibodies in the memory cell are returned as the optimized weight to the neural network. The data sets that used is from the University of California UCI website and the results are evaluated with three criteria for accuracy, precision and recall and it is compared and evaluated with some other algorithms.

Keywords: breast cancer, data mining, perceptron neural network, artificial immune system.

I. INTRODUCTION

The main purpose of using data mining algorithms in the medical field is using data to help doctors to make better decisions. In recent years, many studies have been done in the field of diagnose breast cancer and many algorithms have been successful in this regard. But the significant point is that the success rate of these algorithms depends on many factors and agents and one method can't be selected as the best method.

Factors such as the number of variables, size of the database, low number of missing data and access to the right data, increase the chances of success in data mining and make results of the algorithms are closer to success. Algorithms used in breast cancer have provided different results in terms of accuracy, sensitivity and specificity.

Point of subscription between evolutionary optimization algorithms with the neural networks topic, after designing the network structure is the learning process that ends with an optimization problem. Using the common gradient method used in the neural network, we use evolutionary optimization methods to determine the neural network weights, which are the same as neural network training.

In this paper, combination of perceptron neural network and immune System for diagnosis of breast cancer are introduced. The main objective of this article is to improve the results of breast cancer diagnosis and is evaluated by three criteria of precision, accuracy and recall.

II. BASIC CONCEPTS

A. MLP neural network

The neural networks are a subset of artificial intelligence techniques that are nowadays have become commonplace in a wide range of applications for solving many issues, including Communicator memory, optimization, prediction, diagnosis and control [1]. The structure and function of the ANN imitates the human brain and is composed of a number of simple structural components, but with a complex relationship known as the neuron. Each set of these neurons called a layer. A neural network usually consists of three input, hidden, and output layers.

The input layer only receives the information and acts as an independent variable, the output layer acts as an associated variable, and the number of its neurons depends on the number of variables, but unlike the input and output layers, the hidden layer is the main part of the neural network processing section which can include multiple layers and various neurons. The best method for determining the number of hidden neurons is the trial and error [2].

[i] Neural Network Combination Artificial Immune System

UGC Approved Journal
IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2017

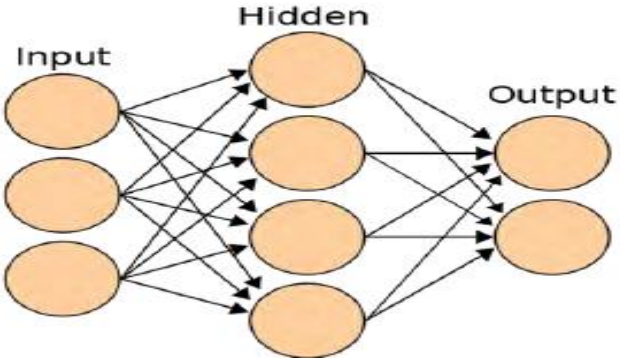

Fig1. Neural network structure

Each network training by taking examples. Train is a process that ultimately leads to learning. Network learning is performed when the communication weights between layers change so that the difference between predicted and calculated values been in acceptable limit.

Having achieved these conditions, the learning process has been realized. These weights represent memory and network knowledge. The trained neural network can be used to predict outputs proportional to the new set of data [3].
In this paper to benign or malignant breast cancer detection used from combination of Multilayer Perceptron Network (MLP) with Artificial Immune System. It should be noted that in order to optimize the network weights, the Train section of the Neural Network is performed with the artificial immune system. The weight correction process is performed for all layers, and the best weight is returned as the output to the main grid.

TABLE1. SPECIFICATIONS OF THE NEURAL NETWORK USED

Number of input layer neurons	9
number of hidden layers	1
Number of hidden layer neurons	9
Number of output layer neurons	2
Type of Function	Sigmoid
Type of Training	Error Back Propagation
Mean squared error	0.03
number of iteration	100

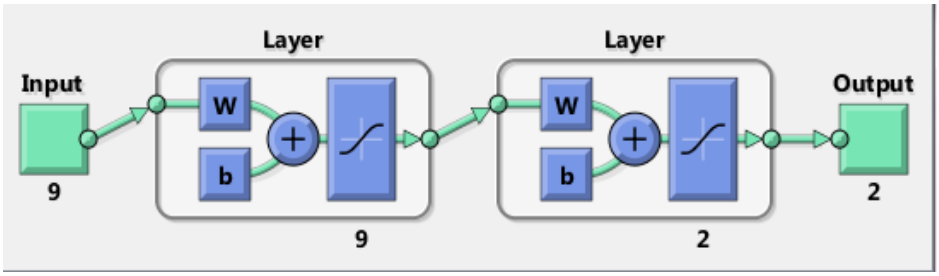

Fig2. Neural network structure in MATLAB environment

B. Artificial immune system

Artificial immune system is one of the branches of computational intelligence in computer science. AIS referred to compatibility systems that from the idea of the human body's mechanism of defence and by inspiring the processes of the natural immune system of living organisms to counteract pathogens, they have presented algorithms for solving complex computer problems. Today, many artificial systems use this idea and choose their model based on natural systems [4, 5].

One of the algorithms of the immune system developed for optimization problems and is the method of this paper for training the neural network is the ClonalG algorithm. This algorithm, after improving the memory antibodies, continues the process of algorithm addition, adding some antibodies randomly to the population of antibodies, adding a portion of the memory antibodies to increase the rate of convergence to the population of antibodies. [6]

UGC Approved Journal
IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2017

Perhaps the secret of the success of AIS algorithms is the choice of cloning. After the pathogen enters, those immune cells that identify the pathogen begin to produce and reproduce. Among the produced identifier cells, a batch is selected and stored as memory cells in order to achieve a faster and stronger immune response in subsequent collisions with similar pathogens or similar structures. During the production process, these cells are affected by a high rate mutation operator. The stages of selection and mutation are called dependence evolution. in the immune system, cells are proliferated through cell division and do not play a role in the production of new cells in the process of mating, which is why all cells produced have a structure that is quite similar to their parent cell. The immune system training process is that the immature T-cells enter the thymus and exposes T- cells to a series of Self Antigens. That is means it trains the immune system with its own antigens. T-cells that kill their own antigens are eliminated in a process, and only cells enter the blood stream that able to identify the antigens themselves [7].

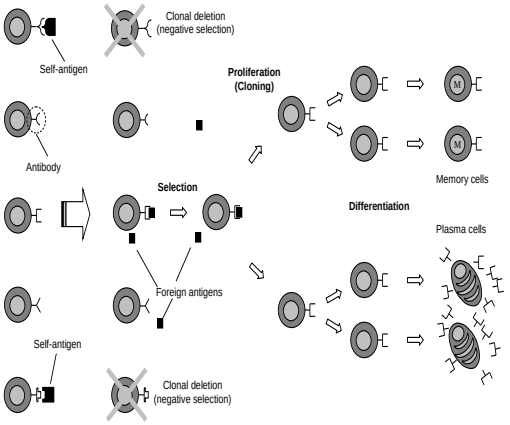

Fig3. process of colony formation in immune system

Clonal approach is a theory to explain how immune system responses the detected non-self-identified antigen patterns. According to this theory, only the cells that are active on the invasive antigen are replicated. Briefly, when a receptor detects a B cell (antibody) detects a non-identical antigen with a definite affinity, it is selected for reproduction and generates a large number of antibodies. During the production process, the generations suffered a high mutation rate and are measured by selection pattern based on the invasive antigen, in order to have a high degree of association with the standard antigen. The whole process of mutation and maturity selection is called affinity or immune response.
In addition, to differentiate between antibody-producing cells, the activated cell that has a higher affinity is selected as the memory cell, which has a longer life span. These cells have a higher priority than new cells in responding to the new pattern.
The key feature of clonal selection is the relation of direction and direction. That is, the reproduction rate of each cell depends on its association with the selection antigen, and a higher affinity leads to the production of more of this antibody, and the mutation applied to each antibody, by the relationship between the antigen and the antibody, the ratio of the image has it. That is, higher affinity leads to less mutation, and vice versa.
After the activation of non-sexual reproduction, lymphocytes begin, which is the receptor of these new lymphocytes, such as the major lymphocytes encountered with the antigen. Thus, the development of the major lymphocyte clonal occurs, and we make sure that lymphocytes that are active for the antigen are produced on a large scale. In this paper, our focus is on training neural networks and modifying network weights with artificial immune algorithms.

III. A REVIEW OF RELATED WORK

- Bichen et al. Used diagnostic of breast cancer in a feature extraction in 2014 using a combination of K-means and support machine algorithms. The proposed method improved accuracy to 97.38. In this study, the WDBC database from the University of California was used [8].
- Peng et al., (2016), explore a semi-monitoring algorithm for learning to reduce the required labels. In the proposed method, the K weight of the nearest neighbour algorithm and the clonal algorithm were used to detect breast cancer. In this research, the UCI database and the machine learning tank were used. Extensive tests performed and evaluated datasets showed effectiveness and efficiency. The proposed algorithm is a promising automatic detection method for breast cancer [9].
- Also, in 2014, Bichen Zheng worked through feature selection through a combination of backup vector machine and K-means algorithm. They named their hybrid algorithm K_SVM. They used the k-means algorithm to distinguish benign or malignant tumour patterns individually and use a support vector machine (SVM) to categorize new

UGC Approved Journal
IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2017

tumors using 10-fold validation. This hybrid method has a precision of 97.28%. This article uses the WDBC dataset. [10]

- In 2017, MovagharNejhad used a combination of two decision tree and genetic algorithms to detect benign and malignant tumours of breast cancer. The combination model has a accuracy of 96.1 [11].
- In 2017, Banihashem used a combination of two LVQ neural network and immune system algorithm to improve detection breast cancer. In this combination algorithm, to improve LVQ network weights used artificial immune algorithm. The accuracy of the LA-VQIS algorithm is 97.8 [12].

IV. PROPOSED METHOD

In this paper, we present a new approach to the NNCAIS algorithm for diagnosis of breast cancer by combining two perceptron neural network algorithm and artificial immune system algorithm. We want to model a physician's knowledge of a program that detects malignant data from benign by taking 9 attributes. In fact, our focus is on a specific application of the artificial immunity algorithm, namely, data mining. In fact, we've been focusing on a series of data routines that are typically performed with synthetic safety. And the main part of our work is on training the neural network by an artificial safety algorithm. To improve the diagnosis, the neural network training uses an artificial immune algorithm and is mapped to the antigen space by invoking a safety training program in the training of the neural network. Finally, the model produced in this article is simulated in MATLAB software. The final model has a precision of 97%, an accuracy of 98%, and a 99% recall, and the results are shown as Roc curve analysis.

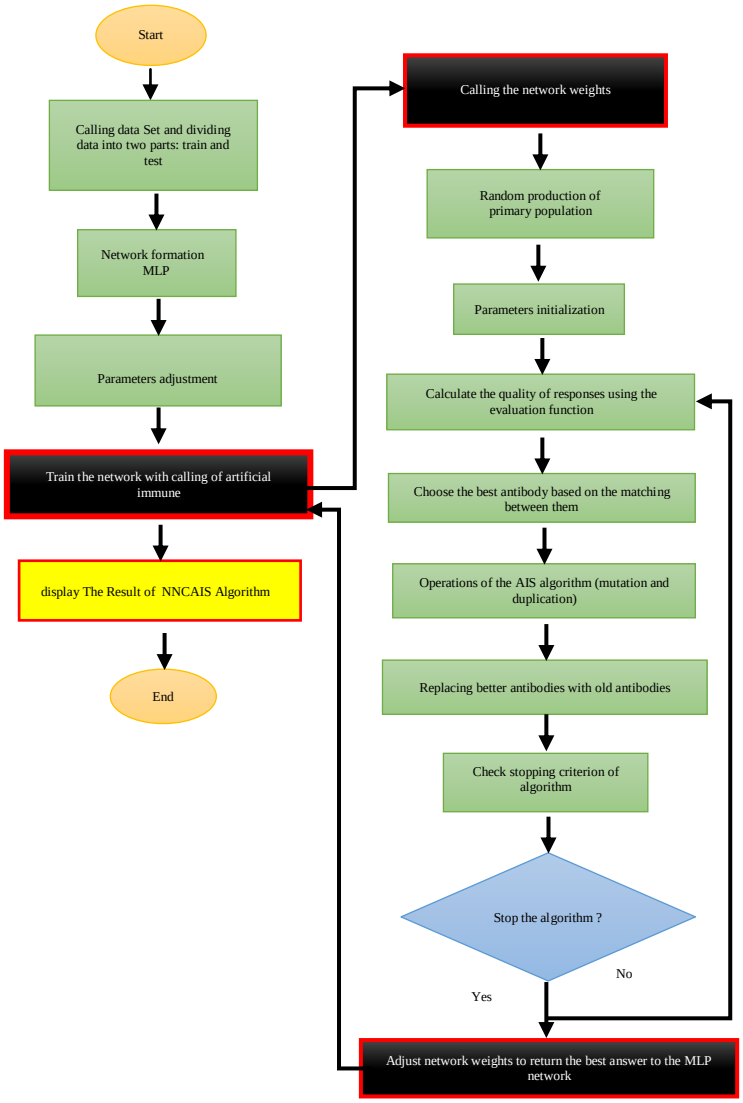

Fig4. Flowchart of Proposed Algorithm

UGC Approved Journal
IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2017

A. Introduction of the dataset

In this paper, the data collection available at the UCI University of California's Learning Machine Tank is used. The total number of records is 699 samples taken for the diagnosis of benign and malignant tumours, which is a great help in medical diagnosis. These treatments consist of 9 characteristics. Measured variables can be seen in Table 2 [1].
Each sample has a benign or malignant label. Of the 683 samples, 444 samples have benign labels and 239 have malignant labels. Properties values are integers between 1 and 10.

TABLE 2: DATA SET CHARACTERISTICS

Attribute	Id number
Clump Thickness	1-10
Uniformity of Cell Size	1-10
Uniformity of Cell Shape	1-10
Marginal Adhesion	1-10
Single Epithelial Cell Size	1-10
Bare Nuclei	1-10
Bland Chromatin	1-10
Normal Nucleoli	1-10
Mitosis	1-10

B. The proposed NNCAIS algorithm

In this paper, we focus on training the neural network and modifying the network weights with an artificial immune algorithm. That is, we want to improve the diagnosis of benign and malignant tumours by combining the MLP algorithm and the artificial immune algorithm. We compare the results of the NNCAIS algorithm with some other algorithm and evaluate the accuracy, precision and recall.
In the training section of the neural network, by calling the Train_Ais function, initially as the network input, the generated network data is taken, then the number of weights and biases in the first layer and the hidden layer of the neural network is determined, and ultimately, based on these weights, the initial solutions are constructed randomly, each of which represents the weights and network bias. The best solution found by the artificial-immune algorithm lies in the memory cell and converted to the neural network by the function defined in the program, which is done by setting up the components of the memory cell as weights and bias The network is ran and finally the network, whose weight is optimized by an artificial immune algorithm, is referred to as the output to the main program and improves the weight of the neural network and, as a result, the accuracy of the diagnosis of the opioid sample is improved. The steps are described below.

I. Antibody is created, i.e. antibodies are the same as the primary population, which is supposed to be selected and multiplied from this primitive population by the best antibodies. Initial population creation is considered as the first solution, and other solutions are randomly generated with a probability of 0.5 around the initial solution and randomly. Then, calculates the matching value. In the sense that all of the solutions are evaluated, the values in the current solution are placed in the structure of the MLP neural network. Indeed, the weights of different layers regulate the neural network.
II. A matrix is created with an initial value of 0 that this matrix holds the value of each solution, creating a solution to calculate the error value or the fitness. The generated solution adjusts network weights. Finally, it assigns the input data vector to the network weights and tests and identifies the output class associated with each data. Then, it compares the actual class with the predicted class and calculates the network error (mean squares of errors / MSE). Then it selects a number of antibodies that are most fitted to the antigen.
III. The purpose of the current study is determining those weights which minimize the error. Each time, this function tests the current solution as weights of network, and then it determines the output class of each data. Finally, it calculates the values of network's error and assesses the current solution as neural network weights by using a function to get the value of the network error value.
IV. The network calculates the qualification of the solution. In this way, the function receives a kind of solution as a vector and assigns its values to the weights and biases of network's layers. Actually, it transforms a kind of solution to the educated network. In the following, it arranges those antibodies which are more accordant with the antigens. In the next stage, the selected antibodies will duplication and mutation. The best antibodies would replace with the primary population. There is a memory in this algorithm which keeps the best solution and finds the minimum error. After that,

UGC Approved Journal
IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2017

this solution is transformed to the neural network. The parameters of memory's cell are organized as the weights and biases of neural network.
V. The network which its weights optimized with the artificial immune algorithm will return to the main program.
VI. In the next stage, we give data to the education network in order to predict the classes of data. Finally, we evaluate the three criteria accuracy, precision and recalling for the train and test data.

TABLE 3: RESULTS OF TRAIN DATA FOR THE NNCAIS ALGORITHM

Precision rate	Accuracy rate	Rate of recalling	Rate of algorithm's qualification
98.035	97.540	99.444	0.019

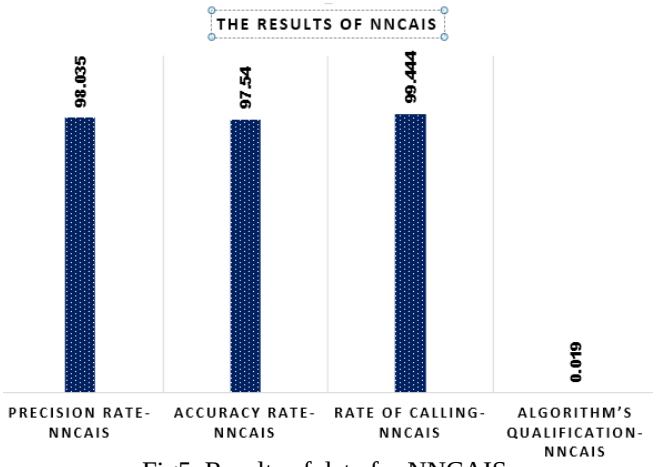

Fig5. Results of data for NNCAIS

RESULTS OF SIMULATION

In this step, the output of each network is saved in different variables which are shown in the following table. Our output contains precision, accuracy, recall and MSE error.

TABLE 4: COMPARE THE RESULTS OF TRAINING DATA FOR THE NNCAIS AND MLP ALGORITHM

Evaluation criteria	Precision rate	Accuracy rate	Rate of recalling	Rate of algorithm's qualification
NNCAIS	98.035	97.540	99.444	0.019
MLP	94.8	96.7	92.6	0.71

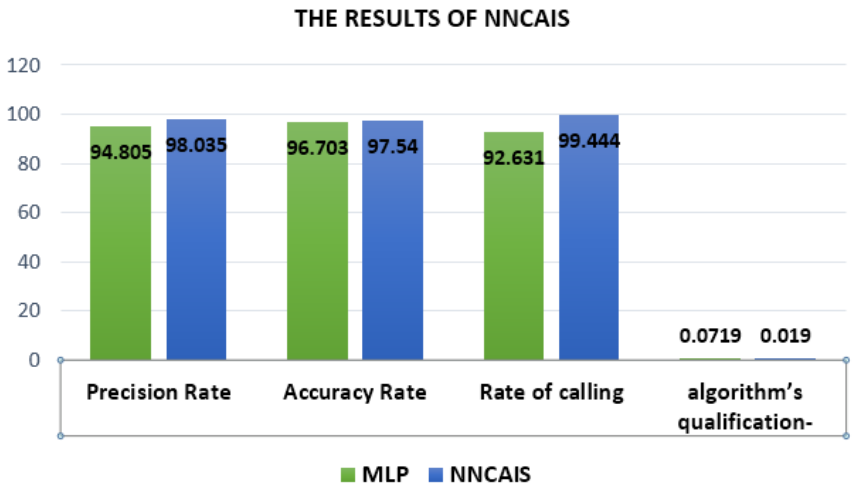

Fig6. Evaluation based on precision, accuracy, recall and MSE error.

UGC Approved Journal
IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2017

- Accuracy

Accuracy means those samples which have been determined well by means of the system. It means that, the numbers of patients with a correct diagnosis were 357 persons in the correct classification. There was also 10 sample were in the inaccurate group. 191 data from the malign data were classified in the accurate group and just 2 samples have not been chosen correctly. Totally, about 97.3% of benign tumors and 99% of malignant tumors have been classified correctly.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad (1)$$

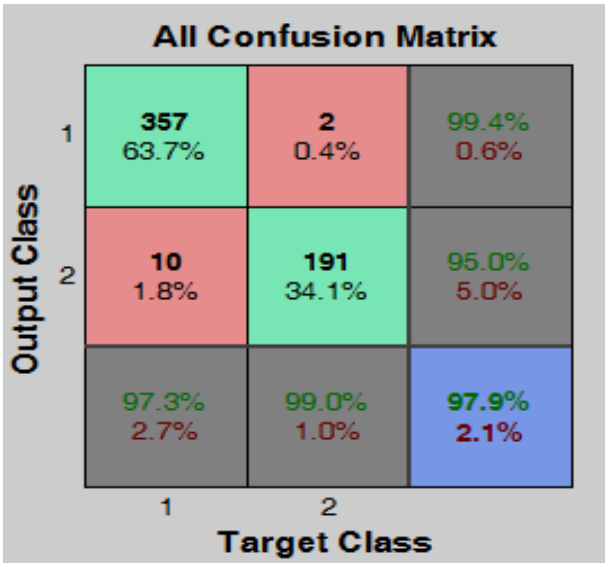

Fig7. Output matrix of the classification's accuracy

- Error rate

The rate of error is dependent on the following function. According to the figure 7, the amount of classification's error is reducing respectively.

$$\text{Error} - \text{Rate} = 1 - \text{Accuray} \quad (2)$$

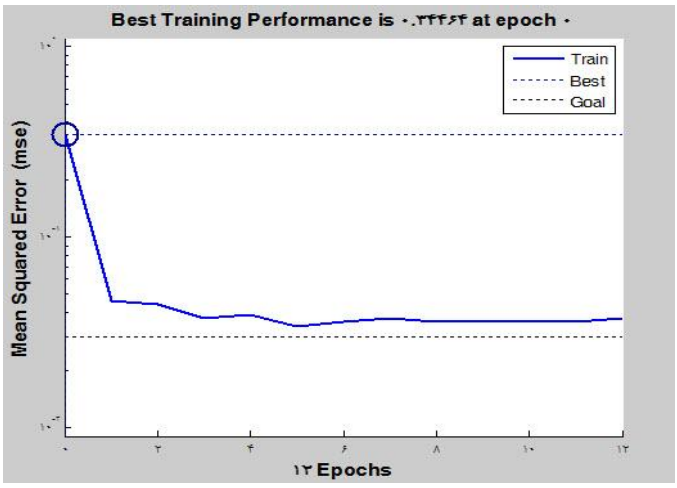

Fig8. The output of MSE chart

- Precision

The precision parameter is equal with Sensitivity. Precision means how much of the data are correct. Figure 8 illustrates the ration positive samples to their total numbers. Figure8 shows the ROC chart. It also indicates the correct classification in the TPR part.

$$\text{percision} = \frac{TP}{Tp + FP} \qquad (3)$$

UGC Approved Journal
IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2017

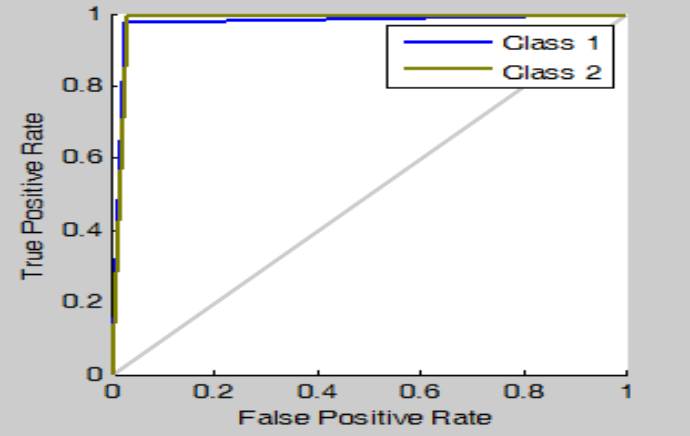

Fig9. Output of receiver operating characteristic curve

- Recalling

Recalling process means how much of the samples have been selected from the integral samples. The ratio of positive samples to the total numbers is called supervise which is obtainable through the following equation

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (4)$$

TABLE 5: COMPARISON OF THE ACCURACY OF THE DIAGNOSIS OF BENIGN AND MALIGNANT TUMORS

Model	Accuracy	Model provider	Year
Proposed NNCAIS Algorithm (the combination of the Neural Network and Artificial Immune System)	97.9	Esmat Bani Hashem	2017
LA-VQIS (the combination of the LVQ and AIS algorithms)	97.7	Esmat Bani Hashem	2017
The combinational method of genetic algorithm and decision tree	96.1	Taranom Movaghar Nejad	2017
K-SVM algorithm (the combinational method of decision vector machine and K-MEANS algorithm)	97.2	Bichen Zheng	2014
Learning algorithm machine SVM	97.1	Hiba Asri	2016
MSAIS algorithm (correction of safety system algorithm with supervisor in diagnosis	97.4	Samaneh Shojaei	2009

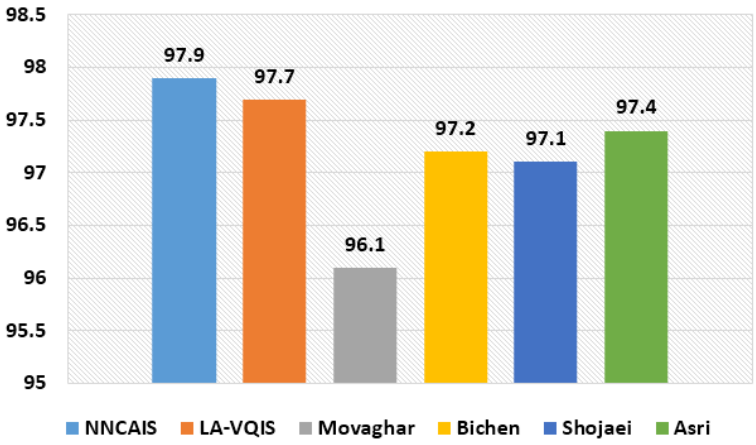

Fig10. Comparison of 5 algorithms based on accuracy in the same conditions and with the same data Set

UGC Approved Journal
IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2017

VI. CONCLUSION

The prediction of the diagnosis, survival and recurrence of breast cancer patients has always been a major challenge for researchers and doctors. Early diagnosis of breast cancer increases the patient's chance of survival. Data mining algorithms can detect samples more accurately and in less time and that's important in treating and preventing cancer. Our focus is on the specific application of artificial immune algorithm and the main part of the research is the training of the neural network by the artificial immune algorithm.
Clonal approach is a theory to explain how immune system responses the detected non-self-identified antigen patterns. According to this theory, only the cells that are active on the invasive antigen are replicated. We have used this property to improve neural network weights and better tumor detection. Finally, the results are calculated and evaluated with three criteria for accuracy, precision and recall. Totally, the results of evaluated simulation NNCAIS with 97.9% accuracy illustrate the effect of NNCAIS method on the data in order to diagnosis the breast cancer. In this way we have implemented and modeled the knowledge of an expert in medicine as a smart accurate system.

REFERENCES

[1] Bichen Zheng, Sang Won Yoon, Sarah S. Lam, “Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms”, Expert Systems with Applications 41 (2014) 1476–1482.
[2] Zheng, B., S.W. Yoon, and S.S. Lam, Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. Expert Systems with Applications, 2014. 41(4): p. 1476-1482.
[3] Asgari. Somayye , Study of Data Ming in Diagnosis of Breast Cancer , 3RD Electronic Conference of New Articles on Technology Science , 2015
[4] F. M. Burnett, "The Clonal Selection Theory of Acquired Immunity", Cambridge University Press, (1959).
[5] Kuo.LungWua, Miin.ShenYangb," Alternative learning vector quantization" , Pattern Recognition 351 – 362,2006.
[6] E. de Bodt, M. Cottrell, P. Letremy, M. Verleysen, "On the use of selforganizing maps to accelerate vector quantization", Neurocomputing 56 (2004) 187–203.
[7] Bichen Zheng, Sang Won Yoon, Sarah S. Lam, “Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms”, Expert Systems with Applications 41 (2014) 1476–1482.
[8] P. Reichardt and M. Gunzer, "The Biophysics of T Lymphocyte Activation In Vitro and In Vivo," in Cell Communication in Nervous and Immune System. vol. 43: Springer Berlin / Heidelberg, 2006, pp. 199-218.
[9] Lingxi Peng, Wenbin Chen, Wubai Zhou, Fufang Li, Jin Yang, Jiandong Zhang, an immune-inspired semi-supervised algorithm for breast cancer diagnosis, Computer Methods and Programs in Biomedicine (2016), http://dx.doi.org/doi: 10.1016/j.cmpb.2016.07.020.
[10] A. Abbas and A. Lichtman, Basic Immunology, Updated Ed. 2006-2007, W.B. Saunders Company, 2007.
[11] Movaghar Nejad, tarannom, Breast Cancer Diagnosis and Identification of Effective Features Using Decision Tree,,2017.
[12] Banihashem. Esmat, Diagnosis of breast cancer by combining the techniques of data mining and artificial immune system , IJCSN International Journal,2017

BIOGRAPHIES

Esmat Banihashem is student of Master of Science in Artificial Intelligence of Electronic Unit of Azad University in Tehran Capital of Iran. This Paper is extracted from thesis of Master of Science in Artificial Intelligence that helping doctors to better diagnosis breast cancer patients.

Touraj Banirostam is Assistant Professor of Computer Engineering Department, Islamic Azad University, Central Tehran Branch 2011-up to now and Head of Computer Engineering Department, Islamic Azad University, Electronic Campus 2013-up to now.