

Using Adaptive Neuro-Fuzzy Inference System for Classification of Microarray Gene Expression Cancer Profiles

Bulent HAZNEDAR^{1*}, Mustafa Turan ARSLAN², Adem KALINLI³

¹Department of Computer Engineering, Engineering Faculty, Hasan Kalyoncu University, 27100 Gaziantep, Turkey

²Department of Computer Technology, Kirikhan Vocational School, Mustafa Kemal University, 31440 Hatay, Turkey

³Department of Computer Engineering, Engineering Faculty, Erciyes University, 38039 Kayseri, Turkey

Received: 20.02.2018

Accepted: 27.04.2018

Published Online: 01.05.2018

Abstract: Microarray is a technology that enables simultaneously analysis of thousands of genes in DNA structure depending on the advances in biochemistry. With this technology, it has become possible to diagnose and treat heredity diseases by analyzing thousands of gene expression levels. This study proposes an artificial intelligence method, Adaptive neuro-fuzzy inference system (ANFIS), to classify cancer gene expression profiles. The findings obtained with the proposed ANFIS approach are compared with the results of statistical methods such as Naive Bayes and Support Vector Machines. In conclusion, although the highest average classification performance achieved with ANFIS is 95.56%, the highest performance achieved with statistical methods are found to be 87.65%.

Keywords: Classification, microarray gene expression, statistical methods, artificial intelligence, fuzzy logic, ANFIS.

1. INTRODUCTION

Interaction of computer technology and molecular biology resulted in a multi-disciplinary science field called "bioinformatics". Bioinformatics is a multi-discipline field because it utilizes biology to analyze the structural properties of genes while at the same time uses computer science to evaluate the data. Determination of the genetic factors underlying the disease with computer support is of interest to the bioinformatics science [1].

Researchers used to execute "a gene principle in an experiment" analysis on a gene in the past. DNA gene chip or otherwise called DNA microarray chips have been manufactured recently. DNA microarray is a gene chip that hosts thousands of genes through the spots inside [2]. With the help of such chips, evaluation and analysis of the genes became possible.

With DNA microarray technology, the gene activities of healthy and diseased cells taken from the same tissue can be compared with each other. Thus, this technology offers great facilities in finding the genes that causes diseases such as cancer. Furthermore, identifying subtypes of cancer related to genes through this technology was made possible [2]. Therefore, classification methods for finding solutions for gene expression problems have great importance.

In the literature, statistical methods are widely used to classify high dimensional gene expression profiles. Ramaswamy et al. [3] preferred Support Vector Mechine (SVM) method in order to classify tumor gene expression. Statnikov and his friends [4] performed a study on classification of microarray gene expression profile by using SVM and k nearest neighborhood (kNN) algorithm. However, the performance of statistical methods may remain limited for certain reasons such as the abundance of the number of genes and mostly non-linear relationship between the genes. Hence, researchers started to practice modern and powerful methods to acquire better results in classifying high-dimensional problems.

In recent years, relevance on the usage of artificial intelligence methods in solving high-dimensional problems has increased and many applications were utilized on the classification problems. Khan et al. [5] studied cancer prediction and classification using Artificial Neural Network (ANN) and gene expression profile. In another study, Loganathan and Girijia [6] trained Adaptive Neuro-Fuzzy Inference System (ANFIS) with the Runge Kutta learning algorithm, model and classified microarray cancer gene expression profiles through the generated classification.

ANFIS integrates best known properties of fuzzy systems and neural networks in its internal structure. The fuzzy systems gain important advantages to the ANFIS model by reducing the search space by distributing the input information over the network and using the back propagation behaviors of the neural networks to find the optimum values of the control parameters for the nonlinear problems.

In this study, artificial intelligence based ANFIS model is utilized to classify various microarray cancer gene expression profiles to in order to compare with the performance of statistical methods. In this context, ANFIS model parameters are optimized by the Backpropagation algorithm (BP) and Hybrid algorithm (HB). The performance of the recommended ANFIS models are compared with those of the statistical methods such as Naive Bayes (NB) and Support Vector Machines (SVM). In the next section, the information about the materials and methods are given. In section 3, proposed approaches are discussed and the results obtained are analyzed.

2. MATERIALS AND METHODS

2.1. Experimental Datasets

Three microarray cancer gene expression profiles are utilized as the data resource in this perspective. These data sets used in common classification were obtained from Rutgers University Bioinformatic database [7].

Bladder Cancer

Dyrskjot et. al [8] created data set 3 subtypes of bladder cancer T1, T2+ and TA extracted patients' bladder tissue in 2003. This data set consists of tumors and 7129 gene expressions were extracted from 40 patients.

Endometrial Cancer

Risinger et al. [9] constructed uterus cancer set from uterus tissues of 42 diseased and healthy people sample. This data set is represented with 8872 microarray gene sequence and contains 3 different types of uterus cancer. While number of PS tumor is 13, of CC tumor is 3 and of E tumor is 19. Seven samples from healthy tissues are extracted for the control group.

Breast Cancer

The dataset constructed by West et al. [10] in 2001 contains ER+ and ER- type tumors taken from the tissues of patients with breast cancer. This dataset consists of 49 samples in total, including 25 ER+ and 24 ER- tumors. gene expressions were extracted from 7149 patients.

2.2. Adaptive Neuro-Fuzzy Inference System (ANFIS)

ANFIS is a representation of Sugeno type fuzzy system as a network structure possessing neural learning ability [11]. It uses both artificial neural network and fuzzy logic [12]. The combination of neural network and fuzzy system is called fuzzy neural network. ANFIS, used for simplifying learning and adaptation processes, is a fuzzy Sugeno model that utilizes adaptive systems. Such an approach makes fuzzy logic more systematic and less experiential [13]. Main purpose of ANFIS, using input-output data clusters, optimizes equivalent fuzzy logic system parameters through a learning algorithm. Parameter optimization minimizes the error value between the estimated output and the target output [14].

ANFIS consists of if-then rules in fuzzy inference system and input/output information pairs. Yet, ANN learning algorithms are used in system training and supervision [15,16]. If x and y are inputs, z is taken as the output, the basic rule structure can be written as:

If $x \in A_i$ and $y \in B_i$ **then** $z_i = p_i x + q_i y + r_i$

A_i and B_i are the labels of the clusters that separate the x and y variable spaces in the forerunner partition in the fuzzy subspaces, respectively. p_i , q_i and r_i are design parameters determined during the training process. z_i is the output value for that rule and is a function of the input variables. For any x , y input pair, the output is the weighted average of the output values z_i of all the rules [13]. The calculation of z output value is given in Equation 1.

$$Z = \frac{\sum_{m=1}^k w_m Z_m}{\sum_{m=1}^k w_m} \tag{1}$$

To summarize the layers of ANFIS model, in the first layer, fuzzification process is applied to input data by using membership functions. In the second layer, the rules are created based on fuzzy logic inference system. In the third layer, each node from the rule layer is normalized with weighted average. In the fourth layer, fuzzy results are converted into numerical values. In the fifth layer, the single output value is produced by summing up output values of all nodes. The possible ANFIS architecture for the first-order Sugeno fuzzy model based on two fuzzy rules is illustrated in Figure 1.

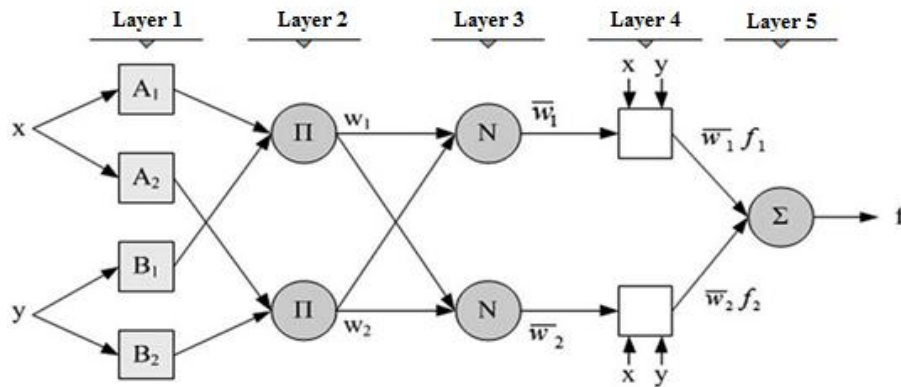


Figure 1. Adaptive Neuro-Fuzzy Inference System [15]

Rule 1: **If** $x = A_1$ and $y = B_1$ **then** $f_1 = (p_1x + q_1y + r_1)$

Rule 2: **If** $x = A_2$ and $y = B_2$ **then** $f_2 = (p_2x + q_2y + r_2)$

p_i , q_i and r_i are equation constants for each rule [14]. As can be seen in the Figure 1, ANFIS consists of 5-layer forward feed artificial neural network architecture. The main task of this architecture adjusts all of the adaptable parameters to simulate ANFIS output and training data. Training data set is introduced to neural network and the network is trained with

the help of any training algorithm. It is aimed to determine the minimum error function between estimated and target output [16].

In this study, BP and HB algorithms are used to optimize ANFIS network parameters. BP algorithm was initially proposed by Rumelhart and his friends in 1986 and it is one of the most used algorithms among ANN training models. The discovery of such algorithm is one of the most important developments for ANN [17].

The training mechanism of BP algorithm depends on the iterative gradient descent method that minimize the error between the estimated and target outputs. In this algorithm, errors are tried to be reduced by adjusting the weights. In the training rule, the error calculated on the network output is used to calculate new values of the weights. The update formula for the input parameters in the backward direction is given in Equation 2.

$$\Delta\alpha = -\eta \frac{\partial E}{\partial \alpha} \quad (2)$$

Here, α is any input parameter, η is rate of training, E displays the error value on the network output [18].

HB algorithm is used to optimize ANFIS network parameters. This algorithm is formed from the combination of “gradient descent” and “least squares” methods. The gradient descent method is used for regulating nonlinear input parameters and least squares method is used for regulating linear output parameters [19-21]. With the help of HB algorithm, input membership function parameter and output parameters are updated and optimum values are obtained [22].

3. EXPERIMENTAL STUDY

This study aims to make classification by using microarray gene expression cancer profiles. Before the classification, the feature selection method is utilized on all cancer gene expression profiles and values are normalized between range at [0-1]. The reference sources and gene numbers of the data sets are illustrated in Table 1. 70% of randomly selected data from the data set in order to analyze the accuracy of classification models are used for the training and the rest of data for testing models.

Table 1. Datasets and gene counts

Dataset	Source	Gene Count
Bladder Cancer	[8]	44
Endometrial Cancer	[9]	53
Breast Cancer	[10]	41

genfis2 function in matlab programming platform is used in order to identify type of membership function in ANFIS models, the number of membership function and number of the rules used in the classification process. This function computes the intensity of each point in the datasets with the *radius* parameter as a parameter and accumulates dataset close to each other in the set. The number of rules in ANFIS models and membership function are calculated according to cluster centers and cluster numbers. In this study, the value of the radius in the *genfis2* function used at the end of trials is set to 0.5. The values of rule and membership functions for three different cancer type is calculated as 29, 31 and 10, respectively. Furthermore, after the model is run several times with different parameters, *gaussmf* for membership function and linear function for output are selected. The training coefficient and momentum ratio for BP algorithm are selected as 0.2 and 0.4, respectively. The iteration count for each algorithm is set to 100.

The antecedent and conclusion parameters of ANFIS network are trained by BP and HB algorithms, the performances of the two different models in the classification are compared to those of statistical methods. ANFIS network is run 10 times by using both algorithms. *RMSE* error function is given Equation 3 is used in calculating the error value of the solution. *F* and *F_d* which are used in this error function, represent the output which is obtained by ANFIS and the actual output of the data, respectively. Training errors (*RMSE*) are calculated for each cancer problem. The convergence graph of training errors for 100 iterations are illustrated in Figure 2-7.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (F(i) - F_d(i))^2}{N}} \quad (3)$$

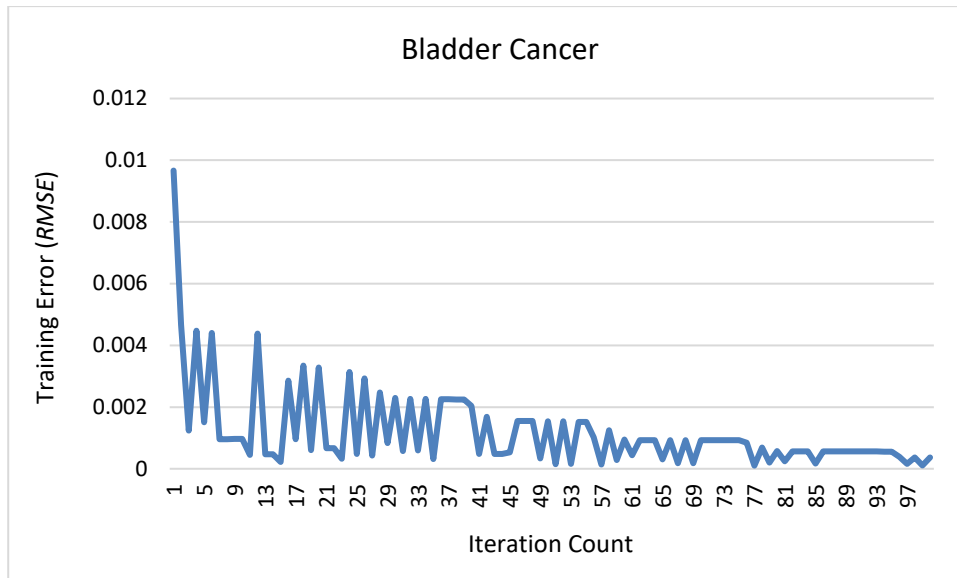


Figure 2. Training error graph of BP-ANFIS model for bladder cancer data

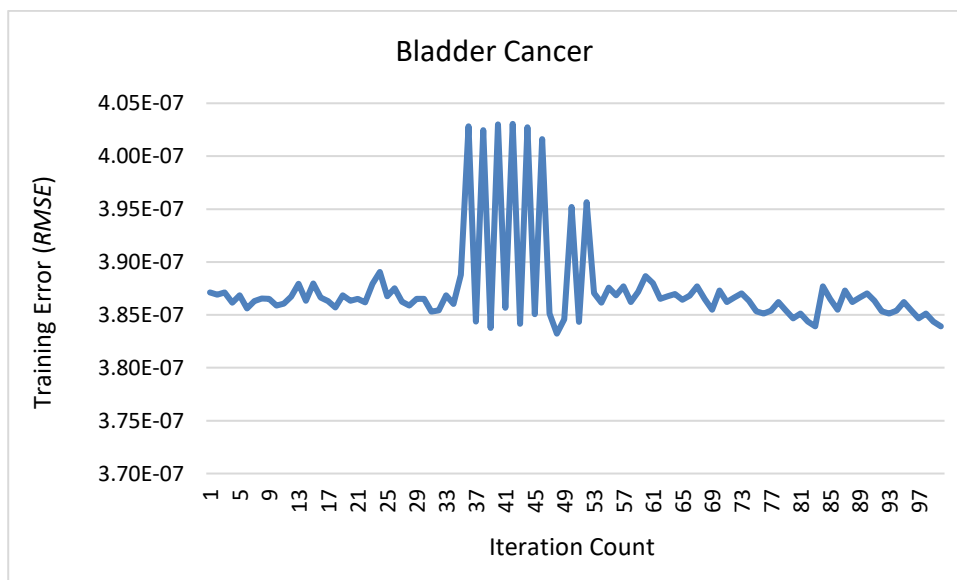


Figure 3. Training error graph of HB-ANFIS model for bladder cancer data

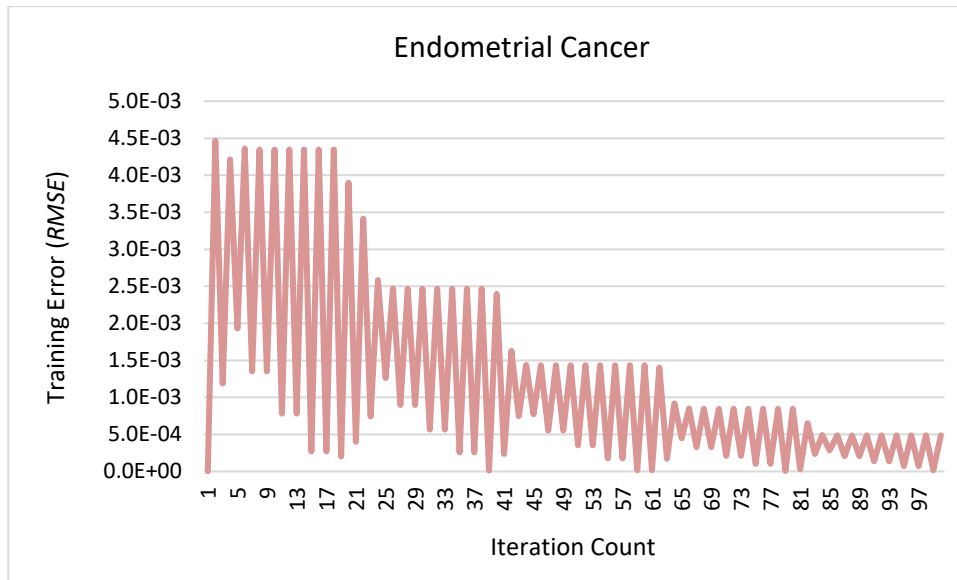


Figure 4. Training error graph of BP-ANFIS model for endometrial cancer data

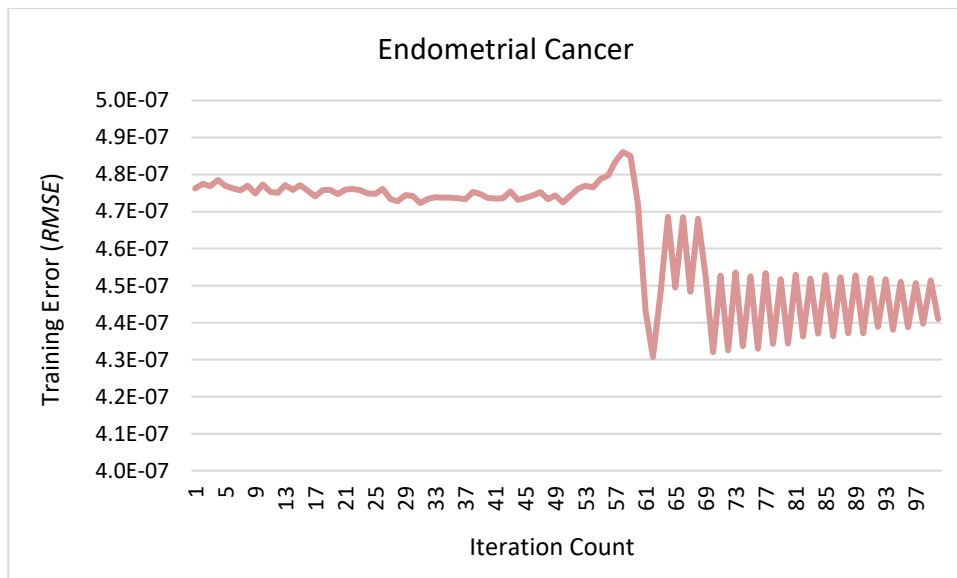


Figure 5. Training error graph of HB-ANFIS model for endometrial cancer data

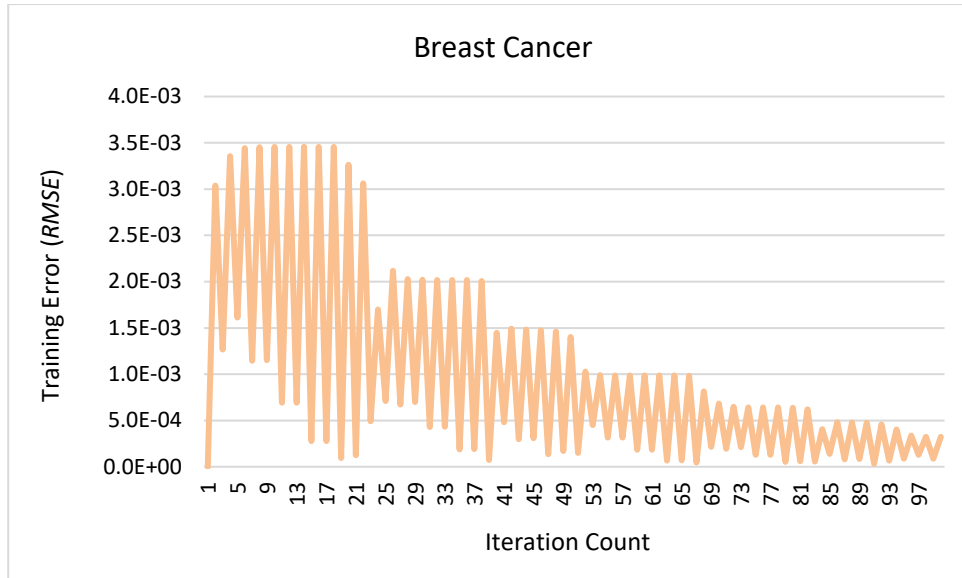


Figure 6. Training error graph of BP-ANFIS model for breast cancer data

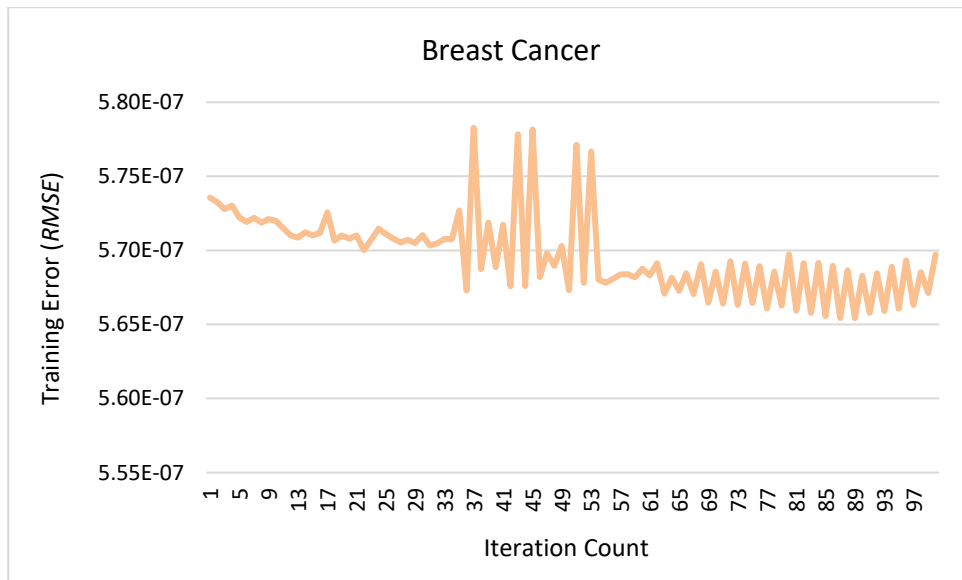


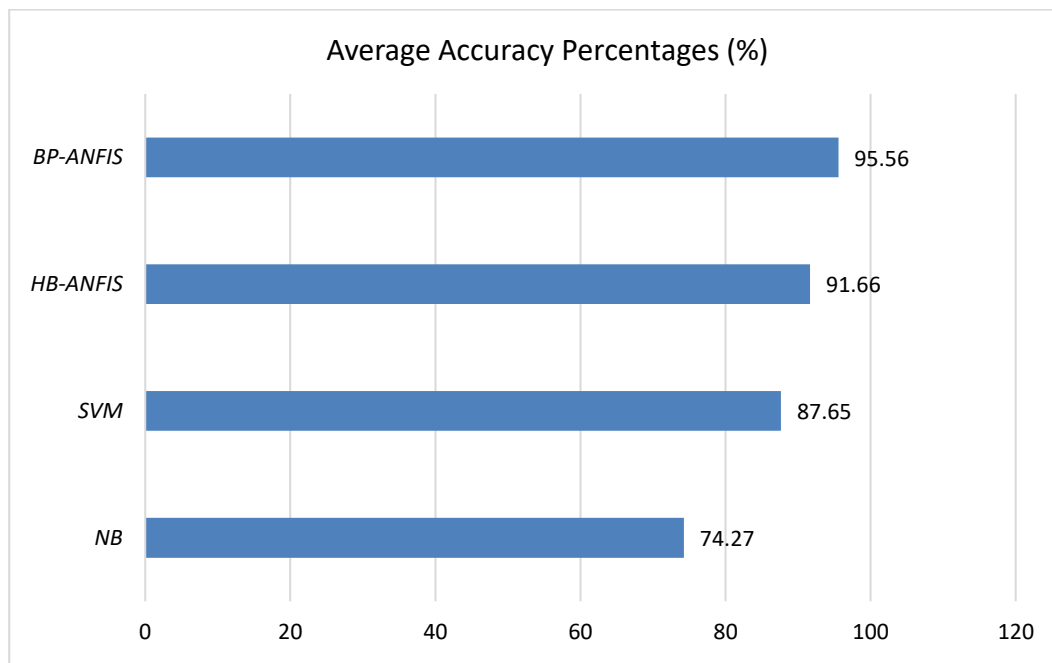
Figure 7. Training error graph of HB-ANFIS model for breast cancer data

As a result of simulation studies, the classification accuracy rate obtained with optimal parameter values of the models on different cancer profiles are given in Table 2. The average success rate in order to classify microarray cancer profiles with different methods is showed in Figure 8.

Table 2. Accuracy percentages of the classification models

Dataset	NB (%)	SVM (%)	BP-ANFIS (%)	HB-ANFIS (%)
Bladder Cancer	83.33	91.66	100	91.66
Endometrial Cancer	46.15	84.62	100	83.33
Breast Cancer	93.33	86.67	86.67	100

According to the accuracy percentages in Table 2, while BP-ANFIS method is found as the most successful method with 100% accuracy rate for classifying of bladder cancer, NB method has the lowest accuracy rate with only 83.33%. Similarly, BP-ANFIS is the most successful method for classifying endometrial cancer with 100% accuracy rate. On the other hand, NB has the lowest performance among classification methods of the endometrial cancer with only 46.15%. Finally, while HB-ANFIS is discovered to be the most successful method for classifying breast cancer with 100%, SVM and BP-ANFIS methods have the lowest performances with 86.67% accuracy rate.

**Figure 8.** Accuracy percentages of models on microarray cancer gene expression profiles

BP-ANFIS method has the highest classification accuracy rate with 95.56% according to the average performance given in Figure 8.

4. RESULT AND DISCUSSIONS

In this study, simulation studies are performed for classifying high-dimensional microarray cancer gene expression. In this context, The performance of ANFIS networks trained by BP and HB algorithms are compared to those of the statistical methods such as NB and SVM. The obtained results showed that ANFIS models could classify high-dimensional cancer gene expression profiles more successfully than statistical methods. Moreover, it has been clearly understood that the proposed ANFIS models can be successfully used in the applications for different high dimensional problems.

CONFLICTS OF INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this paper.

ACKNOWLEDGEMENT

This study was supported by Erciyes University Scientific Research Project Coordination Unit with project number FDK-2016-6371. Also, we thank to TUBITAK Scientific Human Resource Support Department for 2211-A General National Doctorate Scholarship.

REFERENCES

- [1] Polat, M., Karahan, A.G., "A new multi-disciplinary scientific branch: Bioinformatics and its applications in medicine," *SDU Medical Faculty Journal*, vol. 16, no. 3, pp. 41–50, 2009.
- [2] Schulze, A., Downward, J., "Navigating gene expression using microarrays - a technology review," *Nature Cell Biology*, vol. 3, no.8, pp. 190-195, 2001.
- [3] Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Mukherjee, Yeang, C.H., Angelo, M., Poggio, T., "Multiclass cancer diagnosis using tumor gene expression signatures," *Proceedings of the National Academy of Sciences*, vol. 98, no. 26, pp. 15149-15154, 2001.
- [4] Statnikov, A., Aliferis, C. F., Tsamardinos, I., Hardin, D., Levy, S., "A comprehensive evaluation of multcategory classification methods for microarray gene expression cancer diagnosis," *Bioinformatics*, vol. 21, no. 5, pp. 631-643, 2005.

- [5] Khan, J., Wei, J. S., Ringner, M., Saal, L. H., Ladanyi, M., Westermann, F., Meltzer, P. S., "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks," *Nature medicine*, vol. 7, no. 6, pp. 673-679, 2001.
- [6] Loganathan, C., Girija, K.V., "Cancer classification using adaptive neuro fuzzy inference system with runge kutta learning," *International Journal of Computer Applications*, vol. 79, no. 4, 2013.
- [7] Arslan, M. T., "Classification of microarray gene data using artificial intelligence based methods" M.S.c. Thesis, Erciyes University, Kayseri, Turkey, 2016.
- [8] Dyrskjøt, L., Thykjaer, T., Kruhøffer, M., Jensen, J.L., Marcussen, N., Hamilton-Dutoit, S., Ørntoft, T.F., "Identifying distinct classes of bladder carcinoma using microarrays," *Nature genetics*, vol. 33, no. 1, pp. 90-96, 2003.
- [9] Risinger, J.L., Maxwell, G. L., Chandramouli, G.V.R., Jazaeri, A., Aprelikova, O., Patterson, T., Barrett, J.C., "Microarray analysis reveals distinct gene expression profiles among different histologic types of endometrial cancer," *Cancer research*, vol. 63, no. 1, pp. 6-11, 2003.
- [10] West, M., Blanchette, C., Dressman, H., Huang, E., Ishida, S., Spang, R., Nevins, J.R., "Predicting the clinical status of human breast cancer by using gene expression profiles," *Proceedings of the National Academy of Sciences*, vol. 98, no. 20, pp. 11462-11467, 2001.
- [11] Ozcalik, H., Uygur, A., "Efficient modeling of dynamic systems by using adaptive network based fuzzy inference system," *KSU Journal of Science and Engineering*, vol. 6, no. 1, pp. 36-46, 2003.
- [12] Avci, E., Akpolat, Z.H., "Speed control of DC motor by using adaptive network based fuzzy inference system," in *ELECO'2002 Electrical-Electronics and Computer Engineering Symposium*, pp. 193-196, Bursa, Turkey, 2002.
- [13] Ozgan, E., Kap, T., Beycioglu, A., Emiroglu, M., "The prediction of marshall stability of asphalt concrete by using adaptive neuro fuzzy inference system," in *International Advanced Technologies Symposium*, Karabuk, Turkey, 2009.
- [14] Guney, K., Sarikaya, N., "Patch radius calculation for circular microstrip antennas using adaptive network based fuzzy inference systems optimized by various algorithms," in *ELECO'2008 Electrical-Electronics and Computer Engineering Symposium*, pp. 26-30, Bursa, Turkey, 2008.
- [15] Uzundurukan, S., "Determination and modeling of basic parameters that affect the swelling properties of soils," Ph.D. Thesis, Suleyman Demirel University, 2006.

- [16] Hímer, Z., Kovács, J., Benyó, I., Kortela, U., “Neuro- fuzzy modeling and genetic algorithms optimization for flue gas oxygen control,” in *2nd IFAC Workshop on Advanced Fuzzy/Neural Control*, University of Oulu, 2004.
- [17] Rumelhart, D.E., Hinton, G.E., Williams, R.J., “Learning representations by back-propagating errors,” *Cognitive modeling*, vol. 5, no. 3, 1988.
- [18] Ozturk, C., “Training artificial neural networks with artificial bee colony algorithm,” Ph.D. Thesis, Erciyes University, Kayseri, Turkey, 2011.
- [19] Nayak, P.C., Sudheer, K.P., Rangan, D.M., Ramasastri, K.S., “A neuro-fuzzy computing technique for modeling hydrological time series,” *Journal of Hydrology*, vol. 291, no. 1, pp. 52-66, 2004.
- [20] Firat, M., and Gungor, M., “River flow estimation using adaptive neuro fuzzy inference system,” *Mathematics and Computers in Simulation*, vol. 75, no. 3, pp. 87-96, 2007.
- [21] Haznedar, B., “Determine the presence of genetic anomaly in male infertile patients by using artificial intelligence techniques,” M.S.c. Thesis, Erciyes University, Kayseri, Turkey, 2010.
- [22] Ozgan, E., “Effect of the quantity of NAPO_3 to soil particles and prediction of the particle sizes by using anfis method,” in *International Advanced Technologies Symposium*, Karabuk, Turkey, 2009.