



Mustafa Turan Arslan

Mustafa Kemal University, mtarслан@mku.edu.tr, Hatay-Turkey

Derya Arslan

İskenderun Teknik University, derya.oguz.92@gmail.com, Hatay-Turkey

Bülent Haznedar

Hasan Kalyoncu University, bulent.haznedar@hku.edu.tr, Gaziantep-Turkey

DOI	http://dx.doi.org/10.12739/NWSA.2018.13.4.2A0159	
ORCID ID	0000-0001-5498-6571	0000-0002-5283-2306
	0000-0003-0692-9921	
CORRESPONDING AUTHOR	Mustafa Turan Arslan	

TRAINING ANFIS SYSTEM WITH GENETIC ALGORITHM FOR DIAGNOSIS OF PROSTATE CANCER

ABSTRACT

Prostate cancer is one of the most common types of cancer among males as well as causing the most deaths. Early diagnosis of prostate cancer plays an important role in the treatment of the disease. Therefore, microarray technology is widely used in the diagnosis of inherited diseases such as prostate cancer. With this technology, it is possible to obtain more knowledge about cancer by analyzing thousands of gene expressions. However, it is quite difficult to analyze complex relationships among thousands of genes in microarray data. For this reason, high performance artificial intelligence-based classification methods are needed in recent years. In this study, a hybrid method has been proposed for optimizing the parameters of Adaptive Neuro Fuzzy Inference System (ANFIS) with Genetic Algorithm (GA) in order to classify prostate cancer gene expression profiles. The performance of the proposed method is compared with those of ANFIS models trained by different learning algorithms. According to obtained results, the proposed method is more successful than the other methods, with the accuracy of 90.32%.

Keywords: Microarray, Prostate cancer, Classification, ANFIS, Genetic Algorithm

1. INTRODUCTION

Prostate cancer develops as a result of malignant changes in the prostate, an important part of the male reproductive system. The disease is usually one of the most common types of cancer in men at middle and older ages. Although this disease accounts for 11% of all cancers in men, it also accounts for 9% of all cancer-related deaths [1]. Prostate cancer is second most common cause of cancer-related deaths. Approximately, 30.000 men lose their lives each year due to prostate cancer although survival time of patients can be extended with successful operations performed in medical clinical [2]. The cause of prostate cancer cannot be determined for certain. However, it is estimated that genetic and environmental factors play a role in the development of the disease such as age, hormonal changes, some vitamins and having prostate cancer in the family. In the literature, there is lots of study to classify microarray gene expression data. Haznedar et al. [3] studied on classifying Liver cancer microarray gene expression data using Adaptive Network based Fuzzy Inference

How to Cite:

Arslan, M.T., Arslan, D., and Haznedar, B., (2018). Training Anfis System with Genetic Algorithm for Diagnosis of Prostate Cancer, **Technological Applied Sciences (NWSATAS)**, 13(4): 301-309, DOI:10.12739/NWSA.2018.13.4.2A0159.



System (ANFIS) and Genetic Algorithm (GA). According to obtained results, proposed approach performed very well, with an accuracy of 92.14%. Arslan and Kalinli [4] compared the performances of statistical and Artificial Intelligence-based methods by classifying Central Nervous System cancer microarray gene expression dataset. They used commonly known methods to classify such as Multi-Layer Perceptron (MLP), Radial basis Function Network (RBFNetwork), J48 Decision Tree (J48 DT), Support Vector Machines (SVM), One-R algorithms. For classifying the cancer dataset, the best method was found to be MLP with 97.06% accuracy. Arslan and Haznedar [5] classified prostate cancer features predicated on microarray gene expressions employing classification algorithms, called Hybrid-ANFIS, Artificial Neural Network (ANN), k-Nearest Neighbour (k-NN) and J48 DT. They tested the performance of classification algorithm with 5-fold cross validation. They reported 93.18%, 91.18%, 89.22% and 88.24% accuracy for Hybrid-ANFIS, k-NN, ANN and J48 DT, respectively. In another study, Arslan and Kalinli [6] classified prostate cancer data based on microarray gene expression profile with widely used methods namely, SVM and k-NN. The most successful results were obtained by SVM-RFE. They reported 100% and 98.04% accuracy for SVM and k-NN, respectively. Furthermore, the interest of researchers has gradually started to focus on artificial intelligence. ANFIS, which is one of the artificial intelligence methods, is the representation of the fuzzy system as a network structure with neural learning capability. Both artificial neural networks and fuzzy logic are used in the structure. The combination of neural network and fuzzy system is called fuzzy neural network. Such an approach makes fuzzy logic more systematic and less experiential [7]. The main purpose of ANFIS structure is to optimize antecedent and consequent parameters.

It has been known that derivative based algorithms are widely used in the training of ANFIS parameters such as backpropagation algorithm and hybrid algorithms. However, there is a problem which cause the convergence of local minimum in derivative based algorithms. In this context, artificial intelligence-based optimization algorithms have been proposed in recent years to train ANFIS parameters [8-12]. Some of these algorithms are heuristic algorithms such as Genetic Algorithm, PSO and Differential Development Algorithm [7]. In this study, a hybrid approach, consisting of ANFIS and GA, is proposed for the diagnosis of prostate cancer. The GA optimization algorithm is used for the training of all these parameters of ANFIS network. Subsequently, the analysis result of the proposed approach is compared with those of different ANFIS methods. The rest of the paper is organized as follows. Section 2 briefly describes the methods and the data set. We give information about the proposed approach and the results of analysis are presented in the section 3. Finally, the conclusion is given in Section 4.

2. RESEARCH SIGNIFICANCE

Tuning parameter of ANFIS by derivative-based algorithms is commonly used. However, those algorithms consist of a local minimum problem. In this study, a hybrid approach, consisting of ANFIS and population-based GA which is an optimization algorithm, is proposed for the diagnosis of prostate cancer to overcome this type of problems.

3. MATERIAL AND METHODS

The section describes the details of the material and methods used during experimental study, which include the experimental tasks,

dataset description and description of algorithms building the proposed method.

3.1. Prostate Cancer Dataset

This dataset, provided by Singh et al. [13], contains the expression levels of 12533 genes. Each sample is obtained from prostate tissues and is analyzed using Affymetrix microarrays. The dataset consists of 52 of prostate tumor (PR) and 50 of normal prostate (N) samples. It comprises of two classes, namely prostate cancer and normal.

3.2. Genetic Algorithm

Genetic Algorithm (GA) proposed by John Holland [14] in 1970s is designed considering the natural development mechanism of living things [15]. GA is a metaheuristic method for finding optimal results in search and optimization problems. This algorithm can find values that can be better at a certain time rather than searching the whole of the problem space. GA produces solutions that are constantly evolving based on the principle of maintaining the life of the best in nature. It uses a fitness function to determine developing solutions and operators such as mutation, reproduction, crossover, selection to produce new solutions [15]. The basic steps of GA is given below [16].

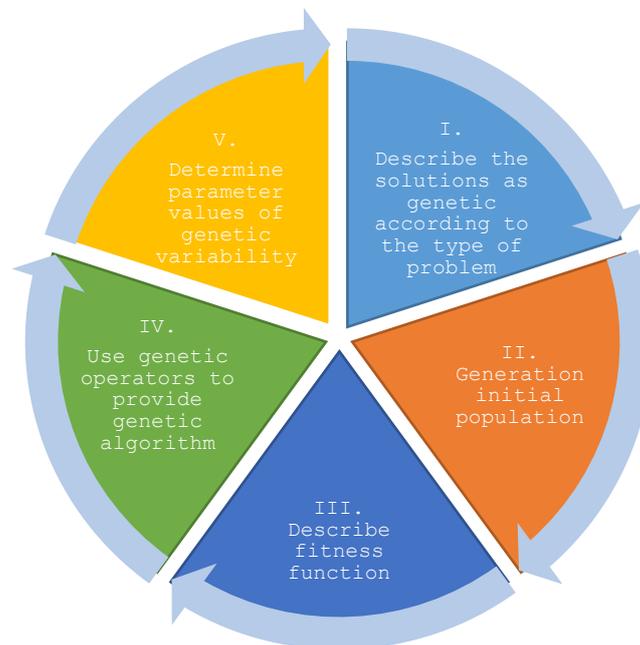


Figure 1. The basic steps of GA

3.3. Adaptive Neuro-Fuzzy Inference Systems

The ANFIS network, which contains the best features of fuzzy systems and neural networks, was described by Jang [17] in 1992. ANFIS contains if-else rules and fuzzy input-output data pairs. This network uses neural network learning algorithms for training. The structure of ANFIS consists of Artificial Neural Networks (ANN) and Fuzzy Logic model. The structure of ANFIS is given in Figure 2.

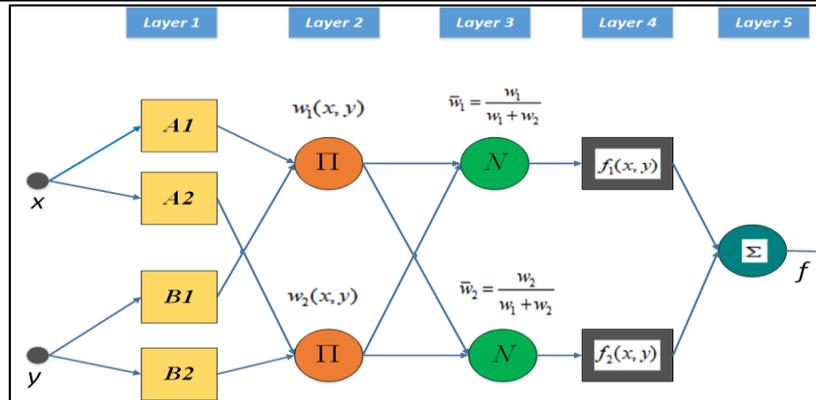


Figure 2. The ANFIS structure that has two inputs and one output

There are basically five layers in the structure of ANFIS. These layers are;

- **Layer 1:** input values are mapped using membership functions in this layer known as the input layer. The generalized bell curve activation function is preferred as membership function in separating input values into fuzzy sets. The values of membership degree vary in the range of (0-1).

$$\mu_{A_i}(x, a_i, b_i, c_i) = \frac{1}{1 + \left[\left(\frac{x - c_i}{a_i} \right)^2 \right]^{b_i}} \quad (1)$$

$\{a_i, b_i, c_i\}$ are called antecedent parameters. The parameters are updated to determine the membership functions while the system is being trained [18].

- **Layer 2:** This layer is known as the rule layer. Each node is in the form of circle in the layer. The nodes represent the number of generated rules in accordance with the Takagi-Sugeno fuzzy inference system. The output of each node is the multiplication of the membership degrees from the input layer [18].

$$O_{2i} = w_i = \mu_{A_i}(x) \cdot \mu_{B_i}(y) \quad i=1,2 \quad (2)$$

- **Layer 3:** This layer is also known as the Normalization layer. Each node in the layer considers the information from the rule layer as input value and calculates the normalized firing strength of each rule.

$$O_{3i} = \bar{w}_i = \frac{w_i}{w_1 + w_2} \quad i=1,2 \quad (3)$$

- **Layer 4:** This layer, known as the purification layer calculates the weighted result values of the rules generated according to the Takagi-Sugeno method [18]. The nodes in this layer are adaptive nodes. The output of these nodes is given in Equation 4.

$$O_{4i} = \bar{w}_i \cdot f_i = \bar{w}_i \cdot (p_i x + q_i y + r_i) \quad i=1,2 \quad (4)$$

The variables (p_i, q_i, r_i) are the result set of parameters for the rule i .

- **Layer 5:** This layer consists of the only node. This layer, which consists of a single node, is also known as the sum layer. The output values of the nodes in the purification layer are summed to determine the actual output value of the ANFIS



structure. The output value of the ANFIS structure is expressed by Equation 8.

$$O_{5i} = f = \sum \bar{w}_i f_i = \frac{\sum w_i f_i}{\sum w_i} \quad i=1,2 \quad (5)$$

4. RESULTS AND DISCUSSION

In this study, microarray gene expression is used for diagnosis of prostate cancer dataset. Before the analysis study, the dataset is normalized to the value range (-1, 1). Processing time is very important in real time applications. For this reason, the analysis process to be performed must be as fast as possible. Thus, reducing the number of features is of importance. In the proposed approach, for the step of feature selection, Correlation based Feature Selection (CFS) is utilized and the data set is represented by 24 features. In order to measure the performances of the ANFIS methods on the data sets, 70% of all the data is used for training and the remaining data is used for testing phase. Due to the large number of parameters in the prostate cancer microarray gene expression data set, a method is designed by using ANFIS. Fuzzy C-Means (FCM) clustering method used to generate fuzzy rules to be able to create ANFIS models with fewer fuzzy rules and membership functions. The number of inputs in the ANFIS model is equal to the number of genes in the data set. The type of membership functions, the number of membership functions and rule are defined as Gaussian, 10 and 10, respectively. ANFIS consists of two parameter types that must be trained, namely antecedent and consequent. The total number of antecedent parameters belong to Gaussian membership functions is equal to the sum of parameters in all membership functions. The consequent parameters are used in the defuzzification layer. In this study, The GA optimization algorithm is used for the training of all these parameters of ANFIS network. The block diagram which shows the training of proposed method is given in Figure 3. In the simulation study, within the experimentation, the parameters utilized as a part of the GA are controlled as characterized in Table 1. A solution space is initialized with the FCM clustering method for the initial values of the ANFIS parameters. Subsequently, the parameters of ANFIS are trained by using the GA algorithm. The root means square error (RMSE) function given in Equation (6) is used for determining the error value of the solution. f and Output Reference Value represent the desired output and actual output, respectively.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (F(i) - F_d(i))^2}{N}} \quad (6)$$

Table 1. The GA parameters applied in the experiment

Iteration Number	100
Population Size	25
Reproduction Rate	0.01
Crossover Rate	0.8
Mutation Rate	0.01

The ANFIS network is also trained by Backpropagation (BP) and Hybrid (HB) learning algorithms. The obtained results are compared with that of proposed method. In this context, the learning rate and momentum rate of the BP algorithm are chosen as 0.2, 0.4, respectively. The number of iterations for the BP and HB is defined as 100. The training error values (RMSE) of the methods for 100

iterations are given in Figure 4. Besides, the test error values (*RMSE*) which show the difference between the actual output of the test data and the desired output, are given in Figure 5.

As shown in Figure 4 and Figure 5, although the least training error value is found with the HB algorithm, the least test error value is obtained with GA. A major difference between training and test error values indicates that ANFIS model is not feasible because of over-learning. On the contrary, a minor difference between training and test error values reveals ANFIS model is more robust and high performance. The results of classification performance achieved with optimal parameter values at the end of simulated studies are given in Figure 6. As can be seen in the figure below, the most successful classification performance belongs to the ANFIS-GA method with an accuracy of 90.32% on prostate cancer.

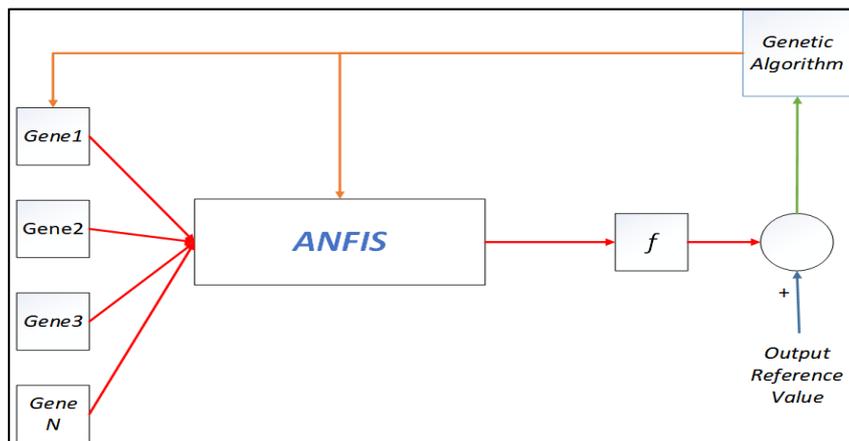


Figure 3. The basic system of proposed method

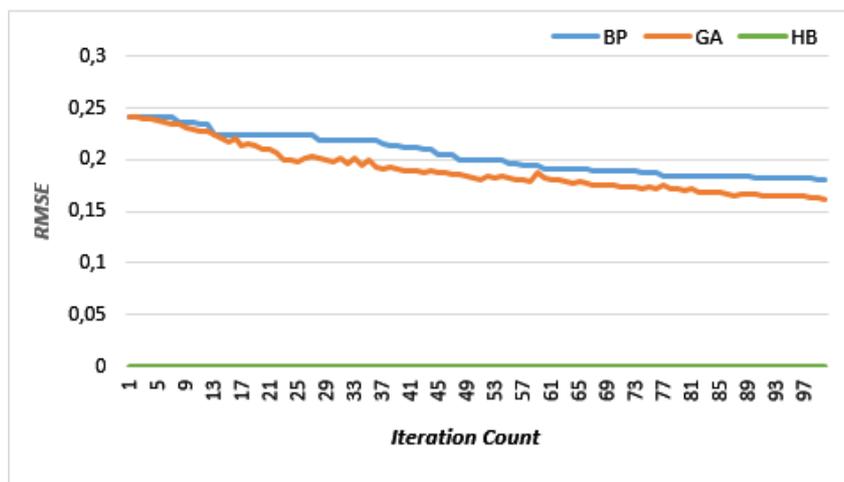


Figure 4. Training *RMSE* error values of the ANFIS models

As expected, GA, a metaheuristic algorithm is more successful than derivative-based algorithms which have constraints such as local minimum, namely BP and HB for training of ANFIS. According to Figure 6, it is shown that prostate cancer can be early diagnosed by the proposed method, with high accuracy.

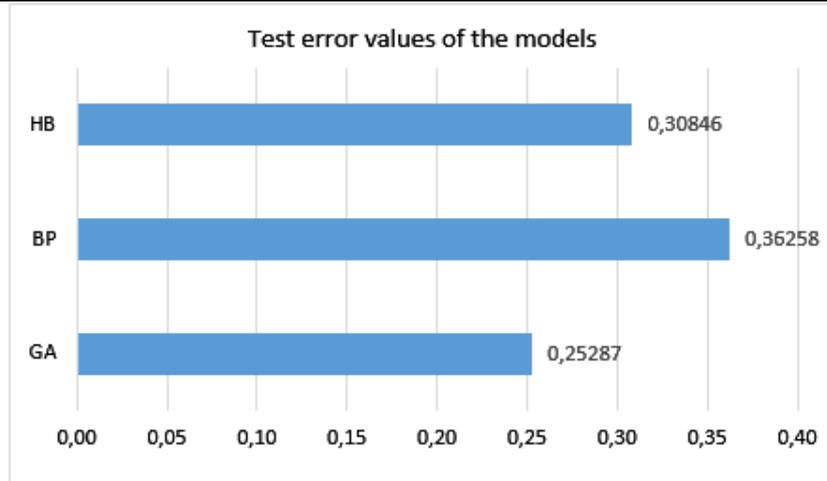


Figure 5. Test *RMSE* error values of the ANFIS models

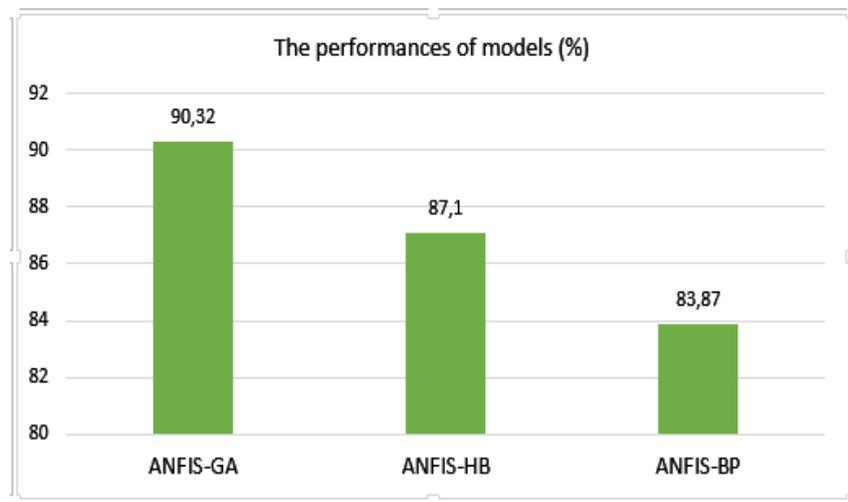


Figure 6. The classification performance of the methods for prostate cancer

The accuracy measure cannot be considered alone for evaluating any method. For this purpose, Sensitivity and Specificity criteria are also calculated to evaluate the performance of the classification methods. These criteria are calculated by using the number of correct positive and correct negative predictions, respectively. Compared to other ANFIS models, the value of Accuracy, Sensitivity and Specificity of ANFIS-GA are the more promising results. The performance of the proposed method is verified with the results given in Table 3. As a result, The ANFIS network can be trained with a low error value using optimization algorithms such as GA. In this context, GA can be used as an alternative method to derivative based learning algorithms for early diagnosis of prostate cancer.

Table 3. The performances of the methods on prostate cancer

	Accuracy (%)	Sensitivity (%)	Specificity (%)
ANFIS-GA	90.32	100.00	85.00
ANFIS-HB	87.10	91.66	84.21
ANFIS-BP	83.87	84.61	83.33



5. CONCLUSION

Diagnosis of cancer predicated on gene expression data is an encouraging research area in the field of data mining. The proposed method tended to the issue of early diagnosis of prostate cancer, one of the world's most genuine health issues. In this paper, a new methodology is presented to diagnose human prostate cancer diseases predicated on the gene expression profiles. Within the proposed methodology, Firstly, CFS is utilized for feature reduction and finally, ANFIS-GA is used for diagnosis of prostate cancer. The presented methodology is evaluated by considering prostate cancer microarray dataset. The results of proposed method are compared with those of two different ANFIS models, namely ANFIS-HB and ANFIS-BP, respectively. It is shown that the proposed approach is very robust and can deliver great results for diagnosis of prostate cancer.

NOTICE

This study was presented as an oral presentation at the I. International Scientific and Vocational Studies Congress (BILMES 2017) in Nevşehir/Ürgüp between 5-8 October 2017.

REFERENCES

- [1] Samli, H., Samli, M., Aztopal, N., Vatansever, B., Sigva, Z.O.D., Dincel, D., and Gunduz, C., (2015). Cytotoxic Effects of Palladium (II) Complex on Prostate Cancer Cells, XIV. National Congress of Medical Biology and Genetics, Muğla, pp:212-213.
- [2] Konac, E., (2015). Does the Future of Prostate Cancer Treatment Lie With Apoptotic Inducers?, XIV. National Congress of Medical Biology and Genetics, Muğla, pp:65-66.
- [3] Haznedar, B., Arslan, M.T., and Kalinli, A., (2017). Training ANFIS Structure Using Genetic Algorithm for Liver Cancer Classification based on Microarray Gene Expression Data. *Sakarya University Journal of Science*, 21(1), 54-62.
- [4] Arslan, M.T. and Kalinli, A., (2016). A Comparative Study of Statistical and Artificial Intelligence Based Classification Algorithms on Central Nervous System Cancer Microarray Gene Expression Data, *International Journal of Intelligent Systems and Applications in Engineering*, 4, 78-81.
- [5] Arslan, M.T. and Haznedar, B., (2016). Classification of Prostate Cancer Gene Expression Profile with ANFIS, 1st International Mediterranean Science and Engineering Congress (IMSEC 2016), Adana, pp:3333-3339.
- [6] Arslan, M.T. and Kalinli, A., (2016). Feature Selection and Classification on Prostate Cancer Microarray Gene Expression Profile, International Conference on Computer Science and Engineering (UBMK), Tekirdağ, pp:331-334.
- [7] Simon, D., (2002). Training Fuzzy Systems with The Extended Kalman Filter, *Fuzzy Sets and Systems*, 132(2), 189-199.
- [8] Seydi Ghomsheh, V., Aliyari Shoorehdeli, M., and Teshnehlab, M., (2007). Training ANFIS structure with Modified PSO Algorithm. 2007 Mediterranean Conference on Control & Automation, Athens, pp:1-6.
- [9] Shoorehdeli, M.A., Teshnehlab, M., and Sedigh, A.K., (2009). Training ANFIS as an Identifier with Intelligent Hybrid Stable Learning Algorithm Based on Particle Swarm Optimization and Extended Kalman Filter. *Fuzzy Sets and Systems*, 160(7), 922-948.
- [10] Turki, M., Bouzaida, S., Sakly, A., and M'Sahli, F., (2012). Adaptive Control of Nonlinear System Using Neuro-Fuzzy Learning by PSO Algorithm, 16th IEEE Mediterranean Electrotechnical Conference, pp:519-523.



-
- [11] Carrano, E.G., Takahashi, R.H.C., Caminhas, W.M., and Neto, O.M., (2008). A genetic Algorithm for Multiobjective Training of ANFIS Fuzzy Networks, *IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence, Hong Kong*, pp:3259-3265.
 - [12] Cus, F., Balic, J., and Zuperl, U., (2009). Hybrid ANFIS-ants System Based Optimisation of Turning Parameters Manufacturing and Processing, *Journal of Achievements in Material and Manufacturing Engineering*, 36(1), 79-86.
 - [13] Singh, D., Febbo, P.G., Ross, K., Jackson, D.G., Manola, J., Ladd, C., Tamayo, P., et al., (2002). Gene Expression Correlates of Clinical Prostate Cancer Behavior, *Cancer Cell*, 1(2), 203-209.
 - [14] Holland, J.H., (1992). Genetic Algorithms, *Scientific American*, 267(1), 66-73.
 - [15] Seker, S., (2007). Vehicle Routing Problems and Genetic Algorithm Approach to the Stochastic Vehicle Routing Problem with Time Windows, *Yıldız Technical University*.
 - [16] Michalewicz, Z., (1996). Genetic Algorithms+Data Structures=Evolution Programs (3 rd edition.), NewYork: Springer-Verlag.
 - [17] Jang, J.S.R., (1992). Self-learning Fuzzy Controllers Based on Temporal Backpropagation, *IEEE Transactions on Neural Networks*, 3(5), 714-723.
 - [18] Gedik, F.A., (2011). Neuro-Fuzzy Approach for Solving Communication Network Problems, *University of Ankara*.