Journal of Chemical and Petroleum Engineering 2020, 54(2): 155-164
DOI: 10.22059/jchpe.2020.264471.1244

RESEARCH PAPER

# Use of ANFIS/Genetic Algorithm and Neural Network to Predict Inorganic Indicators of Water Quality

**Majid Mohadesi, Babak Aghel***

Department of Chemical Engineering, Faculty of Energy, Kermanshah University of Technology, Kermanshah, Iran

Received: 26 August 2018, Revised: 16 September 2020, Accepted: 20 September 2020


## Abstract

The present research used novel hybrid computational intelligence (CI) models to predict inorganic indicators of water quality. Two CI models i.e. artificial neural network (ANN) and a hybrid adaptive neuro-fuzzy inference system (ANFIS) trained by genetic algorithm (GA) were used to predict inorganic indicators of water quality including total dissolved solids (TDS), total hardness (TH), total alkalinity (TAlk), and electrical conductivity (σ). The study was conducted on samples collected from water wells of Kermanshah province through analyzing water parameters including pH, temperature (T), and the sum of mill equivalents of cations (SC) and anions (SA). A multilayer perceptron (MLP) structure was used to forecast inorganic indicators of water quality using the ANN approach. A MATLAB code was used for the proposed ANFIS model to adjust and optimize the ANFIS parameters during the training process using GA. The accuracy of the generated models was described using various evaluation techniques such as mean absolute error (MAE), correlation factor (R), and mean relative error percentage (MRE%). The results showed that both methods were suitable for predicting inorganic indicators of water quality. Moreover, the comparison of the two methods showed that the predicted values obtained from the ANFIS/GA model were better than those obtained from the ANN approach.

**Keywords:**
ANFIS,
ANN,
Genetic Algorithm,
Water Quality

## Introduction

To characterize water quality, it is necessary to assess physical, biological, and chemical variables. In general, to investigate the effects of different processes on the quality of water, many plans examine the water quality to provide the required information for the management of water resources [1,2]. Pollution of water resources with chemicals and excessive nutrients results from contamination with wastewater flow containing degradable organics, domestic effluent, agricultural waste, and nutrients [3,4]. Pollution of water resources has become one of the main threats to public and environmental health, thus it is necessary to continuously monitor water resources [5]. The contaminants hurt water quality parameters such as pH, total dissolved solids (TDS), dissolved oxygen (DO) content, conductivity, temperature, transparency, and a viscosity [6–8]. A variety of chemical and biological assessment methods are available but among them, TDS, conductivity, and hardness are the most important parameters for measuring water quality [9–12]. Modeling has widely been used by researchers to predict water quality index (such as pollutant concentrations) based on current water conditions [13,14]. Meanwhile, various water quality models such as traditional mechanistic approaches and artificial neural networks (ANN) have been successfully applied to accomplish the best practices for predicting

* Corresponding author:
Email: b.aghel@kut.ac.ir (B. Aghel)

water quality [15–17]. However, these models have some limitations, for instance, they require several input parameters that are not easily accessible; thus, the limitations make the process expensive and time-consuming [18]. Nowadays, ANN and adaptive neuro-fuzzy inference system (ANFIS) are used by many researchers to predict and forecast parameters in various areas like water resources quality assessment [19–21]. These methods are highly accepted because they require less data for forecasting and are more preferred than deterministic models; also, unlike other mathematical models, they do not require a complex and explicit description [22,23].

ANFIS and ANN models were used by Areerachakul [24] to forecast the biochemical oxygen demand (BOD) of the Saen Saep canal in Bangkok. The studied variables which were chosen as inputs included chemical oxygen demand (COD), dissolved oxygen, total coliform bacteria (T-coliform), ammonia nitrogen, and nitrate nitrogen. The experimental results revealed that, as compared with the adaptive ANFIS model, ANN model provided a higher correlation coefficient (R=0.7300 vs. R=0.6768) and a lower mean square error (RMSE%=4.53 vs. RMSE%=4.8182). Yan et al. [25] applied ANFIS model for classifying the water quality status of all major river basins in China in terms of several physical and inorganic chemical indicators including dissolved oxygen, chemical oxygen demand, and ammonia-nitrogen. For training and validating the model, they collected 845 samples from 100 monitoring stations. Moreover, ANN was applied to compare the performance of the models. Recently, Masrur Ahmed and Shah [26] applied ANFIS model to estimate BOD of the Surma River in the northeastern region of Bangladesh. Based on their results, ANFIS model was successfully applied to establish a river water quality prediction model and performed better than other conventional conceptual models. The present study aimed to develop ANN and ANFIS techniques for forecasting inorganic indicators of water quality including hardness, total dissolved solids, conductivity, and alkalinity of water wells of Kermanshah province. The models had four input parameters including pH, temperature, and the sum of mill equivalents of cations and anions. The proposed ANFIS model was trained using the genetic algorithm (GA) to obtain an optimized structure with good accuracy and with the minimum number of membership functions. We combined ANFIS with GA to construct a new model with high levels of accuracy and flexibility and low execution time requirements.

## Experimental Method

Kermanshah province is one of the 31 provinces of Iran that is located in the western part of Iran at an approximate latitude of 34.3168 and longitude of 47.0591 with an area of about 24,998 km$^2$. Generally, drinking water in rural areas is obtained from water wells, hence, the water quality parameters are monitored using standard methods introduced by the American Public Health Association (APHA). The data set covered 860 observations at different monitoring stations and included four water quality parameters monitored monthly in 2015. The data were obtained from the Rural Water and Sewage Company of Kermanshah province. The measured parameters were pH, temperature, the sum of mill equivalents of cations and anions, alkalinity, hardness, total dissolved solids, and conductivity; the last four parameters were used to construct a water quality model. Fig. 1 shows the data used in this study. The ranges of the selected input/output parameters are presented in Table 1.

## Developing the Proposed Models

This study aimed to introduce novel models based on ANN and ANFIS/GA structures to predict the inorganic indicators of water quality. The inputs of CI models included the sum of anions

(SA), the sum of cations (SC), temperature ($T$), and pH. The outputs included electrical conductivity ($\sigma$), total dissolved solids (TDS), total hardness (TH), and total alkalinity (TAlk).

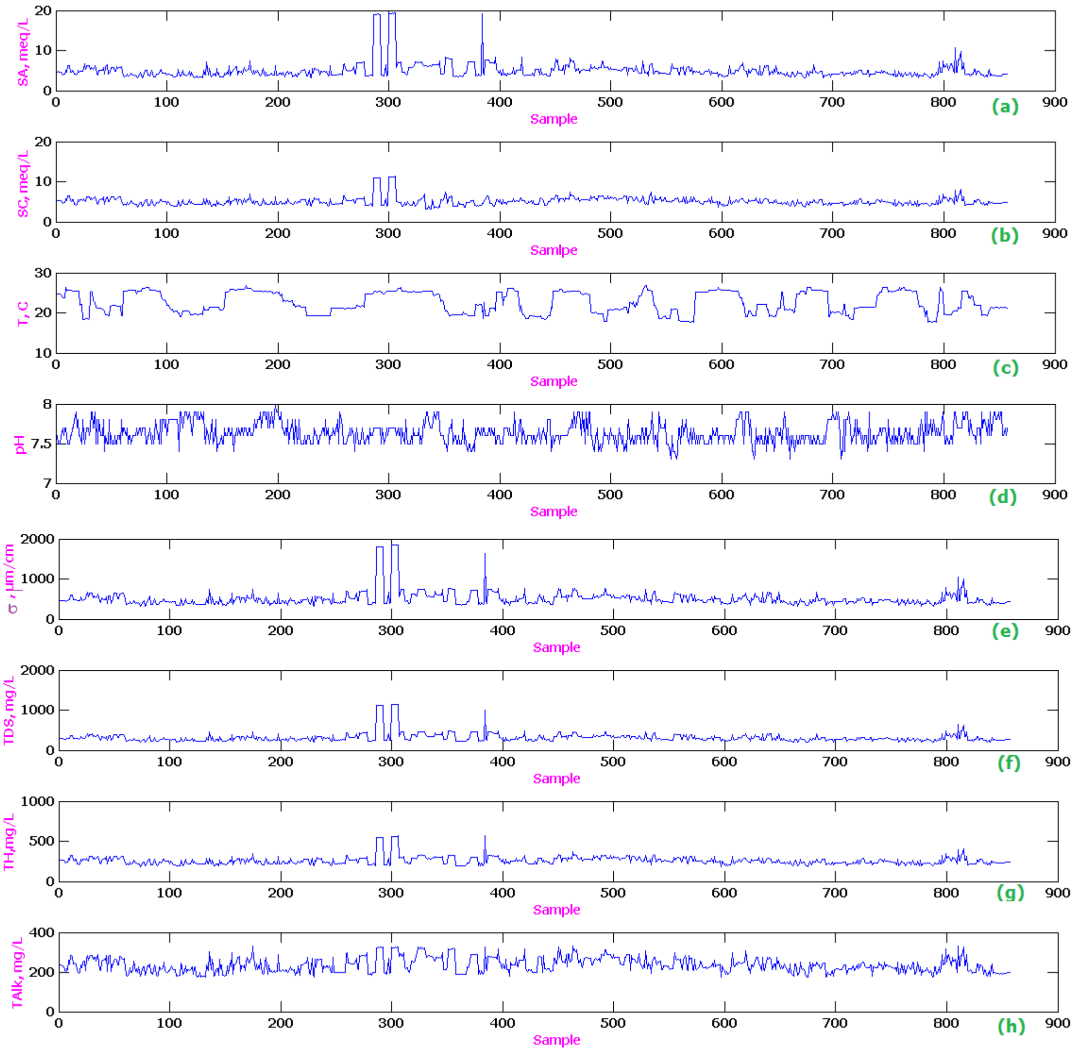


**Fig. 1.** Data used in this study; a) sum of anions, b) sum of cations, c) temperature, d) pH, e) electrical conductivity, f) total dissolved solids, g) total hardness, and h) total alkalinity

**Table 1.** Input and output variables of the models

| Input/Output | Variable | Variable name | Unit | Range |
|---|---|---|---|---|
| Inputs | SA | Sum of anions | meq/L | 3.20-19.41 |
| | SC | Sum of cations | meq/L | 3.16-11.32 |
| | $T$ | Temperature | ℃ | 17.70-26.90 |
| | pH | pH | - | 7.3-8 |
| Output | $\sigma$ | Electrical conductivity | $\mu$s/cm | 321-1849 |
| | TDS | Total dissolved solid | mg/L | 199.02-1146.38 |
| | TH | Total hardness | mg $CaCO_3$/L | 186-570 |
| | Talk | Total alkalinity | mg $CaCO_3$/L | 174-332 |

## Artificial neural network

ANN structure is based on the biological neural network operation [27]. The artificial neuron is the basic processing element of ANN, in which the synapses of biological neurons are modeled as weights. Using the back-propagation algorithm, which is an error-minimization method, the weights can be adjusted. The multi-layer perceptron (MLP) network is one of the widely used ANN structures. As shown in Fig. 2a, MLP has at least three layers; i.e. input layer, an output layer, and one or more hidden layers [28]. Each layer in MLP structure has several neurons. In Fig. 2a $X_1, X_2, \dots, X_n$ are the inputs, $Y_1, Y_2, \dots, Y_m$ are the outputs, n is the number of inputs, and m is the number of outputs. In this figure, the output of $t^{th}$ neuron in the hidden layer is given by the following equation:

$$\theta_t = f\left(\sum_{k=1}^{n} (X_k W_{kt}) + b_t\right) \qquad t = 1, 2, \dots, i \tag{1}$$

where $f$ is the hidden layer activation function (usually tansig function), $b$ is the bias term, and $W$ is the weighting factor. Also, the output of $j^{th}$ neuron in the output layer is given by the following equation:

$$y_i = \sum_{k=1}^{i} (\theta_k W_{kj}) + b_j \qquad j = 1, 2, \dots, m \tag{2}$$

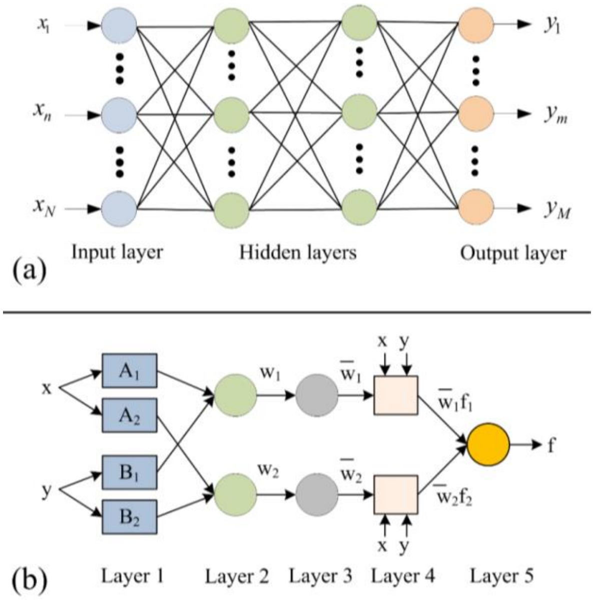


**Fig. 2.** Structure of a) MLP and b) ANFIS

### Adaptive Neuro-Fuzzy Inference System

ANFIS has the advantages of both the fuzzy system and ANN network. It is a fuzzy inference system (FIS) implemented using ANN [29,30]. Fig. 2b shows the ANFIS structure. Each ANFIS structure has five layers described as follow:

Layer 1: Every node in this layer is an adaptive node with a node function given by:

$$O_{1,i} = \mu_{A_i}(x) \ , \qquad i = 1, 2 \tag{3}$$

$$O_{1,i} = \mu_{B_{i-2}}(y) \ , \qquad i = 3, 4 \tag{4}$$

where $i$ is the membership grade of a fuzzy set ($A_1$, $A_2$, $B_1$, $B_2$), and $O_{1,i}$ is the output of the node $i$ in layer 1. Gaussian function is a typical membership function (MF) given by Eq. 5.

$$\mu_A(x) = \exp\left(-\frac{(x-c)^2}{2\sigma^2}\right) \tag{5}$$

In Eq. 5 $c$ and $\sigma$ are called nonlinear parameters (premise parameters).

Layer 2: The fixed nodes in this layer multiply all incoming signals and represent the firing strength of a rule. The outputs of the nodes in this layer are given by Eq. 6.

$$O_{2,i} = w_i = \mu_{A_i}(x)\mu_{B_i}(y) \ , \quad i = 1, 2 \tag{6}$$

Layer 3: The fixed nodes in this layer calculate the ratio of the $i^{th}$ rule's firing strength to the sum of all rule's firing strengths given by:

$$O_{3,i} = \overline{w_i} = \frac{w_i}{w_1 + w_2} \ , \quad i = 1, 2 \tag{7}$$

Layer 4: The outputs of adaptive nodes in this layer is given by:

$$O_{4,i} = \overline{w_i} f_i = \overline{w_i}(p_i x + q_i y + r_i) \ , \quad i = 1, 2 \tag{8}$$

Layer 5: The fixed node in this layer has an output function given by:

$$O_{5,i} = \sum_i \overline{w_i}\, f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \ , \quad i = 1, 2 \tag{9}$$

## Modeling Approach

In general, several parameters influence the inorganic indicators of water quality (electrical conductivity, total dissolved solids, total hardness, and total alkalinity). The effects of the sum of anions, sum of cations, temperature, and pH on these parameters are investigated. Accordingly, accurate models based on ANN and ANFIS structures were presented to model and predict the effect of input parameters on the inorganic indicators of water quality. In these CI models, the input parameters were defined as the sum of anions, the sum of cations, temperature, and pH. Moreover, electrical conductivity, total dissolved solids, total hardness, and total alkalinity were considered as the output parameters of the models.

The data set required to train and test the proposed CI models were obtained from the water wells of Kermanshah province. The total number of samples used to develop the CI models was 860. 70% and 30% of all the samples were used for training and testing, respectively. MATLAB software was used to develop the proposed models. To obtain the best models, different ANN and ANFIS configurations were trained and tested. To obtain the best ANN structure, many different structures tested one to three hidden layers. Also, the number of neurons in each hidden layer changed from 1 to 9, and for each MLP structure, the number of epochs changed from 100 to 550. To obtain the best ANFIS models, input membership function type (Triangular-shaped, Trapezoidal-shaped, Gaussian, etc.), the number of input membership functions (2 to 12), and the number of training epochs (50 to 500) were changed. Then, the GA parameters such as maximum iterations, population size, crossover probability, mutation probability, mutation rate, etc. were determined. Afterward, the GA fitness function was formulated. The optimization problem has the values of the premise and consequent parameters, and the number of membership functions as the decision variables. Finally, ANN and ANFIS structures were trained and tested through testing and training the data. The specifications of the best proposed ANN and ANFIS/GA models are shown in Tables 2 and 3, respectively.

**Table 2.** Characteristics of the best proposed MLP model

| | |
|---|---|
| No. of hidden layers | 2 |
| No. of neurons in the input layer | 4 |
| No. of neurons in the first hidden layer | 5 |
| No. of neurons in the second hidden layer | 8 |
| No. of neurons in the output layer | 4 |
| Learning rate | 0.5 |
| Number of epochs | 200 |
| Learning function | Trainlm |
| Activation function | Tansig |

For the proposed ANFIS models trained by GA, the following parameters were used in GA: crossover percentage=0.7, mutation percentage=0.5, mutation rate=0.1, and selection pressure=8. Trainlm is an ANN training function that updates values of weights and biases based on the Levenberg-Marquardt (LM) algorithm. Also, the learning rate is an important parameter in the process of training MLP networks; it can be changed to confirm that the weights converge fast enough to obtain a response without producing oscillations.

**Table 3.** Specification of the best ANFIS/GA models

| Specification | Proposed ANFIS models | | | |
|---|---|---|---|---|
| | **σ** | **TDS** | **TH** | **TAlk** |
| Type | Sugeno | Sugeno | Sugeno | Sugeno |
| Inputs/outputs | 4/1 | 4/1 | 4/1 | 4/1 |
| No. of MFs for each input | 5 | 3 | 4 | 5 |
| No. of output MFs | 5 | 3 | 4 | 5 |
| Input MF type | Gaussian | Gaussian | Gaussian | Gaussian |
| Output MF type | linear | Linear | linear | linear |
| No. of fuzzy rules | 5 | 3 | 4 | 5 |
| No. of nonlinear parameters | 80 | 48 | 64 | 80 |
| No. of linear parameters | 25 | 15 | 20 | 25 |
| No. of iterations | 1000 | 650 | 600 | 500 |
| No. of populations | 50 | 80 | 80 | 150 |

## Result and Discussion

The comparison between the proposed ANN and ANFIS/GA models and the experimental data for training and testing are shown in Figs. 3 and 4, respectively. Figs. 3a and 4a show the predicted results of ANN and ANFIS/GA models as compared with the experimental data on electrical conductivity. Moreover, the same plots for total dissolved solids, total hardness, and total alkalinity are shown in Figs. 3b and 4b, Figs. 3c and 4c, and Figs. 3d and 4d, respectively. As shown, ANFIS/GA model was more accurate than the ANN model to follow the outputs in both training and testing data.

To show a better comparison between the proposed ANN and ANFIS/GA models, we used four standard error functions i.e. MRE% (mean relative error percentage), RMSE (root mean square error), CF (correlation factor), and MAE (mean absolute error). The following equations define these standard errors:

$$\mathrm{MRE}\ \% = 100 \times \frac{1}{N}\sum_{i=1}^{N}\left|\frac{x_{i_{exp}} - x_{i_{pred}}}{x_{i_{exp}}}\right| \tag{10}$$

$$\mathrm{RMSE} = \left[\frac{\sum_{i=1}^{N}\left(x_{i_{exp}} - x_{i_{pred}}\right)^2}{N}\right]^{0.5} \tag{11}$$

$$\mathrm{CF} = 1 - \left[\frac{\sum_{i=1}^{N}\left(x_{i_{exp}} - x_{i_{pred}}\right)^2}{\sum_{i=1}^{N}\left(x_{i_{exp}}\right)^2}\right] \tag{12}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|x_{i_{exp}} - x_{i_{pred}}\right| \tag{13}$$

where '$x_{i_{exp}}$' and '$x_{i_{pred}}$' are the experimental data and predicted values (ANN or ANFIS), respectively, and $N$ is the number of data. Table 4 shows the overall errors obtained for the proposed models as compared with the experimental data. These errors are obtained for both training and testing data. As shown in Table 4, the proposed ANFIS/GA models are capable to predict the electrical conductivity, total dissolved solids, total hardness, and total alkalinity, outputs better than the proposed ANN model. Fig. 5 shows a better comparison between the experimental data and the ANFIS/GA models used for training and testing the data. As shown in these figures, it is clear that the obtained results using the ANFIS/GA models are close to the experimental data.

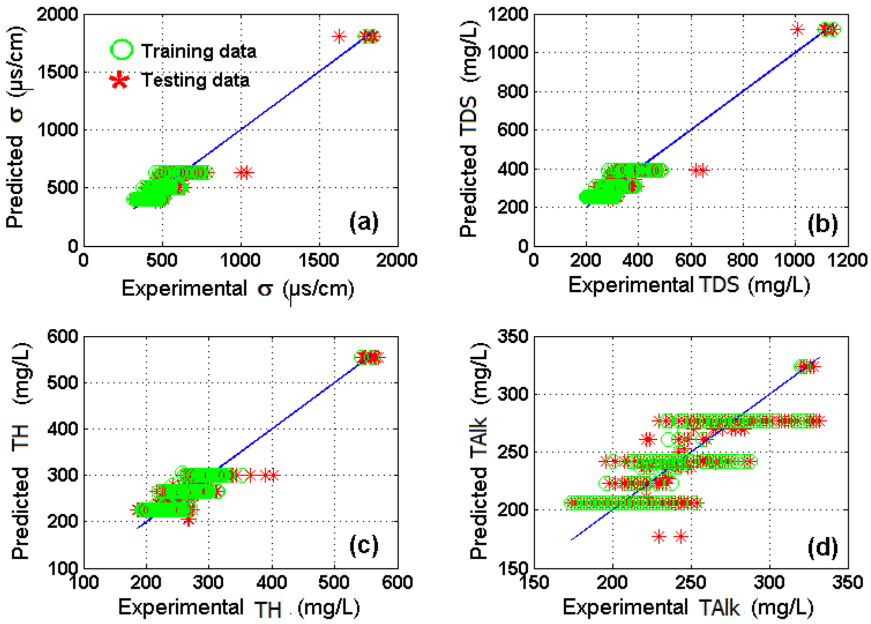


**Fig. 3.** Comparison of ANN model with the experimental data

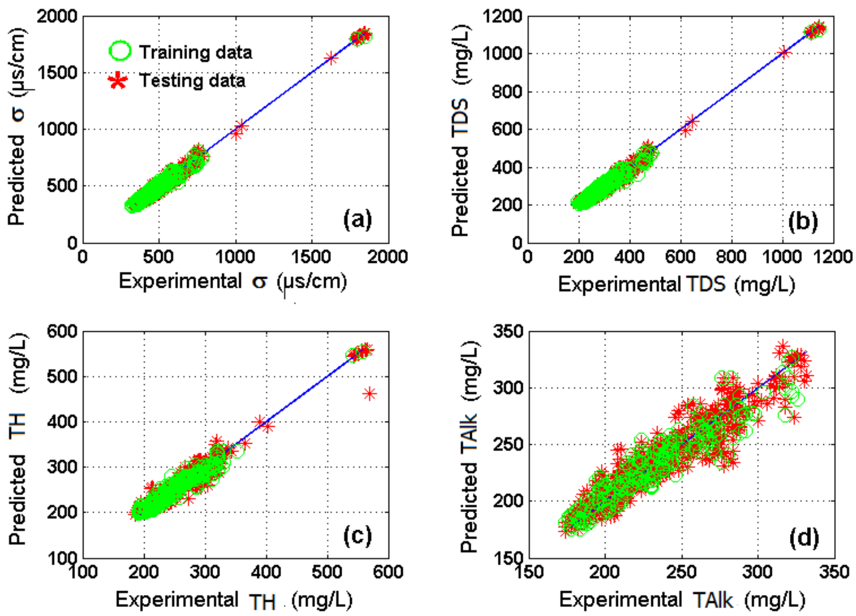


**Fig. 4.** Comparison of ANFIS/GA models with the experimental data

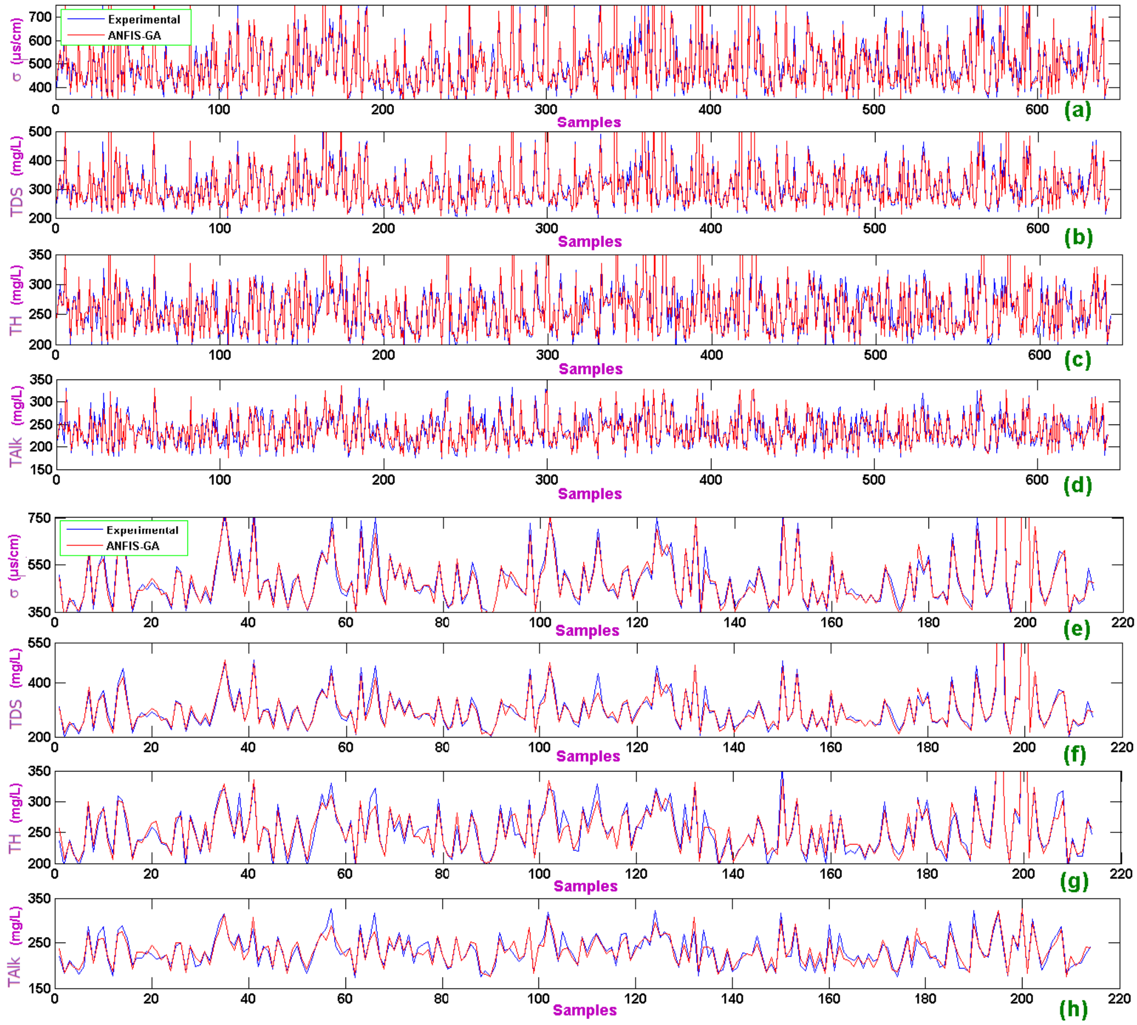


**Fig. 5.** Results obtained from the ANFIS models a, b, c, and d) training data; e, f, g, and h) testing data

**Table 4.** Errors of the best proposed ANN and ANFIS/GA

| Network | Output | Data | Error | | | |
|---|---|---|---|---|---|---|
| | | | MRE% | MAE | RMSE | CF |
| ANN | σ | Training | 8.804 | 44.749 | 59.043 | 0.961471 |
| | | Testing | 9.511 | 46.592 | 59.902 | 0.932712 |
| | TDS | Training | 8.806 | 27.751 | 36.599 | 0.961489 |
| | | Testing | 9.499 | 28.852 | 37.107 | 0.932849 |
| | TH | Training | 5.951 | 15.143 | 18.980 | 0.940413 |
| | | Testing | 6.601 | 16.5490 | 20.321 | 0.898696 |
| | TAlk | Training | 6.522 | 15.229 | 19.308 | 0.851623 |
| | | Testing | 6.934 | 15.9651 | 19.706 | 0.836748 |
| ANFIS/GA | σ | Training | 3.375 | 17.042 | 22.813 | 0.994418 |
| | | Testing | 3.428 | 17.386 | 23.158 | 0.990580 |
| | TDS | Training | 3.426 | 10.754 | 14.332 | 0.994300 |
| | | Testing | 3.528 | 11.062 | 14.978 | 0.989887 |
| | TH | Training | 3.233 | 8.441 | 11.969 | 0.976760 |
| | | Testing | 3.241 | 8.305 | 10.998 | 0.971598 |
| | TAlk | Training | 3.985 | 9.499 | 12.603 | 0.939709 |
| | | Testing | 4.194 | 9.921 | 12.878 | 0.934787 |

## Conclusion

This paper aimed to use an artificial neural network and a novel hybrid computational intelligence model for predicting inorganic indicators of water quality. The proposed ANFIS/GA model combines the genetic algorithm and ANFIS methods. There was a satisfactory level of consistency between the constructed models and the experimental values. The comparison between the results showed that the proposed ANFIS/GA approach was an accurate promising tool with a better result as compared with ANN method. Therefore, ANFIS/GA is a method that can be used to solve more complex scientific and technological problems. Also, the introduced models can provide an insight into some instructions to predict new inorganic indicators of water quality.

## Acknowledgments

The authors would like to acknowledge the financial support of Rural Water and Sewage Company of Kermanshah Province for this research under Grant No. 94/8012.

## References

[1] Khalil B, Ouarda TB, St-Hilaire A, Chebana F. A statistical approach for the rationalization of water quality indicators in surface water quality monitoring networks. Journal of Hydrology. 2010 May 28;386(1-4):173-85.
[2] Behmel S, Damour M, Ludwig R, Rodriguez MJ. Water quality monitoring strategies—A review and future perspectives. Science of the Total Environment. 2016 Nov 15;571:1312-29.Hosseini-Dastgerdi Z, Meshkat SS. An experimental and modeling study of asphaltene adsorption by carbon nanotubes from model oil solution. Journal of Petroleum Science and Engineering. 2019 Mar 1;174:1053-61.
[3] Dimitrovska O, Markoski B, Toshevska BA, Milevski I, Gorin S. Surface water pollution of major rivers in the Republic of Macedonia. Procedia Environmental Sciences. 2012 Jan 1;14:32-40.
[4] Ouyang Y. Evaluation of river water quality monitoring stations by principal component analysis. Water research. 2005 Jul 1;39(12):2621-35.
[5] Park SY, Choi JH, Wang S, Park SS. Design of a water quality monitoring network in a large river system using the genetic algorithm. Ecological modelling. 2006 Dec 1;199(3):289-97.
[6] House MA. Public perception and water quality management. Water Science and Technology. 1996 Jan 1;34(12):25-32.
[7] Wu Q, Xia X, Mou X, Zhu B, Zhao P, Dong H. Effects of seasonal climatic variability on several toxic contaminants in urban lakes: Implications for the impacts of climate change. Journal of Environmental Sciences. 2014 Dec 1;26(12):2369-78.
[8] Delpla I, Benmarhnia T, Lebel A, Levallois P, Rodriguez MJ. Investigating social inequalities in exposure to drinking water contaminants in rural areas. Environmental Pollution. 2015 Dec 1;207:88-96.
[9] Thompson MY, Brandes D, Kney AD. Using electronic conductivity and hardness data for rapid assessment of stream water quality. InWorld Environmental and Water Resources Congress 2010: Challenges of Change 2010 (pp. 3356-3365)..
[10] Kney AD, Brandes D. A graphical screening method for assessing stream water quality using specific conductivity and alkalinity data. Journal of environmental management. 2007 Mar 1;82(4):519-28.
[11] Gergel SE, Turner MG, Miller JR, Melack JM, Stanley EH. Landscape indicators of human impacts to riverine systems. Aquatic sciences. 2002 Jun 1;64(2):118-28.
[12] Aghel B, Rezaei A, Mohadesi M. Modeling and prediction of water quality parameters using a hybrid particle swarm optimization–neural fuzzy approach. International Journal of Environmental Science and Technology. 2019 Aug 1;16(8):4823-32.

[13] Najah A, El-Shafie A, Karim OA, Jaafar O, El-Shafie AH. An application of different artificial intelligences techniques for water quality prediction. International Journal of Physical Sciences. 2011 Oct 2;6(22):5298-308.
[14] Ding YR, Cai YJ, Sun PD, Chen B. The use of combined neural networks and genetic algorithms for prediction of river water quality. Journal of applied research and technology. 2014;12(3):493-9.
[15] Reichert P, Vanrolleghem P. Identifiability and uncertainty analysis of the river water quality model no. 1 (RWQM1). Water Science and technology. 2001 Apr;43(7):329-38.
[16] Chapra SC, Pelletier GJ, Tao H. QUAL2K: a modeling framework for simulating river and stream water quality, version 2.11: documentation and users manual. Civil and Environmental Engineering Dept., Tufts University, Medford, MA. 2008:1-09.
[17] Wool TA, Ambrose RB, Martin JL, Comer EA, Tech T. Water quality analysis simulation program (WASP). User's Manual, Version. 2006;6.
[18] Suen JP, Eheart JW. Evaluation of neural networks for modeling nitrate concentrations in rivers. Journal of water resources planning and management. 2003 Nov;129(6):505-10.
[19] Dahiya S, Singh B, Gaur S, Garg VK, Kushwaha HS. Analysis of groundwater quality using fuzzy synthetic evaluation. Journal of Hazardous Materials. 2007 Aug 25;147(3):938-46.
[20] Chen YH, Chang FJ. Evolutionary artificial neural networks for hydrological systems forecasting. Journal of Hydrology. 2009 Mar 30;367(1-2):125-37.
[21] Icaga Y. Fuzzy evaluation of water quality classification. Ecological Indicators. 2007 Jul 1;7(3):710-8.
[22] Kisi O, Ozkan C, Akay B. Modeling discharge–sediment relationship using neural networks with artificial bee colony algorithm. Journal of Hydrology. 2012 Mar 27;428:94-103.
[23] Nayak PC, Sudheer KP, Ramasastri KS. Fuzzy computing based rainfall–runoff model for real time flood forecasting. Hydrological Processes: An International Journal. 2005 Mar 15;19(4):955-68.
[24] Areerachakul S. Comparison of ANFIS and ANN for estimation of biochemical oxygen demand parameter in surface water. International Journal of Chemical and Biological Engineering. 2012 Apr 26;6:286-90.
[25] Yan H, Zou Z, Wang H. Adaptive neuro fuzzy inference system for classification of water quality status. Journal of Environmental Sciences. 2010 Dec 1;22(12):1891-6.
[26] Ahmed AM, Shah SM. Application of adaptive neuro-fuzzy inference system (ANFIS) to estimate the biochemical oxygen demand (BOD) of Surma River. Journal of King Saud University-Engineering Sciences. 2017 Jul 1;29(3):237-43.
[27] Da Silva IN, Spatti DH, Flauzino RA, Liboni LH, dos Reis Alves SF. Artificial neural network architectures and training processes. InArtificial neural networks 2017 (pp. 21-28). Springer, Cham.
[28] Haykin SS. Neural networks and learning machines/Simon Haykin..
[29] Ram M, editor. Advanced Fuzzy Logic Approaches in Engineering Science. IGI Global; 2018 Sep 14.
[30] Sivanandam SN, Sumathi S, Deepa SN. Introduction to fuzzy logic using MATLAB. Berlin: Springer; 2007 Jan.