Adaptive Neuro-Fuzzy Inference System for intelligent water quality classification in Tilesdit dam from Algeria

Mohamed LADJAL*, Mohamed BOUAMAR, Mohamed DJERIOUI

LASS, Laboratory of Analysis of Signals and Systems,
Department of Electronics, Faculty of Technology, M'sila University, M'sila, Algeria
PB N° 166 Ichebilia, 28000 M'sila, Algeria
medladjal@gmail.com, m.bouamar@gmail.com, mohamed_djerioui@yahoo.fr

Abstract

This paper presents the application of an Adaptive Neuro-Fuzzy Inference System (ANFIS) in classification of water quality status. ANFIS is one of the useful and powerful neural network approaches for the solution of pattern recognition problems in the last decades. This study involved the evaluation and interpretation of surface water quality data in Tilesdit dam from Algeria. It also allowed us to obtain more advanced information about water quality, and to design a monitoring network for this study area. The ANFIS which is a technique for pattern classification has been widely used in many application areas such as water quality monitoring. This method is a binary classification technique, but in some cases, such as pattern recognition, we need more than two classes. A multi-class problem using ANFIS is a typical example for solving the mentioned problem. In this work, four physicochemical parameters in 4 seasons during the period 2009-2011, located at Tilesdit dam, were selected for this study, such as pH, Temperature, Conductivity and Turbidity to supervise water quality. Up to 95 % of the data could be correctly classified using ANFIS model. Its performance is more competitive when compared with artificial neural networks. Furthermore, the results demonstrated that the proposed procedure has a great potential in water quality monitoring.

Keywords— Water quality, Multi-class problem, ANFIS, ANN, Tilesdit dam.

1. Introduction

Water Quality is a major concern around the world. It is affected by a wide range of natural and human influences. The most important of the natural influences are geological, hydrological and climatic, since these affect the quantity and quality of water available. Their influence is generally greatest when available water quantities are low and maximum use must be made of the limited resources; the objective of water quality monitoring is to obtain quantitative information on the physical, chemical, and biological characteristics of water via statistical sampling [1]. The quality of surface water is one of the irreplaceable strategic resources for human health, ecological systems and social development [2]. Reliable information about the spatial distribution of open surface water is critically important in various scientific disciplines [3]. Water quality monitoring plays a vital role in environmental management and decision-making and it provides a scientific basic and considerable interest for rational utilization and protection of water resources in most countries of the contemporary world. However, traditional methods of monitoring are used with an inability due to complex relationships between monitoring variables and water quality status. We can say that there is no general method accepted so far. Several applications of evaluation and classification of water quality based on such methods are developed [4] [5]. Two of them are ANFIS and ANN which are effective methods for classification and machine learning systems compared to other methods [6-12]. In this study, it has been taken into consideration the successfulness of ANFIS and ANN in classification applications.

In this paper, ANFIS is proposed to perform the water quality classification data from the Tilesdit dam (Algeria). This technique is a binary classifier but in some cases, like pattern recognition, we usually have more than two classes. The problem is regarded as a multi-class classification based on three classes of water quality (Class I: upper, Class II: middle, Class III: lower). The multi-class problem using ANFIS or ANN is chosen as classifiers in this case. Its roles is to separate the data in three distinct classes.

The purpose of the present study is to develop a model based on ANFIS, evaluate its applicability to assess and classify water quality status, and compare its performance with ANN. The results are compared to get the best performance of classification process with respect to recognition rate.

2. Study area and water quality data description

The study area (Tilesdit dam) is built 122 km east of Algiers (Algeria) (Figure 1). It is geographically located in the town of Bechloul 20 km southeast of the wilaya of Bouira (Figure 2). This dam is located between the following coordinates: 35° 13' 22''N 4° 14' 23''E. It's characterized by a semi arid climate were is the climate is hot and dry in summer, cold and rainy in winter and an average precipitation of about 440-660 mm/year.

We seek to apply our approach for surface water quality monitoring using four physicochemical parameters supplied by the measurement sensors of the station. Our knowledge of the treatment process is limited to the recorded data from the station during the three years 2009-2011.

In production station, seventeen water parameters were collected in three years (2009-2011). More quality descriptors parameters of the surface water were daily measured, in addition to laboratory tests which are carried out every week at all treatment levels. Immediately after sampling, pH, Temperature (T°), Conductivity (C) and Turbidity (TU) were measured in the field using the sensors directly installed in the station. These parameters are measured continuously at 3 times/day and at any level of the treatment process (raw water, water decanted, filtered water and treated water). Descriptive statistics of the physicochemical compositions of the surface water samples are given in Table 1.

Fig. 1. Study area [Google Earth].

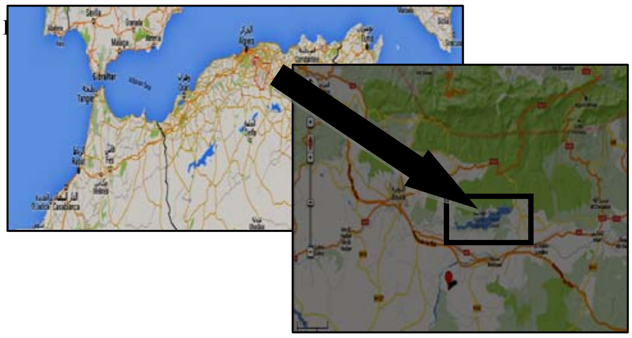

Fig. 2. Map showing the region under study - Tilesdit dam [Google Maps].

Table 1. Descriptive statistics of the analyzed physicochemical parameters.

Parameters	Minimum	Maximum	Mean	Standard deviation
pH	7,15	8,30	7,567	0,25
C	414,00	624,00	585,393	36,278
T°	9,70	24,20	16,13	3,483
TU	1,320	23,81	3,835	2,392

3. Proposed Methods

Water quality monitoring consists of data acquisition, signal processing and decision. The classification process of water quality is carried out by using ANFIS compared to ANN.

3.1. ANFIS technique

The Adaptive Neuro-Fuzzy Inference System (ANFIS) defined by J.-S. Roger Jang in 1992 is a class of adaptive networks that are functionally equivalent to fuzzy inference systems [13]. The ANFIS used in the study, is a fuzzy inference model of Sugeno type, and is a composition of ANNs and fuzzy logic approaches [13][14]. The model identifies a set of parameters through a hybrid learning rule combining the back-propagation gradient descent and a least squares method. It can be used as a basis for constructing a set of fuzzy IF-THEN rules with appropriate membership functions in order to generate the previously stipulated input-output pairs [15]. The Sugeno fuzzy inference system is computationally efficient and works well with linear techniques, optimization and adaptive techniques. As a simple example, we assume a fuzzy inference system with two inputs x and y and one output f. The first-order Sugeno fuzzy model, a typical rule set with two fuzzy If-Then rules can be expressed as [10]:

Rule 1 : If x is A_1 and y is B_1, Then $f_1 = p_1x + q_1y + r_1$

Rule 2 : If x is A_2 and y is B_2, Then $f_2 = p_2x + q_2y + r_2$

The resulting Sugeno fuzzy reasoning system is shown in Fig. 3. It illustrates the fuzzy reasoning mechanism for this Sugeno model to derive an output function (f) from a given input vector [x, y]. The corresponding equivalent ANFIS architecture showed in Fig 4, is a five-layer feed forward network that uses neural network learning algorithms coupled with fuzzy reasoning to map an input space to an output space.

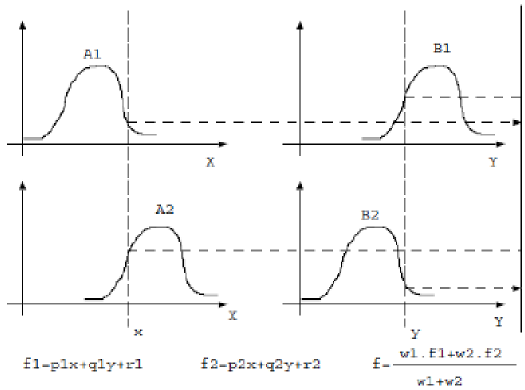

Fig. 3. Sugeno fuzzy model.

Figure 4 demonstrates the ANFIS architecture with its layers. $O_{l,i}$ is the output of the i^{th} node of layer 1. Every node i in this layer is an adaptive node with a node function [16]:

$$\begin{aligned} O_{1,i} &= \mu_{A_i}(x) \quad \text{for} \quad i = 1,2, \text{ or} \\ O_{1,i} &= \mu_{B_{i-2}}(x) \quad \text{for} \quad i = 3,4 \end{aligned} \tag{1}$$

x (or y) is the input node i and A_i (or B_{i-2}) is a label associated with this node. $O_{1,i}$ is the membership grade of a fuzzy set (A_1, A_2, B_1, B_2). Typical membership function is defined by [17] :

$$\mu_{A_i}(x) = \frac{1}{1 + \left| \frac{x - c_i}{a_i} \right|^{2b_i}} ; \forall i \tag{2}$$

Where a_i, b_i, c_i are the parameter set.

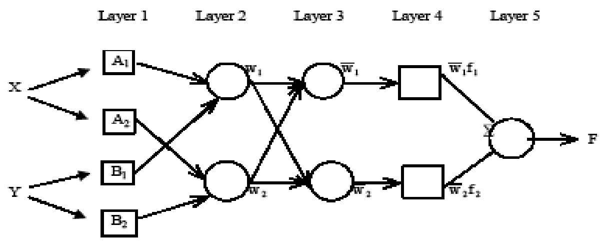

Fig. 4. ANFIS architecture with two inputs and an output

The output is the product of all the incoming signals [17]:

$$O_{2,i} = w_i = \mu_{A_i}(x\ \mu_{B_i}(y), \quad i = 1,2 \qquad (3)$$

The i^{th} node calculates the ratio of the i^{th} rulet's firing strength to the sum of all rulet's firing strengths [18]:

$$O_{3,i} = \overline{w_i} = \frac{w_i}{w_1 + w_2}, \quad \mathrm{i} = 1,2 \qquad (4)$$

Outputs are called normalized firing strengths. Every node *i* in this layer is an adaptive node with a node function [18]:

$$O_{4,i} = \overline{w_i}\, f_i = \overline{w_i}(p_x + q_i y + r_i) \qquad (5)$$

Where $\overline{w_i}$ is the normalized firing strength from layer 3 and $\{p_i, q_i, r_i\}$ is the parameter set of this node. The single node in this layer is a fixed node labeled sum, which computes the overall output as the summation of all incoming signals [19]:

$$Overalloutput = O_{5,i} = \sum_i \overline{w_i}\, f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \qquad (6)$$

The ANFIS can be trained by hybrid learning algorithm presented by Jang [13][14]. In the forward pass the algorithm uses least square method to identify the consequent parameters on the layer 4. The errors are propagated to the backward and the premise parameters are updated by gradient descent.

3.2. ANN technique

Several models based on the artificial neural networks were developed and applied in the field of water. Some recent studies showed the potential effectiveness of this approach [20]. Nonlinear classification of data is operated by using the Multi-layer perceptron (MLP). Figure 5 shows the example of a network used in our application with four inputs data.

The supervised training by the network consists in determining the weights which minimize the whole training data, the differences between the desired output values y_r and the computed output values y_c. The minimum of the following quadratic criterion is to be found:

$$C_w = \frac{1}{N}\sum_{i=1}^{N}(y_{r_i} - y_{c_i})^2 \qquad (7)$$

N represents the examples number of the training dataset.

The technique traditionally employed to carry out the supervised training is the back propagation of the error algorithm. Initially, the nonlinear optimization method of the gradient is used. This is method known to have an oscillatory behavior near to the solution. Actually, the methods known as 2nd order (based on a *Hessian* approximation) are rather preferred; they provide however much better results. Among the most known, we used the method of *Levenberg-Marquardt* in this application increasingly due to the better performance and learning speed with a simple structure [19].

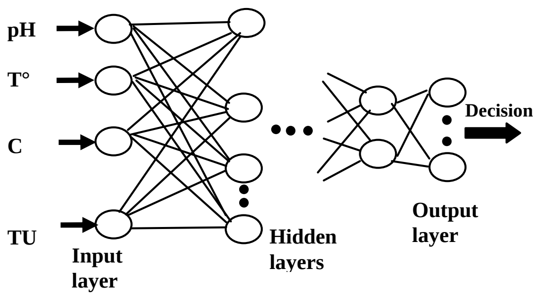

Fig. 5. Multilayer perceptron, MLP

4. Results and discussion

According to the Environmental Quality Standards of water, three classes of water quality have been considered: (Class I: upper, Class II: middle, Class III: lower). In order to proceed with the tests, training and test sets constituted of real data relating to the various qualitative water statuses are used. We applied a real dataset of 400 samples, all constituted of the 4 physicochemical parameters: pH, Temperature (T°), Conductivity (C) and Turbidity (TU) selected as input variables. The decision on the water status depending on these descriptors that are more easily measured continuously in the field using the sensors directly installed in each steps of the production station of study area. Figure 6 shows an evolution of these different descriptors used for the water quality classification.

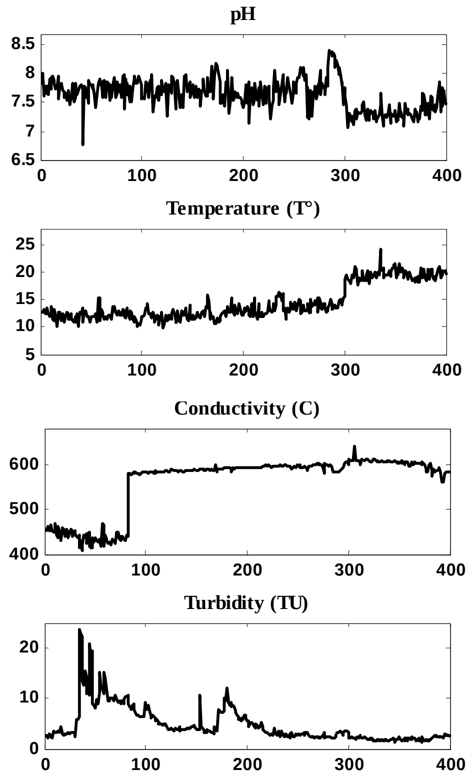

Fig. 6. Evolution of the water quality descriptors.

4.1. ANFIS development

The goal of ANFIS is to generalize the relationship between water quality status and input variables of the form:

$$Y = f(T°, pH, C, TU) \quad (8)$$

There are no fixed rules for developing an ANFIS model [15]. In this study, the water quality data (total of 400 samples) were divided into two data sets using ANFIS: training set and testing set. The performance of ANFIS models were evaluated according to statistical criteria such as correlation coefficient (R^2), and root mean square error (RMSE) [15][17][19].

$$R^2 = \frac{\sum_{i=1}^{N}(Y_r - \overline{Y}_r)(Y_c - \overline{Y}_c)}{\sqrt{\sum_{i=1}^{N}(Y_r - \overline{Y}_r)^2 (Y_c - \overline{Y}_c)^2}} \quad (9)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_r - Y_c)^2} \quad (9)$$

where, Y_c is the computed value, Y_r is the desired value; $\overline{Y_c}$ is the average of computed values, $\overline{Y_r}$ is the average of desired values. The correlation coefficient is a commonly used statistic and provides information on the strength of linear relationship between the desired and the computed values. The RMSE statistic indicates a model's ability to predict a value away from the mean [15].

4.2. Classification results

The ANFIS used in this study (Fig. 7) contained eight rules, with two membership functions being assigned to each input variable. Different membership functions types including *Generalized bell*, *Gaussian*, *Trapezoidal* and *Triangular* were tested. The number of linear and non-linear parameters to be optimized is displayed in Table 2. Optimum parameters were found once checking data error reached the minimum. The performance of ANFIS models with different membership functions are also given in Table 2.

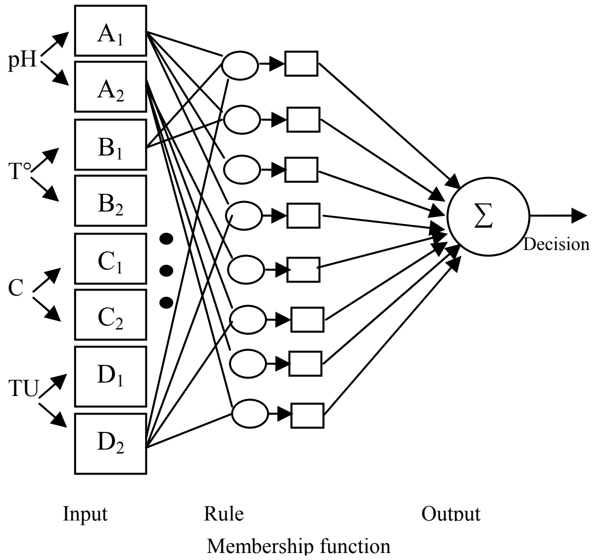

Fig. 7. The ANFIS architecture used in this study.
(The connections from inputs to layer 3 are not shown)

As seen in Table 2, for the training dataset, R^2 and RMSE with ANFIS model were in the ranges from 0.8760 to 0.9501 and 0.2821 to 2.4500, respectively. For the testing data set, the same indices were in the ranges from 0.8830 to 0.9873 and 0.4630 to 3.4300, respectively. Obviously, the performance of model 3 is better than those of other models. As a result, the best fitting model was obtained with the FIS composed by two Gaussian membership function and the correctly classified points of this model is 193.

Table 2. Performance of ANFIS models with different membership functions

Phases	Parameters	Membreship functions				
		Generalized Bell	Gaussian	Two Gaussian	Trapezoidal	Triangular
	Linear	81	81	81	81	81
	Non Linear	36	24	48	48	36
Training data set	R^2	0.9501	0.9051	0.8972	0.8760	0.9350
	RMSE	0.3403	0.3622	0.2821	2.3340	2.4500
Testing data set	R^2	0.9433	0.9143	0.9873	0.8830	0.8840
	RMSE	0.4630	0.4930	3.3010	2.9340	3.4300
Correctly classified points		193	193	193	193	193

In order to assess the ability of the ANFIS model relative to that of an artificial neural network (ANN) model, an ANN model with *sigmoid* selected as activation function, has been tested to determine the adequate number of neurons and hidden layers. It was constructed using the same input parameters to the ANFIS model. The variation of number of neurons and number of hidden layers for ANN model is performed to show the excellent characteristic and its performance in classification process. A *Levenberg Marquardt* algorithm has been employed for training, and the hidden neurons were optimized by trial and error. This algorithm is retained to have carried out much improved results compared to the other well known training algorithms. The final ANN architecture consists of seven hidden neurons. The ANN model has been trained using the same training dataset as used for the ANFIS. The performances of ANFIS and ANN in terms of the performance indices are presented in Table 3.

In this table, the performance of the ANFIS model is showed better than the ANN model either for training and testing data set in terms of R^2, RMSE and recognition rate (*Rec_rate*) of correctly classified points. The ANFIS shows an improvement over the ANN in terms of correctly classified points and classification rate in testing phase. When ANN was asked for predicting the water quality status for the testing data set composed by 200 samples, only 85% were correctly classified. However, the ANFIS correctly classified the 96.5% of the testing points, demonstrating its higher generalization skills. These results are equivalent to the results obtained by Han Yan and *al*. [15].

Table 3. Comparative performance of classification models of ANFIS and ANN.

Phases	Parameters	ANFIS	ANN
Training dataset	R^2	0.9501	0.8987
	RMSE	2.3340	0.4435
Testing dataset	R^2	0.9873	0.8988
	RMSE	2.9340	0.3704
	Correctly classified points	193	170
	Rec_rate	96.5%	85%

5. Conclusion

In this work, we have presented the application of ANFIS and ANN multi-class models dedicated for an intelligent system of water quality classification. The study area was the Tilesdit dam from Algeria. The results obtained have shown that ANFIS model can achieve high performance in classification in terms of the recognition rates. Applicability of ANFIS approach in the field of water quality status assessment and classification was investigated and well justified. Five models with different membership functions were constructed and trained by this approach. Selecting the proper parameters values, we could build a classification model with high performance and accuracy. Comparing the performance of models, the ANFIS model with Two Gaussian membership function had the best performance and was selected as the best fitting model. The highest value of correlation coefficient was obtained from the ANFIS model. As a result, the model can correctly predict 96.5% of the water quality status, which demonstrated satisfactory results of this approach. This model performed better than ANN model and can generate output value in continuous form which makes water quality assessment more comprehensible. In general, a continual enrichment of the database is practically essential because it depends on many climatic and geographic parameters. This obviously concerns the number and the type of samples to be used in the training dataset. Indeed, a number of input variables, means a limited number of sensors. In this case, a multi-sensor monitoring system operating with several inputs for many zones and regions can be perfectly justified.

The accuracy of the system decision can be improved by exploiting the combination data fusion method that merged the power of each classifier to derive a more powerful classifier. Our future work consists also of using soft-sensors in the presence of the chemical parameters that cannot be measured continuously. The robustness of the technique with respect to noise is also must be performed. It remains to note that the sensitivity of the domain and unforeseen threats, require greater efforts to maximize the immunity of the system and make more improvements to minimize the risks that incurred the public health. Finally, the present application shows a promising alternative for intelligent water quality monitoring in the future.

6. References

[1] G. Shoba, G. Shobha, "Water Quality Prediction Using Data Mining techniques: A Survey", *IJECS. International Journal Of Engineering And Computer Science*, vol., no. 6, pp. 6299-6306, ISSN: 2319-7242, June, 2014.

[2] Yi Wang, Peng Wang, Yujun Bai, Zaixing Tian, Jingwen Li, Xue Shao, Laura F. Mustavich, Bai-Lian Li, "Assessment of surface water quality via multivariate statistical techniques: A case study of the Songhua River Harbin region, China", *Journal of Hydro-environment Research*, vol 7, pp. 30-40, 2013.

[3] W. F. Zhou, Bingfang Wu, "Assessment of soil erosion and sediment delivery ratio using remote sensing and GIS: a case study of upstream Chaobaihe River catchment, north China", *International Journal of Sediment Research*, vol. 6, no. 23(2), pp.167-173, 2008.

[4] Sirilak Areerachakul and Siripun Sanguansintukul, "Classification and regression trees and MLP neural network to classify water quality of canals in Bangkok, Thailand", *IJICR*, *International Journal of Intelligent Computing Research*, vol. 1, no. 1/2, pp. 43-50, 2010.

[5] Han Yan, Zhihong Zou and Huiwen Wang, "Adaptive neuro fuzzy inference system for classification of water quality status", *Journal of Environmental Sciences*, vol. 22, no. 12, pp. 1891-1896, 2010.

[6] K. Polat, Ş. Yosunkaya, and S. Güneş, "Pairwise ANFIS approach to determining the disorder degree of obstructive sleep apnea syndrome", *Journal of Medical Systems*, vol. 32, no. 5, pp. 379–387, 2008.

[7] S. Issac Niwas, P. Palanisamy, R. Chibbar, and W. J. Zhang, "An expert support system for breast cancer diagnosis using color wavelet features", *Journal of Medical Systems*, vol. 36, pp. 3091–3102, 2011.

[8] H. Ankishan and F. Ari, "Snore-related sound classification based on time-domain features by using ANFIS model", IEEE INISTA'11, in Proceedings of the International Symposium on Innovations in Intelligent Systems and Applications, Istanbul, Turkey, June 2011, pp. 441–444.

[9] J.-S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system", *IEEE Transactions on Systems, Man and Cybernetics*, vol. 23, no. 3, pp. 665–685, 1993.

[10] H. Ankışhan and D. Yılmaz, "Comparison of SVM and ANFIS for Snore Related Sounds Classification by

Using the Largest Lyapunov Exponent and Entropy", *Computational and Mathematical Methods in Medicine*, vol. 2013, Article ID 238937, 2013

[11] J. Senthilnath, Shivesh Bajpai, S.N. Omkar, P.G. Diwakar, V. Mani, "An approach to multi-temporal MODIS image analysis using image classification and segmentation", *Advances in Space Research*, vol. 50, pp. 1274–1287, 2012.

[12] N. Sarani, J. Soltani, S. Sarani and A. Moasheri, "Comparison of Artificial Neural Network and Multivariate Linear Regression Model to Predict Sodium adsorption ratio (SAR) (Case Study: Sistan River, Iran) ", ICEES'2012, in International Conference on Chemical, Ecology and Environmental Sciences, Bangkok, march 17-18, 2012.

[13] J.S.R. Jang, "ANFIS: adaptive-network-based fuzzy inference system*", IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 3, pp. 665–685, 1993.

[14] J.S.R. Jang, C.T. Sun and E. Mizutani. (1997). *Neuro-fuzzy and soft computing*, PTR, Prentice Hall.

[15] H. Yan, Z. Zou, H. Wang, "Adaptive neuro fuzzy inference system for classification of water quality status", *Journal of Environmental Sciences*, vol. 22, no. 12, pp. 1891–1896, 2010.

[16] A. Devi B V, S. S. Priyadharsini, "Diagnosis of Neuromuscular Disorders Using Softcomputing Techniques", *IJSCE*, *International Journal of Soft Computing and Engineering*, vol. 3, no. 5, ISSN: 2231-2307, Nov. 2013.

[17] W-C. Wang, K-W. Chau, C-T. Cheng, L. Qiu, "A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series", *Journal of Hydrology*, vol. 374, pp. 294–306, 2009.

[18] H. Tabari, O. Kisi, A. Ezani, P. H. Talaee, "SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment", *Journal of Hydrology*, vol. 444–445, pp. 78–89, 2012.

[19] Z. He, X. Wen, H. Liu, J. Du, "A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region", *Journal of Hydrology*, vol. 509, pp. 379–386, 2014.

[20] X. Chen, Y. S. Li, Z. Liu, K. Yin, Z. Li, O. W. Wai and B. King, “ Integration of multisource data for water quality classification in the Pearl River estuary and its adjacent coastal waters of Hong Kong,” *Continent. Shelf Res*, vol 24, no. 16, pp. 1827-1843, 2004.