Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 8 (2): 309-315, 2008
ISSN 1812-5654

Application of Artificial Neuro-Fuzzy Logic Inference System for Predicting the Microbiological Pollution in Fresh Water

[1,2]S. Bouharati, [1]K. Benmahammed, [2]D. Harzallah and [3]Y.M. El-Assaf
[1]Laboratory of Intelligent Systems, Department of Electronic, Faculty of Engineering, UFA Setif-Algeria
[2]Microbiology Laboratory, Department of Biology, Faculty of Science, UFA Setif-Algeria
[3]Faculty of Engineering, American University of Sharjah, UAE

Abstract: The classical methods for detecting the micro biological pollution in water are based on the detection of the coliform bacteria which indicators of contamination. But to check each water supply for these contaminants would be a time-consuming job and a qualify operators. In this study, we propose a novel intelligent system which provides a detection of microbiological pollution in fresh water. The proposed system is a hierarchical integration of an Artificial Neuro-Fuzzy Inference System (ANFIS). This method is based on the variations of the physical and chemical parameters occurred during bacteria growth. The instantaneous result obtained by the measurements of the variations of the physical and chemical parameters occurred during bacteria growth-temperature, pH, electrical potential and electrical conductivity of many varieties of water (surface water, well water, drinking water and used water) on the number *Escherichia coli* in water. The instantaneous result obtained by measurements of the inputs parameters of water from sensors.

Key words: Testing water, *E. coli* detection, fuzzy expert system, hybrid intelligent system, artificial neuro-fuzzy logic inference

INTRODUCTION

The classical microbiological analysis in water pollution relies highly upon the operator's experience. The bacteriological quality of water, gagged by the presence of coliform bacteria, is easily tested in the laboratory. But in medical and industrial microbiology is now a need for «Rapid methods» that will shorten the time between receipt of specimen or samples and the issue of a report. The mechanical, electrical or electronic procedures alone or in combination, can be removing the need for direct human action (Collins *et al*., 1995). Conductance or impedance measurements are used e.g., for rapid measurement of total microbial activity in many products. An alternative to the traditional direct conductimetry has become available, namely the indirect conductimetry in which the evolution of pH and electrical potential from culture media as a result of growth of micro-organisms can be studied (Owens *et al*., 1989). Some microbiological tests based on the indirect method by the measures of the physical and chemical variations during bacteria growth in water. The level of bacteria control is regarded as one of the most important factors in environmental protection and in hygienic public health. In this study, modelling methods are used for predicting behaviour of a complex system, such as *Escherichia coli* bacteria which indicators of contamination concentration in water, which have a high non-linearity and a complex structure. The data used in this study were Temperature (T), pH of the water, Electrical Potential (EP) and Electrical Conductivity (EC). The difficulties are embedded in these measurements especially in the values of parameters and their origins. The Artificial Neuro-Fuzzy Inference Systems (ANFIS) can be used to model nonlinear relationships. ANFIS have been employed to learn numerical data recorded from sensory measurement. After being trained, ANFIS store knowledge in numerical weights and biases that are often regarded as a black box scheme. ANNs can be translated to symbolic/linguistic rules extracted are then used as a knowledge base for a fuzzy expert system. (Fhayung and Gary, 2003). Hybrid intelligent architecture tends to be more appropriate in applications that require both numerical computation for higher generalization and symbolic/linguistic reasoning for explanation. it is found that hybridization between symbolic/linguistic and numerical representation can achieve higher accuracy compared to either one alone (Tan, 1997; Taha and Ghosh, 1999; Wernier and Sun, 2000).

Neuro-fuzzy system: One of the major challenges of microbiological detection in drinking water is the

Corresponding Author: S. Bouharati, Microbiology Laboratory, Department of Biology, Faculty of Science, Setif University, Setif 1900, Algeria

detecting of bacteria in an unstructured environment, in particular where the physical and chemical properties of the bacteria are not known *a priori*. The resultant uncertainty makes it difficult to detect nature. Various techniques have been applied to solve this problem. Some approaches are analytic and cannot be easily implemented in real-time applications. Also, the analytic approach cannot be used if variables such as the physical's weight are unknown. To overcome this situation, other approaches using fuzzy detectors have been developed using a number of different sensors to measure the physical variables. The fuzzy control can be used in conjunction with powerful automatic learning methods (i.e., neural networks) as a neuro-fuzzy system (Brown and Harris, 1994a; Kecman, 2001a).

In this study, we explore these issues using a very simple four-fingered gripper as an experimental system.

MATERIALS AND METHODS

Fuzzy logic modeling: In traditional set theory, something either belongs to a set or does not depending on whether it fits the definition for that set. In fuzzy set theory, something can partially belong to a set. For example, let us assume, for illustrative purposes only, that the variable 'number of bacteria' has a range of values where safe is considered to be $<10^{+4}$ mL^{-1}, whereas low is below this range. and 'contaminated' above this range. In traditional set theory, a number of bacteria in water of 10^{+3} mL^{-1} would be classified as safe, whereas 10^{+4} mL^{-1} would be contaminated (Fig. 1). With fuzzy sets, a value for a variable can partially belong to asset and have a degree of membership anywhere between zero and one (i.e., $0\leq\mu\leq1$) (Fig. 2). Relationships among fuzzy sets are expressed as a series of 'if-than' rules to form a rule base. One of the most widely known fuzzy logic modeling algorithms is tat of Sugeno-Yasukawa. The following is a general description of the basic steps in the process, using a microbiological modeling example (Beth *et al.*, 2002).

Assemble input-output dataset: The first step is to collect all the variables (e.g., origin, temperature, pH, electrical potential and electrical conductivity of the water) and output (e.g., the number of bacteria). In our study, 1000 samples of water are used. The data were studied in Microbiological Laboratory of the Setif University Algeria. They were collected from different varieties of water (well water -250 samples; surface water -250 samples; used water -250 samples and drinking water -250 samples). The physical and chemical changes, pH, Electrical Potential (EP), Electrical Conductivity (EC) were measured at different Temperatures (T) according the

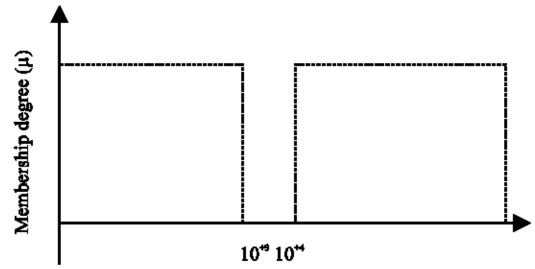

Fig. 1: A value for a variable either belong to a set or does not belong to a set

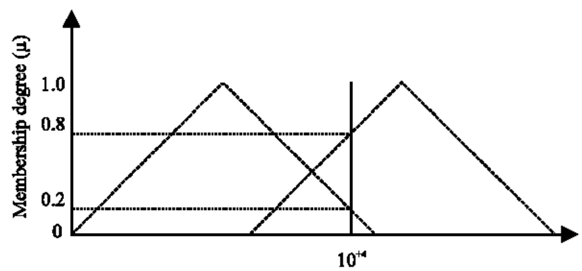

Fig. 2: A value for a variable can partially belong to a set and have a degree of membership anywhere zero and one

Table 1: Variations of the output LogN according each output T°; pH; EP; EC and the origin of water

Origin	T°	pH	EP	EC	Log N
1	15	8.45	0.85	3.01	4.80
1	24	6.98	0.04	4.32	6.90
2	16	6.42	0.30	11.01	5.12
2	26	7.46	0.50	15.51	9.94
3	24	8.11	0.61	2.10	4.10
3	32	6.91	0.02	4.25	5.72
4	12	8.20	0.51	2.08	3.11
4	32	7.41	0.07	2.60	3.56

1: Surface water, 2: Used water, 3: Well water and 4: Drinking water

number of bacteria (LogN). The bacterial analysis concerned the *E. coli* bacteria (fecal coliforms) measured by the classical methods-NPP in each case (Table 1).

Membership function: The membership function for the input variables is shown in Fig. 3. (Temperature as example). The values of k in the figure for all inputs ($k = k_1, k_2,\dots k_5$) are equal to the mean values of each time series (Chen *et al.*, 2000). Using these mean k values, the data for the inputs were classified into three linguistic categories: Low, middle, high. The data for the output was classified into three linguistic categories: safe, contaminated, more contaminated Fig. 4.

Cluster the output: The values of the output (response) are fuzzy clustered using for examples the fuzzy c-means

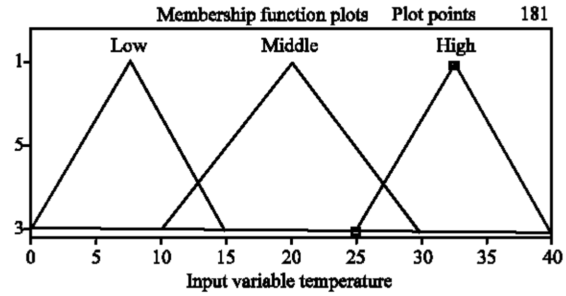

Fig. 3: The input discourse universe was classified linguistic categories: low, middle, high

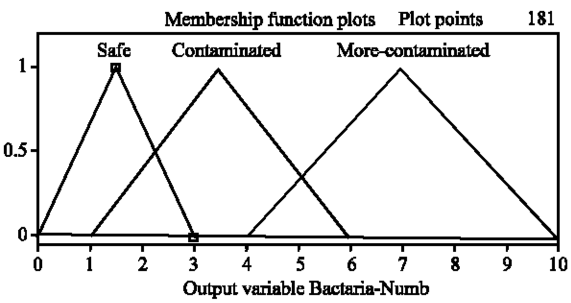

Fig. 4: The output discourse universe was into three classified into three linguistic categories: Safe, Contaminated, More contaminated

algorithm (Bezdek, 1981). The number of fuzzy sets created determines the number of rules in the 'if-than' rule base. One rule is used for each fuzzy set of the output.

Fuzzy system inference model: Basically, any fuzzy logical model constitutes three parts: the fuzzy membership function, fuzzy decision rules and fuzzy reasoning (Chen *et al*., 2000). In this study, five inputs are chooses: X_1 is the origin of water, X_2 is the temperature of water, X_3 is the pH, X_4 is the electrical potential and X_5 is the electrical conductivity. The output of the fuzzy logic model is Y, which is the number of *Escherichia coli* in water. By using fuzzy sets, we can formulate fuzzy if-then rules, which commonly used in our daily expressions. We can use a collection of fuzzy rules to describe a system's behavior; this forms the fuzzy inference system, or fuzzy controller if used in control systems.

In this study, we are going to use fuzzy inference system exclusively, where the output equation of each rule is a linear equation. First we find the membership grades of the IF parts of the rules; the heights of the dashed line represent these values. Since the pre-conditions in the IF part are connected by AND, so we use multiplication to find the firing strength of each rule.

The physical and chemical changes, pH, Electrical Potential (EP), Electrical Conductivity (EC) were measured at different Temperatures (T) are chosen as input of the system. The number of bacteria (LogN) is chosen as the corresponding output. Both the different levels of input and output are defined by specific membership function for the fuzzy sets. The models have a multivariable system with N input variables and M output variable.

Fuzzy rules: The rules determined by the choice of the fuzzy membership function are defined for each input variable. In general form, each fuzzy rule is written as were A_1 and A_2 are the fuzzy sets that describe the nature of the inputs, such as small, medium, or high.

Examples 1: If Origin is (1) and T is low and pH is medium and EP is low and EC is high THAN LogN is safe.
...Also.

The linguistic control rules of this system are given by: (Li-Xin Wang, 1997).

IF X_1 is $X_1(1)$ and X_2 is $X_2(2)$ and ... X_n is $X_n(1)$ THAN Y_1 is $Y_1(1)$.

Neural network model: The study consisted of training and testing of the ANN for prediction of outcome in collected water samples. The data on all measured parameters were prospectively collected in a computer database (Microsoft Excel). ANN was applied to provide a nonlinear relationship between inputs variables (Temperature, pH, electrical potential, electrical conductivity and origin of water) and the output variable (Log. N).

In present study, a network with one hidden layer was selected. The output layer comprised a single neuron corresponding to the value of the dependent variable to be fit to or predicted the number of bacteria. The model used was standard, three layer, back-propagation, neural network with N input nodes, L hidden nodes and K output nodes, which can be mathematically expressed as:

$$O_{pk} = f_1\left(\sum_{j=1}^{L} W_{jk}^{o} f_2\left(\sum_{i=1}^{N} W_{ij}^{h} x_{PI} + b_1^{j}\right) + b_2^{k}\right)$$
$$\forall k \in 1, 2, \ldots k,$$

Were, O_{pk} is the output from the kth node of the output layer of the network for the Pth vector (data point). X_{pi} the inputs to the network for the Pth vector (data point). W^{o}_{jk} the connection weight between the fth node of the hidden layer and the kth node of the output layer

(Fig. 5). W^h_{ij} the connection weight between the ith node of the input layer and the jth node of the hidden layer, b^l_1 and b^l_2 are bias terms and $f_1(.)$ and $f_2(.)$ are activation functions. the activation function used was a logistic sigmoid function which produces output in the range 0-1. Logistic sigmoid function has a form as given below:

$$f(x) = \frac{1}{1+e^x}$$

Suitable activation function for the hidden units, $f_2(.)$ is needed to introduce non-linearity into the network, which gives the power to capture non-linear relationship between input and output. For the output units, the selection of the activation function $f_1(.)$, is based on the distribution of the target values (Brion *et al.*, 2002).

The Fig. 5 describes the topology of a five-input, one-hidden and one-output (5-1-1 in neural network terminology) neural network model. The four inputs are the Temperature (T°), pH, Electrical Potential (EP), Electrical Conductivity (EC), Origin of the water (1) surface water, (2) used water, (3) well water, (4) drinking water. W_{ij} and W_{jk} are weights, which represent the link between the inputs and the output. The weights contain all the information about the network. Therefore, the objective is to train the network to find a series of weights that yield an output signal that has a small error relative to the observed output (Chen *et al.*, 2000). The results of the variations of bacteria variation in different varieties of water using ANN system are shown in Fig. 6.

Neurofuzzy systems: Neurofuzzy systems combine the advantages of fuzzy systems, such as transparent representation of knowledge and those of neural networks, which deal with implicit knowledge that can be acquired by means of learning (Harris *et al.*, 2002; Brown and Harris, 1994B; Kecman, 2001b). They mimic human decision processes, in that they manage imprecise, partial, vague or imperfect information. Also, they are able to resolve conflicts by collaboration and aggregation. Moreover, they have self-learning, self-organising and self-tuning capabilities, with no requirement for prior knowledge of relationships of data although they can use this if it is available. They have fast computation using fuzzy number operations. As do fuzzy systems, neurofuzzy systems embody rules (defined in a linguistic way) so that system operation can be interpreted in a transparent fashion (Fig. 7).

The number of rules is given by $P = \prod_I N_i$ where i is the number of inputs and N_i is the number of antecedent fuzzy sets for input i while M is the number of consequent fuzzy sets. In addition, $Q = \Sigma I\ N\ I$ and $R = \Sigma^{i-1}_{k=0} N_k$. Figure 7, shows the implementation of a neurofuzzy system as a feed forward network with three layers, namely: the fuzzification layer, the fuzzy rule layer and defuzzification layer. The mapping between the first two

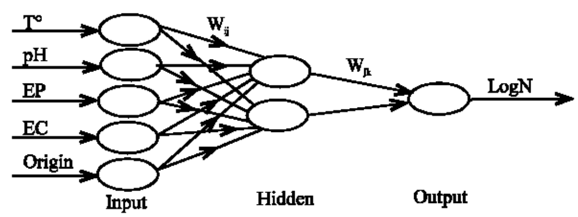

Fig. 5: The topology of a five-input, one-hidden and one-output

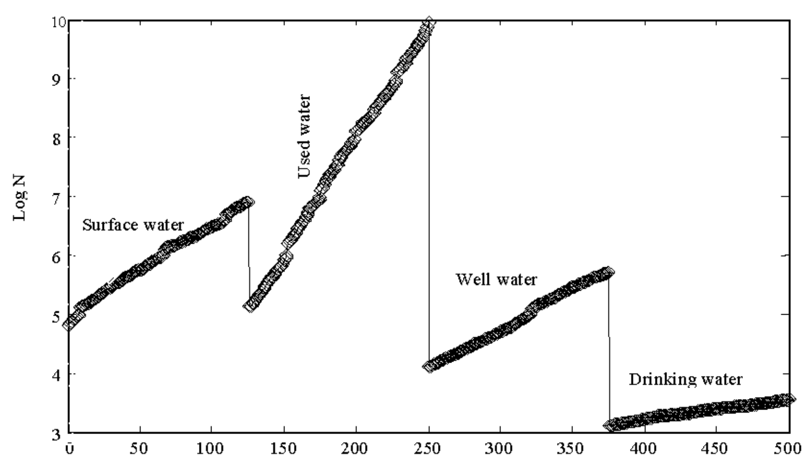

Fig. 6: The bacteria number variations in different varieties of water. The values of testing steps are totally confused with the training values

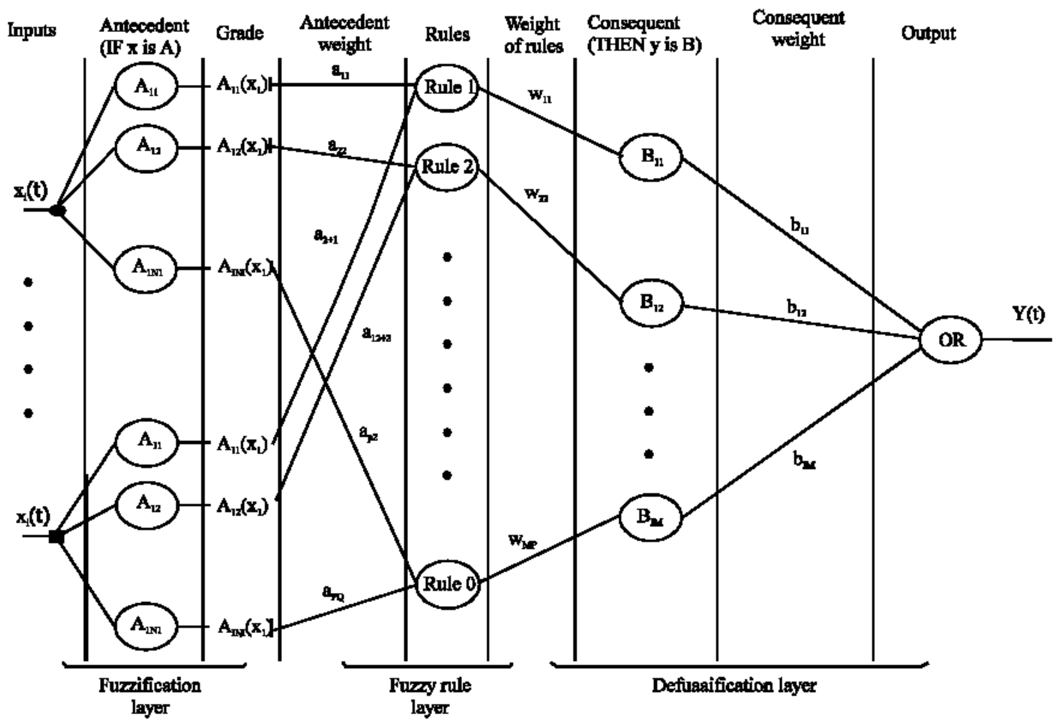

Fig. 7: Architecture of a typical neurofuzzy system

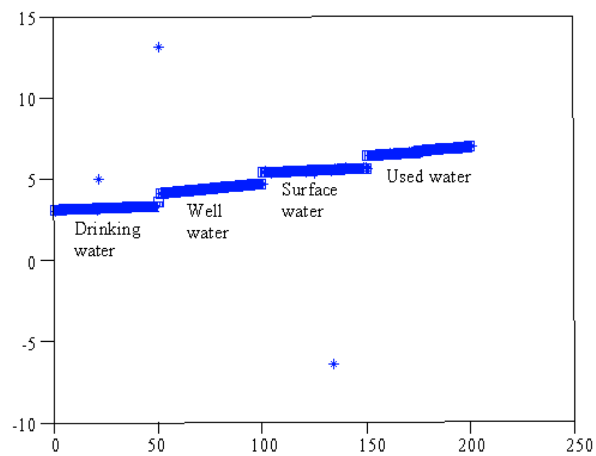

Fig. 8: The bacteria number variations in different varieties of water. The values of testing steps are totally confused with the training values even some input values are outside of the specified input range

layers is nonlinear and linear between the last two layers (Jantzen, 1998). The fuzzification layer consists of two components: the measurements x_i, which are the input signals from the system under control and the environment and the computation of the antecedent membership function value, $\mu A_{ij}(x_i)$ termed the grade of membership. The fuzzy rule layer then calculates the rule firing strength, which indicates how well the conditions in the antecedent are satisfied. An input x fires a rule to a degree, $w \in [0,1]$. The calculation of the rule firing strength is performed using the membership function values for fuzzy sets used in the rule antecedent. The product combiner is the fuzzy implication method, used because it preserves the original shape of the fuzzy set and it is differentiable. Finally, the defuzzification layer converts the values delivered by the fired rules at the previous layer into analogue output values.

The concepts of fuzzy sets and membership functions: By using fuzzy sets, we can formulate fuzzy if-then rules, which are commonly used in our daily expressions. We can use a collection of fuzzy rules to describe a system's behavior; this forms the fuzzy inference system, or fuzzy controller if used in control systems. In particular, we can apply neural networks, learning method in a fuzzy inference system.

In this study, we are going to use first-order Sugeno fuzzy inference system exclusively, where the output equation of each rule is a linear equation.

This slide explains the basic architecture of an ANFIS. We have to graphical representation of the process of fuzzy reasoning. For each operation of fuzzy reasoning, we can put it into a node in an adaptive network. For instance, given two input x and y, first we need to find the membership grades; this is represented by the four blocks in the first layer and each of them generates a membership grades. The firing strengths are computed by nodes in the second layer. The results of the variations of bacteria variation in different varieties of water using ANFIS system are shown in Fig. 8.

CONCLUSIONS

The Artificial neural network system shows that the output-number of bacteria in water corresponding to any variation in the input parameters. The result of the ANN program so far, is a numeric and symbolic terms of number of bacteria in water, using the input data from sensors and output data in the universe of discourse (safety; contaminated; more contaminated). if the number of bacteria in output is around 5 (LogN = 5), the water is contaminated; if it is less than 4, the water is safety; if it is more than 6 it is more contaminated.

In acute microbiological detection in fresh water, there are based on simple and easy-to-handle concepts like laboratory methods. By the classical methods of water contamination detection would be a time-consuming job and a qualify operators. The bacteria growth in water produces many variations in physical parameters. One of the problems in water quality modeling is the vagueness in the values of pollutant sources and of biochemical coefficients, arising either form natural randomness in time and space or from indirect measurements and limited number of samples. In this study, we used different water varieties with different contamination degree measured by the classical methods. The data were analyzed by the neuro-fuzzy logic modeling technique in an attempt to predict the microbiological pollution in fresh water. With the neuro-fuzzy modeling, we can represent imprecise data and produce imprecise output in the form of fuzzy numbers. For any predictive instrument to be useful in making a triage decision, an important feature is that only data are readily available from sensors. From the results obtained by this study, appear to be a useful tool for future water-testing on pathogenic risk identification, quantification and development of early warning systems for fresh water quality. There room for improvement in parameter selection. We emphasize that our FIS is not meant to replace or to substitute for an experienced microbiologist; or laboratory investigations; on the contrary, we envisage that the FIS should be viewed as a decision aid for the busy emergency-department, microbiologist, particularly in times of epidemics or water diseases.

The Fuzzy logic inference system shows that the output-number of bacteria in water corresponding to any variation in the input parameters. The result of the Fuzzy program so far, is a numeric and symbolic terms of number of bacteria in water, using the input data from sensors and output data in the universe of discourse (safety; contaminated; more contaminated). if the number of bacteria in output is around 5 (LogN = 5), the water is contaminated; if it is less than 4, the water is safety; if it is more than 6 it is more contaminated.

REFERENCES

Beth, A., B.A. Sproule, C.A. Naranjo and B. Türksen, 2002. Fuzzy pharmacology: Theory and application. Trends Pharmacol. Sci., 23 (9): 412-417.

Bezdek, C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York.

Brion, G.M., T.R. Neelakantan and S. Lingireddy, 2002. A neural-network-based classification scheme for sorting sources and ages of fecal contamination in water. Water Res., 36: 3765-3774.

Brown, M. and C. Harris, 1994a. Neurofuzzy Adaptive Modelling and Control. Hertfordshire, Prentice Hall International. UK.

Brown, M. and C.J. Harris, 1994b. Neurofuzzy Adaptive Modelling and Control. Hemel Hempstead, Prentice Hall. UK.

Chen, D.G., N.B. Haregreaves, D.M. Ware and Y. Liu, 2000. A Fuzzy logic model with genetic algorithm for analyzing fish stock-recruitment relationships. Can. J. Fish. Aquat. Sci., 57: 1878-1887.

Collins, C.H., M. Patricia and J.N. Lyne, 1995. Microbiological methods. Butterworth-Heinmann Ltd. Linacre House, Jordan Hill, Oxford OX2 8DP, pp: 121.

Fhayung, M. and G.Y. Gary, 2003. Combined numerical and linguistic knowledge representation and its application to medical diagnosis. IEEE Transactions on Systems, Man and Cybernetics Part A: Syst. Hum., 33 (2): 206-232.

Harris, C.J., X. Hong and Q. Gan, 2002. Adaptive Modelling, Estimation and Fusion from Data: A Neurofuzzy Approach. Berlin and Heidelberg, Germany: Springer-Verlag.

Jantzen, J., 1998. Neurofuzzy modelling. Technical University of Denmark, Lyngby, Denmark.

Kecman, V., 2001a. Learning and Soft Computing. Cambridge, MA: MIT Press/Bradford Books.

Kecman, V., 2001b. Learning and Soft Computing: Support Vector Machines. Neura Networks and Fuzzy Logic Models. Boston, MA: MIT Press.

Li-Xin Wang, 1997. Modeling and control of hierarchical systems with fuzzy systems. Automatica J. IFAC., 33 (6): 1041-1053.

Owens, R., D.S. Thompson and A.W. Timmerman, 1989. Indirect conductimetry; A novel approach, the conductometric enumeration of microbial pollution. Lett. Applied Microbiol., 9: 245-249.

Taha, T.A. and J. Ghosh, 1999. Symbolic interpretation of artificial neural networks. IEEE Trans. Knowl. Data Eng., 11: 448-463.

Tan, H., 1997. Cascade ARTMAP: Integrating neural computation and symbolic knowledge processing. IEEE Trans. Neural. Networks, 8: 237-250.

Wernier, S. and R. Sun, 2000. An Overview of Hybrid Neural Systems. In Hybrid Neural Systems, Wernier, S. and R. Sun (Eds.). Berlin, Germany-Springer-Verlag, pp: 1-13.