

Detection of waterborne bacteria using Adaptive Neuro-Fuzzy Inference System

Farhan Mohammad Khan^{1*}, Smriti Sridhar², Rajiv Gupta³

¹Ph.D. Research Scholar, Department of Civil Engineering, BITS Pilani, Pilani campus.

²MS Student, Department of System Engineering, University of Virginia.

³Senior Professor, Department of Civil Engineering, BITS Pilani, Pilani campus

Abstract. The detection of waterborne bacteria is crucial to prevent health risks. Current research uses soft computing techniques based on Artificial Neural Networks (ANN) for the detection of bacterial pollution in water. The limitation of only relying on sensor-based water quality analysis for detection can be prone to human errors. Hence, there is a need to automate the process of real-time bacterial monitoring for minimizing the error, as mentioned above. To address this issue, we implement an automated process of water-borne bacterial detection using a hybrid technique called Adaptive Neuro-fuzzy Inference System (ANFIS), that integrates the advantage of learning in an ANN and a set of fuzzy if-then rules with appropriate membership functions. The experimental data as the input to the ANFIS model is obtained from the open-sourced dataset of government of India data platform, having 1992 experimental laboratory results from the years 2003-2014. We have included the following water quality parameters: Temperature, Dissolved Oxygen (DO), pH, Electrical conductivity, Biochemical oxygen demand (BOD) as the significant factors in the detection and existence of bacteria. The membership function changes automatically with every iteration during training of the system. The goal of the study is to compare the results obtained from the three membership functions of ANFIS: Triangle, Trapezoidal, and Bell-shaped with 35 = 243 fuzzy set rules. The results show that ANFIS with generalized bell-shaped membership function is best with its average error 0.00619 at epoch 100.

1. Introduction

In the present situation, people are struggling to obtain access to clean water due to significant infectious risks linked with the ingestion of contaminated water with human or animal feces [1]. Micro-organisms or pathogens cause some of the primary diseases because they may live, reproduce, and disperse in water systems [2]. Approximately 1.7 billion children in developing countries under the age of five died due to diarrhea, mainly by drinking contaminated water that has been reported by the World Health Organization (WHO) in 2011. Besides, 525,000 children deaths a year in 2018 in world-wide because of the poor water quality, sanitation, and hygiene conditions, mainly through infectious diarrhea. In the world, 1.9 billion people use water, which is faecally contaminated [3]. Drinking water comes from surface and groundwater sources. Due to damaged sanitation and dwindling resources that make the availability of safe water almost inaccessible because of the bacterial and chemical contaminations [4]. With rapid research on the detection of water-borne pathogens to curb diseases, the adequacy of the standards of drinking water quality to prevent water-borne illnesses is still not met [5]. According to the micro-biological drinking water quality standards established by the WHO, EPA, and IS: 10500 [6], bacteria shall not be detectable in 100 ml of

water sample [1, 17]. Artificial neural networks (ANN) is a mimic of the biological neural network of the visual cortex in the brain. The brain consists of a densely interconnected set of information-processing units called neurons. Information is stored and processed in the brain by the involvement of every neuron, which subsequently helps in human learning. Similarly, an ANN model trains itself for learning by connecting to different nodes [7]. The ANFIS incorporates the self-learning ability of neural networks with the linguistic expression function of fuzzy inference. ANFIS is a multilayer feed-forward network in which each node performs a particular function on receiving signals and has a set of parameters about this node. Similar to ANN, ANFIS is capable of converting unseen inputs to their respective outputs by learning the rules from previously observed data [8]. An ANFIS model is capable of adjusting the parameters better in any series and takes into consideration all the edge cases in a rule-viewer interface. ANN model may not take the probabilistic values, but using ANFIS, we can make a set of rules for the same. While ANFIS integrates with both fuzzy inference systems and ANNs, it is useful in solving non-linear and complex problems within a frame [9] Hybrid based methods like ANN and fuzzy or ANFIS-GA (Genetic Algorithm), can prove to be extremely useful in dealing with missing data [10]

* Corresponding author: p20180411@pilani.bits-pilani.ac.in

Hybrid models significantly increase the accuracy of estimation, especially in non-linear problems [11]

2. Related work

Bouharati, S et al. [8] proposed a method for the detection of micro bacterial pollution in freshwater using an ANFIS. The model produced instantaneous results by the measurement of the physical and chemical properties of the sensors. ANFIS based methods are based on the concept of Fuzzy set theory, which states that a variable can partially belong to a set and can have a membership value between 0 and 1. In this paper, three parameters were selected as input, i.e., pH, temperature, electrical potential, and the output was the number of bacteria. The next step was to create a membership function for the input variables; using these variables we can classify them into different levels like low, middle, or high. The number of rules depicted the number of fuzzy sets created. The author revealed the use of an artificial neural network model of three layers trained and tested on the collected water samples. ANFIS based model was used because it combines the advantages of fuzzy systems with transparent knowledge representation and those neural networks which deal with implicit knowledge that can be acquired using learning. Kamali and Binesh [12] used ANN and ANFIS to study the diffusion of water through nanotubes using the molecular dynamics data. They concluded that ANFIS outperformed ANN.

Similarly, Azeez et al. [13] compared the performance of ANN and ANFIS in the triage of emergency patients using various vital signs of patients as input parameters. Chandaran, U. D et al. [14] explained the detection of sulphate reducing bacteria (SRB) using ANFIS, which can be crucial in curbing the corrosion of iron material in the system. The author used three parameters: Voltage, Temperature, and humidity for training the model. The membership functions were taken to be trapezoidal and bell-shaped. The ANFIS model used three inputs, which finally gives the output as either 1 or 0. The predicted results were obtained by the input parameters and the number of epochs was taken as 20. Lastly, the model was tested with testing data up to 250 epochs. The author compared the results and the best membership function was given by trapezoidal shape. Keshavarz, Z et al. [10] explained the application of ANFIS based method in determining the compressive strength of concrete. The model used 150 different concrete specimens with various mix design parameters. Five different concrete mix parameters, i.e., cement, water to cement ratio, gravel, sand, and micro-silica, were considered as the parameters. For results, two of the soft computing methods: ANN and ANFIS, were selected to detect the compressive strength of concrete. The results were computed in MATLAB, where the concrete mix parameters were used as input variables and the compressive strength of concrete was used as an output parameter. In order to compare the ANN and ANFIS based methods, the author used parameters like the R squared coefficient of both the models. The higher values

of the coefficient of determination would indicate the better capability of the model in predicting the specific studied characteristics. Calp, M. H [11] proposed a hybrid model for the estimation of the regional rainfall amount. The proposed model focused on providing efficient management of water resources by estimating the amount of rainfall that can occur in the region. While creating the model, the MATLAB package program was used and regression values (R) or mean squared error (MSE) were taken into account. The error rate was obtained as 0.9920, 0.9840 and 0.0011, respectively, for the model. The author concludes by stating that this hybrid model is an important support tool for estimating the amount of annual rainfall and ensuring the effective management of water resources.

3. Methodology

In the first part of this section, we explain the design of a fuzzy expert system based on membership functions. Subsequently, an ANFIS model will be introduced based on the fuzzy rule base. In this study, we have assigned three fuzzy sets to each water quality parameter, namely desirable, undesirable, and highly undesirable. The individual membership function is assigned to each parameter, as described in Table 1.

Table 1. Groups defined for water quality parameters

Parameters	Range		
	Un desirable	Desirable	Highly Undesirable
Temperature	0-10	5-41	35-48
Dissolved oxygen	0-5	2-14	10-20
pH	0-7.5	6.5-8.5	8-14
Conductivity	0-300	200-1000	800-1000
BOD	0-2	1-5	3-8

3.1 Data Input

A series of experimental data is obtained from open government data platform India [17]. There are 1992 experimental results collected from 2003-2014. The detection of bacterial presence is based on five basic water quality parameters, i.e., Temperature, Dissolved Oxygen (DO), pH, Electrical conductivity, Biochemical oxygen demand (BOD). There are five inputs and one output for each set of data. The output, '1' represents the existence of bacteria, and '0' indicates absence. These five inputs are used to model the ANFIS system, which gives an output that gives output in between 1 and 0. The computation of data for ANFIS is conducted using MATLAB 2019a. However, the ANFIS training algorithms are embedded in the MATLAB's fuzzy logic toolbox [16]. There are four steps for computation. The first step is data input and normalization. The total experimental data after normalization are divided randomly into 70% for training, 15% for testing, and 15% for validation. The next step is to assign membership

functions for data, then to train the input using the ANFIS training function. Finally, the predicted result can be obtained by inputting the parameters. Figure 1 shows the structure of the ANFIS model.

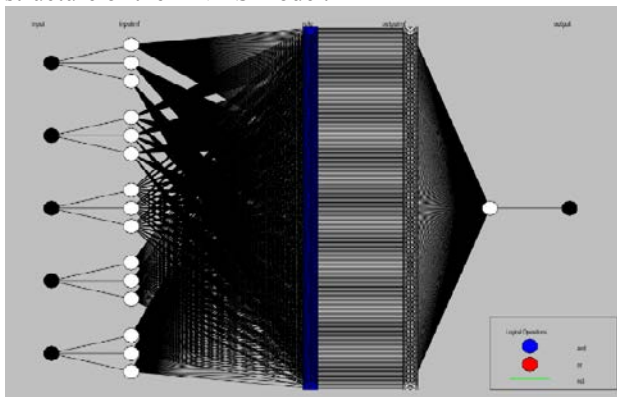


Figure 1. Structure of the ANFIS model

3.2 Assigning membership function

In this study, three types of membership functions (Triangular, Trapezoidal, Bell-shaped) are tested in order to get the best function for the prediction model. The interval contains three fuzzy sets as: “Desirable”, “Undesirable” and “Highly undesirable”. Input parameters of the Fuzzy Inference System (FIS) are: Temperature, Dissolved Oxygen, pH, Conductivity, BOD. Temperature accepts values in [0 48] °C. Mathematical equations of membership expressions for temperature are shown from Eq. (1, 2, and 3), Dissolved Oxygen accepts values in [0 20] ppm. Mathematical equations of membership expressions for DO are shown from Eq. (4, 5, and 6). pH accepts values in [0 14]. Membership equations of membership expressions for pH are shown from Eq. (7, 8, and 9). Conductivity accepts values in [0 1000] μS/cm. Membership equations for conductivity are shown from Eq. (10, 11, and 12). Biochemical demand (BOD) accepts values in [0 8] ppm. Membership functions for BOD are shown from Eq. (13, 14, and 15)

1 Temperature

Undesirable:

$$\mu_U(x) \Rightarrow \begin{cases} 1 & \text{if } x < 5 \\ \frac{10-x}{5} & \text{if } x < 10 \text{ and } x \geq 5 \end{cases} \quad (1)$$

Desirable:

$$\mu_D(x) \Rightarrow \begin{cases} \frac{x-5}{23} & \text{if } x < 23 \text{ and } x \geq 5 \\ 1 & \text{x} = 23 \\ \frac{41-x}{23} & \text{if } x < 41 \text{ and } x \geq 23 \end{cases} \quad (2)$$

Highly Undesirable:

$$\mu_H(x) \Rightarrow \begin{cases} \frac{x-35}{13} & \text{if } x < 48 \text{ and } x \geq 35 \\ 1 & \text{x} \geq 48 \end{cases} \quad (3)$$

2 Dissolved Oxygen (DO)

Undesirable:

$$\mu_U(x) \Rightarrow \begin{cases} 1 & \text{if } x < 2 \\ \frac{2-x}{3} & \text{if } x < 5 \text{ and } x \geq 2 \end{cases} \quad (4)$$

Desirable:

$$\mu_D(x) \Rightarrow \begin{cases} \frac{x-2}{6} & \text{if } x < 8 \text{ and } x \geq 2 \\ 1 & \text{x} = 8 \\ \frac{14-x}{6} & \text{if } x < 14 \text{ and } x \geq 8 \end{cases} \quad (5)$$

Highly Undesirable:

$$\mu_H(x) \Rightarrow \begin{cases} \frac{x-10}{10} & \text{if } x < 20 \text{ and } x \geq 10 \\ 1 & \text{x} \geq 20 \end{cases} \quad (6)$$

3 pH

Undesirable:

$$\mu_U(x) \Rightarrow \begin{cases} 1 & \text{if } x < 6.5 \\ \frac{7.5-x}{1} & \text{if } x < 7.5 \text{ and } x \geq 6.5 \end{cases} \quad (7)$$

Desirable:

$$\mu_D(x) \Rightarrow \begin{cases} \frac{x-6.5}{1} & \text{if } x < 7.5 \text{ and } x \geq 6.5 \\ 1 & \text{x} = 7.5 \\ \frac{8.5-x}{1} & \text{if } x < 8.5 \text{ and } x \geq 7.5 \end{cases} \quad (8)$$

Highly Undesirable:

$$\mu_H(x) \Rightarrow \begin{cases} \frac{x-8}{3} & \text{if } x < 8 \text{ and } x \geq 11 \\ 1 & \text{x} \geq 11 \end{cases} \quad (9)$$

4 Conductivity

Undesirable:

$$\mu_U(x) \Rightarrow \begin{cases} 1 & \text{if } x < 200 \\ \frac{300-x}{100} & \text{if } x < 200 \text{ and } x \geq 300 \end{cases} \quad (10)$$

Desirable:

$$\mu_D(x) \Rightarrow \begin{cases} \frac{x-200}{400} & \text{if } x < 600 \text{ and } x \geq 200 \\ 1 & \text{x} = 600 \\ \frac{1000-x}{400} & \text{if } x < 600 \text{ and } x \geq 1000 \end{cases} \quad (11)$$

Highly Undesirable:

$$\mu_H(x) \Rightarrow \begin{cases} \frac{x-800}{200} & \text{if } x < 1000 \text{ and } x \geq 800 \\ 1 & \text{x} \geq 1000 \end{cases} \quad (12)$$

5 BOD

Undesirable:

$$\mu_U(x) \Rightarrow \begin{cases} 1 & \text{if } x < 1 \\ \frac{2-x}{1} & \text{if } x < 1 \text{ and } x \geq 2 \end{cases} \quad (13)$$

Desirable:

$$\mu_D(x) \Rightarrow \begin{cases} \frac{x-1}{3} & \text{if } x < 3 \text{ and } x \geq 1 \\ 1 & \text{x} = 3 \\ \frac{5-x}{2} & \text{if } x < 3 \text{ and } x \geq 5 \end{cases} \quad (14)$$

Highly Undesirable:

$$\mu_H(x) \Rightarrow \begin{cases} \frac{x-3}{5} & \text{if } x < 8 \text{ and } x \geq 3 \\ 1 & \text{x} \geq 8 \end{cases} \quad (15)$$

3.3 Predicted output for ANFIS

The testing data is used to check the capability of the model. For output prediction, we compare the results obtained from three membership functions (triangle, trapezoidal, and bell-shaped). The error rate and parameters for every single membership function are plotted. The resulted FIS is tested using testing data for a hundred epochs. We tabulate the error analysis for each membership function type.

4. Results and discussions

a) Triangular as Membership Function

By setting the number of membership functions to three for input data and using triangular function, the

parameters for each input's membership function are tabulated for epoch 1-50. Figure 2 and Figure 3 show the initial and final membership functions of the input data derived by training via the triangular function. Table-2 shows the label of results that are used and their representative of the membership functions. The proposed ANFIS is shown in Figure 1.

Table 2. Labels of results showing membership function (Where in=input, mf=membership function)

Labels of results	Temperature	Dissolve oxygen	pH	Conductivity	BOD
Desirable limits	in1mf1	in2mf1	in3mf1	in4mf1	in5mf1
Undesirable limits	in1mf2	in2mf2	in3mf2	in4mf2	in5mf2
Highly undesirable limits	in1mf3	in2mf3	in3mf3	in4mf3	in5mf3

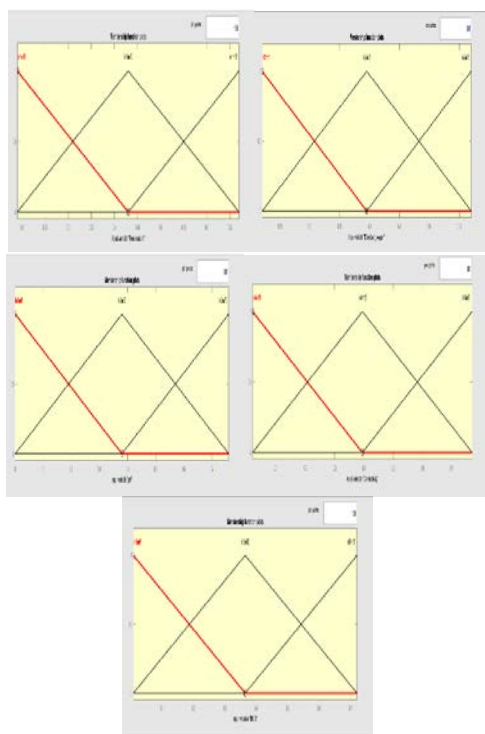


Figure 2: Initial membership function of Temperature, DO, pH, conductivity, BOD (Triangular).

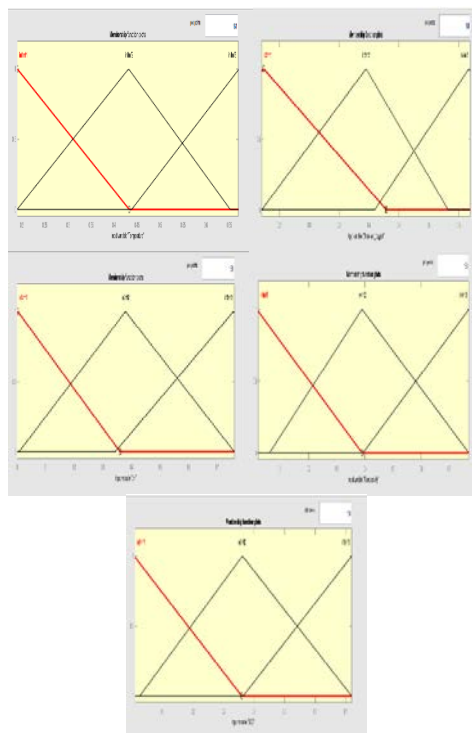


Figure 3. Final membership function of Temperature, DO, pH, conductivity, BOD (Triangular).

b) Trapezoidal as Membership Function

By setting the number of membership functions to three for input data and using trapezoidal function, the parameters for each input's membership function are recorded and tabulated for epoch 1-20. Figure 4 and Figure 5 show the initial and final membership functions of the input data derived by training via the trapezoidal function.

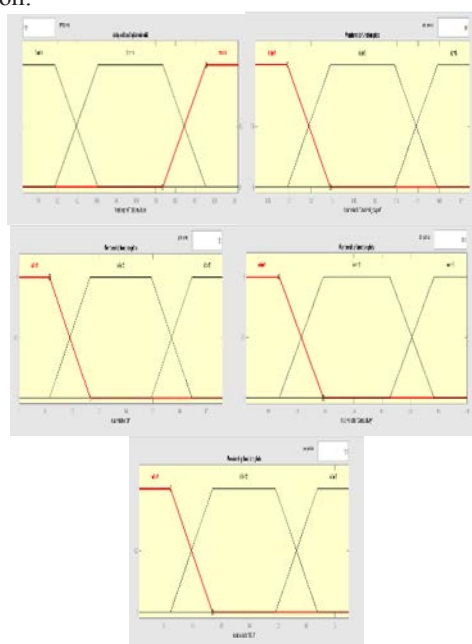


Figure 4. Initial membership function of Temperature, DO, pH, conductivity, BOD (Trapezoidal).

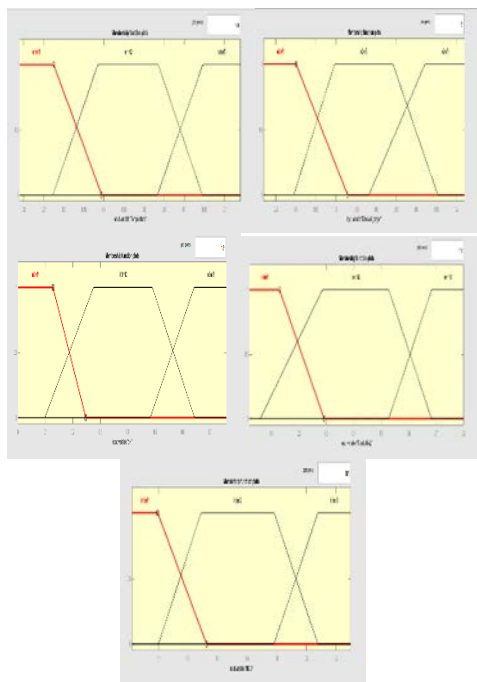


Figure 5. Final membership function of Temperature, DO, pH, conductivity, BOD (Trapezoidal).

c) Generalized bell-shaped as Membership Function
 By setting the number of membership functions to 3 for input data and using a generalized bell function, the parameters for each input's membership function are recorded and tabulated for epoch 1-100. Figure 6 and Figure 7 show the initial and final membership functions of the input data derived by training via the trapezoidal function.

Figure 6. Initial membership function of Temperature, DO, pH, conductivity, BOD (Bell-shaped).

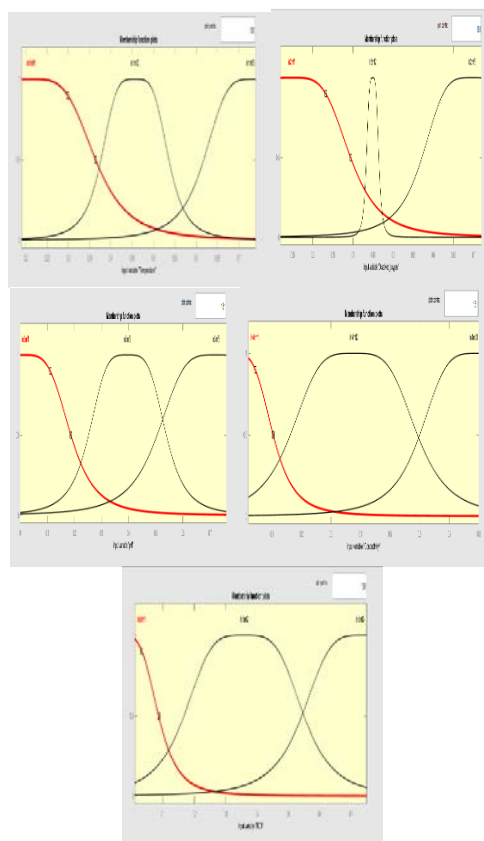


Figure 7. Final membership function of Temperature, DO, pH, conductivity, BOD (Bell-shaped).

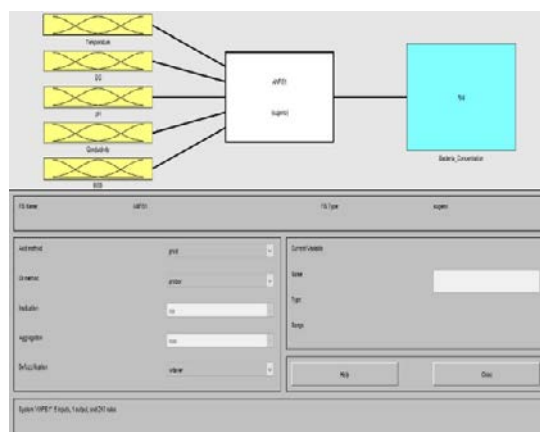
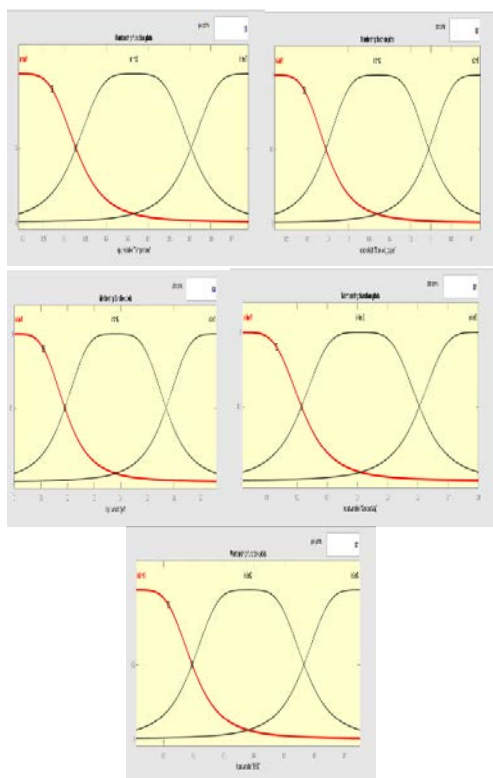


Figure 8. Proposed ANFIS for the detection of bacteria in terms of temperature, dissolved oxygen, pH, conductivity and BOD as the model inputs.

Error rate

The training error of three membership functions are tested and shown in Figure 9. For each type of membership function, the error rate after every single epoch has been recorded. The rate is tabulated in Table 3.

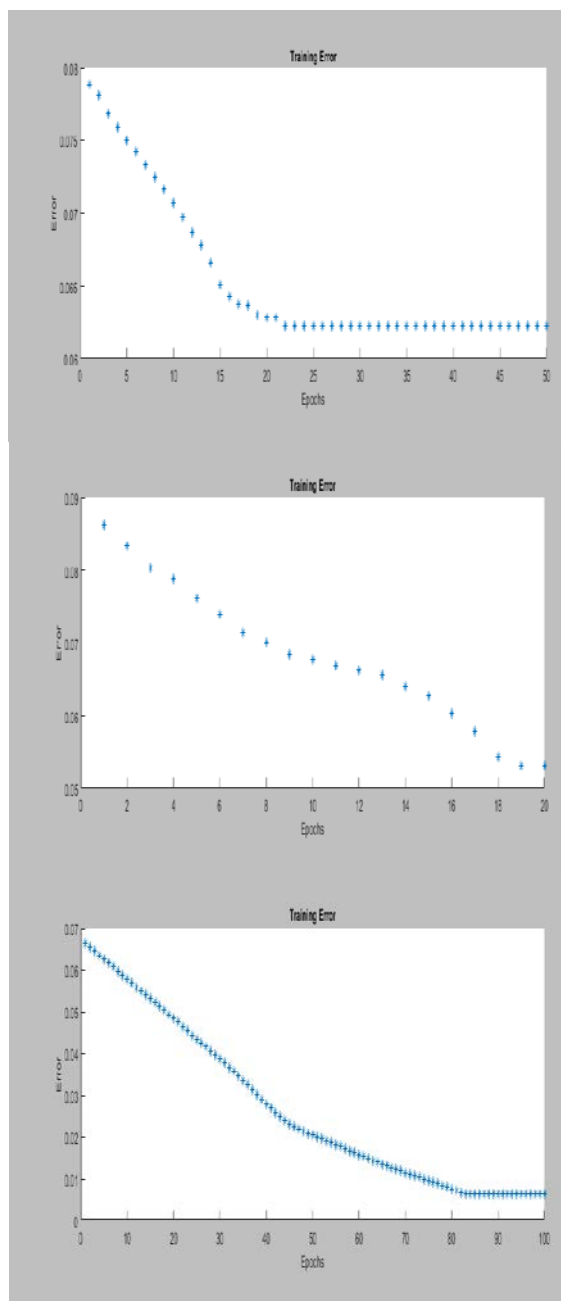


Figure 9. ANFIS training error of triangle, trapezoidal and bell-shaped as membership function

Table 3. Error rate

Epoch	Error		
	Triangle	Trapezoidal	Bell-shaped
1	0.0789	0.07678	0.06645
10	0.0707	0.06321	0.05775
20	0.0628	0.05665	0.04838
50	0.0623	0.05665	0.02036
100	0.0623	0.05665	0.00619

From Table 2, it can be seen that the error rate of using bell-shaped membership function is lesser from the starting if compared to triangular and trapezoidal function. The difference becomes more noticeable when epochs equal to 100. The error rate is only 0.006 (bell-shaped), which is lesser than 0.06 (triangular) and 0.05 (trapezoidal), and the results show that the bell-shaped function always gives the least error.

5. Conclusions:

An ANFIS model for the detection of bacteria in drinking water sources has been developed with 243 fuzzy set rules, and the predictive ability of the model is compared with three membership functions. The membership function changes automatically with every iteration during the model training. The results show that ANFIS with a generalized bell-shaped membership function is the most suitable membership function to model bacterial detection. The least error obtained at epoch 100 is 0.00619 by applying a bell-shaped function, through the testing data verification. ANFIS with bell-shaped membership function gives precisely the same output as experimental output.

References

1. Edition F 2011 Guidelines for drinking-water quality *WHO chronicle* **38(4)** 104.
2. Payus C, Haziqah N, Basri N, and Wan V.L 2018 Faecal bacteria contaminations in untreated drinking water (Groundwater Well and Hill Water) from Rural Community Areas 215.
3. Diarrhoeal disease *World Health Organization* 2017 Available Online at <http://www.who.int/mediacentre/factsheets/fs330/en>.
4. Cabral J.P.S 2010 Water microbiology Bacterial pathogens and water *International Journal of Environmental Research and Public Health* **7** 3657.
5. Payment P, Richardson L, Siemiatycki J, Dewar R, Edwardes M and Franco E 1991 A randomized trial to evaluate the risk of gastrointestinal disease due to consumption of drinking water meeting current microbiological standards *American journal of public health* **81(6)** 703.
6. Indian standard drinking water specification (Second Revision) *IS 10500:2012*.
7. Negnevitsky M 2005 *Artificial intelligence: a guide to intelligent systems*. Pearson education.
8. Bouharati S, Benmahammed K, Harzallah D and El-Assaf YM 2008 Application of artificial neuro-fuzzy logic inference system for predicting the microbiological pollution in fresh water *Journal of Applied Sciences* **8(2)** 309.
9. Okwu MO, & Adetunji O 2018 A comparative study of artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS) models in distribution system with nondeterministic inputs *International Journal of*

- Engineering Business Management* **10**
1847979018768421.
10. Keshavarz Z and Torkian H 2018 Application of ANN and ANFIS models in determining compressive strength of concrete *Soft Computing in Civil Engineering* **2(1)** 62.
 11. Calp MH 2019 A Hybrid ANFIS-GA Approach for Estimation of Regional Rainfall Amount. *Gazi University Journal of Science* **32(1)** 145.
 12. Kamali R and Binesh AR 2013 A comparison of neural networks and adaptive neuro-fuzzy inference systems for the prediction of water diffusion through carbon nanotubes. *Microfluidics and nanofluidics* **14(3-4)** 575.
 13. Azeez D, Ali MAM, Gan KB and Saiboon I 2013 Comparison of adaptive neuro-fuzzy inference system and artificial neural networks model to categorize patients in the emergency department *SpringerPlus* **2(1)** 416.
 14. Chandaran UD, Halim ZA and Sian LK 2012. Study on sulfate reducing bacteria detection using Adaptive Neuro-fuzzy Inference System In 2012 *IEEE International Conference on Circuits and Systems (ICCAS)* 59.
 15. World Health Organization 1993 *Guidelines for drinking-water quality* World Health Organization.
 16. MathWorks: Fuzzy Inference System Modeling – MATLAB & Simulink – MathWorks India, available at, last access: 14 October 2019.