A ClassicalWay of Finding Water Pollution by Using Artificial Neural Network

Venkata Rao, K
School of Computing Science and Engineering
VIT University, Vellore-632014, TamilNadu, India.

Swarnalatha P.
School of Computing Science and Engineering
VIT University, Vellore-632014, TamilNadu, India.

ABSTRACT

Vellore district is located within Tamil Nadu, India. The Vellore district waters were grouped into four types, namely four: class I, class II, class III and class IV. In this study, we aimed to introduce a method for classifying waters in the study area using some parameters such as Mg, pH, NO_3, SO_4,Ca,Cl, F and major ions by means of Artificial Neural Network (ANN) method. Preparing the data set for ANN method we are taken water samples from the area Vellore district. In each output value,the known water is represented by 1 and others by 0. A test data set of 15 samples are taken, these values are known buttheir group are unknown was prepared. And these inputvalues were run in ANN model in order to see how thewaters were grouped.

The advantages of artificial neural networks can beexploited to solve this problem. The most common ANNarchitecture is multilayeredperceptrons (MLP), which was used inthis study. For this solution, the ANN model using Back propagation(BP)algorithm has been successfully implemented.These resultsshow that thegroup in which the waters in the study area fall can bedetermined with high accuracy by some parameters ofwater the ion content of water. And these groups are formed by using Indian Standard Specifications for Drinking Water IS:10500.

Keywords:ANN, BP, water quality parameters, Vellore district, training samples.

1. Introduction

Water quality classification assessment is an important project in environmental science. Present situation is resolving of water pollution become more serious. Now a days, there are several methods to classify the water quality such as Gray correlation analysis method,gray clustering method, integrated pollution index method, Fuzzy Comprehensive Evaluation, Unascertained measure. Each and every method has its own characteristics.All these traditional methods are not able to solve the complicated nonlinear relationship between evaluations indicators and the grade of water quality. The traditional methods emerged some irrationality. Because there are many factors affecting the water quality, and the factors are nonlinear relations with water quality. Traditional assessment cannot meet the demands for evaluating precision. An ANN could simulate any non-linearity mapping relations in theory and the results are also proved more close to the reality. ANNs are loosely based on the neural structure of the brain which provides the ability to learn from the input data and then apply this to unknown data, in effect they can generalize and associate unknown data. Just due to the excellent characteristics, it has been applied widely in the field like pattern recognition, etc. And we can say, these ANNs are selforganizing, self-teaching, nonlinear and can deal with the systems which are difficult to describe with traditional methods[1][5][6][10]. In this model of ANN on water quality assessment, the network was trained by learning sample formed by Indian Standard Specifications For Drinking water IS:10500 [2]. After achieving convergence, the network could be used in assessing of the environment quality. Here the main objectives of this study is to construct model for classifying the water quality at Vellore district, TamilNadu, India by using the BP neural network.

2. Materials and Methods

Data, ground water quality standards and methodology are described in this section.

2.1 Site Description and Data

Vellore district is located within Tamil Nadu, India. The total length is 6,077 square kilometres. Here the population about 3,928,106 [3]. Figure1 shows the Vellore district map [4]. Especially a result of Vellore is having Golden Temple, VIT University and Vellore Fort.The rapid growth of tourism, industries, high-rise, low-rise buildings, andother infrastructures, had a significant effect onthe drinking water quality. In order to improve water quality, classification different levels of water quality becomes a major concern. The understanding of different levels of water quality can be utilized in water management and treatment systems.

In this study, water quality standards taken from Indian standard specifications for drinking water IS:10500 [2]. And we collected the mineral contents of water in Vellore district from department of water pollution control board TamilNadu.There are 50samples of data. Each sample consists of 7 parameters namely: Mg,pH,NO_3,SO_4, Ca, Cl, and F. The classifications of ground water quality are based on Indian Standard Specifications for Drinking Water IS:10500 [2].

2.2. Indian Standards For Drinking Water [2]

There are many parameters influencing the ground water quality. We are selected Seven parameters for the classifying the water. Indian standards for drinking water as in table 1 [2].

Table 1: Indian standards for drinking water [2].

pollutants	Desirable Limit	Permissible limit in the absence of alternate source
pH(mg/l)	6.5-8.5	No relaxation
Mg(mg/l)	30	100
NO_3(mg/l)	45	100
SO_4(mg/l)	200	400
Ca(mg/l)	75	200
Cl(mg/l)	250	1000
F(mg/l)	1.0	1.5

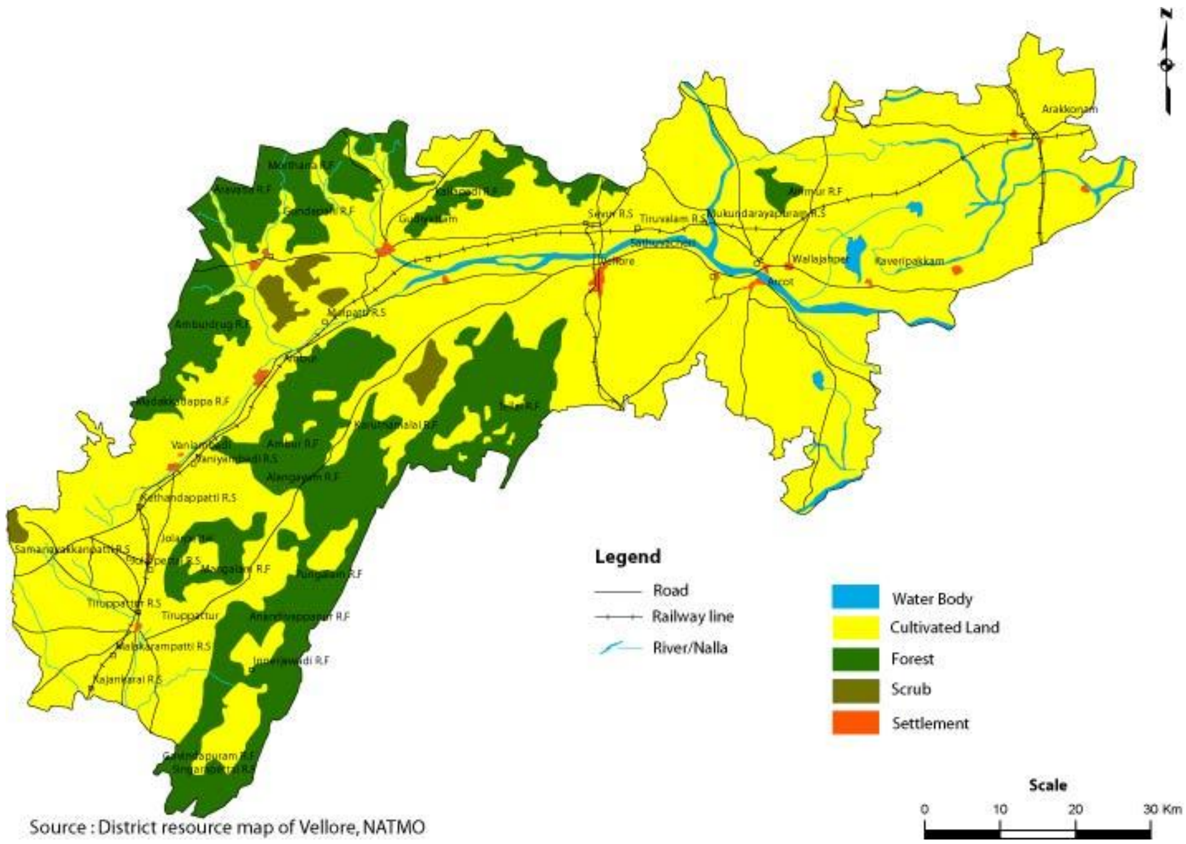

Figure 1:The map showing the Vellore District [4]

In Vellore, we classified the ground water quality as in Table 2. We are divided into four classes: class I, extraclean fresh ground water resources use forconservation that are not necessary to pass throughwater treatment processes and require only ordinaryprocesses for pathogenic destruction and ecosystemconservation where basic organisms can breednaturally; class II, medium clean fresh ground water resourcesuse for consumption, but are passed through anordinary treatment process before use; class III,fairly clean fresh ground water resources use forconsumption, but requires special water treatmentprocesses before use; and class IV, the sources whichare not within class I to class III and are used fornavigation.

Table 2: Ground water quality standards

Pollutants index	Class			
	I	II	III	IV
pH(mg/l)	6.5-8.5	6.5-8.5	6.5-8.5	6.5<
Mg(mg/l)	<30	30-60	60-100	>100
NO_3(mg/l)	<45	45-70	70-100	>100
SO_4(mg/l)	<200	200-300	300-400	>400

Ca(mg/l)	<75	75-130	130-200	>200
Cl(mg/l)	<250	250-600	600-1000	>1000
F(mg/l)	<1.0	1.0-1.5	1.0-1.5	>1.5

2.3 ANNs Methodology

ANNs are computer programsdesigned to process the information like human brain. ANNs gain their knowledge byfinding the patterns and relationships in data and learn through experience, not from programming. Basically ANNs consists three layers namely input layer, hidden layer layers and output layer. The inputs and weight are giving to ANN by using input layer. The ANN process the data by using the hidden layer or layers. There can be a any number of hidden layers. Figure 2 shows the BP ANN. The output of the ANN is produced by using the output layer. We can use a neural network to represent a nonlinear mapping between input and output vectors. Among the popular signal-processing technologies neural network are using widely [7][10]. Each neuron has weighted inputs, summation function, transfer function and one output. The behaviour of a neural network is determined by the transfer functions of its neurons, by the learning rule, and by the architecture itself. The weights are the adjustable parameters and, in that sense, a neural network is a parameterized system. The weighted sum of the inputs constitutes the activation of the neuron. The activation signal is passed through a transfer function to produce the output of a neuron. Transfer function introduces nonlinearity to the network. During training, the inter-unit connections are optimized until the error in predictions is minimized and the network reaches the specified level of accuracy. Once the network is trained, new unseen input information is entered to the network to calculate the output for test. ANN represents a promising modeling technique, especially for data sets having non-linear relationships that are frequently encountered in engineering [1][5][6][7][10].In each neuron, a specific mathematical function called the activation function accepts input from previous layers and generates output for the next layer. Figure2 shows the general form of multilayer perception [8]. In the experiment, the output function used is the sigmoidal transfer function which is defined as in equation (1) [9].

$$O_j(\bar{x},\bar{w}) = \frac{1}{1+e^{A_j(\bar{x},\bar{w})}} \qquad (1)$$

Where $A_j(\bar{x},\bar{w}) = \sum_{i=0}^{n} x_i w_{ji}$ where x_i is the inputs and w_{ji} is the weights.

The sigmoidal function is very close to one for large positive numbers, 0.5 at zero, and very close to zero for large negative numbers. This allows a smooth transition between the low and high output of the neuron (close to zero or close to one). We can see that the output depends only in the activation, which in turn depends on the values of the inputs and their respective weights [9].

Now, the goal of the training process is to obtain a desired output when certain inputs are given. Since the error is the difference between the actual and the desired output, the error depends on the weights, and we need to adjust the weights in order to minimize the error [9].

We can define the error function for the output of each neuron:

$$E_j(\bar{x},\bar{w},d) = \left(O_j(\bar{x},\bar{w}) - d_j\right)^2 \qquad (2)$$

We take the square of the difference between the output and the desired target because it will be always positive, and because it will be greater if the difference is big, and lesser if the difference is small. The error of the network will simply be the sum of the errors of all the neurons in the output layer: [9]

$$E(\bar{x},\bar{w},\bar{d}) = \sum_j (O_j(\bar{x},\bar{w}) - d_j)^2 \qquad (3)$$

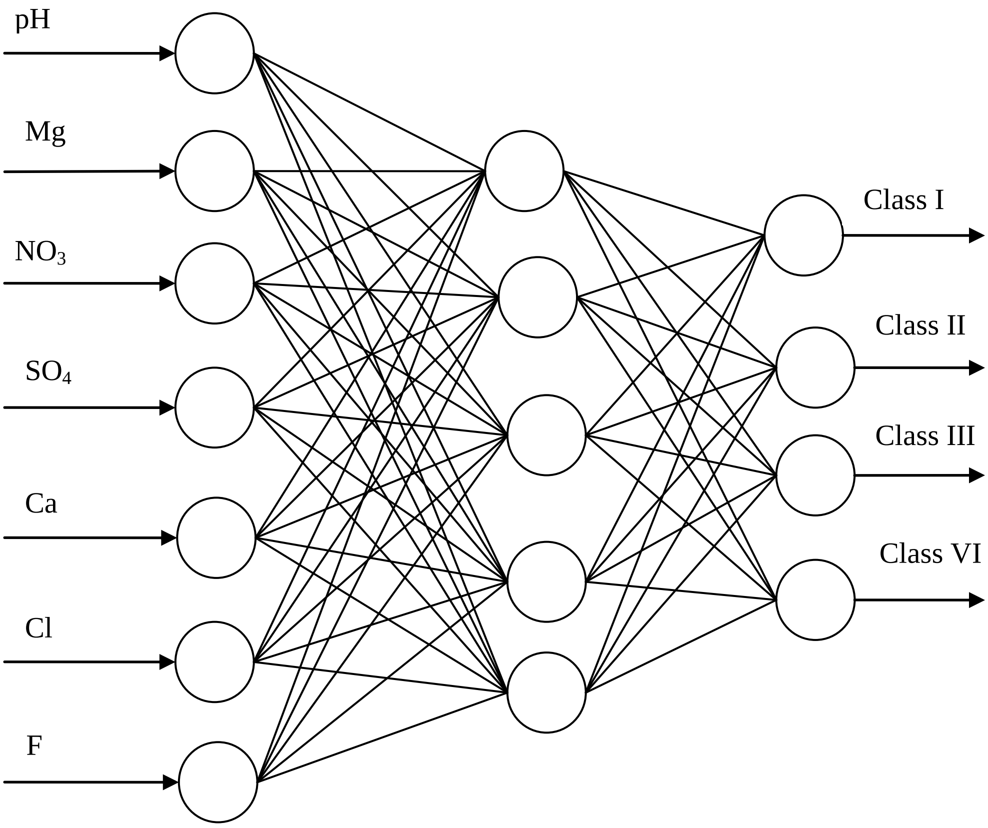

Figure 2: General form of multilayer perceptrons [7]

2.4 Back propagation algorithm

The back propagation algorithm now calculates how the error depends on the output, inputs, and weights. After we find this, we can adjust the weights using the method of *gradient descendent.[9]*

$$\Delta w_{ji} = -\eta \frac{\partial E}{\partial w_{ji}} \qquad (4)$$

This formula can be interpreted in the following way: the adjustment of each weightji(Δ w_{ji}) will be the negative of a constant eta (η) multiplied by the independence of theprevious weight on the error of the network, which is the derivative of E in respect to w_i.The size of the adjustment will depend on η, and on the contribution of the weight to theerror of the function. This is, if the weight contributes a lot to the error, the adjustment willbe greater than if it contributes in a smaller amount. (4) is used until we find appropriateweights (the error is minimal). So, we "only" need to find the derivative of E in respect to w . This is the goal of the BP algorithm, since we need to achieve this backwards. First, we need to calculate how much the error depends onthe output, which is the derivative of E in respect toO_j (from (2)) [9].

$$\frac{dE}{dOj} = 2(O_j, d_j) \qquad (5)$$

And then, how much the output depends on the activation, which in turn depends on the weights (from (1)):

$$\frac{dO_j}{dw_{ji}} = \frac{dO_j}{dA_j}\frac{dA_j}{dw_{ji}} = \mathbf{O_j(1 - O_j)x_i} \qquad \mathbf{(6)}$$

And we can see that (from (5) and (6))

$$\frac{dE}{dw_{ji}} = \frac{dE}{dO_j}\frac{dO_j}{dw_{ji}} = 2(O_j - d_j)O_j(1 - O_j)x_i \quad (7)$$

And so, the adjustment to each weight will be (from (4) and (8)):

$$\Delta w_{ji} = -2\eta(O_j - d_j)O_j(1 - O_j)x_j \quad (8)$$

We can use (8) as it is for training an ANN with two layers. Now, for training thenetwork with one more layer we need to make some considerations. If we want to adjustthe weights (let's call them v_{ik}) of a previous layer, we need first to calculate how the errordepends not on the weight, but in the input from the previous layer. This is easy, we wouldjust need to change x_i with w_{ji} in (6), (7), and (8). But we also need to see how the error ofthe network depends on the adjustment of v_{ik}. So:[9]

$$\mathbf{\Delta} v_{ik} = -\boldsymbol{\eta}\frac{dE}{dv_{ik}} = -\boldsymbol{\eta}\frac{dE}{dx_i}\frac{dx_i}{dv_{ik}} \qquad (9)$$

Where

$$\frac{dE}{dw_{ji}} = \mathbf{2}(O_j - d_j)O_j(1 - O_j)w_{ji}\mathbf{(10)}$$

And, assuming that there are inputs u_k into the neuron with v_{ik} (from (6)):

$$\frac{dx_i}{dv_{ik}} = x_i(1 - x_i)v_{ik} \qquad (11)$$

If we want to add yet another layer, we can do the same, calculating how the error depends on the inputs and weights of the first layer. We should just be careful with the indexes, since each layer can have a different number of neurons, and we should not confuse them.

For practical reasons, ANNs implementing the backpropagation algorithm do not have too many layers, since the time for training the networks grows exponentially. Also, there are refinements to the backpropagation algorithm which allow a faster learning.

3. Classifying Water Quality In Vellore District As a Case Study

In the ANN model, the inputs are Mg (mg/l), pH, NO_3 (mg/l), SO_4 (mg/l), Ca (mg/l), Cl (mg/l) and F(mg/l). Four output values were used. The outputs are Class I, Class II, Class III and Class IV water. In each output values the known waterrepresented by 1 and others by 0.

In this ANN model, out of 75 data sets generated, 50 datasets were used for training (Table 2) and the rest were used for testing (Table 3) [11] the network. A set of random valuesdistributed uniformly between -0.1 and +0.1 was used toinitialize the weights of the networks. HoweverAfter several trials, it was found that threehidden layers network achieved the task with highaccuracy.

Table 2:Data set for training [11]

Rank	SITENAME	Input							output Class			
		pH	Mg	NO_3	SO_4	Ca	Cl	F	I	II	III	IV
1	Vengalapuram	7.28	267.95	276	304	16	1304	1.64	0	0	0	1
2	Kavalur1	7.18	94.85	2	12	24	163	1.34	1	0	0	0
3	Alangayam	7.48	226.14	16	150	20	631	0.99	0	1	0	0
4	Vaniyambadi1	7.34	211.55	140	264	20	680	1.65	1	0	0	0
5	Sangilikuppam	7.22	238.3	174	240	24	305	1.21	1	0	0	0

6	Thottalam	7.17	350.14	350	352	28	709	1.02	0	0	1	0
7	Anaikkattu	7.07	37.72	12	34	26	64	1.00	1	0	0	0
8	Arcot2	7.35	53.53	86	140	36	354	1.41	1	0	0	0
9	Vellore1	8.11	59.71	24	3	114	574	0.83	1	0	0	0
10	Vellore2	7.35	82.69	59	152	20	127	1.50	1	0	0	0
11	Ranipet1	7.21	306.4	57	490	48	2183	1.61	0	0	1	0
12	Banavaram	7.06	72.99	16	60	40	191	1.12	1	0	0	0
13	Sholingar1	6.82	40.16	13	28	34	99	0.47	1	0	0	0
14	Perambattu1	7.6	54.33	18	68	90	96	0.03	1	0	0	0
15	Perambattu2	7.66	105	5	130	38	340	3.20	1	0	0	0
16	Alankuppam1	7.3	241.68	24	162	307	1761	0.02	0	1	0	0
17	Alankuppam2	7.46	153.35	30	200	144	759	1.68	0	1	0	0
18	Vengalapuram	7.6	238.36	25	369	72	1660	1.74	0	0	1	0
19	Vokkanampet	7.25	170.53	20	160	280	1319	1.28	0	1	0	0
20	Kodiyur	7.58	75.48	5	50	86	238	1.10	1	0	0	0
21	Kavalur2	7.3	73.25	4	48	58	114	0.26	1	0	0	0
22	Kavalur3	7.8	40.19	2	30	64	191	1.48	1	0	0	0
23	Natrampalli	7.56	94.9	0	200	62	532	2.28	1	0	0	0
24	Alangayam1	7.4	88.01	0	66	194	341	0.28	1	0	0	0
25	Alangayam2	7.36	79.17	5	130	122	415	1.40	1	0	0	0
26	Vaniyambadi1	7.8	136.3	6	72	120	568	0.46	0	1	0	0
27	Vaniyambadi2	7.61	176.35	5	320	66	837	2.70	0	0	1	0
28	Sangilikuppam1	7.8	105.33	12	126	147	210	0.02	0	1	0	0
29	Sangilikuppam2	7.43	96.2	0	220	138	319	1.32	1	0	0	0
30	Kalavai	6.9	12.19	1	4	30	67	0.09	1	0	0	0

Table 3:Data set for training [continue][11]

31	Odugathur	7.4	180.12	80	180	168	980	1.26	0	1	0	0
32	ambur	7.37	34.33	0	220	240	957	1.12	1	0	0	0
33	Thottalam	7.45	214.07	10	360	96	723	1.42	0	0	1	0
34	Machampattu	7.59	105	30	160	116	695	1.12	0	1	0	0
35	Anaikkattu1	7.3	74.57	4	58	43	114	0.12	1	0	0	0
36	Anaikkattu2	7.5	36.53	2	70	50	138	1.28	1	0	0	0
37	P.Agaram	7.66	87.62	10	240	80	539	1.76	0	0	1	0
38	Arcot1	7.45	24.38	0	164	56	419	0.96	1	0	0	0
39	Arcot2	7.84	19.57	10	260	96	610	2.00	1	0	0	0
40	Pallikonda	7.5	113.66	2	132	94	355	0.03	0	1	0	0
41	Devadanam1	8.1	75.64	8	524	220	298	0.31	1	0	0	0
42	Devadanam2	7.3	3.87	5	90	180	390	0.82	1	0	0	0
43	Vellore3	7.3	148.51	2	318	168	383	0.16	0	0	1	0
44	Vellore4	7.55	19.63	10	340	146	369	1.68	1	0	0	0
45	Ranipet1	8	150.99	3	492	208	1775	0.36	0	1	0	0
46	Ranipet2	7.43	136.41	20	500	208	2340	2.06	0	1	0	0
47	Pernampet	7.58	18.29	4	40	46	124	1.20	1	0	0	0
48	Veppaneri	7.45	69.37	2	80	64	238	1.28	1	0	0	0
49	Lalpet1	7.6	22.47	0	42	80	128	0.25	1	0	0	0
50	Lalpet2	7.55	9.86	10	120	110	230	0.90	1	0	0	0

Table 4:Data set for testing [11]

Rank	SITENAME	Input							Output Class			
		pH	Mg	NO_3	SO_4	Ca	Cl	F	I	II	III	IV
1	Vengalapuram	7.61	176.35	5	320	66	837	2.70	0	0	1	0
2	Kavalur	7.8	105.33	12	126	147	210	0.02	0	1	0	0
3	Alangayam	7.43	96.2	0	220	138	319	1.32	1	0	0	0
4	Vaniyambadi1	6.9	12.19	1	4	30	67	0.09	1	0	0	0
5	Sangilikuppam	7.4	180.12	80	180	168	980	1.26	0	1	0	0
6	Thottalam	7.37	34.33	0	220	240	957	1.12	1	0	0	0
7	Anaikkattu	7.45	214.07	10	360	96	723	1.42	0	0	1	0
8	Arcot2	7.59	105	30	160	116	695	1.12	0	1	0	0
9	Vellore1	7.3	74.57	4	58	43	114	0.12	1	0	0	0
10	Vellore2	7.5	36.53	2	70	50	138	1.28	1	0	0	0
11	Ranipet1	7.66	87.62	10	240	80	539	1.76	0	0	1	0
12	Banavaram	7.45	24.38	0	164	56	419	0.96	1	0	0	0
13	Sholingar1	7.84	19.57	10	260	96	610	2.00	1	0	0	0
14	Kodiyur	7.5	113.66	2	132	94	355	0.03	0	1	0	0
15	Ammur-EW	8.1	75.64	8	524	220	298	0.31	1	0	0	0

4. RESULTS AND CONCLUSIONS

The Vellore ground water is classified into four groups, namely class I, class II, class III and class IV. The classification is impossible employing only chemical analysis results of any water in the ground.However, in this study, it has been shown this can be possible employing ANN. For this purpose, we tried different MLP training algorithms, and among them back propagation algorithm has given good results. This paper adopts the BP neural network to evaluate the water quality at Vellore district.

Some conclusions can be got by our case study:

(1) BP neural network can be used to evaluate the water quality. The results are objective.

(2) The result of evaluation through BP neural network has high precision.

(3) The assessment result can provide reference to the water environment protection and plan.

5. Acknowledgment

We would like to thank Dr Venkata Krishna P, Dr Ganapathy P, Prof .VenkatesanM, and DrSantha Kumar, VIT University, for their support in completing this paper.

6. References

[1] Miao Qun; Li Yue; Hai Yang; He Hui; Zhang Xiaomei; The 2nd international conference on " Back Propagation Neural Network in the Water Quality Evaluation Qingdao Dagu River ".Bioinformatics and Biomedical engineering, 2008.ICBEE 2008.

[2] BIS Indian Standard Specification for Drinking Water "IS:10500",1992.

[3] Population information accessed from "http://www.census2011.co.in"

[4] Map accessed from "http://www.mapsofindia.com"

[5]International Journal of the Physical Sciences Vol. 6(1), pp. 43-51, 4 January, 2011.
Available online at http://www.academicjournals.org/IJPS

[6] Ali FerhatBayram, SeyfettinSinanGültekin and Ali RızaSöğüt "Specifications of thermal waters and their classification on the base of neural network method: Examples from Simav geothermal area, Western Turkey". International Journal of the Physical SciencesVol. 6(1), pp. 43–51, 4 January, 2011.

[7] L.Fausett, *"Fundamentals of Neural Networks Architectur Algorithms and Applications"*, Pearson Prentice Hall, USA, 1994.

[8] SirilakAreerachakul, SiripunSanguansintukul, "Classification and Regression Trees and MLP Neural Network to Classify Water Quality of Canals in Bangkok, Thailand" International Journal of Intelligent Computing Research (IJICR), Volume 1, Issue 1/2, March/June 2010

[9] Accessed from http://unam.academia.edu/CarlosGershenson/Papers/226665/Artificial_Neural_Networks_for_Beginners

[10] Li-hua Chen, and Xiao-yun Zhang, "Application of Artificial Neural Network to Classify Water Quality of the Yellow River", *Journal 0f Fuzzy Information and Engineering*, Springer-Verlag,pp. 15-23, Jan 2009.

[11]Accessed from cpcbenvis.nic.in

Authors profile:

Author A. SwarnalathaPurushotham is an Assistant Professor (Selection Grade), in the school of computing sciences and engineering, VIT University, at Vellore, India. She is pursuing her Ph.D degree in Intelligent Systems. She has published more than 20+ papers in international journals/international conference proceedings. She is having 11 years of teaching experiences. She is a member of ACM, IEEE(90747121),IACSIT, CSI, IACSIT,IEEE (WIE). Her current research interest includes Image Processing, Remote Sensing and Artificial Intelligence.

Author B. Kondeti venkataraois pursuing M-tech degree, in computer scienc and engineering, VIT University, at vellore, India.