International Journal of Engineering Research and General Science Volume 3, Issue 1, January-February, 2015
ISSN 2091-2730

# Application of ANN model for the prediction of Water Quality Index

SUJANA PRAJITHKUMAR[1] , Dr. SHWETA VERMA[2], B V MAHAJAN[3]

[1]Lecturer, Rajarshi Shahu College of Engineering ,Tathawade

[2]Assistant Professor, DY Patil College of Engineering, Akurdi, shweta_verma1@rediffmail.com, Phn No.: +91-9405580629

[3]Assistant Professor, Rajarshi Shahu College of Engineering ,Tathawade

**Abstract** - Water is the prime requirement for the existence of life. The unbridled exploitation of water for irrigation, drinking and industrial purposes has caused a drastic decline of the quality and availability of water. The over-exploitation of limited resources has not only caused a perceptible decline in the water table, but also resulted in the enormous increase of pollutants concentration. The ever growing population exerts a great pressure on this resource. The never ending growth of population and ill-planned exploitation of the water resource created a situation, where the very survival of man has become endangered. Assembling the various parameters of the water quality data into one single number leads an easy interpretation of data, thus providing an important tool for management and decision making purposes. The purpose of an index is to transform the large quantity of data into information that is easily understandable by the general public. Water quality index exhibits the overall water quality at a specific location and specific time based on several water quality parameters. WQI is a set of standards used to measure changes in water quality in a particular river reach over time and make comparisons from different reaches of a river. A WQI also allows for comparisons to be made between different rivers.

The present study is concentrated on the prediction of WQI using ANN model. Pavna river originates south of Lonavala from western ghatts and flows a total of nearly 60Km to meet the Mula river in Pune, Maharashtra, India. Pavna river flows through Pune city covering Pimpri Chinchwad Area. Pimpri –Chinchwad area is developing area and due to industrialization and high population growth the river is getting polluted. The data for the prediction of water quality index were collected from the water treatment Plant, Nigdi – sector 21.The collected data were for the point Ravet Intake.(2001-2014)(monthly).In the present study it is found that the recurrent neural network give better results as compared to the radial multilayer neural network. R value for the multilayer neural net work is 0.960 .

**Keywords—** Water Quality Index (WQI), Artificial neural network, recurrent neural network, multilayer neural network , Pavna River ,conjugate function

## 1. INTRODUCTION

Water is the prime requirement for the existence of life and thus it has been man's endeavour for the immemorial to utilize the available resources. The growing population exerts a great pressure on this resource. The increased population growth and ill planned exploitation of the water resources created a situation, where the very survival of man has become endangered[1].

Interpretation of complex water quality data is difficult to understand. Also it is difficult to communicate during decision making process. Assembling the various parameters of the water quality data into one single number is essential for decision making purpose. The purpose of an index is to transform the large quantity of data into information that is easily understandable by the general public. Water quality index exhibits the overall water quality at a specific location and specific time based on several water quality parameters. The index result represents the level of water quality in a given water basin, such as lake, river or stream.[2] WQI is a set of standards used to measure changes in water quality in a particular river reach over time and make comparisons from different reaches of a river. A WQI also allows for comparisons to be made between different rivers. This index allows for a general analysis of water quality. It also helps for the analysis of water many levels that affect a stream's ability to host life and to check whether the overall quality of water bodies poses any potential threat to various uses of water. A computer programming using MATLab had been used for training and testing of the ANN .Once the a number of training trials are completed ,the best neural network model was generated. The maximum number of epochs was chosen by a trial and error approach. Trial and error method was used to find the most suitable network model for the WQI analysis using the different ten parameters. The result of WQI was valid in respect of experimental and ANN model.

In recent years the trend has been to use statistical method instead of traditional and domestic methods to forecasting WQI. Lee Yoot Khan et al conclude that the modular neural network was found to be the most suitable model for the termination of the WQI[2].A A Mansur et al predicted the dissolved oxygen in Surma river using fed forward neural network and radial basis function neural network[3]. Both the methods provided better results. Sundarambal et al concluded that the ANN model can be used even for sea water quality forecasting[4]. In this study the modular neural network and radial basis function models of ANN are used to forecast the WQI of pavna river at ravet intake.

## 2. STUDY AREA

The Pavna river in Pune, Maharashtra, India is selected as the study area for the ANN application. Pavna river originates south of Lonavala from western ghatts and flows a total of nearly 60Km to meet the Mula river in Pune. Pavna river flows through Pune city covering Pimpri Chinchwad Area. Pimpri –Chinchwad area is developing area and due to industrialization and high population growth in thease area the river is getting polluted. The study area is Ravet intake from where water is collected and treated in water treatment plant in Nigidi.

## 3. METHODOLOGY

There are different methods for the determination of water quality Index. In this particular study the methods used are given below:

1. Weightage Rating method
2. Prediction of WQI by using ANN

Weigtage Rating Method:

Factors which have higher permissible limits are less harmful because they can harm quality of river water when they are present in very high quantity. So weightage of factor has an inverse relationship with its permissible limits.

Therefore Wi α(1/ Sn)Or (1)
Wi = K/Sn
Where, K = constant of proportionality
Wi = unit weight of all chemical factor
Vs or Sn = Standard alue of $i^{th}$ parameter
Values of K were calculated as: $K = \frac{1}{\sum_1^7 1/_{Vs}}$

$$\sum_1^7 \frac{1}{Si} = \left(\frac{1}{SiPH}\right)+\left(\frac{1}{Si(EC)}\right)+\left(\frac{1}{Si(TDS)}\right)+\left(\frac{1}{Si(DO)}\right)+\left(\frac{1}{Si(nitrates)}\right)+\left(\frac{1}{Si(Alkalinity)}\right)+\left(\frac{1}{Si(Total\ Hardness)}\right)+\cdots\dots\dots\dots\dots.+\left(\frac{1}{Si(Calcium)}\right)+\left(\frac{1}{Si(calcium)}\right) \quad (2)$$

The weightage of all the factors were calculated on the basis of the above equation.

WQI =∑ Wi x Qi/∑Wi (3)

Wi x Qi = Wi(pH) x Qi(pH) + Wi(EC) + Qi(EC)+ Wi(TDS) xQi(TDS) + Wi(DO) x Qi(DO) + . ............+ Wi(Calcium) x Qi (Calcium) (4)

Qi = 100[(Va-Vi)/(Vs-Vi)] (5)

= Rating scale

Va =average measured values of water sample at study area.
Vi =standard value of $i^{th}$ parameter
Vs =ideal value for pure water (0 for all parameters except pH and DO)
∑Wi - total unit weight of all chemical factors. Using the water quality index, all the samples were categorized into the following five classes: excellent (0 - 25), good (26 - 50), moderately polluted (51 - 75), severely polluted (76-100) and unfit for human consumption(above 100) based on their suitability.[5]

TABLE 1 The permissible values of various pollutants for drinking water(expressed in mg/l except ph and ec recommended by Indian Standards and CPCB Standard[2]

| Sl No | Parameters | CPCB/IS | IS (10500) |
|---|---|---|---|
| 1 | PH | 6.5-8.5 | 6.5-8.5 |
| 2 | EC | <300** | --- |
| 3 | Turbidity | - | 5 |
| 4 | TDS | <500 | <500 |
| 5 | Total Alkalinity | 200 | 200 |
| 6 | Total hardness | 300 | 300 |
| 7 | DO | 6 | - |
| 8 | Chloride | 250 | 250 |
| 9 | Nitrate | 20 | 45 |
| 10 | Calcium | 75 | 75 |

## Pre-processing of data

At the initial stages of the project, real environmental data (fortnight)on the condition of Pavna river over a period of time 10 years beginning from 2001 to 2014 were acquired from water treatment plant sector 21 Nigidi, Pune. These make a total of 97. Out of these data 70 datas are set for training and remaining 27 data are set for testing. Trial and error method is tried for the different combinations of the data.

## Training of data

The network architecture for WQI consists of input nodes and output node. In the present study it consists of ten input nodes and one output node. The parameters of choices as input must have an influence on the desired output. For the current project the ten water pollution parameters are selected to be as input nodes. These input and output nodes are used to produce ANN model. The input variables are DO concentration, concentration of Nitrates, concentration of Calcium, Total Suspended Solids concentration, concentration of Total hardness, concentration of Alkalinity, concentration of Chlorides, Total Hardness concentration. All the eight parameters are measured in milligram per litre. The remaining three parameters are pH, turbidity and conductivity.

## Artificial neural network

Artificial Neural networks are non-linear mapping structure which is inspired by the observed process of natural networks of biological neurons in brain. It consists of simple computational units called neurons. These are highly inter connected. ANNs become very popular nowadays because of their wide range of applicability and the ease with which it can treat complicated problems.[6]

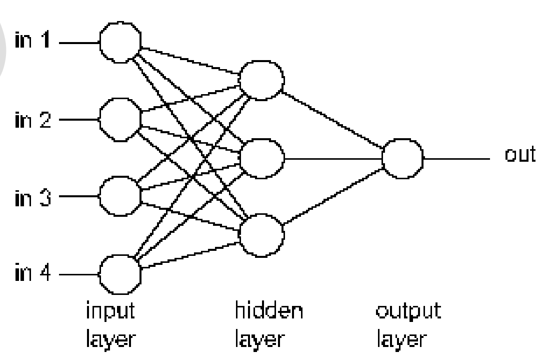


Fig 1 The schematic representation of a artificial neural network

### 3.1 Development of ANN Model

ANNs are constructed with layers of units, and hence it is termed as multilayer ANNs. The first layer of multilayer ANN is known as input units. In statistical nomenclature input units are known as independent variables. Last layer is called output units. In statistical nomenclature these are known as dependent variables or response variables. All the other units in the model is known as hidden units and it includes hidden layers. There are two function to achieve the behaviour of a unit in a particular layer, which normally are the same for all units within the whole ANN.Fig 2 shows the mathematical representation of Neural Network.

Input function and Output function

Input into a node is a weigted sum of outputs from nodes connected to it. The input function is given by the equation:

$Net_i = \Sigma W_{ij} X_j + \mu_i$ (6)

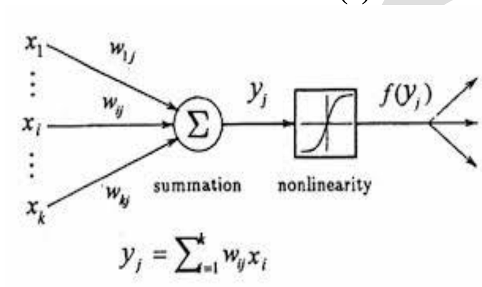


fig 2 Mathematical representation of neural network

Where Neti describes the result of the net inputs $x_i$ (weigted by the weights $w_{ij}$)impacting on unit i.Also wij are weights connecting j to neuron i ,x j is output from unit j and $\mu_i$ is the threshold for neuron i, Threshold term is baseline input into to a node in absence of any other input. If a weight $w_{ij}$ is negative it is termed inhibitory because it decreases net input ,otherwise it is called excitatory. Each unit takes its net input and applies an activation function to it. A number of nonlinear functions have been used in the literature as activation function. The threshold function is useful in situation where the input and outputs are binary encoded.most common choice in activation function is sigmoid function[6] ,such as

$g(netinput) = [1+ e^{-netinput}]^{-1}$ (7)

$g(netinput) = tanh(netinput)$ (8)

### 3.2 Multilayer percepton

The network was constructed by interconnecting several neural objects called as components. Multilayer perceptrons (MLPs) are layered feedforward networks typically trained with static backpropagation. These networks have found their way into countless applications requiring static pattern classification. Their main advantage is that they are easy to use, and that they can approximate any input/output map. The key disadvantages are that they train slowly, and require lots of training data (typically three times more training samples than network weights).Fig 3 shows the schematic diagram of Multilayer perceptron Network

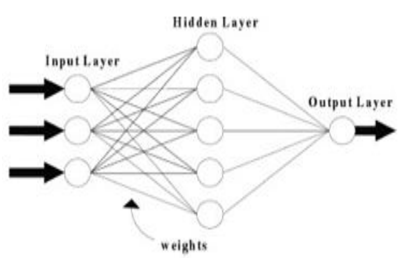


Fig 3 Schematic diagram of Multilayer perceptron Network

International Journal of Engineering Research and General Science Volume 3, Issue 1, January-February, 2015
ISSN 2091-2730

### 3.3 Recurrent Neural Network

A feed forward neural network contains one or more hidden layers and at least one feed back loop is known as recurrent network. In recurrent neural network the out put network will fed back may be a self feedback. The output of neutron is fed back to its own input. Schematic diagram of the recurrent network is given in fig 4 shows the schematic diagram of Recurrent Neural Network.

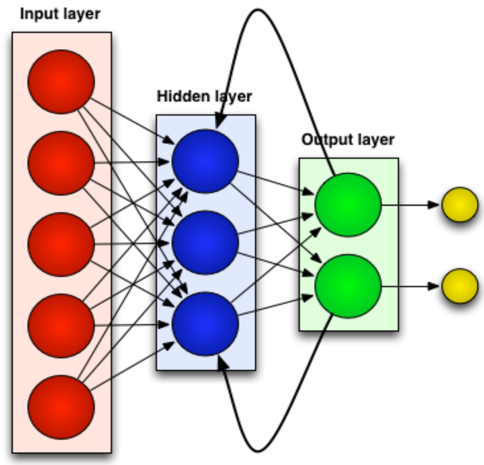


*Fig 4Schematic diagram of recurrent neural network*

## 4. RESULT AND DISCUSSION

The ANN models were trained using number of neurons in the hidden layers and for training and testing. After number of trials by changing the length of data set for training ,testing and also changing the number of neurons in the hidden layer the maximum value of R was obtained. The Multilayer perceptron consists of sigmoid as transfer function and conjugate as learning rule. In case of recurrent neural network TanAxon as transfer function and LevenbergMarqua is used as a learning rule. The number of epochs was set to 1000 to 5000 throughout the trial and error process. The number of ANN models were developed using Multilayer Neural Network and Recurrent Neural Network. The hidden layer is taken to be between 1-2.The training and testing data set an epochs were varied as shown in table 2 in order to achieve better accuracy.

Among the selected two ANN models the developed ANN model with Recurrent Neural Network model stimulated the WQI of pavna river with great accuracy when compared with Multilayer neural network architecture. The maximum value of R was 0.960, MSE=62.763and MAE=5.030 as shown in table 2 . In this case the value of R obtained is quite high (0.960) and reliable for short term prediction. The observed and modelled WQI values in whole array using Multilayer neural network and Recurrent Neural Network is illustrated in fig 5a and 5b and fig.6a and fig 6b which indicates that the modelling has been quite successful.

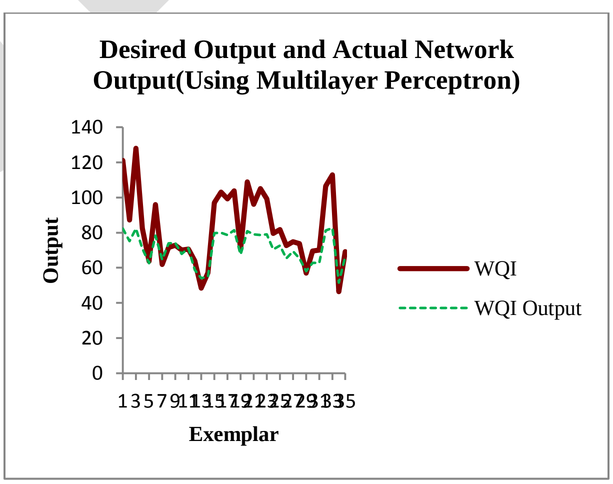


Fig5.a The observed and modelled WQI values for Multilayer Neural Network

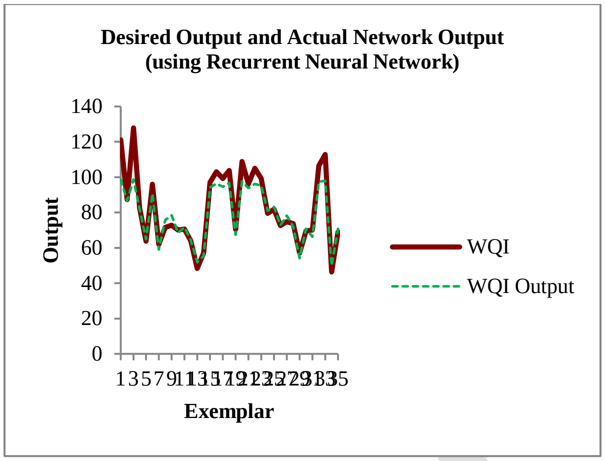


Fig5.b The observed and modelled WQI values in recurrent neural Network

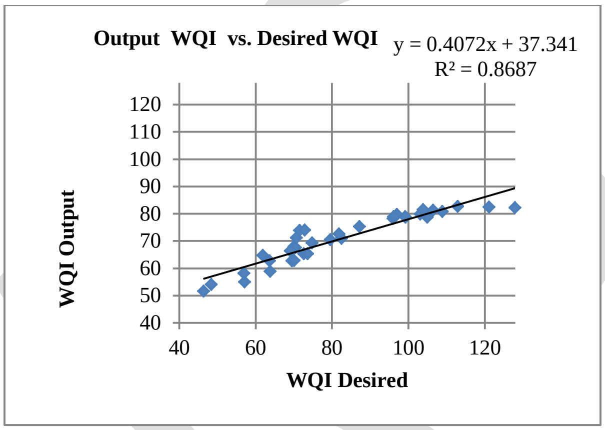


Fig 6a Scatter plot of observed verses modelled WQI for Multilayer Neural Network

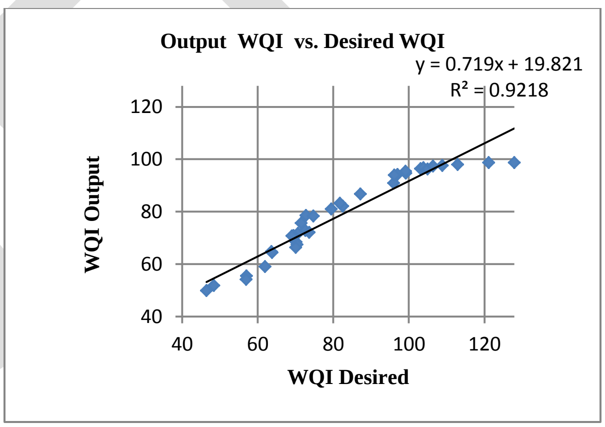


Fig 6.b Scatter plot of observed verses modelled WQI for recurrent neural Network

TABLE 2 SUMMARY OF ANN MODEL

| Model | Training data% | Testing data% | Epoch | R | MSE | MAE |
|---|---|---|---|---|---|---|
| r perceptro n | 70 | 30 | 2000 | 0.761 | 247.562 | 12.019 |
| | 70 | 30 | 1000 | 0.891 | 237.959 | 12.219 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | 50 | 50 | 5000 | 0.891 | 148.960 | 10.452 |
| | 60 | 40 | 1000 | 0.932 | 292.604 | 12.716 |
| | 60 | 40 | 5000 | 0.896 | 154.784 | 9.794 |
| Recurrent neural Network | 70 | 30 | 5000 | 0.700 | 392.234 | 15.994 |
| | 70 | 30 | 1000 | 0.917 | 107.103 | 7.799 |
| | 70 | 30 | 1000 | 0.934 | 79.610 | 6.737 |
| | 50 | 50 | 5000 | 0.920 | 78.623 | 7.013 |
| | 60 | 40 | 1000 | 0.944 | 167.329 | 10.744 |
| | 60 | 40 | 5000 | **0.960** | 62.763 | 5.030 |

The Recurrent Neural Network with all input parameters are found to be the most appropriate model for WQI prediction with high correlation co-efficient of 0.960 and a mean square error (MSE) value of 62.763.

## 5. CONCLUSION

In this model ANN models were developed to predict the WQI in Pavna river at Ravet intake. The proposed model shows efficiency in forecasting the WQI in water bodies. The result showed that the modular network model prepared by different ten water quality parameter provided high R(0.960) value. It has been observed that the WQI of Pavna river can be predicted using both with acceptable accuracy using Recurrent Neural Network .

## REFFERENCES:

[1] Maheshkumar .Akkaraboyina [1],prof B.S.N.Raju[2],A Comparative study of water quality indices of river Godavari, *Journal* IJRD- eISSN : 2278-067X, pISSN : 2278-800 Volume 2, Issue 3 (July 2012), PP. 29-34

[2] Lee Yoot Khauan,Noraliza Hamzah ,and Rozita Jailani, Prediction of Water quality Index (WQI)based on Artificial Neural Network.(ANN) *Journal* 2002 IEEE 157-161

[3] A A Masrur Ahemed ,Prediction of dissolved oxygen in Surma River influenced by biochemical oxygen demand and chemical oxygen demand using Artificial Neural Network,*Jounal of King Saud UNIVERSITY*,-ENG.SCI(2014)xxx,xxx-xxx

[4] Sunarambal Palani, Shie-Yui Liong,Pavel Tkalich,An ANN application for water quality forecasting, *Marine Pollution Bulletin*56(2008)1586-1597

*[5]* Girish kumar Jha ,Artificial Neural Networks ,*Indian Agricultural Research Institute .New Delhi.*

[6] Maheshkumar .Akkaraboyina, Prof B.S.N.Raju "A comparative study of water quality indices of river Godavari,"

[7] S. Kalavathy1, T. Rakesh Sharma1, P. Sureshkumar2* Water Quality Index of River Cauvery in Tiruchirappalli district, Tamilnadu, *Journal* ARCH. ENVIRON. SCI. (2011), 5, 55-61

[8] Papita Das Saha ,R .Sengupta ,Jhuma Saha ,P K Banerjee Assessment of water quality characteristics of River GAnga At Kolkata Region ,India using Water quality Index and ANN simulation method. *Journal* ARCH...ENVIRON.SCI.(2012),6,34-41

[9] Ani1 K. Jain K.M. Mohiuddin Artificial Neural Network :A tutorial , ,0018-962/96@1996 IEEE

[10] Mohd Fahmi Mohd Nasir , Hafizan Juahir, Nuraainaa Roslan,"Artificial Neural Networks Combined with Sensitivity Analysis as a Prediction Model for Water Quality Index in Juru River", MALAYSIA, , IJEP Vol.1 No. 3 2011 PP.1-8

[11] Ashwani Kumar, Water quality index for assessment of water quality of river Ravi at Madhopur (India) Standard methods for the examination of water and waste water by APHA