Water quality modelling of the River Ganga using artificial neural network with reference to the various training functions

Anil Kumar Bisht [1] ✉, Ravendra Singh [2], R. Bhutiani [3], Ashutosh Bhatt [4] and Krishan Kumar [5]

Abstract

The River Ganga (2,525 km long) is the largest River basin in India, covering 26.2 percent of India's total geographical area and recently granted living entity status by the court. It is the holiest River and also among the dirtiest in the world. That's why it is mandatory to maintain its water quality (WQ). Though, monitoring and assessment of WQ of a River is a very challenging task. In this research work, Soft Computing (SC) based popular and commononly used Artificial Neural Network (ANN) technique has been used for modelling the WQ of the Ganga River by developing a prediction model based on six different training functions. Five sampling stations along this River stretch were selected from DEVPRAYAG to ROORKEE in the Uttarakhand state of India. The monthly data sets of five water quality parameters temperature, pH, dissolved oxygen (DO), biochemical oxygen demand (BOD) and total coliform (TC) for the time period from 2001 to 2015 have been taken. The feed forward error back propagation neural network method has been used to develop the WQ-prediction model by conducting various experiments following a neural network structure of 5-10-1, 0.1 as a training goal and various training functions. Using the Mean square error (MSE) statistical method the prediction performance of the developed model was evaluated. The model developed with traincgp (Conjugate Gradient with Polak-Ribiere Restarts) comes out to be the worst one (MSE=0.786) while the other model with trainlm (Levenberg-Marquardt backpropagation) rule proved to be the best one (MSE=0.163) among others. Consequently, it is found that ANNs are capable of predicting WQ of the River Ganga with acceptable results.

Introduction

Water is the precious environmental-gift given by nature to the mankind. It is the need of current time that we all must be concern to conserve both the water quantity and water quality so that the present generations would be able to transfer this ecological gift to their coming generations. Now-a-days, there is a vast discussion all around on this present scenario related with the biggest challenge of water. Rivers are the main source of water and water quality is one of its primary characteristics. Water quality measurements defined on the basis of a number of physical, chemical, and biological

parameters Rankovi *et al.* (2010). However, a gradual deterioration in the water quality of Rivers have been reported for the last 2- 3 decades. This drastic situation was caused due to heavy population growth, heavy industrialization and heavy urbanization with unmanaged human activities. The Ganga River is considered as Holy River in India. Moreover, it is contributing towards the country's economic growth as well. But, day-by-day it is also getting polluted with a deterioration in its WQ. The water management requires quick, reliable, efficient and technical solutions for water related challenges like that of WQ (Document Prepared by Union Ministry of Science and Technology Government of India) (2009). Keeping all this in mind, many researchers have contributed their effort towards this specific problem. One of the soft computing techniques i.e. ANN has been applied successfully to model a variety of complex non-linear problems especially

Author's Address

[1&2] Deptt.of CS & IT, MJP Rohilkhand University, Bareilly, U.P., India
[3] Deptt. of Zoology & Env. Sc. Gurukula Kangri Vishwavidyalaya Haridwar, U.K., India
[4] Deptt.of CSE, BIAS, Bhimtal, U.K., India
[5] Department of Computer Science, Gurukula Kangri Vishwavidyalaya Haridwar, U.K, India.
  E-mail:anilbisht_20@yahoo.com

the ecological ones because ANNs are capable of learning from examples. As a result, recently few researches have been done in the field of WQ assessment and prediction using ANN techniques. Various deterministic models have been used for developing the WQ models. But the statistical accuracy of such models is not up-to the mark because the ecological systems are highly non-linear in nature Sarkar *et al.* (2015). Besides this, the limited availability of water quality data is a major issue Palani *et al.* (2008). The ANN-based modelling approaches are data-driven, nonlinear, flexible, required no explicit knowledge about the actual system under consideration with minimal human involvement Yan (2012). The models developed using ANNs are computationally very quick, require only few input data and parameters. Therefore the application of ANNs in the province of machine learning fascinated a lot of attention. In this present work, we have developed an ANN based model for modelling the WQ of the River Ganga in the Uttarakhand state along the Haridwar stretch by considering a long time-series authentic quality data. The performance of the developed model was determined using mean square error measure. The rest of the paper is organized in five sections. Section 2 discussed the related work. Section 3 presents the material and methods used in this work. In section 4 we have discussed the experiment methodology and results. Finally section 5 concludes the present work.

Related Work

Recently, the application of artificial neural networks, ANNs have been increased tremendously in modelling the complex and non-linear problems especially in the ecological domain due to their capability to identify and learn correlated patterns between input and output data sets. Many researches have been done all around using ANNs for modelling the ecological problems.Rankovi *et al.* (2010) proposed their work in the paper titled as "Neural network modelling of dissolved oxygen in the Gruza reservoir, Serbia". Their aim was to develop a feedforward neural network (FNN) model to predict the dissolved oxygen. Yan (2012) proposed his research work titled as "Toward Automatic Time-Series Forecasting Using Neural Networks". An automatic ANN Modelling scheme was suggested for time series forecasting (TSF)

based on the generalized regression neural network (GRNN), requiring minimal human intervention, achieving computationally efficient scheme for a large number of series and to have good overall forecasting performance. Author concluded that as compared to statistics-based forecasting techniques, neural network approaches have various advantages like ANNs are data driven, nonlinear, flexible, requiring no explicit underlying model and hence applicable to complicated problem domains. Khashei *et al.* (2010) have published their paper on "An artificial neural network (p,d,q) model for time series forecasting". They focused on ANNs as flexible computing frameworks and universal approximators that can be applied to a wide range of time series forecasting problems with a high degree of accuracy. In the first stage, an auto-regressive integrated moving average model was used to generate the necessary data, and then a neural network was used to determine a model for capturing the underlying data generating process and predict the future, using preprocessed data. They concluded that their proposed model can be an effective way in order to yield more accurate model.

Sarkar *et al.* (2015) presented her work on "River Water Quality Modelling using Artificial Neural Network Technique". The authors used the ANN technique to estimate the dissolved oxygen (DO) for the River Yamuna in Mathura city, Uttar Pradesh, India. Water quality data for the years 1990 to 1996 used in this study. The feed forward error back propagation neural network technique has been applied. Monthly data sets on flow discharge, temperature, pH, biochemical oxygen demand (BOD) and dissolved oxygen (DO) at three locations have been used for the analysis. It was found that the ANN approach turned out to be an efficient approach for water quality Modelling. Palani *et al.* (2008) discussed their work on "An ANN application for water quality forecasting". In this paper, a study of ANN Modelling to predict and forecast temperature, salinity, DO, and chlorophyll-a in Singapore coastal waters was presented. The results demonstrated the ANN's great potential to simulate water quality variables. Authors concluded that ANN model may provide simulated values for desired locations at which measured data are unavailable yet required for water quality models. They also highlighted that

ANN modelling is a useful tool that optimizes monitoring networks by finding essential monitoring stations thus leading to cost reduction. Thair *et al.* (2014) presented their research on "Prediction of Water Quality of Euphrates River by Using Artificial Neural Network Model (Spatial and Temporal Study). In this study, ANN was used for prediction and forecasting the monthly Total Dissolved Solid (TDS) parameter in water using Feed Forward Back Propagation (FFBP). The authors stated that self-learning and automatic abstracting capability of ANN technique may reduce the time of modelling the complex systems. Finally they concluded that ANNs outperform conventional statistical techniques in data analysis and modelling in forecasting which have limitations of strict assumptions of normality, linearity, variable independence and dimensionality.

Rani *et al.* (2014) presented their research work as "Predicting Reservoir Water Level Using Artificial Neural Network". Authors presented an Artificial Neural Network (ANN) approach for forecasting of reservoir water level using Feed Forward Back Propagation in an effective and timely way that can also help in disaster monitoring, response and control. The authors explained that the traditional methods of time series analysis used by the researchers such as autoregressive method of Box-Jenkins (AR), auto-regressive moving average (ARMA), auto-regressive integrated moving average (ARIMA), etc. but they suffer with the problem of stationary and linearity and gives only reasonable accuracy. While the ANN concept was introduced as an efficient tool of Modelling and forecasting since two decades.

Talib *et al.* (2012) presented their work on "Prediction of Chemical Oxygen Demand in Dondang River Using Artificial Neural Network". They applied the neural networks to study the factors that affect water quality of Dondang River. Total nine different variables considered as input at three stations different sites.The Multi-Layer Perceptron (MLP) was used as an ANN structure. MLP works through back-propagation method. The prediction was made for Chemical Oxygen Demand (COD). They come to the result that neural networks can be applied to study the factors that affect water quality of Dondang River. Banejad *et al.* (2011) published their paper on "Application of an Artificial Neural Network Model to Rivers

Water Quality Indexes Prediction – A Case Study". In this work the authors presented an empirical multi-layer perceptron neural network to estimate water quality indexes (BOD, DO) in Morad Big River, Iran. The back-propagation algorithm based upon the generalized delta rule was used for training. The results show the identified ANN's great potential to simulate water quality variables. Schleiter *et al.* (1999) discussed their work as "Modelling water quality, bioindication and population dynamics in lotic ecosystems using neural networks" .They have performed analysis of modelling water quality, bio-indication of chemical and long-term population dynamics of aquatic insects using neural networks model for the River Lahn (Germany).They used multi-layer-perceptrons with the Backpropagation (BP) algorithm. The results show that ANNs can successfully be applied in the analysis of the prediction of system behavior. Finally they concluded that the quality of the neural networks based models greatly depends on the quality of the data base due to their ability of learning from examples.

Bisht *et.al.* (2017) proposed their work titled "ANN based prediction model for estimating the WQ of the River Ganga". ANN model has been developed by using the one kind of training function trainlm (Levenberg-Marquardt backpropagation) with goal=0.01, 5-10-1 as network architecture over various values of learning rate. This model when applied showed the least mean square error of 0.041 and thus providing a higher precision of 95.9% for predicting the water quality of the River Ganga.

Material and Methods
Study area and Data
The national River of India i.e. the Ganga River which is also the the largest River basin in India has been selected. It comprises of 26.2% of the country's land mass (8, 61,404 Sq. km). This River lies between East longitudes 73°3′ and 89°0′ and North latitudes of 22°30′ and 31°30′, covering an area of 1,086,000 sq km. About 79% area of Ganga basin lies in India and remaining area lies in Nepal and Bangladesh. The Ganga River basin covers 11 states as Uttarakhand, U.P., M.P., Rajasthan, Haryana, Himachal Pradesh, Chhattisgarh, Jharkhand, Bihar, West Bengal and Delhi (Status Paper by: Ministry of Environment and Forests

Government of India (2009). At Devprayag confluence of River Bhagirathi and River Alaknanda forms River Ganga. It traverses a total length of 2525 km originated from the Uttarakhand state across the Uttar Pradesh, Bihar, Jharkhand and West Bengal, and finally falls into the Bay of Bengal. The map of Ganga River basin in India is displayed in figure 1. Five sampling stations along this River were selected from the Uttarakhand state.

The monthly data sets of five water quality parameters temperature, pH, dissolved oxygen (DO), biochemical oxygen demand (BOD) and total coliform (TC) for the time period from 2001 to 2015 have been used for the this research work analyzed by Limnology & Ecological Modelling Laboratory, Department of Zoology and Environment Science, Gurukul Kangri Vishwavidhyalaya, Haridwar, Uttarakhand, India.
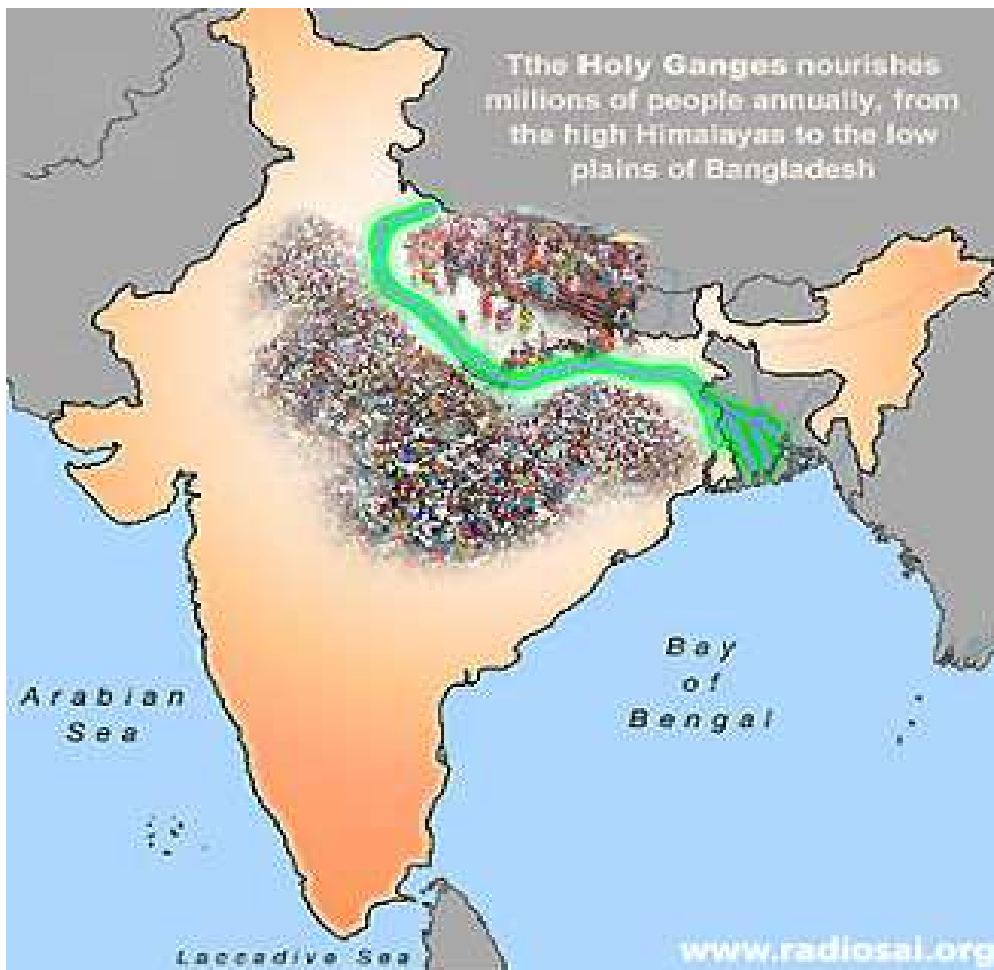

Figure 1: The Ganga River Basin

Artificial Neural Networks (ANNs)

An Artificial Neural Network is an information processing model that is stirred by the functioning of biological nervous systems i.e. brain in order to process information Abraham *et al.* (2001). The first ANN model was given by Mc Culloch and Pitts in 1943. They took nearly 20 years over how a human brain thinks and feel. Basically ANNs are inspired from biological neuron. Neural networks process information in a same fashion as the human brain perform. The network is composed of a large number of highly interconnected processing elements called neurons working in parallel to solve

a specific problem. Neural networks learn by example. They cannot be programmed to perform a specific task. An ANN is configured for a specific application, like pattern recognition problem or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons.
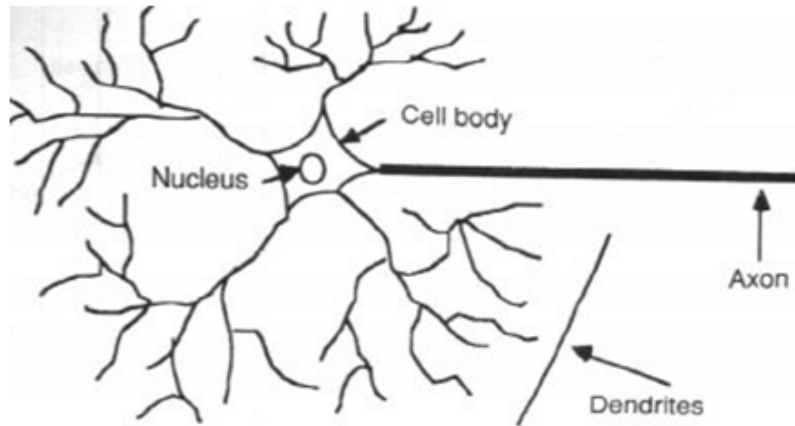
Figure 2: Components of a biological neuron (Source: NEURAL NETWORKS by C. Stergiou and D. Siganos)

the equilibrium was achieved. The concentration of phenols and chlorophenols left unadsorbed in aqueous solutions were determined by using Agilent Cary 60 UV-Vis spectrophotometer at wavelength 274, 280, 285 and 270 nm for 2-chlorophenol, 4-chlorophenol, 2,4-dichlorophenol andphenol,respectively. All the experiments were carried out in duplicate and the mean value was used in all cases. The equilibrium adsorption In the human brain, a typical neuron collects signals from others through many of fine structures called *dendrites*. The neuron sends out spikes of electrical activity through a long, thin stand known as an *axon*, which splits into thousands of branches as shown in the figure 2. At the end of each branch, a structure called a *synapse* converts the activity from the axon into electrical effects that inhibit or excite activity from the axon into electrical effects that inhibit or excite activity in the connected neurons. When a neuron receives excitatory input that is sufficiently large compared with its inhibitory input, it sends a spike of electrical activity down its axon. Learning started by changing the effectiveness of the synapses so that the influence of one neuron on another changes.

The artificial neuron is supposed to do the computing in the same way. The elementary mechanism of the artificial neuron is the same as in the human neuron and it is designed to pass messages on if the sum of the inputs is greater than a certain threshold. The basic structure of ANN is as shown in the figure 3. Neural networks can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques.
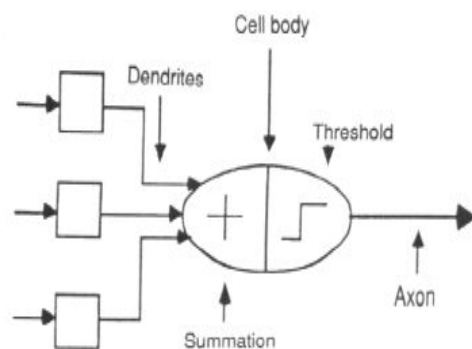

Figure 3: The basic structure of neural network (Source: NEURAL NETWORKS by C. Stergiou and D. Siganos)

Results and Discussion

The methodology for the development of ANN based water quality model involves selection of stations, collection of data and then applying the technique of ANN. In connection to this, five different stations in the stretch from Devprayag to Roorkee, Uttarakhand, India have been considered. The monthly data for defining the WQ of the River Ganga characterized by five water quality parameters temperature, pH, dissolved oxygen (DO), biochemical oxygen demand (BOD) and total coliform (TC) was collected over the time period 2001 to 2015. This record of 900 data is partitioned using 60%:40% partition strategy out of which 60% data was used for the purpose of training the neural network and remaining 40% for testing the trained model. We have performed all the experiments in MATLAB. The ANN-based program NN-tool has been developed using MATLAB 2010. For this, the feed forward error back propagation method has been used by conducting 90 experimental investigations following various training parameters. In the Neural Network model by using MATLAB, there were several training functions. We have implemented our system by considering six different functions as: traingd (Gradient descent backpropagation), traingda (Gradient descent with adaptive learning rate backpropagation), trainrp

(Resilient backpropagation), traincgp (Conjugate Gradient with Polak-Ribiere Restarts), tainscg (Scaled Conjugate Gradient backpropagation) and trainlm (Levenberg-Marquardt backpropagation). A multilayer perceptron (MLP) type neural network architecture was used by creating the neural network configuration of 5-10-1 (5 neurons in input layer, 10 in hidden layer, and 1 in output layer) showing 5 input parameters, 10 neurons in hidden layer and 1 output variable, WQ. Tan Sigmoid and Linear transfer functions were used in configuring the neural network. The training goal is set to the value of 0.1. The model was developed by initiating the training process applying various parameters like dataset, learning rate, goals and neurons. Once training was completed by achieving the stopping criteria, the developed model then tested on testing data portion. The prediction performance of the developed model is calculated by means of mean square error (MSE) statistical method. It is the average of squares of errors or deviations. The model which achieved the higher prediction accuracy is concluded as the best ANN-model for predicting the WQ of the River Ganga. The prediction performance of the worst and the best ANN model among all other training rules is displayed in the table1.
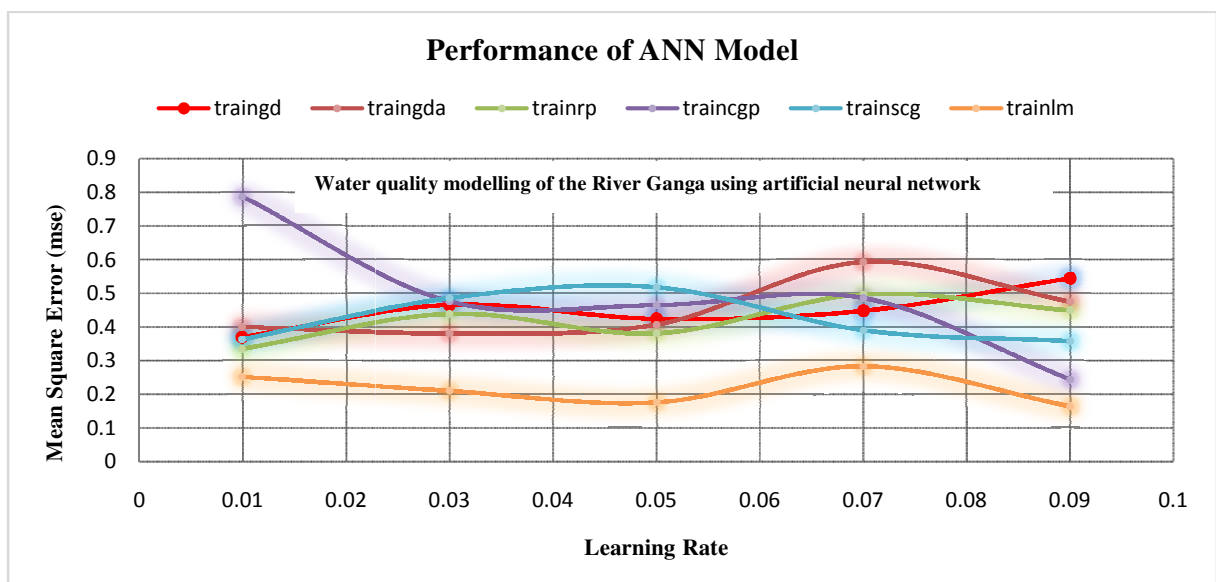
Figure 4: Performance of the developed ANN model

Table 1: Prediction performance parameters of the worst and best ANN model

Learning Rate		0.01	0.03	0.05	0.07	0.09
Goal	Training Function					
0.1	traincgp	0.786	0.475	0.464	0.485	0.243
	trainlm	0.251	0.209	0.175	0.282	0.163

On applying various training rules the performance of the developed model was evaluated using the value of MSE. Among all rules, the model developed with traincgp comes out to be the worst one (MSE=0.786 [max. among others] at learning rate 0.01) while the other model with trainlm rule proved to be the best one as shown in the table 1 as well as in figure 4. This model when applied showed the prediction error of 0.163 at learning rate of 0.09 [least among all] and thus providing an accuracy of 83.7 % which comes out to be superior to all other developed models for predicting the water quality of the River Ganga. The figure 4 showed the effect of learning rates and training rules on the performance of ANN-model. Thus the WQ model developed using trainlm proved to be the best one among all.

Conclusion

In this present work, several research papers targeted for WQ modelling using ANN. Moreover artificial neural network has been applied successfully in order to model the water quality of the Ganga River in the stretch from Devprayag to the Roorkee in the Uttarakhand state of India. For the time period from 2001 to 2015 the monthly data sets of five water quality parameters temperature, pH, dissolved oxygen (DO), biochemical oxygen demand (BOD) and total coliform (TC) have been taken for the this research work. In developing the ANN model for the WQ prediction, the feed forward error back propagation method was used to develop model for conducting many experiments following many training parameters by adopting the neural network configuration of 5-10-1, keeping training goal at 0.1 and various learning rates using six different training rules. The experimental results proved the importance of ANN in modelling and predicting the complex and non-linear relationships among the input and output parameters. The model developed with traincgp (Conjugate Gradient with Polak-Ribiere Restarts) comes out to be the worst one (MSE=0.786) while the other model with trainlm (Levenberg-Marquardt backpropagation) rule proved to be the best one (MSE=0.163). This model when applied showed an accuracy of 83.7 % which comes out to be superior to all other developed models for predicting the water quality of the River Ganga. Consequently, this study concludes that computational model developed during the study using ANNs concept is found satisfactory. it was found that ANNs are capable of predicting WQ of the River Ganga with an accuracy of 83.7% using the best model as per the present case. This current study and work indicates the significance of ANNs and would definitely stirred up new opportunities for the future work.

References

Abraham, A., and Nath, B, 2001. "A neuro-fuzzy approach for modeling electricity demand in Victoria",*Applied Soft Computing* 1, Elsevier, 127–138.

Banejad,H., and Olyaie, E., 2011. "Application of an Artificial Neural Network Model to Rivers Water Quality Indexes Prediction – A Case Study", *Journal of American Science, 7(1).*

Bisht, A.K., Singh, R., Bhutiani, R. and Bhatt, A., 2017. Artificial Neural Network Based Prediction Model Forstimating the Water Quality Of the River Ganga. *IEEE*, (Communicated).

Khashei, M., and Bijari, M., 2010. "An artificial neural network (p,d,q) model for timeseries forecasting", *Expert Systems with Applications* 37,Elsevier, pp. 479–489.

Ministry of Environment and Forests Government of India National River Conservation Directorate, 2009. *Status Paper On River Ganga,* "State of Environment and Water Quality".

Palani, S., Liong, S.Y., and Tkalich, P., 2008."An ANN application for water quality forecasting",*Marine Pollution Bulletin* 56 1586–1597, Elsevier.

Rani, S., and Parekh, F.,2014. "Predicting Reservoir Water Level Using Artificial Neural Network", *International Journal of Innovative Research in Science, Engineering and Technology*, Vol. 3, Issue 7

Rankovi,V., Radulovic, J., Radojevic, I., Ostojic, A., and Comic,L., 2010. "Neural network Modelling of dissolved oxygen in the Gruza reservoir, Serbia", *Ecological Modelling* 221, 1239–1244, Elsevier.

Sarkar, A., and Pandey, P., 2015. "*River Water Quality Modelling using Artificial Neural Network Technique*", International Conference on Water Resources And Ocean Engineering (ICWRCOE), Elsevier**.**

Chleiter, I,M., Borchardt, B., Wagner,R.,Dapper,T., Schmidt, K.D., Schmidt,H.H, and Werner,H., 1999. "Modelling water quality, bioindication and population dynamics in lotic ecosystems using neural networks", ***Ecological Modelling*** 120, Elsevier, pp. 271–286.

Talib, A.,and Amat, M.,I., 2012. "Prediction of Chemical Oxygen Demand In Dondang River Using Artificial Neural Network", ***International Journal of Information and Education Technology***, Vol. 2, No. 3.

Thair S.K., Hameed, A. M. J., and Ayad S. M., 2014. "Prediction of Water Quality Of Euphrates River By Using Artificial Neural Network Model (Spatial and Temporal Study), ***International Research Journal of Natural Sciences*** Vol.2, No.3 ,pp.25-38.

Yan, W., 2012. "*Toward Automatic Time-Series Forecasting Using Neural Networks*", IEEE Transactions on Neural Networks and Learning Systems, VOL. 23, NO. 7.

Union Ministry of Science and Technology Government of India, 2009. *Technology Mission: "Winning, Augmentation and Renovation*" Technology Mission: WAR for Water, Plan Document Prepared by, On the directive of Supreme Court of India Order on Writ Petition (C) No 230 of 2001.

.