International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 02, 2021

Time Series Analysis Of Water Quality Prediction Model For Rivers Using Machine Learning Approach

Anil Kumar Bisht[1],Ravendra Singh[2], Rakesh Bhutiani[3], Ashutosh Bhatt[4]

[1,2]Computer Science & Information Technology, Mahatma Jyotibha Phule Rohilkhand University, Bareilly, U.P. India
[3]Deptt. Of Environment Science, Gurukul Kangri University, Haridwar, Uttarakhand, India
[4]Deptt. Of Computer Science, Uttarakhand Open University, Uttarakhand, India

Email : [1]bishtanil20@gmail.com, [2]rsiet2002@gmail.com, [3]rbhutiani@gmail.com, [4]ashutoshbhatt123@gmail.com

Abstract: Water Quality (WQ) modeling and forecasting are very challenging for water management bodies due to the complex and nonlinear relationship between the parameters responsible for determining water quality. The main focus of this paper is development and time series analysis of the water quality prediction model of the Ganges River based on one of the significant machine learning (ML) approach known as Artificial Neural Network (ANN). The impact of one of the critical configuration parameter of neural network known as learning rate was analyzed. The proposed prediction model based on an artificial neural network (ANN) consists of different sets of experiments performed by comparing twelve different training functions against the variation in learning rates. A total of 360 experiments have been conducted on the dataset collected over the period 2001 to 2015 with five stations along the Ganges River in the state of Uttarakhand, India. All experiments have been conducted in MATLAB-software. The ANN-based program is written in Matlab's NN-Toolbox. As input parameters, we have used temperature, pH, dissolved oxygen (DO), biochemical oxygen demand (BOD), and total coliform. The water quality standard set by the Central Pollution Control Board of India has been used. The performance of the developed model has been calculated based on Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE). Trainlm training function-based artificial neural network models indicate higher predictive accuracy when compared to other models developed using the remaining eleven training functions when the learning rate is set to 0.04. In conclusion, ANN has the ability to efficiently predict water quality of rivers and the learning rate has a greater impact in the development of such predictive models. So, it is required to be tuned very carefully. Overall, the machine learning approach, ANN proved to be successful in the time series analysis of WQ prediction model.

Keywords: **Water Quality, Artificial Neural Network, Machine Learning, Forecasting, Learning Rate, Training Function**

1. INTRODUCTION

Water quality (WQ) refers to the chemical, physical and biological characteristics of water, according to its recorded uses and well-defined standards [1]. In general, the main sources of water are rivers. However, the quality of river water is deteriorating day by day. The Ganges River is the largest river valley in India but its water quality has been gradually declining. Water quality management has an important responsibility to control water pollution in rivers [2]. Therefore, the WQ issue of the Ganges river requires time to adopt some technical solutions which in turn helps the water management authorities in their decision support system (DSS). Apart from this, the Supreme Court of India has also directed the Government of India to find technical solutions to water related challenges. In recent times, artificial neural network (ANN) has emerged as a very effective tool regarding pattern recognition, regression and function estimation.

Unlike statistics-based methods, ANN-based model are data deployed, flexible, non-linear in nature and do not require the description of the process under consideration [3] as well as their results are highly satisfactory and accurate. Many researchers used ANN as an effective tool for modeling and predicting water quality for rivers abroad [4], [5], [6], [7] and [8]. The author [9] proposed a technical approach that integrates mobile application and web technology into water quality scientific data so that water management decisions can be easily adopted by the authorities concerned and also available in the public sector. In recent years, Indian scientists have also applied various statistical approaches to water quality modeling [10], [11], [12], [13], [14] and [15] of Indian rivers.

There are few studies related to modeling the water quality of various Indian rivers like Gomti [16], Yamuna [17], and Mahanadi [18] using ANN's approach. In addition, this instrument has also been reported in the prediction of the WQ-index for Parakai Lake, Tamil Nadu [19]. Thus, ANNs prove to be an effective and efficient tool for water quality modeling and evaluation because they have presented incredible forecasting performance. However, still no research has been conducted to forecast the WQ of the Ganges River using the techniques of ANNs. Therefore, taking all these facts into account, an application of ANN is planned for the development of the model applied for modeling here and predicts the WQ of the Ganges River through Matlab software. The development of such forecasting models is of exceptional importance for biologists and concern authorities because they will be able to take necessary and retrospective actions according to the predicted level of water pollution. This will provide cost reduction involved in monitoring by identifying and planning control measures for highly polluted stations [2].

Learning rate is one of the important configuration parameter in the designing of any neural network. It is the size of steps taken in the weight space in order to reach to the global minima that is the minimum point in the error surface. It is very critical to decide it’s appropriate value as it is responsible for controlling the magnitude of the weight adjustment to be done in any multilayered neural network while adopting the back propagation neural network algorithm (BPNN) for training and learning purpose that is based on gradient descent approach. Authors [20] applied the trainlm (Levenberg-Marquardt backpropagation) training algorithm only which shows the least prediction error of 0.041. So, here in this paper, trainlm function is compared with other functions and then after conducting a set of

International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 02, 2021

various experiments the overall impact of learning rate on all the functions is analyzed. Generally, trial and error principle is used to determine its optimal value. Thus, objective in this paper is to find the impact of learning rate in developing the predictive model for water quality of rivers and accordingly analyze its performance.

2. MATERIALS AND METHODS

2.1 Sampling Stations

Five sampling stations along the banks of the Ganges in the state of Uttarakhand, India, 'Devprayag' (Station A), 'Rishikesh' (Station B), 'Haridwar' (Station C), 'Jwalapur' (Station D) and' Roorkee (Station E) has been selected for the study as shown in Fig. 1. The 15-year dataset collected at these five sampling stations has been used to perform experiments.

Fig.1. The map based view of all the five stations taken for study (Source: Google Earth)

2.2 Data Used

Monthly experimental time series water quality samples used here from 2001 to 2015 were collected and examined by the laboratory of "Limnology and Ecological" modelling, Gurukul Kangri, University, Haridwar, Uttarakhand, India. The dataset includes five different WQ factors as temperature, pH, dissolved oxygen (DO), biochemical oxygen demand (BOD), and total coliform.

2.3 ANN based Model Development via Training Phase

2.3.1.Artificial Neural Network

A neural-network created artificially to mimic the biological brain. It is composed of artificial neurons specified to perform precise tasks such as pattern recognition, classification, prediction, clustering, and so on. An artificial neuron is also required to mimic the same process as natural neurons. Its basic phenomenon is similar to that in the case of biological neurons, messages can be transmitted only if and only net information exceeds the prescribed limit. Thus ANNs are computational models consisting of various artificial neurons that

behave as processing nodes. These nodes communicate with each other by directing signals over a large amount of weighted connections.
A weight value is associated with every input to an artificial neuron. As presented in fig. 2 Each input X1, X2, X3, X4, X5 is multiplied by their respective assigned weights W1, W2, W3, W4, W5 in order to obtain the product of all inputs and their linked weights which is then have to feed in the summation function unit to get their amount and then this net value is transferred as input to the transfer function unit. [21]. In this computational model, the artificial neuron resembles the same structure as the biological neuron. Mathematically, the net input is computed as:

(Let net input be Yin),

Yin= x1 w1+ x2w2+ x3w3 +…+ xnwn

(Where n stands for total number of inputs)

$= \sum_{i=1}^{n} xiwi$

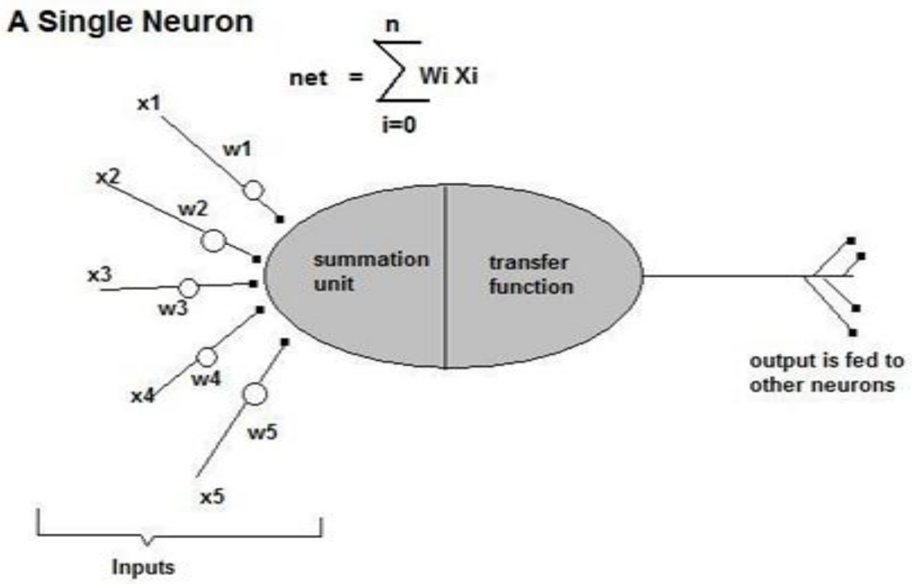

Fig. 2. Mathematical Model of an Artificial Neuron [21]

Various experiments are performed in the MATLAB software by initiating the training phase to implement the ANN-based water quality model. Different types of neural network models are developed by adopting four different strategies that vary in neural network training parameters.

In, this proposed model, there are 5 input parameters (i.e. independent variables as temperature variables, pH, DO, BOD and TC) and one output parameters (i.e. dependent variables called WQ). A dataset segmentation strategy of 60%: 40% has been adopted to develop the ANN model, which is divided into two parts of the original dataset of WQ: The first training set comprises 60% (i.e. a total of 540 records) used for model development and the second set includes 40% (i.e. a total of 360 records) for testing the developed model.

Training dataset records are grouped from 1 to 540, while the test data set contains the remaining records from 541 to 900.

To begin the experiment, an <input, output> pair is created with two variables "in" and "out" for representing input data and output data, respectively. Consider the specimen of the training dataset as specified in the table 1 assuming total number of records are 5. Now, as per the data partition strategy adopted i.e. of 60%:40% the two variables are defined for input and output data. The input variable 'in' consists of the 60% training data defining the input parameters:

in**= [**14.5 14.8 15.2 **;** 7.42 7.68 7.7 **;** 12.6 11.8 11**;** 1.8 2 2.4**;** 94 70 110**];**

Whereas, the output variable 'out' contains the testing or target data for the purpose of training. In respect of 60% input dataset the corresponding first 60% target dataset is defined labelling the one output parameter WQ like:

out= [2 2 2];

Here, the remaining 40% of target dataset of same WQ parameter has remained unused as it has to be forecasted by the developed model which was trained using 60% of <input, output> training combination of the training dataset ANN. Now, in_new variable is defined which is initialized to the 40% of the testing dataset corresponds to the remaining training or input dataset as:

in_new= [15 16.5; 8.03 8.08; 11.5 11.2; 1.8 2; 110 120];

As here, the input dataset and the equivalent output dataset are designated and arranged for the specimen dataset, in similar fashion the desired ANN based WQ model is developed for the research work over the available dataset consisting of 900 records.

Table 1: Specimen for Training Data-Set (Assuming 5 records in all)

S.No.	1.	2.	3.	4.	5.
Temp °C	14.5	14.8	15.2	15	16.5
pH	7.42	7.68	7.7	8.03	8.08
DO (mg/ L)	12.6	11.8	11	11.5	11.2
BOD(mg/ L)	1.8	2	2.4	1.8	2
Total Coliform /100ml	94	70	110	110	120
WQ	2	2	2	2	2

Table 2 contains all the parameters used during the training phase for the development of the NN model, along with their respective values. To apply the backpropagation training algorithm to develop the ANN model in Matlab, a total of twelve training functions are used. Using Matlab, there are two options to train any model based on ANN. One is the command interface and the other is the graphical interface. Here, the command interface method is used.

Table 2: Training Parameters and their values to develop Artificial Neural Network models

Dataset Partition strategy	60% 40% where 60%: training data and 40%: testing data (i.e. to be forecasted) out of total dataset of size 900.
Learning rate	0.01 0.02 0.03…….0.1
Goal	0.01
No. of Neurons/nodes	10
Training Function	'traingd', 'traingdm', 'traingdx', 'traingda', 'trainrp', 'traincgf', 'traincgp', 'traincgb', 'trainscg', 'trainbfg', 'trainoss', 'trainlm'
Epoch	50000
Hidden Layer	1
Total Models	360
Model Testing	Every model is trained thrice and the corresponding best result has been accepted. As a result, the total number of tested models becomes=360

The following fig. 3 displays the pseudo code used for training the artificial neural network model.

```
tic
in=[14.5 14.8 15.2 ; 7.42 7.68 7.7; 12.6 11.8 11; 1.8 2 2; 94 70 110];
out=[2 2 2];
[in, ins] = mapminmax(in);
[out, outs] = mapminmax(out);
net= newff ([minmax(inn)], [10 1],{'tansig',' purelin'},'trainlm');
net.trainParam.show= 500;
net.trainParam.lr= 0.01;
net.trainParam.epochs= 50000;
net.trainParam.goal= 0.01;
net= train(net, inn, outn);
an= sim(net, inn);
a= mapminmax('reverse', an, outs);
toc
```

Figure 3: Pseudocode for training the ANN model

As seen above using this code block, the proposed water quality ANN model is trained by following a variety of variations in various training parameters in order for the development of an optimal NN model. The training procedure of neural network is now started which is to be followed by testing procedure at later stage. Once the network met the "performance goal", then the developed (trained) model is to be tested by running the code block in Matlab NN toolbox.

The testing code block is as highlighted in the fig. 4.

```
in_new= [17.1 16.5 ; 8.81 8.7; 11 12.1; 3.6 3 ; 160 120];
in_new= mapminmax ('apply',in_new,ins);
anew= sim (net,in_newn);
anew= mapminmax ('reverse', anewn, outs);
```

Figure 4: Pseudocode for testing the developed ANN model

Once the code for training get executed in the command window of the Matlab's neural network toolbox, the training of network started as presented in the fig. 5.

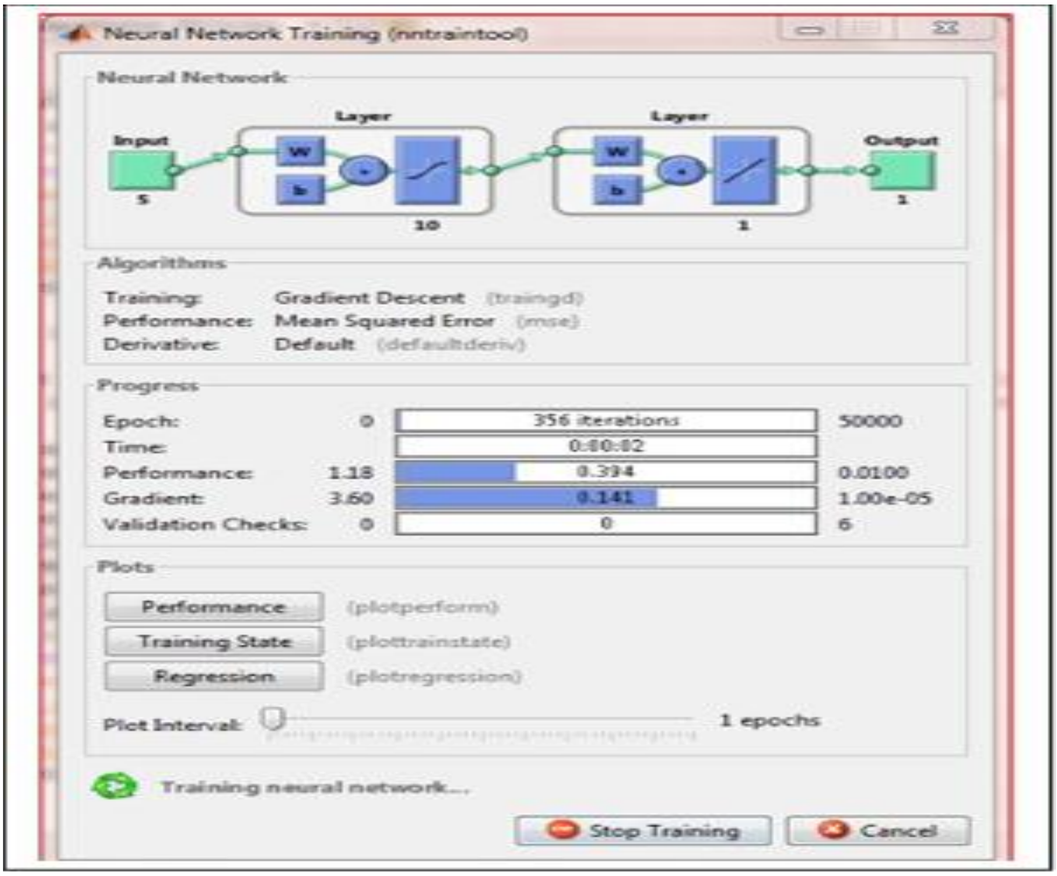

Figure 5: Window showing Neural Network Training in progress

The back-propagation method was applied for developing the prediction model for water quality by performing various experiments using a neural network structure of 5-10-1 which means there are five nodes in the input of ANN architecture representing input parameters, the hidden layer comprised of ten nodes and the output for the target variable is represented by one node. There are two main possibilities regarding the output which is to be appeared in the output window after completion of the training as displayed in the fig. 6 and fig. 7. In first case the outcome becomes “maximum epoch reached”. It indicated that the defined no. of epochs in the code have been exhausted i.e. followed during the training phase (like 50000 fixed in this research work).

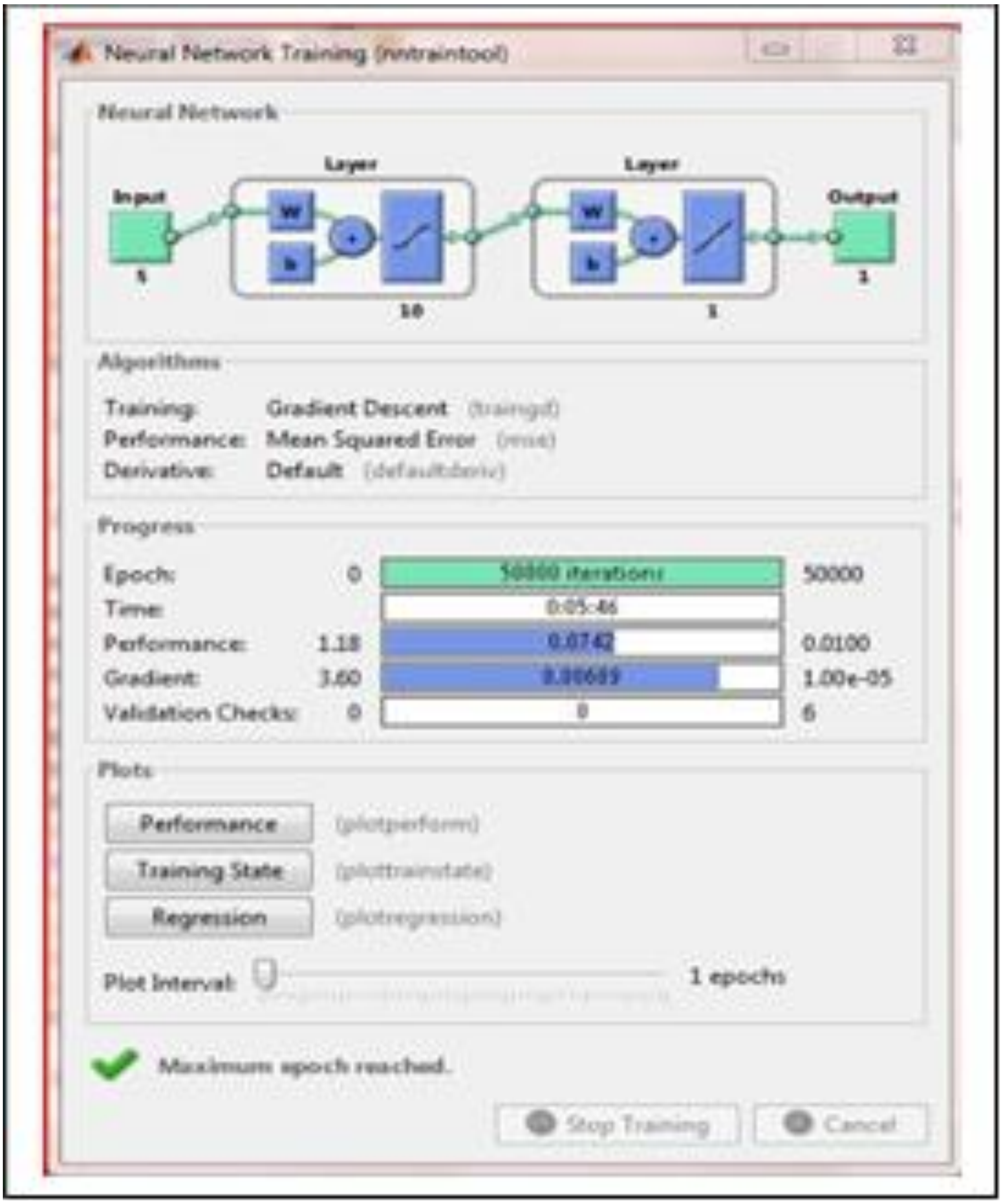

Figure 6: Window showing one of the possible result of the NN Training

This means, 50000 times the complete dataset is presented to the learning machine during the training but even then the desired performance (error) goal has not been achieved. As a consequence, the training is to be stopped. While in the second case the set performance goal now satisfied.

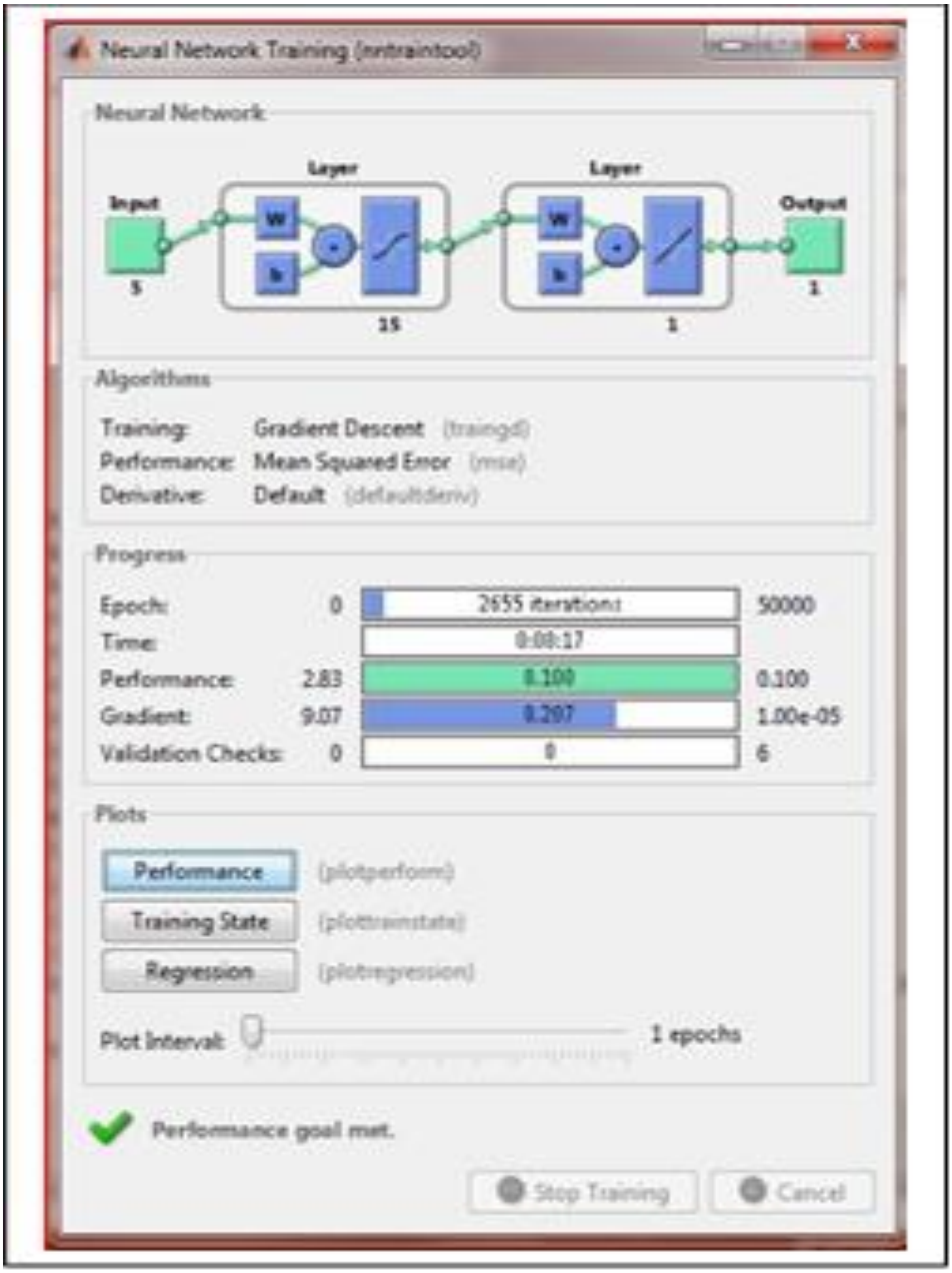

Figure 7: Window showing another possible result of the NN Training

2.4 Testing the Performance of the Trained (Developed) Model via Testing Phase

When the output of training is obtained as "performance goal met", then another Pseudocode is to be executed as presented in fig. 8 to test the developed model. This code is pasted in the command window of the NN toolbox and then execute the same after typing the name of the network as “anew”. Following this the output is displayed in the command window as presented in the figure 4.6. This output is occupied from the NN tool box’s command window and kept at somewhere from where it will be accessed as and when required for further processing. . The main purpose to retrieve this stored output is to calculate the error

```
Columns 1 through 9

2.0324    1.9946    1.4506    1.5571    1.5285    3.5779
4.6868    3.0816    2.2744

Columns 10 through 18

3.3595    1.1536    1.2682    1.6521    1.5542    1.9574
1.0600    1.0299    2.4029

Columns 19 through 27

3.6912    3.2538    2.1810    1.9560    1.8815    1.9962
1.8173    1.6365    2.2917

Columns 28 through 36

1.8531    3.2555    4.5122    4.1020    5.0966    3.0080
1.3670    1.0723    2.4201

----------------------------------------------------
----------------------------------------------------

Columns 343 through 351

2.1261    2.0481    2.2417    2.0792    1.2720    2.1483
2.2217    2.1567    2.1372

Columns 352 through 360

2.0231    2.0105    2.0269    1.9593    1.9014    1.9464
1.9874    2.0946    2.1799
```

Figure 8: Testing output displayed in the command window of NN toolbox

Then only the undertaken concept of training called supervised training or learning is fully achieved where the computed or forecasted target is matched with the actual target. This difference (if any) returned as error like “Mean Square Error”, “Mean Absolute Percentage Error” etc. Our objective is that this error must be as lowest as possible in order to achieve a maximum accuracy from the developed model. But if this error is not lowest then variations are required to be imposed in the training factors like number of neurons, learning-rate, goal, training-functions, and size of training set.

3. Experiments, Results and Discussion

In this chapter, ANN dependent application is proposed for developing the WQ prediction model for the Ganga River. A number of experiments are conducted involving various combinations of different ANN parameters. All these experiments are implemented via MATLAB software. An ANN based program was formed in NN Toolbox of matlab. The feed forward error back-propagation method was used by conducting a total of 360 experimental investigations following various training factors which resulted in development of total 360 various models. There are several training functions which can be used in the development of Neural Network models by using MATLAB. Total twelve different training functions are used for the implementation purpose. In this way, various kinds of neural network models are developed by adopting the variation in learning rates based methodology.

The proposed ANN based models developed using varying Learning Rate from 0.01 to 1.0, keeping Goal fixed at=0.01, Dataset Partition Strategy=60%:40%, Number of Nodes=10 and 12 Training Functions. There are 5 input parameters and one corresponding output parameter termed WQ. For developing the ANN models the dataset partition strategy of 60%:40% is adopted where the original dataset of WQ is divided into two parts: first training set comprised of 60% (i.e. total 540 records) used for model development and second set is for testing the developed model which contains 40% (i.e. 360 records). However, for implementing the fourth approach a lot of variation is adopted in the data partition policy.

Therefore, by performing a grouping of different experiments, the effect of variations in the crucial parameter of ANN i.e. the learning rate, over the performance of the developed water quality models is analyzed. After a critical analysis of all these factors an optimal ANN based WQ prediction model for the Ganga River is identified.

3.1 Proposed Approach for model development by varying "Learning Rate"

In this scenario, the goal is kept constant at=0.01, the number of nodes are also kept constant at 10 with undertaking the dataset partition strategy of 60%:40% and considering the twelve Training Functions. The experiments are repeated with adopting a variation over the learning rate from 0.01 to 1.0. On applying various training functions the prediction efficiency of the different developed models was calculated utilizing the value of MSE, RMSE and MAPE. For every training function the training is performed three times by running its code and consequently the prediction performance is computed every time and the best value is selected against each function. The function which provides the ideal value (least error or highest accuracy) among the all other functions is to be selected as the preeminent for the model development.

Following this process, some models developed using specific functions have not attained the 'performance goal" criteria and hence such models are of no use as their performance cannot be evaluated. However, the functions using which the models attained the "performance goal" criteria are to be evaluated next. The prediction accuracy of all the developed models with respect to MSE is presented in the table 3. Thus, after conducting a total of 360 experiments, the same number of models are developed for the WQ prediction of the Ganga River. When the learning rate =0.04, the trainlm function provides the best model as compared to the models developed by other training functions.

Table 3: MSE Result of keeping Goal=0.01, Dataset Partition Strategy=60% : 40%, Number of Nodes=10 with varying learning rate											
Learni	→	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1

ng Rate (lr)		1	2	3	4	5	6	7	8	9	
Goal	Training Function	MSE	MSE	MSE	MSE	MSE	MSE	MSE	MSE	MSE	MSE
0.01	traingd	NIL	NIL	NIL	NIL	NIL	NIL	NIL	NIL	NIL	0.155
0.01	traingdm	NIL	NIL	NIL	NIL	NIL	NIL	NIL	NIL	NIL	0.187
0.01	traingdx	NIL	NIL	NIL	NIL	NIL	NIL	NIL	NIL	NIL	0.105
0.01	traingda	NIL	NIL	NIL	NIL	NIL	NIL	NIL	NIL	NIL	0.209
0.01	trainrp	NIL	NIL	NIL	NIL	NIL	NIL	NIL	NIL	NIL	NIL
0.01	traincgf	0.324	0.061	6.436	0.369	NIL	NIL	NIL	NIL	NIL	0.767
0.01	traincgp	0.062	0.060	0.120	0.052	0.244	0.085	0.060	0.057	0.087	0.605
0.01	traincgb	0.074	0.059	5.499	0.050	0.366	0.366	0.055	0.062	4.008	0.110
0.01	trainscg	0.057	0.058	0.125	0.089	0.089	0.063	0.060	0.076	0.071	0.125
0.01	trainbfg	0.138	NIL	NIL	6.866	NIL	NIL	NIL	NIL	NIL	0.056
0.01	trainoss	0.257	0.057	0.057	0.070	0.302	0.067	0.067	0.283	*0.080*	0.117
0.01	trainlm	0.056	0.051	0.053	0.041	0.060	0.047	1.532	0.060	0.054	0.053

The prediction performance of the best model was determined at learning rate=0.04 as shown in the fig. 9.

International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 02, 2021

				CATEGORY-I (lr=0.04, goal=0.01, n=10)									
		First Time Training				Second Time Training				Third Time Training			
S.N	Training Function	Epoches1	MSE1	RMSE1	MAPE1	Epoches2	MSE2	RMSE2	MAPE2	Epoches3	MSE3	RMSE3	MAPE3
1	traingd	50000	NIL	NIL	NIL	50000	NIL	NIL	NIL	5000	NIL	NIL	NIL
2	traingdm	50000	NIL	NIL	NIL	50000	NIL	NIL	NIL	5000	NIL	NIL	NIL
3	traingdx	50000	NIL	NIL	NIL	50000	NIL	NIL	NIL	5000	NIL	NIL	NIL
4	traingda	50000	NIL	NIL	NIL	50000	NIL	NIL	NIL	5000	NIL	NIL	NIL
5	trainrp	50000	NIL	NIL	NIL	50000	NIL	NIL	NIL	5000	NIL	NIL	NIL
6	traincgf	3956	0.369	0.60763	12.07974	50000	NIL	NIL	NIL	4821	1.20488	1.09767	15.8648
7	traincgp	3833	0.0842	0.29017	4.074124	2907	0.0521	0.228325	3.31190964	5737	0.0822	0.28671	4.34723
8	traincgb	1543	0.05	0.22465	3.588618	50000	NIL	NIL	NIL	1110	0.63922	0.7951	11.393
9	trainscg	5559	0.5207	0.72161	16.30756	6226	0.0889	0.29822	5.60631348	3271	6.0815	2.46607	58.4706
10	trainbfg	320	102.06	10.1023	84.65821	50000	NIL	NIL	NIL	170	6.8664	2.62039	12.5508
11	trainoss	7665	2.1996	1.4831	36.9845	6385	0.09868	0.314141	7.44981323	6726	0.0701	0.26485	5.15638
12	trainlm	46	0.0967	0.31101	6.990161	9	0.0413	0.203226	2.13131098	28	0.43353	0.65843	13.7121

Figure 9: Prediction Performance of best model at learning rate=0.04

The influence of learning rates and training rules on the performance of ANN-model is shown in fig. 10.

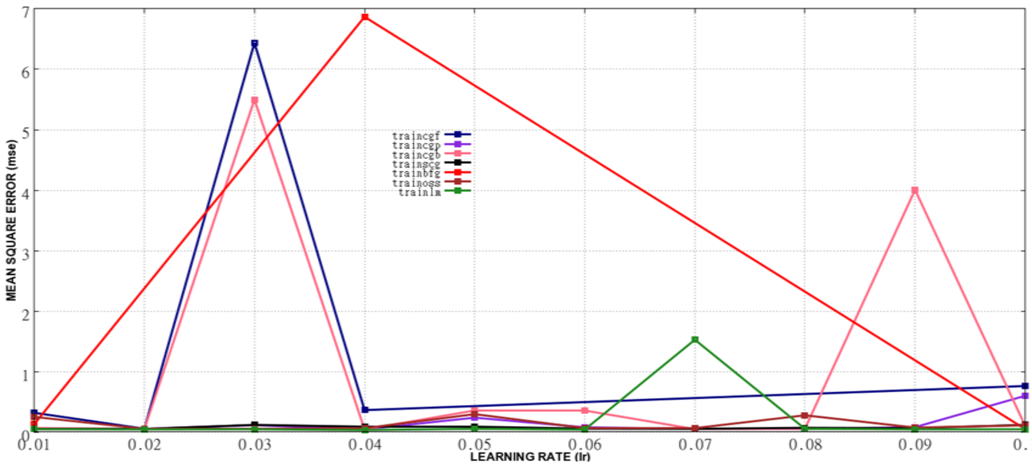

Figure 10: Performance of the developed ANN model in 1st approach

The analysis of results firstly concluded that the learning rate parameter has a greater effect in development of ANN based prediction and forecasting models. As it can be presented in the table 3 where on keep changing the values of learning rate for any of the training functions the outcome that is the mean square error (MSE) has also changed. Therefore, this can be concluded that the error or the cost function is much affected by a little variation in the learning rate. When the learning rate was fixed at 0.04 the model developed using this setting of parameter with fixed goal at 0.01, number of nodes fixed at 10 and adopting the 60%:40% data partition strategy, the least error in the form of MSE is achieved as 0.041 and accordingly the highest prediction accuracy of 95.9% is obtained in comparison to other developed models. Secondly, the result also showed that the training function trainlm is best for generating the above said model as in majority of cases it results in best accuracy as compared to the others.

4. CONCLUSION

In this paper, an efficient analysis of time series water quality prediction model for the river Ganga based on machine learning approach termed as ANN has been performed successfully. In order to implement this approach, an impact of learning rate while developing novel predictive models for water quality of rivers has been analyzed effectively. Learning rate is one of the vital configuration parameter in the designing of any neural network based prediction and forecasting solution. It is very critical to decide it's appropriate value as it is responsible for controlling the magnitude of the weight adjustment to be done in any multilayered neural network while adopting the back propagation neural network algorithm (BPNN) for training and learning purpose that is based on gradient descent approach. So, this is required to be tuned very carefully. Generally, trial and error principle is used to determine its optimal value. The proposed prediction model based on artificial neural network (ANN) developed by comparing twelve different training functions against the variation in learning rates. A total of 360 experiments have been conducted on the dataset collected over the period 2001 to 2015 with five stations along the Ganges River in the state of Uttarakhand, India. All experiments have been conducted in MATLAB-software

The analysis of results firstly concluded that the learning rate parameter has a greater effect in development of ANN based prediction and forecasting models. Various values of learning rate has been taken for analyzing its effect for different training functions ranges from 0.01, 0.02, 0.03 and so on till 0.1. When the learning rate was fixed at 0.04 the model developed using this setting of parameter with fixed goal at 0.01, number of nodes fixed at 10 and adopting the 60%:40% data partition strategy, the least error in the form of MSE is achieved =0.041 for trainlm algorithm and accordingly the highest prediction accuracy of 95.9% is obtained in comparison to other developed models. This implementation of the present approach also concluded that among the twelve training functions, the trainlm function is most suitable for developing the WQ prediction model for the Ganga River and also the best model is developed when setting the learning rate parameter at 0.04. Finally it has been concluded that that learning rate parameter of neural network design has a significant impact in development of such prediction models with respect to their performances. Therefore, it should be tuned very carefully for getting the better results. Finally, the machine learning approach based on ANN proved to be highly efficient in developing and analyzing the water quality prediction model for the rivers.

5. REFERENCES

[1] N. M. Gazzaz, Mohd Kamil Yusoff, Ahmad Zaharin Aris b, Hafizan Juahir, Mohammad Firuz Ramli, "Artificial neural network Modelling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors", Marine Pollution Bulletin, ELSEVIER, 2012.

[2] Sundarambal Palani, Shie-Yui Liong, Pavel Tkalich, "An ANN application for water quality forecasting", Marine Pollution Bulletin 56 1586–1597, ELSEVIER, 2008.

[3] Weizhong Yan, "Toward Automatic Time-Series Forecasting Using Neural Networks", IEEE Transactions on Neural Networks and Learning Systems, VOL. 23, NO. 7, JULY 2012.

[4] Ehsan Olyaie, Hamid Zare Abyaneh, Ali Danandeh Mehr, "A comparative analysis among computational intelligence techniques for dissolved oxygen prediction in Delaware River", ELSEVIER, Geoscience Frontiers xxx (2016) 1-11.

[5] Chou-Ping Yang, Chi-Ying Hsieh, Yu-Min Wang, Wen-Ping Hsiao, "Using Artificial Neural Network for Reservoir Water Quality Analysis in Taiwan", 978-1-61284-459-6/11 ©2011 IEEE.

[6] Yi Wang & Tong Zheng & Ying Zhao & Jiping Jiang, Yuanyuan Wang, Liang Guo, Peng Wang, "Monthly water quality forecasting and uncertainty assessment via bootstrapped wavelet neural networks under missing data for Harbin, China", Environ Sci Pollut Res (2013) 20:8909–8923, DOI 10.1007/s11356-013-1874-8.

[7] Ahmad Gholamalizadeh Ahangar, Jaber Soltani, Alireza Shakeri Abdolmaleki, "Predicting Mn concentration in water reservoir using Artificial neural network (Chahnimeh1 reservoir, Iran)", International Journal of Agriculture and Crop Sciences, IJACS/2013/6-20/1413-1420, ISSN 2227-670X ©2013 IJACS Journal.

[8] Maria J. Diamantopoulou, Dimitris M.Papamichail, Vassilis Z. Antonopoulos, "The Use of a Neural Network Technique for the Prediction of Water Quality Parameters", Operational Research, An International Journal / Vol.5, No. l / January-April 2005, pp. 115-125.

[9] Weigang Yan, Mike Hutchins, Steven Loiselle, Charlotte Hall, "An Informatics Approach for Smart Evaluation of Water Quality Related Ecosystem Services", Ann. Data. Sci. DOI 10.1007/s40745-015-0067-3, © Springer-Verlag Berlin Heidelberg 2016.

[10] Sukarma Thareja, Siddhartha Choudhury, Priyanka Trivedi, "Assessment of Water Quality of Ganga River in Kanpur by Using Principal Components Analysis", Advances in Applied Science Research, 2011, 2 (5):84-91, ISSN: 0976-8610, CODEN (USA): AASRFC.

[11] Shamshad Ahmed, Iqbal H Khan, B.P. Parida," Performance of Stochastic Approaches Forecasting river Water Quality", pp. 4261-4266, 2001.

[12] Liya Fu, You-GanWang (2012). Statistical Tools for Analyzing Water Quality Data, Water Quality Monitoring and Assessment, Dr. Voudouris (Ed.), ISBN: 978-953-51-0486-5, InTech, Available-from: http://www.intechopen.com/books/water-quality-monitoring-and-assessment/statistical-tools-for-analyzing-water-quality-data.

[13] Kanchan Bahukhandi Deoli, Satendra, NA Siddiqui, Rajesh Singh, Shefali Arora, Prasenjit Mondal, "Assessment of Major Ion Chemistry and Spatial Variation of Water Quality of Ganga River of Uttarakhand, India", Journal of Chemical and Pharmaceutical Research, 2018, 10(7): 145-150, ISSN : 0975-7384, CODEN(USA) : JCPRC5.

[14] D.R. Khanna, R. Bhutiani, Bharti Tyagi, Prashant Kumar Tyagi, Mukesh Ruhela, "Assessment of water quality of River Ganges during Kumbh mela 2010", Environment Conservation Journal 13(3) 165-169, 2012, ISSN 0972-3099 (Print) 2278-5124 (Online).

[15] C.Gajendran, A.Jesumi, "Assessment of Water Quality Index in Cauvery River Basin: A Case Study on Tiruchchirappalli District, Tamil Nadu, India", Universal Journal of Environmental Research and Technology, Euresian Publication © 2013 eISSN 2249 0256, Volume 3, Issue 2: 137-140.

[16] Kunwar P. Singh, Ankita Basant, Amrita Malik, Gunja Jain, "Artificial neural network modeling of the river water quality—A case study", Ecological Modelling, Volume: 220, Issue 6, 2009 888–895, ELSEVIER, doi:10.1016/j.ecolmodel.2009.01.004, ISSN 0304-3800.
[17] Archana Sarkar, Prashant Pandey "River Water Quality Modelling using Artificial Neural Network Technique", International Conference on Water Resources And Ocean Engineering (ICWRCOE), ELSEVIER, 2015.
[18] Nibedita Guru, Ramakar Jha, "Simulation of BOD-DO Modeling in Mahanadi River System lying in Odisha using ANN, India", IOSR Journal Of Environmental Science, Toxicology And Food Technology (IOSR-JESTFT) ISSN: 2319-2402, ISBN: 2319-2399. Volume 2, Issue 4, PP 52-57, Jan. - Feb. 2013.
[19] Sahaya Vasanthi S., Adish Kumar S., "Application of Artificial Neural Network Techniques for Predicting the Water Quality Index In The Parakai Lake, Tamil Nadu, India", Applied Ecology and Environmental Ecology And Environmental Research 17(2):1947-1958, ISSN 1589 1623 (Print) ISSN 1785 0037 (Online) DOI: http://dx.doi.org/10.15666/aeer/1702_19471958,2019, ALOKI Kft., Budapest, Hungary.
[20] A.K. Bisht, R.Singh, R. Bhutiani, A. Bhatt, "Artificial Neural Network Based Prediction Model for Estimating the Water Quality of the River Ganga", 3rd International Conference On Advances in Computing, Communication and automation, (ICACCA 2017), IEEE, organized by Tula's Institute Engg. And Management College, Dehradun.Uttarakhand, India, during 15-16 September 2017.
[21] https://www.codeproject.com/Articles/14342/%2FArticles%2F14342%2FDesigning-And-Implementing-A-Neural-Network-Library.