AnEstimating Model for Water quality of river Ganga using Artificial Neural Network

Yamini Soni, Vikas Sejwar

Abstract*——In given propose paper we have worked on the water quality of river ganga, not only for management of water resources, but also for the prevention of water pollution, the water quality forecast has a more practical significance. To evolvesuitableideals for the water quality (WQ) insidethe water physiquesobtainingcontaminant samples & then to confirm that these standards are encountered, this is the environmentalWQ management program have goal. In the realistic standard setting, the institutional capacity of the basin's water science, environmental, & the land usagesituations, possibleusages of getting water bodies, &the determination & implementation of WQ ideals has been kept in mind. In this paper, an efficient Machine learning algorithmwas modeled. A feed forward error back propagation neural network is implemented with different training functionsnamely trainlm(Levenberg-Marquardtbackpropagation), trainb(Batch training with weight & bias learning rules), trainr(Random order incremental training w/learning functions) & trainbr(Bayesian regularization). Five sampling stations along Ganga River stretch were selected from DEVPRAYAG-to-ROORKEE city inside the Uttarakhand state of the India.These states are Bihar, Uttarakhand, Delhi, UP& West Bengal. The hill rivers of the Uttarakhand are Alkananda, Bhagirathi, Mandakini.We have used the above given training functions at different-different learning rate (i.e., 0.01,0.03,0.05,0.07&0.09) to measure classification rate& use the mean square method to measure the performance of the model. Results indicate that the proposed algorithm gives best estimating model and generate less mean square error (i.e., 0.004)and accuracy is 99.5% at 0.09 learning rate with trainbr training functionin respect to othertraining function.

***Keywords*— Artificial Neural Network, Mean Squared Error Training Functions,Water Quality, Prediction of Water Quality.**

I. INTRODUCTION

WQ can be considered as the chemical, physical& biological physiognomies of the water,that could be utilized to forecast the WQ. WQhelps,for determining the attentiveness of chemicals which is extant in the water. In the field of WQ severalinstances have been considered. In the areas of city, the techniques of water purification are utilized to eliminateinjuriouspollutants from the water before beingdisseminatedfor the usage of homes & other activities.

WQ is reliant on the ecology along with the use of human like sewage, industrial pollution, wastage of water & very significant water misuse which leads to low level of the water [1]. To forecast the WQ, the analysts haveconsideredseveraltechniques of data mining. Currently, an interest has been shown in the study of the comprehensive idea of artificial neural networks termed as (ANNs), which provides an interesting alternative deviceto the modeling of WQ& forecast [2]. To categorize & predict the WQ, several techniques are used,which reduces time.

Revised Manuscript Received on July 05, 2019.

Yamini Soni, Department of CSE & IT, Madhav Institute of Technology & Science, Gwalior, India.

Vikas Sejwar, Department of CSE & IT, Madhav Institute of Technology & Science, Gwalior, India.

Using the techniques of machine learning, the data is composed& then mined from the huge datasets &categorize the quality [3].

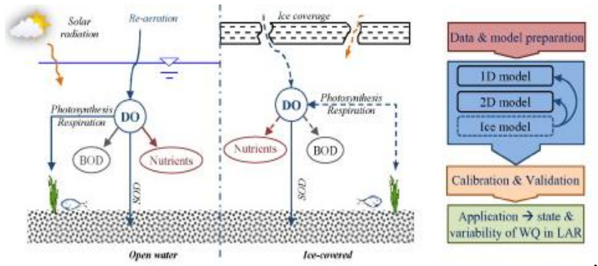

Fig. 1. An integrated framework for Water Quality

In this paper, the ANN-based modeling approach is used with various training functions to predict the WQ of river-Ganga, as ANN methods are data supple, enforced no clear knowledge about the real system underneaththe attention with the minimum participation of humanYan (2012). Performance of the established model with various training functions was determined using mean square error measure. Also, we have defined various literatures about related work in section II. In section III, described about proposed work that has been done. After this, we have discussed about experiment and results part in section IV. Andin section V, we have concluded this paper.

II. LITERATURE SURVEY

Hamza Ahmad Isiyaka et al. [2018]In this research work, the authors examines&introduces a decrease inquantity of WQ monitoring places, constraints&builds up the best i/pmixture for the modelling of WQ utilizing the artificial neural network termed as (ANN)& multivariate statistical procedure. Physical chemical parameters of 14 WQ wereacquired for8locations of monitoring tothe 8-years(i.e.,2006–2013). The Hierarchical agglomerative cluster analyses termed as (HACA) classifies8 locations of monitoring in the2importantgroups. The principal component analysis termed as (PCA) is responsibleover 82% of the overallfluctuation&qualities the pollution sources to significant anthropogenic exercises, run-off surface &parental rocks enduring. Moreover, the percentage contribution of the sensitivity analyses of toxinshas shown that the dissolved oxygen termed as (DO) as veryimportant parameter liable for contaminationthe percent of (66.3%), pursuethroughthe ammonia nitrogen percentage of (14.4%), the chemical oxygen demand percentage of (9.4%) &biochemical oxygen demand percentage of (5.3%). Outcomes for the source typeupshot wereallocatedto 39% enduring of rock, the activities of anthropogenic is 25%, the run-off of surface is 20%, the waste of faucal is 11%, the human & natural aspectsare 3.4% &the river bank erosion is 1.4%. Furthermore, the mixture models of 3 i/p (the 1, 2 & 3 model) were established to recognizea best,which can forecastthe water quality index termed as (WQI) at mostly higheraccuracy. The Model-2 utilizing themainfactor scores before the rotation of varimax looks to have best forecastabilityon8 hubs with the determination coefficient (R2) = 0.999 & the root mean square error termed as (RMSE) = 0.159. Such discoveriesexplainanenviron metrics-modelling procedure usageforexposingWQ patternsfor the decision making through the government & the investors [4].

A.K. Bisht et al. [2017]offered an examination of the currentinvestigation work which has been usedto assess & model the WQ of several rivers utilizing the method of artificial neural networks termed as (ANN). This was discovered that ANN methodbecame an effectivemethod for the modelling &predicting of WQ. All of suchanalysis work have actuallystimulatednovelscientiststhrough introducing novel data-set, parameters &optimummethod on the basis of ANNs to work in this area [5].

Yafra Khan et al. [2016] implemented a prediction model using artificial neural networks and time-series analysis with ANN-NARA to support of components of water quality. This model gives better accurate results on such performance measures like regression, root mean-squared error and mean squared error [6].

Shakibaeinia et al. [2016] Thecoordinated deterministic arithmetical modelling systemhas been established&connectedtolasting&completethe simulation of a state,& variation of mainWQ component (spatial & temporal) both into the open types water& the ice coated situations into the place of low Athabascariver also named as (LAR). A background is depend on the1D & 2D hydrodynamic &the models of WQoutwardlyjoinedin the account of cold-term goods with the ice river of 1D process model.The models are validated/calibrated utilizing the existingcalculated data &implementedto thedissolved oxygen termed as (DO) &the nutrients (that is, phosphorus& nitrogen) simulation.The outcomes demonstrate the impact of the winter season ice spread over lessening DO focus, & a changing fleeting pattern tothe DO & supplements amid the summer season epochsinclude the considerable contrasts into fixation among the primary channel & thedeluge fields. Sucharithmetical types framework may be reason for upcomingstudies which is based on WQ scenario in LAR [7].

SalisuYusaf et al. [2015] According to the machine learning algorithms, classifying quality of water as well as developing a suitable classification model for analysis. Before comparing the proposed model to other classification algorithm and model, it is analyzed the proposed model so that important features can be determined which are presented to classify Kinta River (Perak Malaysia) water quality. Finally, overall observations prove that lazy typesarchetypalenforcing the algorithm of K-star to classify WQ is agreatest algorithm [8].

Ming Hu et al. [2014]In this research paper, authors implements the self-adaptive with GA-aided multi-objective ecological reservoir operational model termed as (SMEROM) &put it inthe management of WQ inside the Xiangxi river nearby to 3valleysartificial lake, china. SMEROM assimilatesthe statistical WQ models, the operations of multi-objective artificial lake,&the self-adaptive GA insidetheall-purposebasis. Between them, statistical models of WQin the River of Xiangxi has beendesigned to manage the connection between the reservoir task& WQ, that are installed in the SMEROM constraints. Varioustarget functions, with augmenting the hydropower generation, minimalizingthe loss of deluge control, limiting rate of the deluge hazard, boosting normal outstanding limit of the deluge control & expanding aadvantage of delivery, is viewed as at the same time to get far reaching advantage among nature, society &economy. Weight process is hired to change many objectives for the samepurpose. For solvingcomposite SMEROM, a better self-adaptive GA has been hiredbythe addition offake binary crossover & themutation of self-adaptive. For determining the benefit of evolved SMEROM archetypal, resolution is compared to the ecological reservoir operation by the conventional reservoir operations & the practical operation in the year of 2011, in terms of WQ, operation of reservoir & the objective function values.Outcomesdisplay that most of the benefits in environmentalconductare better as compare to traditional / real-world operations,without the advantageof hydropower &forfeiturethe flood control profit. It is asthe control of flood & ecologicalfortification are measuredsensibly in environmental operation [9].

Pei Zhao et al. [2013]After the completion of the dam, the authors explored the changes of WQ in key channel of river Yangtze, along with control their relations with the water level fluctuation termed as (WLF), which was measuredthroughthe annual encroachment processes. And the air was trainedthrough flood. At long last, mass parity budget &the integrativeWQ indexing termed as (WQI) approachwas conducted forexplaining thetotalWQ status since the finishing of dam. The outcomesdisplayed that the TGR opening water (Yichang) showedthe high pH & the COD Mn values &the low focuses of the dissolved oxygen termed as (DO) & the ammonia nitrogen represent as (NH 3 -N) compare to thecove water (Zhutuo). The temporary changes in the WQ parameters showedthe sametendencies for outlet & inlet water. The parameters of WQdisplayedeveryone a -ve correlations at thelevel of water, due to which various effects of damage on water quality were exposed [10].

S. Wechmongkhonkon et al. [2012]The authors haveestablishedthe MLP neural network, which is hired to scrutinize & distribute WQ of Levenberg-Marquardt algorithm, Dusit district canals of Bangkok, Thailand. The outcomes of the MLP includevery higherprecision,by the assistance of which cost& time is minimalized [11].

III. PROPOSED METHODOLOGY

A. Data

The model is implemented considering the national river of the India country- named Ganga River, that is as wellconsidered massive river basin in the country of India. Therange of 2,525 km(i.e., 1,569 mi) the streamflows inwestern Himalayas inside thecountry of India Uttarakhandstate,&tidesthe southern&the easternbya Gangetic unadorned of the Northern side India. This stream lies between the Himachal Pradesh, Eastern longitudes 73°30 &89° 0 &the Northern latitudes 22°30 & 31°30, coverthe 1,086,000 square kmpart. The ganga basin has about 79% area in the country of India & the residual areas are inside the Nepal & Bangladesh. River Ganga basin is covering the 11 states, these states areU.P., Uttarakhand, M.P., Haryana, Rajasthan, Chhattisgarh, Jharkhand, West-Bengal, Bihar& theDelhi [12]. The figure-2 shows Ganga river basin map in the India country.

The monthly data to defineGanga's WQcharacterized by 5parameters of WQi.e., pH, temperature, dissolved oxygen termed as (DO), biochemical oxygen demand termed as (BOD) & totalcoliform termed as (TC) was composedon the period of time 2002-to-2016. This record of 654 data is which is partitionedusing 60%-40% for training and testing resp.

Fig. 2. The Ganga River Basin

B. Artificial Neural Network (ANN)

Artificial neuron is fundamentally an engineering types method of the biological neuron. This has a tool with several i/p's& single o/p. The ANN is made up of a huge number of easy processing elements which are associated with each of the other &also layered.

A neuron which is in ANN is stimulatedthroughthe biological neuron which is existing in our nervous system. In the biological neuron there are key cell bodies named Soma (the elliptical types structure), & the axon & several dendrites.

Everyneuron can be either in a state 1st one is firing / 2nd one is rest. Existing synapse is a reedygap between the axonof a single neuron & the dendrites of the alternativeneuron. The synapseson dendrites /skin of neuron givesthe signs tothe natural neurons. As soon asthe received signssurpassthe particular threshold, the neuron become active& emits thesignby the axon. This signcantriggerother neuronsthruthe synapse.

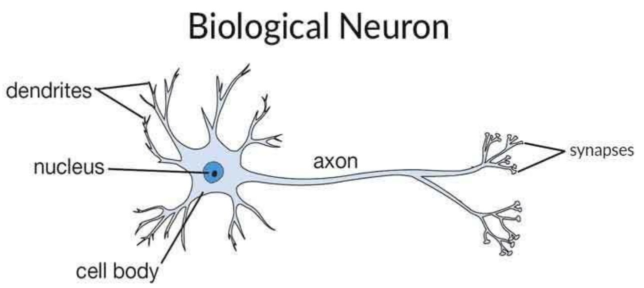

Fig. 3. Basic Structure of Neural network

Effectiveness of synapse changes causes to learn to start. Basicsystem of artificial neuron is similar toneuron of human &this is formedfor passing the messages on if amount of i/p's ismore than particular threshold. The neural networks arecapableofremovingthe patterns &findings the trends from extremely complicated& non-linear issues in the effectiveway than with resolutiongiventhroughanothertechnologies. The evolved archetypal on the basis of ANN approach nowperformsthe form of an "expert" system, that can workto make logic-baseddecision-makingworkthroughscrutinizing the given data. Biggest benefit of the neural network is that the nature of the relationship has the capabilityto archetypalthe compositethe non-linear relationship except the prioriassumptions.

C. Proposed Methodology

Step1: Load the dataset.
Step2: Get only mean value of Temperature, Dissolved Oxygen, pH, B.O. Demand, Total Coliform.

$$Mean = \frac{Min + Max}{2}$$

Step3: Clean the dataset by filling empty spaces with zero value.
Step4: Calculate the water quality index (WQI).

$$\text{WQI} = \frac{\sum W_i Q_i}{\sum W_i}$$

Where,
Qi = quality rating,
WI = Unit weight

Quality rating for each parameter was calculated by using the following equation

$$Q_i = \frac{(W_{Actual} - W_{Ideal})}{(W_{Stabdard} - W_{Ideal})} \times 100$$

Where,
Q_i = Quality rating of ith parameter for a total of n water quality parameters.
W_{Actual} = Actual value of the water quality parameter obtained from laboratory analysis
W_{Ideal} = ideal value of that quality parameter can be obtained from the standard tables.
W_{Ideal} for pH = 7 and for other parameters it is equating to zero and DO
W_{Ideal} = 14.6 mg / L
$W_{Standard}$ = Recommended WHO standard of the water quality parameter.

Calculation of Unit weight (Wi):

$$W_i = \frac{K}{S_i}$$

Where,
Wi = Unit weight for nth parameter,
Si = Standard permissible value for nth parameter
K = proportionality constant, For the sake of simplicity, K is assumed as 1.

Step5: Train the feed forward Neural Network (configuration:Tan Sigmoid and Log sigmoid) using trainbr, trainb, trainr function. The training goal is set to the value of 0.1.
Step6: Obtain Mean Square Error Value.

IV. EXPERIMENTAL RESULT & ANALYSIS

Inside the section of result analysis, experiment ofimplemented work performed by using MATLAB 2018a.For this research work, we have utilized the MATLAB Neural Network Toolbox. Theprogram which is based on ANN is established inthe too named as the toolbox of MATLAB Neural Network. Inside themodel of neural networktermed as (NN) through utilizing the MATLAB tool, herevarious trainingfunctions. Here, the neural network istrained using trainbr (Bayesian regularizationBackpropagation).Log sigmoid and tan sigmoid are the two-transfer function that are used in the propose model. The configuration of neural network is taken as the 5-10-1 layered structure i.e. the 5 i/p parameters presenting the 5 i/p nodes in neural network architecture, the 10 neurons in hidden layer &1o/p nodepresenting the o/p,which is shown in figure 4.

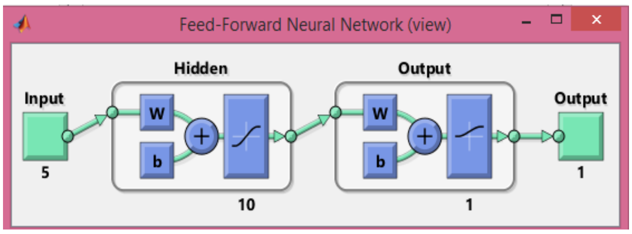

Fig.4.Architecture of ANN Model (5-10-1)

Forecastbehavior of theevolvedarchetypal is calculated by the mean square error named as (MSE) statistical method. This is anaverage of the squares of errors / deviations. Also, neural network is trained by taking different training functions i.e.trainrnamed as (random order incremental training with learning functions) and trainb named as (Batch training with weight & bias learning rules). Comparison results shows that the neural network is best trained with trainbr function with minimum MSE (0.004608) at 0.09 learning rate (LR). Table I shows the performance prediction parameter of the neural network in which they are best trained in.Thus, this model gives an accuracy of 99.5% which comes out to be superior to allother developed models for forecastingWQ of the River Ganga.

In our research work, the predictive model based on water quality at river ganga provide their best accuracy. Training function trainbr provide its maximum accuracy with the learning rate 0.09 & goal 0.1. So, we show the best results in this paper.

Fig.5. shows the confusion matrix of training function trainbr. This represent the 99.5% correct classification.Aconfusion matrix is the summary of the outcomes of prediction on theclassification issue.The number of correct & incorrect forecasts are summarizingincludethe count-values &the broken down through everyclass. The confusion matrix demonstrates methods when our classification archetypal is confused when thisdoesforecasts.This not only provides insights into errors being createdthroughthe classifier neverthelessis also more important about the error'stypes being done.

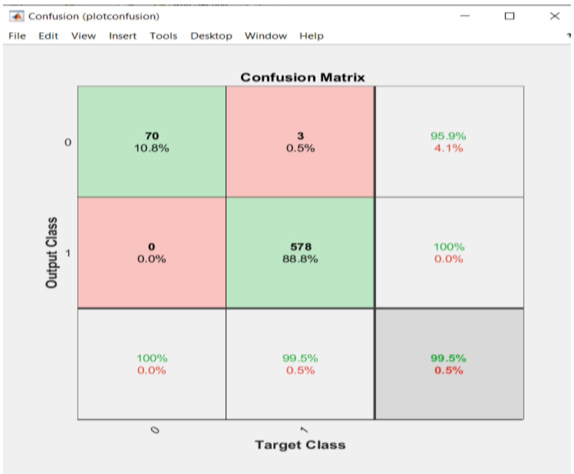

Fig.5.Confusion Matrix

Fig.6.shows the performance graph of training function trainbr at 0.09 learning rate, produce by NN toolbox.Usually,error decreasesafterward more of the trainingepochs, neverthelesscanbegin to intensification on the data-set of validation because n/wbegins over-fitting the data of training. Withinthe default types setup, the training haltsafterwardthe 6 successiveintensifications in the validation error, & best training performance is the 0.079956 at epoch 5 is occupied from epoch with minimum validation error.

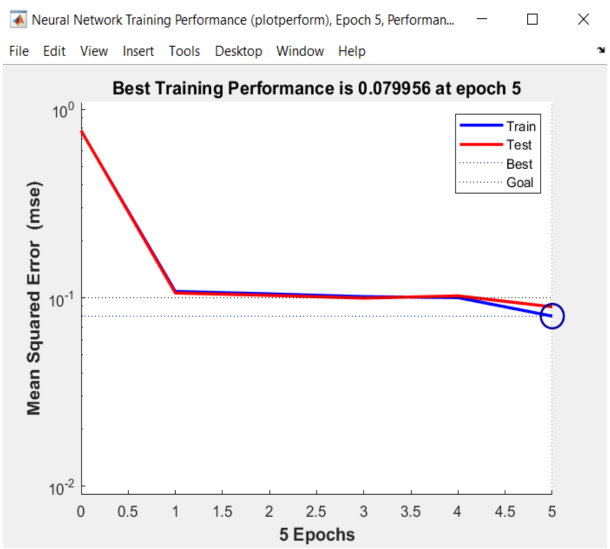

Fig.6.Performance Graph

Fig.7. represent the training state of training

function trainbr at 0.09 learning rate which produce by the neural network toolbox.The training state displaysus some of the other types trainingstatistics. The gradient is a gradient value of back propagation over every iteration in the logarithmic scale 5e-7, that means reachinglowest of the local minimum of theneeded function.Validation fails are the iterations oncethe MSE of validation booststheir value. Many failuresmeanthe overtraining.The MATLAB toolautomatically halts training after the 6 fails in a row. Mu is the training gain it must be between 0.8-1 in neural network, itapproximates the inverse of the hessian matrix which is very complicated function. This algorithm has become converged if sum squared error termed as (SSE) & the sum squared weights termed as (SSW) are comparativelystableon many iterations. As soon asithappens,weneed to push "Stop Training" key in training window.

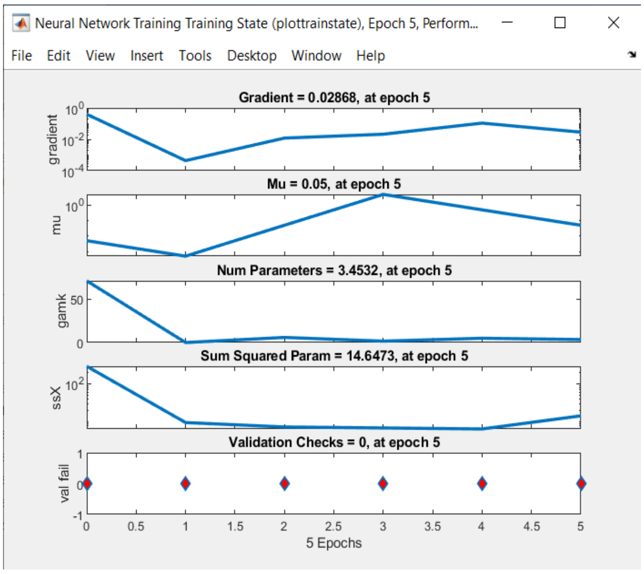

Fig.7.Training State

Fig.8. represent the error histogram with 20 bins if training function trainbr at 0.09 learning rate. This shows how exactlyour trained archetypal fits a dataset. If R-squared value is nearbythe 1 (good), then thissignifies that model's prediction is very nearbythe real data-set. If it is zero (bad) thisdisplays that ourarchetypaltotally fails in creatingtheexactforecast. Give theextra verification of the performance of network. Thisspecifiesthe outliers.Blue, red&orange color bars signifythe training-data, testing-data &0error respectively. The most of the data fall on training data, test bar represent the data falls on testing set, & 0 error bar represent the minimum error rate occur on each instance which lies near to 0.

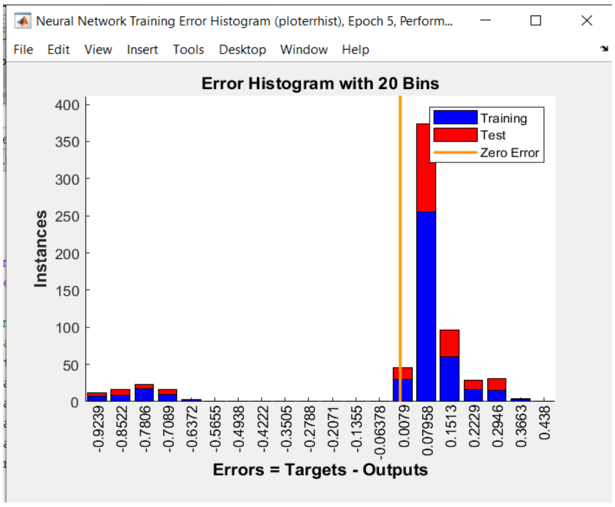

Fig.8.Error Histogram

Fig.9. represent the regression of training function trainbr at 0.09 learning rate.The plot ofregression provide relations between network &targets the o/p's. If training wasfaultless, then o/p's of network & targets would be preciselythe same, neverthelessrelationship in practice is hardlycorrect. The three plots denote training,test& average of training & testing dataset. In each of the plot, dashed line signifiescorrect result – outputs = targets. Solid types line signifies best fit linear regression line between the o/p's& the targets. The value of R is thesign of relationsbetween o/p's & the targets. If the value of R = 1, this shows that here a precise linear relationship between the o/p's& the targets. If the value of R is nearbythe0, then here no linear relationship between the o/p's& the targets.

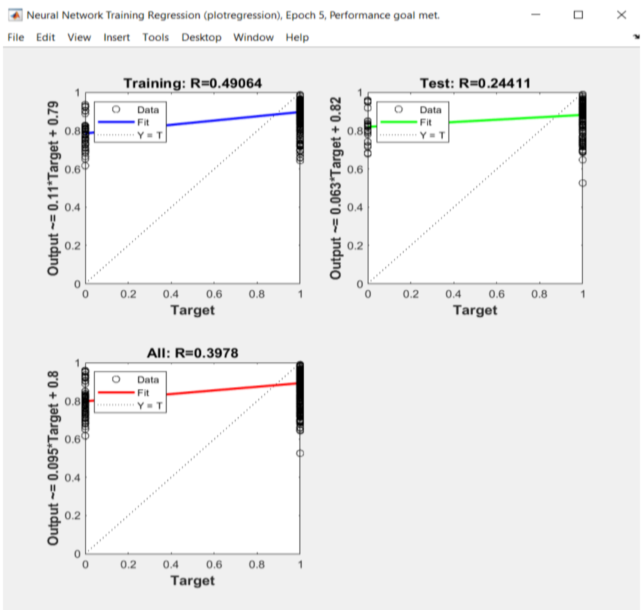

Fig.9.Regression

Table I. Performance prediction parameter of ANN (goal=0.1)

Learning rate	**0.01**	**0.03**	**0.05**	**0.07**	**0.09**
Training function	**Mean Square Error**				
trainlm	0.251	0.209	0.175	0.282	0.163
trainr	0.112	0.163	0.121	0.107	0.087
trainb	0.124	0.127	0.173	0.117	0.109
trainbr	0.107	0.094	0.083	0.097	0.004

The following graph fig.10. shows the performance graph of neural network with different training function trained at different learning rate.

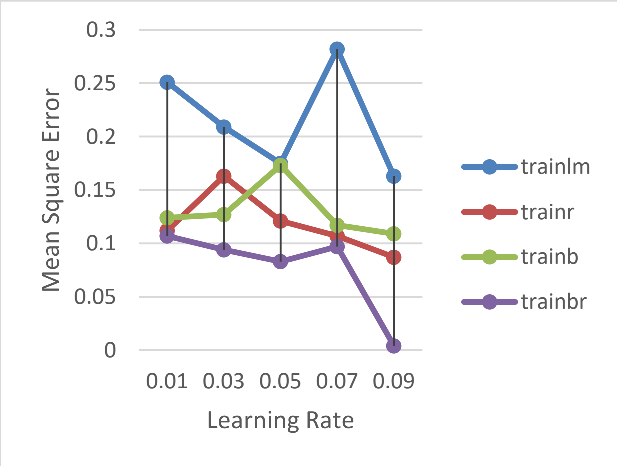

Fig. 10. Performance of ANN model

V. CONCLUSION

The management of water sources includes monitoring &WQ managingalong with the WQ monitoring &managing. The several models can help in forecasting the WQ effects of another land & water management policies and practices. It is done by using MATLAB tool. Proposedcomputational model is developed using ANNs concept. It was found that ANNs are able of forecastingthe WQ of the Ganga River with thetrainbr function at learning rate set to 0.09 giving an accuracy of 99.5% using the best model as per the present case.

REFERENCES

1. W. Chine, T. Wang, L. Chen, C. Kou, "Artificial Neural Networks for Water Quality Prediction in a Reservoir", 2009 Second International Workshop on Computer Science and Engineering.
2. J. Lu, T. Huang, "Data Mining on Forecast Raw Water Quality from Online Monitoring Station Based on Decision-making Tree", 2009 Fifth International Joint Conference on INC, IMS and IDC.
3. Priya Singh and Pankaj Deep Kaur, "Review on Data Mining Techniques for Prediction of Water Quality", International Journal of Advanced Research in Computer Science, 8 (5), May-June 2017,396-401.
4. Hamza Ahmad Isiyaka, Adamu Mustapha, Hafizan Juahir and Philip Ogbonnia Phil-Eze, "Water quality modelling using artificial neural network and multivariate statistical techniques", Modeling Earth Systems and Environment, 2018, pp. 1-11.
5. A.K. Bisht, Dr. Ravendra Singh, Rakesh Bhutiani and Ashutosh Kumar Bhatt, "An Investigation of Assessment and Modelling the Water Quality of Rivers Based on Artificial Neural Networks-An Initiative towards the River Ganga", International Journal on Future Revolution in Computer Science & Communication Engineering, Volume: 3 Issue: 8, 17 – 23, August 2017.
6. Y. Khan, C. S. See, "Predicting and Analyzing Water Quality using Machine Learning: A Comprehensive Model", IEEE, 2016.
7. Shakibaeinia, A., Kashyap, S., Dibike, Y. B., & Prowse, T. D. (2016). An integrated numerical framework for water quality modelling in cold-region rivers: A case of the lower Athabasca River. Science of The Total Environment, 569-570, 634–646.
8. S. Y. Muhammad, M. Makhtar, A. Rozaimee, A. A. Aziz, A. A. Jamal, "Classification Model for Water Quality using Machine Learning Techniques", International Journal of Software Engineering and Its Applications Vol. 9, No. 6 (2015), pp. 45-52.
9. Ming Hu, Guo H. Huang, Wei Sun, Yongping Li, Xiaowen Ding, Chunjiang An, Xiaofei Zhang and Ting Li, "Multi-objective ecological reservoir operation based on water quality response models and improved genetic algorithm", Journal Engineering Applications of Artificial Intelligence,Volume 36 Issue C, November 2014, Pages 332-346.
10. Pei Zhao & Xiangyu Tang & Jialiang Tang & Chao Wang, 2013. "Assessing Water Quality of Three Gorges Reservoir, China, Over a Five-Year Period From 2006 to 2011," Water Resources Management: An International Journal, Published for the European Water Resources Association (EWRA), Springer;European Water Resources Association (EWRA), vol. 27(13), pages 4545-4558.
11. S. Wechmongkhonkon, N.Poomtong, S. Areerachakul, "Application of Artificial Neural Network to Classification Surface Water Quality", World Academy of Science, Engineering and Technology International Journal of Environmental and Ecological Engineering Vol:6, No:9, 2012, pp. 574-578.
12. Status Paper On River Ganga, "State of Environment and Water Quality," National River ConservationDirectorate Ministry of Environment and ForestsGovernment of India, August, 2009.
13. Anima upadhyay and Chandrakala M,"Water Quality Index of Ganga River Water, Rishikesh, Uttarakhand, India.", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 5 Issue XI November 2017,PP-2876-2880.

AUTHORS PROFILE

Yamini Sonihas got B.E in Information Technology from Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal in the year of 2017. She is a PG research scholar in Department of CSE/IT, Madhav Institute of Technology and Science, Gwalior, India. She has one paper publish in refereed journal.

Vikas Sejwarhas got B. E in Information Technology from Madhav Institute of Technology and Science, Gwalior, India, in 2006 and M. TECH. in Information Technology from School of Information Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal in 2008. He is presently working as an Assistant Professorin Madhav institute of Technology and Science, Gwalior, India andhis research experience of 10 years. He has published 23 research papers in different refereed journals and 07 papers presented in different IEEE conferences.He is member of IEEE, AIENG and CSI Societies.