Application of ANN for Water Quality Index

Rajiv Gupta, A N Singh, and Anupam Singhal

***Abstract*—Attempt has been made to create a Water Quality Index (WQI) based on artificial neural network (ANN) and globally accepted parameters. Several methods to measure WQI are available in the research and ambiguity problems exist where all the sub-indices of WQI are acceptable but overall index is not acceptable. In this study, we have tried to develop the WQI based on the WHO (world Health Organization) parameters (Dissolved Oxygen, pH, Turbidity, E. Coli and Electric Conductivity). The results also reveal changes in ANN based result from various input neural network model and its parameters. Even within same model, changes occur with variation in parameter. Based on the statistical parameter of regression value, the parameter and network model would be selected. With the dataset created for this study have shown the Cascade network is best for predicting the WQI.**

***Index Terms*—Artificial neural network, cascade network, water quality index, who parameters.**

I. Introduction

The quality of water is defined in terms of its physico-chemical and biological parameters WHO [1]. Ascertaining its quality is crucial before use for various intended purposes such as potable water, agricultural, recreational and industrial water uses etc. For effective planning and management of groundwater resources, groundwater vulnerability assessment is significant. [2] Traditional approaches to assessing water quality are based on a comparison of experimentally determined parameter values with existing guidelines. In many cases, the use of this methodology allows proper identification of contamination sources and may be essential for checking legal compliance. However, it does not readily give an overall view of the spatial and temporal trends in the overall water quality in a watershed. One of the difficult tasks faced by water expert is to transfer interpretation of complex environmental data into information that is understandable and useful to technical and policy makers as well as the common people. A number of attempts have been made to produce a methodology that meaningfully integrates the data sets and converts them into handy information.

Since 1965, when Horton [3] proposed the first water quality index (WQI), a great deal of consideration has been given to the development of 'water quality index' methods with the intent of providing a tool for simplifying the reporting of water quality data. The WQI concept is based on the comparison of the water quality parameter with respective regulatory standards. The development process of a water quality index can be generalized in four steps [4]: (i) Selecting the set of water quality variables of concern: Parameter selection; (ii) Transformation of the different units and dimensions of water quality variables to a common scale: Developing sub-indices; (iii) Weighting of the water quality variables based on their relative importance to overall water quality: Assignment of weights; and (iv)Formulation of overall water quality index: Aggregation of sub-indices to produce an overall index.

Conventional methods for water quality assessment do not consider the uncertainties involved either in measurement of water quality parameters or in the limits provided by the regulatory bodies. Application of fuzzy rule based optimization model is illustrated with twenty groundwater samples from Sohna town of Southern Haryana, India by Bhupender *et al.* [5]. Intrinsic uncertainties and subjectivities of environmental problems have been increasingly dealt by using computation methods based on artificial intelligence. A new index, called fuzzy water quality index (FWQI) is developed to correct perceived deficiencies in environmental monitoring, water quality classification and management of water resources in cases where the conventional, deterministic methods can be inaccurate or conceptually limited. [6]. Contamination of groundwater has become a major concern in recent years. Since testing of water quality of all domestic and irrigation wells within large watersheds is not economically feasible. Overall goal of work done by Dixon *et al.* [7] is to improve the methodology for the generation of contamination potential maps by using detailed land use / pesticide and soil structure information in conjunction with selected parameters from the DRASTIC model. A groundwater vulnerability and risk mapping assessment is done by Nobre *et al.* [8]. A modified version of the DRASTIC methodology was used to map the intrinsic and specific groundwater vulnerability of a 292 km^2 study area. Numerical modeling was performed for delineation of well capture zones, using MODFLOW and MODPATH.

The fluoride vulnerability map is produced for four villages of Jind district of Haryana state, India by Meenakshi *et al.* [9]. A systematic calculation of correlation coefficients among different physico-chemical parameters was performed. The analytical results indicated considerable variations among the analyzed samples with respect to their chemical composition. [9]

Forecasting the ground water level fluctuations is an important requirement for planning conjunctive use in any basin. Nayak *et al.* [10] investigated the potential of artificial neural network technique in forecasting the groundwater level fluctuations in an unconfined coastal aquifer in India. In general, the results suggest that the ANN

Manuscript received May 25, 2019; revised July 12, 2019.

Rajiv Gupta, A N Singh and Anupam Singhal are with the Dept. of Civil Engineering, BITS-Pilani, Rajasthan, India (e-mail: rajiv@pilani.bits-pilani.ac.in).

doi: 10.18178/ijmlc.2019.9.5.859

models are able to forecast the water levels up to 4 months in advance reasonably well. Yilmaz *et al.* [11] proposed an index model for quality evaluation of surface water quality classification using fuzzy logic. Patric *et al.* [12] worked on the Chillan river for quality condition and performed PCA for WQI. On the basis of the results from a Principal Component Analysis (PCA), modifications were introduced into the original WQI to reduce the costs associated with its implementation. Available water quality indices have some limitations such as incorporating a limited number of water quality variables and providing deterministic outputs. Reza Nikoo *et al.* [13] presents a hybrid probabilistic water quality index by utilizing fuzzy inference systems (FIS), Bayesian networks (BNs), and probabilistic neural networks (PNNs). Their results showed that the average relative error in the validation process of the trained BN is 7.8% and both the trained BN and PNN can be accurately used for probabilistic water quality assessment.

Modeling of groundwater vulnerability, reliably and cost effectively for non-point source (NPS) pollution at a regional scale remains a major challenge. In recent years, Geographic Information Systems (GIS), neural networks and fuzzy logic techniques have been used in several hydrological studies. The strength of this method is that it offers a means of dealing with imprecise data, therefore, is a viable option for regional and continental scale environmental modeling where imprecise data prevail. Neuro-fuzzy techniques integrated with a GIS lends itself to a sophisticated overlay and index model and have the potential for facilitating modeling groundwater vulnerability at a regional scale as it is capable of dealing with imprecise data. Author suggested that due to the potentially subjective nature of the neural networks, fuzzy logic and neuro-fuzzy methods, caution should be taken while using them. [14]. Sundarambal *et al.* (2008) work, artificial neural networks (ANNs) was used to predict and forecast quantitative characteristics of water bodies. Author(s) suggested the true power and advantage of ANNs method lie in its ability to (1) represent both linear and non-linear relationships and (2) learn these relationships directly from the data and provide easy way to model them. They concluded that a trained ANN model may potentially provide simulated values for desired locations at which measured data are unavailable yet required for water quality models. [15]. Lee *et al.* (2002) investigated the effectiveness of artificial neural network models for predicting the Water Quality Index for rivers in Malaysia. The network was trained with reference to seven major parameters for the determination of the Water Pollutant Index, Water Quality Index and Water Quality Class, for rivers in Pahang and Selangor. Artificial neural network models with different learning approaches, such as back propagation neural network, modular neural network and radial basis function network, are considered and adopted to model the Water Quality Index. In his work, it is *observed* that the RMS error was between 0.11 to 0.12 and accuracy of this *ANN* prediction is found to be 92.4 percent to 99.96 percent [16]. Three layer perception neural network was used for determining WQI in Kinta river of Malaysia and the correlation was 0.977 using feed forward method [17], [18]. They [19] designed a ANN model to predict WQI using land use areas as predictors and used the statistical data of 10 year of land use and water quality in the area. Numerous physical, chemical, and biological parameters are involved in the environment–organism relationship. Thus to analyze better, use of self organizing map, WQI, and PCA are proposed to obtain the classification and pollution status in water samples [20]. Delbari *et al.* [21] proved that the spatial correlation always exists between the samples and physico chemical analysis. Rajib *et al.* [22] lead to water quality improvement with reduction in surface runoff, sediment, nitrate and total phosphorus loadings. The WQI values can even range over 5000 [23], so the classification of WQI in safe and non-safe categories varies from case to case. The approach adopted by Li *et al.*, [24] combine particle swarm optimization, chaos theory, self-adaptive strategy and back propagation artificial neural network to evaluate the water quality.

II. Methodology

Methodology is presented in Fig. 1 and Fig. 2. Fig. 1 indicates the application of ANN to predict the WQI, whereas Fig. 2 represents the methodology to develop database for different parameters.

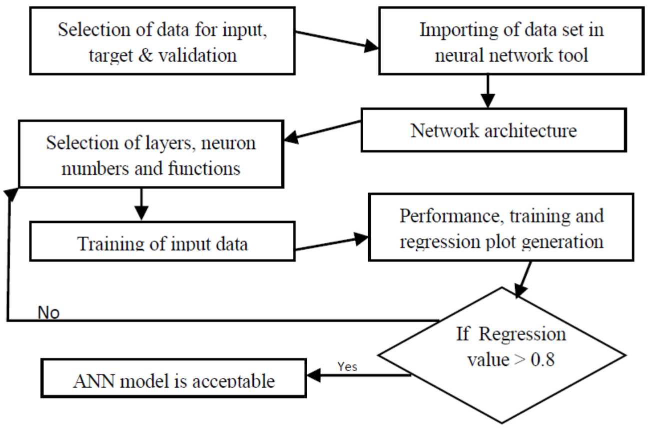

Fig. 1. Methodology to apply ANN for WQI.

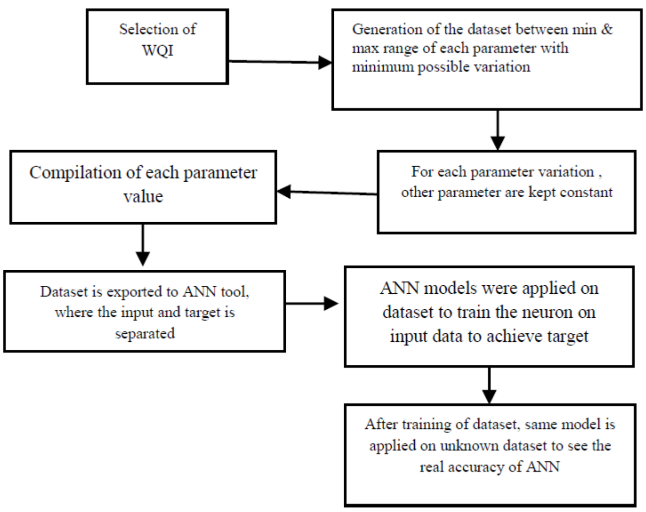

Fig. 2. Methodology for developing WQI database.

TABLE I: WQI Parameter Used in WQI Estimation with ANN

pH	DO	Turbidity	E Coli	EC	Limit
(1-14)	(0-12)	(0-10)	(0-100)	(0-100)	range
(6.5-8.5)	(4.0-12.0)	(0-5)	0	(4-30)	safe limit

The dataset developed for WQI consisting the 5 basic parameters and their ranges are given in Table I.

pH indicates the sample's acidity but is actually a measurement of the potential activity of hydrogen ions (H^+) in the sample. All organisms work and survive in a specific pH range. DO makes the water tasty. Turbidity indicates the clarity of water. Higher the clarity of water, turbidity will be less and taste of it is good. If any of the E. Coli is present in water then it can cause problem in digestive system of humans. Electric conductivity is a measure of presence of various chemical ions and higher EC indicate more ions in water.

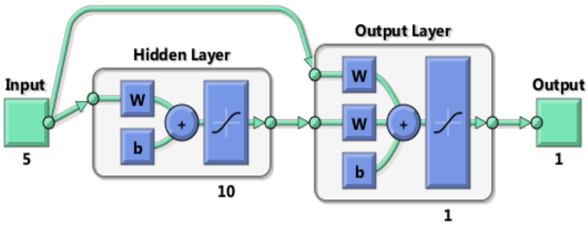

Fig. 3a. Neural network design for cascade forward network.

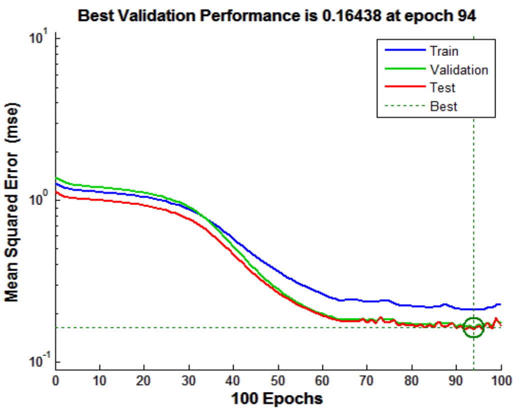

Fig. 3b. Performance plot for cascade forward network.

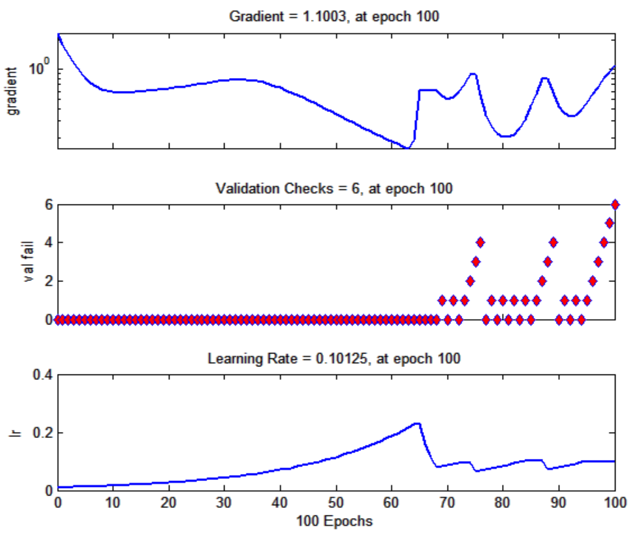

Fig. 3c. Training state plot for cascade forward network.

III. Result and Discussion

The database for the 5 parameters viz., pH, DO, Turbidity, E. Coli, EC and TDS is created. In database, the known limits for the range (min –max) in which the parameters can vary above & below the permissible limit should be considered as safe. The coding with respect to permissible ranges of the parameter is assigned. Values lies in permissible range will be considered as safe and values lying either side will be considered as unsafe.

Cascade Forward network and Fast forward Network have been studied for the dataset and modeled. Their comparative outcome is given in Table II. The design of neural network, performance plot, training plot and regression plot for Cascade and Fast forward network are given in Fig 3a – Fig. 3d and Fig. 4a – Fig. 4d respectively.

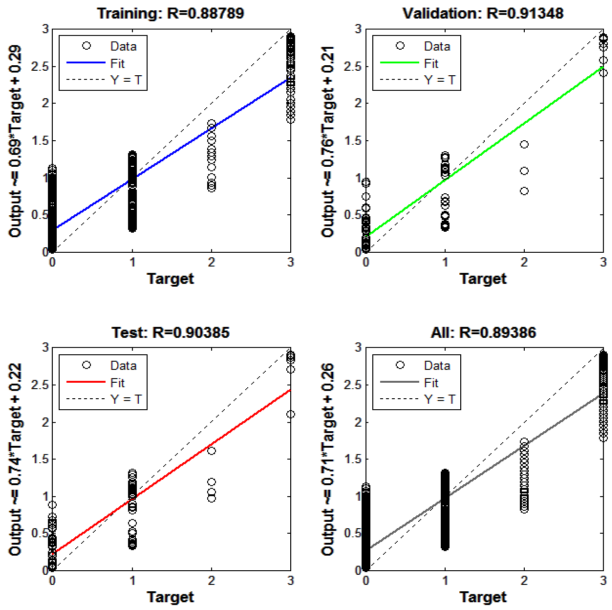

Fig 3d. Regression plot for cascade forward network.

TABLE II: OUTPUT OF CASCADE AND FAST FORWARD NETWORK

Parameters	Cascade network	Fast forward network
Gradient	1.1003	0.27645
Val fail	6	6
Lr	0.10125	0.3481
Training R	0.88789	0.85165
Validation R	0.91348	0.87359
Test R	0.90385	0.8535
All R	0.89386	0.85613

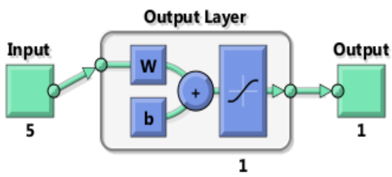

Fig. 4a. Neural network design for fast forward network.

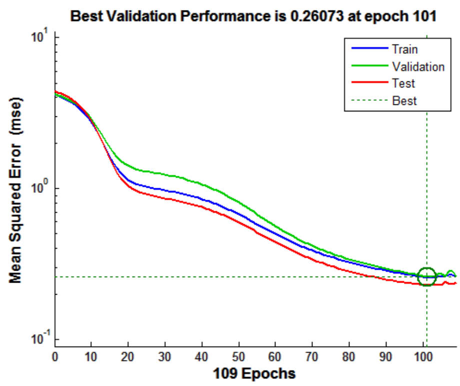

Fig. 4b. Performance plot for fast forward network.

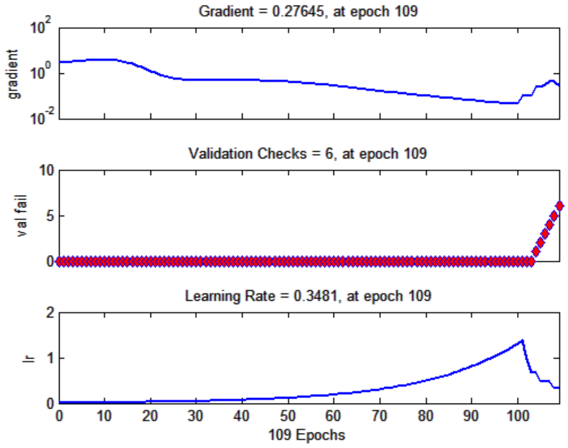

Fig. 4c. Training state plot for fast forward network.

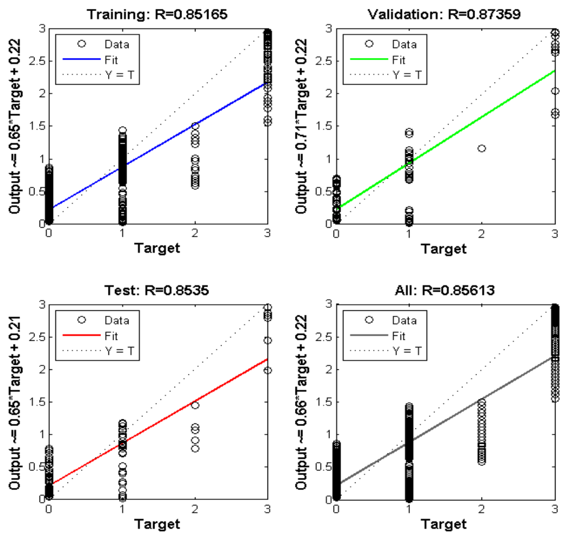

Fig. 4d. Regression plot for fast forward network.

The modeling efforts showed that the optimal network architecture was Cascade forward backprop. The best WQI predictions were associated with the gradient descent with adaptive learning rate, Traingda as training algorithm & Learngdm as performance function; a learning rate of 0.101; and a gradient coefficient of 1.1. The WQI predictions of this model had significant, positive, high correlation (r=0.89) with the measured WQI values, implying that the model predictions explain around 89% of the variation in the measured WQI values.

The fast forward backprop network also showed the WQI prediction capabilities. In it Traingda as training algorithm & Learngdm as performance function were used. It had a learning rate of 0.348 & gradient coefficient of 0.276. It had high correlation (r=0.85) with measured values but slightly less than Cascade forward model.

The cascade model had achieved the learning rate of more than 0.2 at 65 epoch while fast forward model achieved rate of more than 1.5 at epoch 100. This shows that the learning rate is less in cascade model but the R is high while fast forward model achieved the higher learning rate but R value is less than cascade.

During the neural network implication, three classes were developed based on rating given to their values provided.

- Class A: Rating given 1 - This indicate the values are less than desirable limit
- Class B: Rating given 2 - This indicate the values are in permissible limit of use
- Class C: Rating given 3 - This indicate the value are above the permissible limit

The condition applied in training of the neuron to assign the quality index is given in table3. Class “B” category of water suggests that it is safe for direct use and no treatment is required. While in Class “C” category of water need treatment based on its chemical and biological characteristic before used by human. Class “A” category of water could be used directly in absence of alternate source but with some food supplement. The outcome of WQI based on ANN would be predicted as:

- WQI=1, Below desirable limit (Poor water quality)
- WQI=2, Desirable water quality (excellent water quality)
- WQI=3, Above permissible limit (Rejected water quality)

We have tested the nine model present in neural network tools in Matlab R2012. These models are Self organizing network, Probabilistic network, Perceptron network, NARX network, LVQ network, Elman backprop network, Competitive network, Cascade forward backprop and Fast forward backprop network. Out of these nine networks, only four networks have accepted the data. Out of these four, only two network have completed the process of generating regression plots. So we can say that only the two network model of cascade and fast forward really work on data. Further for these two network type various combination have tried within the model and found that only a single combination in each model have given high regression value.

With the selected network type, the input and target data were fixed. With this we had used fourteen different training function, two adaptation learning function, three performance function and three transfer function. We had found with our dataset that in transfer function out of three (TANSIG, LOGSIG, PURELIN) only TANSIG gives acceptable result. In performance function (MSE, MSEREG, SSE) only MSE give the acceptable results. Under performance function (LEARNGDM, LEARNGD) LEARNGDM gives the best and in training functions (TRAIN CGF, TRAINBGF, TRAINBR, TRAINCGB, TRAINCGP, TRAINGD, TRAINGDM, TRAINGDA, TRAINGDX, TRAINLM, TRAINOSS, TRAINR, TRAINRP, TRAINSCG) TRAINGDA gives the best results. Number of layer chosen to one and neuron numbers in the layer is ten.

TABLE III: CHANGES ASSOCIATED WITH VARIATION OF TRAINING MEMBERS

Network	**Parameter**	**Output with**	**90 training members**	**180 training members**	**560 training members**
Cascade	TF: TRAINGDA ALF:LEARNGDM PF: MSE TF: TANSIG	Gradient Learning rate All R	0.6293 0.15947 0.854	0.72872 0.1239 0.85184	1.1003 0.10125 0.89386
Feed	TF: TRAINGDA ALF:LEARNGDM PF: MSE TF: TANSIG	Gradient Learning rate All R	0.42883 0.20353 0.83962	0.42827 0.11584 0.86755	0.27645 0.3481 0.85613

The result of variation for two models is shown in Table

III and it is reflected that: (a) In cascade model-

As the number of training members increases the gradient also increases, learning rate decreases and value of R increases; (b) In Feed model- As the members increased in training, the gradient decreases, learning rate first decrease then increased and value of R first increase then decreased. Cascade model has shown the pattern of increasing R with increase in member numbers while Feed haven't shown, thus cascade with associated variable are selected for final training of dataset.

When layers are changed from one to five and ten in network, the gradient decreased in 5 layer and again increased in 10 layer. It even crossed the gradient of single layer. Learning rate decreased respectively with increase in layer numbers. The value of regression also increased then decreased respectively with 5 layer network and 10 layer networks.

IV. Conclusion

ANN could help in saving resources in predicting the WQI of the large dataset, as ANN model had the training dataset and it would be easy to modify the training dataset in the ANN tools. The network could be trained with reference to 5 major parameters for the determination of the water quality index, for Indian sub continent. The data used as a training set could be used globally. The Cascade model have shown its ability to predict the WQI when the five parameters defined by WHO, would be used. The variation in network parameters can alter the result but indicated that the higher the number of member in training dataset, the regression value would be higher with learning rate and gradient. The 5 layer network model give the maximum regression. The ANN could have ability to simplify the complexity of interpretation of WQI and the index could be universal because only 3 categories will be displayed and out of them only one class represent the accepted WQI and rest of two would be rejected. The few problems still remain with ANN but could be solved out in future.

References:

[1] World Health Organization, 2017, *Guidelines for Drinking-Water Quality*, 4th ed., World Health Organization, 2017.

[2] S. K. Garewal, A. D. Vasudeo, V. S. Landge, and D. G. Aniruddha, "A GIS-based Modified DRASTIC (ANP) method for assessment of groundwater vulnerability: A case study of Nagpur city," *Water Quality Research Journal*, vol. 52, no. 2, pp. 121-135, India, 2017.

[3] R. K. Horton, "An index number system for rating water quality," *J. Water Pollu. Cont. Fed*., vol. 37, no. 3, pp. 300-305, 1965.

[4] H. Boyacioglu, "Development of a water quality index based on a European classification scheme," *Water SA*, vol. 33, no. 1, pp. 101-106, 2007.

[5] B. Singh, S. Dahiya, S. Jain, V. K. S. Kushwaha, "Use of fuzzy synthetic evaluation for assessment of groundwater quality for drinking usage: A case study of Southern Haryana," India, *Environ Geol*., vol. 54, pp. 249–255, 2008.

[6] A. Lermontov, L. Yokoyama *et al*., "River quality analysis using fuzzy water quality index: Ribeira do Iguape river watershed," Brazil, *Ecological Indicators*, vol. 9, pp. 1188–1197, 2009.

[7] B. Dixon, "Groundwater vulnerability mapping: A GIS and fuzzy rule based integrated tool," *Applied Geography*, vol. 25, pp. 327–347, 2005.

[8] R. C. M. Nobre, O. C. Rotunno Filho, W. J. Mansur, M. M. M. Nobre, and C. A. N. Cosenza, "Groundwater vulnerability and risk mapping using GIS, modeling and a fuzzy logic tool," *Journal of Contaminant Hydrology*, vol. 94, pp. 277–292, 2007.

[9] V. K. Garg, A. Malik *et al.*, "Groundwater quality in some villages of Haryana, India: focus on fluoride and fluorosis," *Journal of Hazardous Materials*, vol. 106B, pp. 85–97, 2004.

[10] P. C. Nayak, Y. R. S. Rao, and K. P. Sudheer, "Groundwater level forecasting in a shallow aquifer using artificial neural network approach," *Water Resources Management*, vol. 20, pp. 77–90, 2006.

[11] Y. Icaga, "Fuzzy evaluation of water quality classification," *Ecological Indicators*, vol. 7, pp. 710–718, 2007.

[12] P. Debels, R. Figueroa, R. Urrutia, R. Barra, and X. Niell, "Evaluation ofwater quality in the chilla´ n river (central chile) using physicochemical parameters and a modifiedwater quality index," *Environmental Monitoring and Assessment*, vol. 110, pp. 301–322, 2005.

[13] M. R. Nikoo, R. Kerachian *et al.*, "A probabilistic water quality index for river water quality assessment: A case study," *Environ Monit Assess*, vol. 181, pp. 465–478, 2011.

[14] B. Dixon, "Applicability of neuro-fuzzy techniques in predicting ground-water vulnerability: A GIS-based sensitivity analysis," *Journal of Hydrology*, vol. 309, pp. 17–38, 2005.

[15] S. Palani, S.-Y. Liong, and P. Tkalich, "An ANN application for water quality forecasting," *Marine Pollution Bulletin*, vol. 56, pp. 1586–1597, 2008.

[16] L. Y. Khuan, N. Hamzah, and R. Jailani, "Prediction of Water Quality Index (WQI) based on Artificial Neural Network (ANN)," in *Proc. 2002 Student Conference on Research and Development*, Shah Alam, Malaysia

[17] N. M. Gazzaz, M. K. Yusoff, A. Z. Aris, H. Juahir, and M. F. Ramli, "Artificial neural network modelling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors," *Mar Pollut Bull*., vol. 64, no. 11, pp. 2409-2420, Nov. 2012.

[18] N. M. Gazzaz, M. K. Yusoff, M. F. Ramli, H. Juahir, and A. Z. Aris, "Artificial neural network modeling of the water quality index using land use areas as predictors," *Water Environ Res*, vol. 87, no. 2, pp. 99-112, Feb. 2015.

[19] K. Sinha and P. D. Saha, "Assessment of water quality index using cluster analysis and artificial neural network modeling: A case study of the hooghly river basin," *West Bengal*, India, vol. 54, no. 1, pp. 28-36, 2015.

[20] V. Sheykhi, F. Moore, and A. Kavousi-Fard, "Combining self-organizing maps with WQI and PCA for assessing surface water quality – a case study," *International Journal of River Basin Management*, Kor River, Southwest Iran, vol. 13, no. 1, pp. 41-49, 2015.

[21] M. Delbari, M. Amiri, and M. B. Motlagh, "Assessing groundwater quality for irrigation using indicator kriging method," *Appl Water Sci.*, vol. 6, pp. 371-381, 2016.

[22] M. A. Rajib, A. Laurent, and M. Paul, "Modeling the effects of future land use change on water quality under multiple scenarios: A case study of low-input agriculture with hay/pasture production," *Sustainability of Water Quality and Ecology*, vol. 8, pp. 50-66, 2016.

[23] R. Felaniaina, R. N. N. Jules, M. Zakari, H. R. Eddy *et al*., "Assessment of surface water quality of bétaré-oya gold mining area (east-cameroon)," *Journal of Water Resources and Protection*, vol. 9. pp. 960-984, 2017.

[24] M. Li, W. Wei, B. Chen *et al*., "Water Quality Evaluation Using Back Propagation Artificial Neural Network Based on Self-Adaptive Particle Swarm Optimization Algorithm and Chaos Theory," Computational Water Energy and Environmental, Engineering, vol. 6, pp. 229-242, 2017.

Rajiv Gupta completed his B.E. and M.E. in civil engineering from BITS-Pilani, India in 1983 and 1989 respectively. In 1995, he was awarded with PhD in fluid structure interaction from BITS-Pilani, India. Presently, he has been working as a senior professor at the Dept. of Civil Engineering at BITS-Pilani, India. In his last 3 decades of teaching and research, he has authored a number of books and course development material and published more than 150 research papers in India and abroad. His fields of interests are energy-water conservation, GIS and RS and concrete technology. He is executing various projects related to water domain of worth approx. USD 1M.

Arun N. Singh completed his B.Sc. in geology, physics and mathematics from HNB Garhwal University, Srinagar, India in 2002. He has completed M.Sc. in Remote sensing & GIS application from Bundelkhand University, Jhansi, India in 2005. In 2011, he had done a certificate course from ITC

Netherlands, University of Twenty in applied geophysics.

He completed his Ph.D. from the Dept. of Civil Engineering, BITS-Pilani, India, where his work is concentrated on characterization of groundwater and finding the source of contamination of groundwater in the Shekhawati region of Rajasthan.

Anupam Singhal completed B.E. in civil engineering and M.E. in environment engineering from IIT-R, Roorkee, India in 1992 and 1994 respectively. In year 2006, he was awarded with PhD from IIT Roorkee in Hazardous Waste Management. He has diverse exposure in environmental engineering projects. He has worked as an environmental engineer in industry & consultancy organization for eight years. He has designed many treatment plants for Water & Wastewater and Air pollution control systems and consultant in industry and consultancy organization. He has also carried out many EIA studies. Presently, he is working as an associate professor in the Dept. of Civil Engineering at BITS-Pilani.