JOURNAL GEOLOGICAL SOCIETY OF INDIA
Vol.97, September 2021, pp.1063-1072

Evaluation of Water Quality Based on a Machine Learning Algorithm and Water Quality Index for Mid Gangetic Region (South Bihar plain), India

Amar Nath Gupta*, Deepak Kumar and Anshuman Singh
Department of Civil Engineering, National Institute of Technology Patna - 800 005, India
E-mail: *amarg.pg17.ce@nitp.ac.in*

ABSTRACT

Water quality index (WQI) is an indicator of the quality of any ground water storage in the form of a single number representing a combination of different water quality parameter. Different parameters like that pH, total dissolved solids (TDS), electrical conductivity (ECE), nitrate, sulphate, total hardness, calcium hardness, magnesium hardness, etc. are critical to assess the WQI. Additionally, the precision in the prediction of this parameter affects the quality of the result. In this research, Extreme Learning Model (ELM) and three hybrid variants of the same model, namely, RBF-ELM, Online Sequencing-ELM (OS-ELM), Biogeography-based optimization-ELM (BBO-ELM) were tested for the prediction of WQI for ground water quality. A time series river water quality dataset was used to develop and test the models. The performance of the proposed models are evaluated using various fitness indices such as, the correlation of coefficient (r), root mean square error (RMSE), Kling-Gupta Efficiency (KGE), the index of agreement (d). Based on the comparisons, BBO-ELM was indicated as a possible alternative or substitute to assist the water quality assessment for the groundwater and can be readily applied an efficient data-driven methodology. BBO-ELM emerged as the better generalized hybrid model for calculating WQI.

INTRODUCTION

Quality of water is of vital importance to a healthy ecosystem. Ground water is a prime source of water for individuals, small towns and cities in developing and underdeveloped nations. In these countries, there is an absence of elaborate treatment systems required for surface water sources. Since, there is a lack of dissolved oxygen and water has to travel through layers of soil during its extraction, its chemical composition is different from surface water. Groundwater stored underground has a very stable yield. It is generally found in an enclosed but pervious formation called as aquifer (or aquitard). In these places, the water quality is qualitatively in good condition. In general, ground water resources have been exploited through its uncontrolled extensive extraction exacerbated by slow recharge of groundwater. The continuous and comparatively high rate drawdown and low rate and non-perennial recharge ultimately causes the qualitative and quantitative deterioration of groundwater and destruction of the groundwater systems. Hence, the groundwater quality evaluation is of prime importance for identifying the pollutants affecting the water quality the most, understanding the overall groundwater quality, utilizing the groundwater resource reasonably and at the last find a proper way to restore the groundwater environment (Vasanthavigar et al. 2010). The quality of groundwater has been vastly monitored and investigated by several scientists worldwide (Vasanthavigar et al. 2010; Jain et al. 2010; Tyagi and Sharma, 2014) So, the areas which have the condition of ground water level depletion should be kept under investigation. This paper discusses over the water quality existing in a certain area. Several approaches to assess the water quality index (WQI) have been developed in many previous studies like (Tyagi et al. 2017; Akkoyunlu and Akiner 2012), to generalize the quality of water with a simple expression proposed by various researchers. The status of water quality can be computed using different water quality parameters and converting them to a single index called as water quality index (WQI). A single number for water quality enables policy makers to develop a general idea about the water quality for a region and making it easily understandable to public. Horton developed the WQI in the 1970s (Horton, 1965). The standard procedure to develop a WQI includes following steps - selection of parameters, obtaining the standard values as prescribed by concerning agencies i.e. the Bureau of Indian Standards (BIS) and World Health Organization (WHO), the assignment of weights, development of sub-indices and aggregating the weighted sub-index values to give WQI (Abbasi and Abbasi, 2012). Finally, it gives a conclusive statement on the basis of individual test results obtained. The correlation of various sources of uncertainties such that pH, TDS, ECE, nitrate, sulphate, total hardness, calcium hardness, magnesium hardness, etc. are critical to assess the water quality index reliably.

The spatial and temporal series variation in water quality parameters creates uncertainty in the true assessment of water quality index (Cotter et al. 2003). Statistical-cum-reliability based analysis of WQI problems have been carried out in many studies. Reliability-based WQI seems to have gained importance in the recent past (Gazzaz et al. 2012; Yidana and Yidana, 2014). In this context, the spatial variations of various water quality parameters are transformed as WQI has been analyzed, spatially interpolated and mapped and reliability analyzed through models. The increase in ground water contamination, quality analysis for drinking water and water for irrigation purposes and distribution have been studied by a number of researchers all over the world (Adimalla and Li, 2019; Li and Li, 2019; Abd El-Aziz, 2017; Nagaraju et al., 2016; Khan and Jhariya, 2017; Haghiabi et al. 2018). Li et al. (2008) has done explicit calculations for groundwater quality status in the Taiwan province where Blackfoot disease prevailed using factor analysis (FA) and found that over-extraction of groundwater has led to salination and arsenic pollution. Mohebbi et al. (2013) analyzed the ground water quality within the region of Iran with the help of Drinking Water Quality (DWQI), and overall status of water quality in the region was found to be good. (Abba et al. 2020) have computed WQI using Evolutionary computational intelligence algorithm coupled with the self-tuning predictive model.

India has vast consumption of ground water amounting to 250 billion m^3 per year extraction of groundwater. This quantity exceeds the quarter of the world's total groundwater extraction. In India, the groundwater quality in the Tamil Nadu region is continuously

0016-7622/2021-97-9-1063/$ 1.00 © GEOL. SOC. INDIA | DOI: 10.1007/s12594-021-1821-0

degrading day by day (Arumugam and Elangovan 2009). Lu et al. (2010) had applied fuzzy synthetic evaluation (FSE) for the ground water quality evaluation, in the region of southern Haryana, India and concluded that drinking water in the area was free from health risks and water quality was observed as good.

Though above-discussed attempts were largely successful in evaluating the quality of groundwater, still there exists some shortcomings and defects in accuracy and objectivity in WQI (Hurley et al. 2012; Kumar and James 2013).

Recent advancements in Extreme Learning Machine (ELM) model development have given a great boost in various fields, especially in rainfall prediction and water quality problems. Huang et al. (2015) improves the interpolation theory, generalization ability and universal approximation capability. ELM has wide application in classification and regression. Other than this, it has better expertise for representational learning, feature selection, clustering and different learning techniques. These algorithms expand their application greatly. The training of ELM makes easier and faster implementation of parallel computation techniques. ELM makes feasible for big data processing and real-time reasoning (Huang et al. 2015).

MATERIALS AND METHODS

Study Area

The Gaya district is considered as the study site (Bihar, India) (Fig. 1) which is geographically extending from the latitude of 24.50-25.10°N and longitude of 84.4-85.5°E and covering 4976 km^2. The area is in the Middle Gangetic Plain in the Southern part of Gaya surrounded by four local river streams, namely the Morhar, the Phalgu, the Paimar and the Dhardhar and is the part of the Ganga basin (CGWB 2013). The entire study area is under the influence of heavy monsoon, hot blasts in summer and westerly blow in winter. Rocky upland; southern boundary is part of the Chotanagpur plateau and the northern part is underlain with erosional landscapes; alluvium sediment deposits (Dhobi-Gaya terrace, Sone-ganga plain). Sone-Ganga plain occurs as the mid-Gangetic plain area forming flood plains of rivers Ganga and Sone consisting of unoxidized Quaternary alluvial plain 80-100 m above the mean sea level. In northern parts, the soil has poor crops while the population is spread sparsely. River sand mining is the common practice in this area (Mineral, 2018) (Water et al. 2009). The climate is with average rainfall of 1105mm and the average temperature of 26.5°C. The annual rainfall of the region varies between 568.5 mm to 1109 mm (CGWB 2013). The younger alluvial soil is the major soil type occurring in the northern and north-west region whereas the southern region has sandy, red and yellow type of soils. Associated rock type occurring in this region are of pre-cambrian age (CGWB 2013). The groundwater level in the area during monsoon period fluctuates between 5 to 10 meters of depth from the ground level (CGWB 2013).

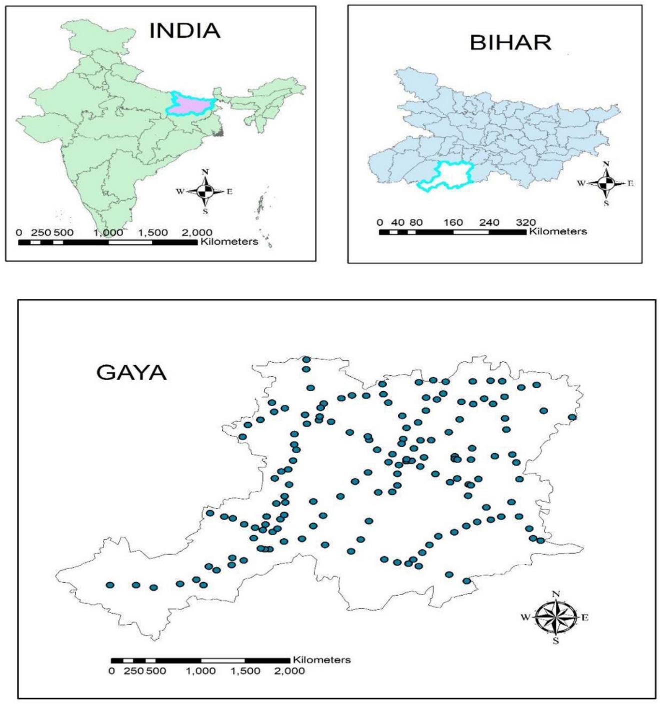

Fig.1. Location map of the study area with sample locations.

Water Sampling

A total number of 156 representative water samples were collected in the Gaya district of Bihar during June 2015 as per standard procedures described in APHA (2005). The location of the sampling area is shown in Fig.1. The Systematic Grab Sampling method was adopted for which the whole study area was divided into 5 km x 5 km grids and one sample was collected from a location inside the grid. Discrete samples collected in this manner were approximately uniformly separated spatial locations. The underground water samples were collected from tube wells, bore wells and hand pumps installed at considerable depths which are in regular consumption to the livelihoods. The depths were confirmed with information from nearby residents and no confirmatory measurements were done. The spatial positions of each source were recorded using GPS (Global Positioning System) device. The location for the water sample collection has been shown in Fig. 1.

Sampling Procedure

One-liter capacity low-density PVC bottles decontaminated with 5M concentrated Nitric Acid followed by repeatedly washing with milli-pore distilled water was used to collect the water samples. Prior to the collection at each site, the sampling bottles were flushed with the water to be sampled for a period of 2-3 minutes to attain the localized uniform representation in the sample. Also, prior to the analysis suspended solids were filtered out through 0.45µm Whatman filter paper (Kumar et al. 2018). Thermo scientific Orion VERSA STAR pH/ISE/RDO/EC/Dissolved Oxygen kit used for insitu measurements of the pH, ECE & TDS. Once the in-situ parameters were recorded bottles were carried back to the laboratory for measurement of other parameters. After analysis of nitrates in the laboratory, samples were preserved with approx. 1 ml of 65 % Nitric Acid in 1 liter of the collected samples and were stored in the dark and 4°C till the analysis were completed. Experimental glassware were acidified for 2 weeks, then all of them were washed with distilled water. The Thermo Scientific Evolution 201 UV-visible spectrophotometer instrument was used for analyzing phosphates, sulfate and nitrate using the colorimetric method. Standard procedures and methods were followed for the examination of water samples as described in (Kumar et al. 2018; Kumar et al. 2019a). Statistical analyses were done for all the samples using R programming software (open source version 3.6.1) (Gradilla-Hernandez et al. 2020).

Physio-chemical Analysis

All the samples were analyzed for their different physical and chemical parameters namely (pH, Electrical conductivity, Conductivity, chlorides, dissolved oxygen (DO), biological oxygen demand (BOD), Total dissolved solids (TDS), total alkalinity, total hardness, Calcium Hardness, Magnesium Hardness, Nitrates, and sulfate) as per methods described by (APHA 2005, WHO).

Calculation of Water Quality Index (WQI)

In this article, Water Quality Index is calculated by Weighted Arithmetic Index Method, proposed by Horton in 1965 using twelve parameters. These parameters are pH, conductivity, chlorides, dissolved oxygen (DO), biological oxygen demand (BOD), Total dissolved solids (TDS), total alkalinity, total hardness and sulfate. The used method to calculate the water quality index is recommended by the World Health Organization (WHO) and Indian Council for Medical Research (ICMR) (**F. J. et al. 2006**, Sharmin Yousuf Rikta et al. 2016). Since Different parameters have their own Unit of measurement, the reciprocal of all the Standard values are added to observe a dimensionless numeric constant as (k).

The numeric value WQI was calculated using the equation (1).

$$WQI = \Sigma_{i=1}^{n} W_i\, Q_i \quad (1)$$

Where, *WQI* indicates water quality index, Q_i is water quality rating and W_i indicates the unit weight for individual water quality parameters.

$$W_n = k/S_n \quad (2)$$

where,

$$\text{k(constant)} = 1/(1/V_{s1} + 1/V_{s2} + 1/V_{s3} \ldots..+ 1/V_{sn}) \quad (3)$$

S_n denotes the number of the standard value for water quality parameters. Further, Water quality is calculated rating is calculated using the equation (4).

$$Q_i = (V_a - V_i)/(V_s - V_i) \times 100 \quad (4)$$

Where, V_a indicates the actual concentration of the water sample, V_i indicates the ideal value (0 for all the water quality parameters except pH (7.0) and DO (14.6 mg. l-1) and V_s is the standard value. If the value of Q_i is zero that signify the absence of individual pollutants, the value between $0 < Q_i < 100$ shows the pollutant present the prescribed limit and $Q_i > 100$ represents the pollutants are above the standard limit. The status of water quality in their final numbers are further categorized in different groupings to give there easily understandable meaning to the common people. If the values are below 50; the water is classified as excellent water, if it varies between 50 to 100, the quality of water is good water; the quality of water is deemed to be poor if the values vary between 100 and 200; the water is of very poor water quality if the values are ranging from 200 to 300; and the water is not suitable for drinking purpose the values exceeds 300 mg/L.

Principal Component Analysis (PCA)

Principal component analysis has been briefly discussed by (Fisher and Mackenzie, 1923). It is applied in the analysis of multivariate data analysis. PCA analysis is in form of a numerical data table where the values are more or less interconnected dependent variables. It performs the function of extracting important information from the data set and represent it in a fresh set of orthogonal variables which is also termed as principal components. It observes the pattern of similarity of the variables and of the observations as points in maps (Abdi and Williams, 2010). PCA has been widely used in water quality index development for parameter selection as discussed in (Tripathi and Singal, 2019; Parinet et al. 2004). In the same way, the PCA is applied and the outputs obtained from the PCA analysis were fed as input for the models to predict and validate WQI.

Extreme Learning Model (ELM)

Extreme learning model is a fast learning single layer feed forward neural network model. Hidden nodes try to learn from the ancestral characters. Huang et al. (2004) Basically the output weights are being learned in a single step necessary for the linear model and applied over the inputs. They can produce better generalized results and can learn faster than back propagated trained networks. Basically, the Neural Network is constructed as a black box which is regularized by ELM to redirect the output pattern after recognizing the input pattern. Firstly, a large number of input data along with its actual output are feed to prepare an efficient and precise pattern called as training. Neural Networks recognizes the inputs all together to develop a similarity pattern. Secondly, it is tested with other input data to predict the output with an accuracy called as testing. That is how this process flows resulting a development of a model for predicting desired output. (Ding et al. 2015).

Biogeography-based Optimization (BBO)

Dan Samon proposed the Biogeography- algorithm in 2008 conceptualized over migration of species (Habitants) from previous habitat to a new habitat, (Roy et al. 2019). The BBO algorithm is procedural explanation of reaching the steady condition of species evaluating the influence of migration and mutation over discreteness of that specific species (Roy et al. 2019). The algorithm understands the spatial and timeline behavioral pattern based on distribution over the area. It describes how individual habitats have arisen to establish their relationship and their distinctiveness compared to others, their characteristics in terms of mutation, immigration and emigration (Li and Li 2019). The emigration and immigration represent the information interaction mechanism in the optimization algorithm.

Radio Based Function (RBF)

It has an extremely fast learning process after which it can produce the generalization performance somewhat nearer to that of SVM in many classification problems and real and artificial benchmarking function approximations (Huang and Slew 2004). Rather than tuning the centers and impact widths, the values may be chosen on the random basis for these parameters. After that output weights of the RBF networks may be analytically calculated. Extreme Learning Machine (ELM) is basically a single layer feedforward neural networks (SLFNs) which is further improvised with additive neurons case to SLFNs with radial basis function (RBF) kernels case-RBF networks (Huang and Siew, 2005a). It arbitrarily assigns the kernels and eliminates the need to tune them, in comparison with support vector machines. It can be used easily at higher speed to provide a more compact network. It can also be used further for the systematic investigation of the arbitrariness of the RBF kernels (Huang and Siew, 2005b). A number of research articles have applied RBF-kernel functions in solving the problem of Water quality prediction, discharge prediction (Heddam and Kisi, 2017).

Online Sequencing (OS)

The capability of online learning capability of Online sequencing (Yadav et al. 2016).

Online sequencing - Originated from Batch Extreme Learning Machine (ELM) algorithm, it has efficient and powerful learning abilities (Guo et al. 2018). Out of many popular online algorithms, it outperforms with faster learning speed. OS-ELM can be applied for system modelling and prediction and time series prediction.(Huang et al. 2006). It is basically a sequential advancement over batch learning algorithm to learn algorithms online based on recursive least squares (RLS) algorithm, called as Online Sequential Extreme Learning Machine (OS-ELM) (Guo et al. 2018). It can update the model equation on the basis of a fresh data entry without increasing the computational cost (Yadav et al. 2016). The tool is very useful in rainfall and discharge forecasting problems (Yadav et al. 2016) and Water quality parameter prediction problems (Goz et al. 2019; Heddam and Kisi, 2017). Potentially ill-conditioned matrix inversion emerges as a drawback of OS-ELM. The stability and performance issues due to ill conditioning or singularity of the auto correction matrix of the hidden layer are reversed by applying Regularization technique (Guo et al. 2018).

MODEL DEVELOPMENT

Parameter Selection and Parameter Reduction using Principal Component Analysis (PCA)

For the parameters observed in the study, it seems that they are correlated with each other as the ions present in water usually gets imparted from the somewhat uniform geology and uniform upstream inputs (Rudwick, 1998). Karl Pearson (1901) invented a statistical tool to convert a possibly correlated variables (numeric values only) into a set of linearly uncorrelated variables using orthogonal transformation called as Principal component analysis (Chen et al. 2011). PCA can be applied for the dimensional reduction of a data set of multivariate nature. In this process it tries to maintains the original structure to the possible maximum extent (Chu et al. 2018). The principal components are the Eigen vectors of the covariance matrix of the original data set. PCA is applied after normalizing the numeric data set and then eigenvalue decomposition of a data covariance or singular value decomposition of a data matrix is performed (Chen et al. 2011). In this way, the previously correlated values get converted to entirely discrete and normalized (to create a dimensionless entities) values so that the principal component analysis (an analog of principal axis theorem), transforms the numeric inputs such that its first principal component has the highest possible variance. This value of variance decreases as we proceed further to the next principal components because each succeeding component has the largest possible variance under the constraint that is orthogonal to the proceeding components. The resulting vectors (each have linearly correlated variables, containing n no. of observations) obviously becomes uncorrelated orthogonal set. The tool is used here to create a set of uncorrelated set of input data to be fed to predictive models. There is an ambiguity in modelling of the spatial data especially with the Water Quality in determining the random component. Hence, the Water quality parameters as the random component to build the models after performing PCA. Thereafter the individual Principle Components are fed as an input for the models considered. The division in the data sets into training and testing differs case to case.

The model structure of ELM can be explained in the given manner, considering j, m, n, i.e. the input layer nodes, output layer nodes and the hidden layer nodes, respectively, and the hidden layer activation function g(x). For N distinct samples $x_i \in R_N \times R_j$, $y_i \in R_N \times R_m$ (i=1,2,3,…4), equation (5) is representation of hidden layer outputs, and equation (6) represents the numerical relationship between output equation (5) and output of the output layer.

$$h = g(ax + b) \tag{5}$$

$$h(x)V = y, \quad i = 1, 2, 3\ldots, N \tag{6}$$

in compact form the equation can be re written as,

$$HV = Y, \tag{7}$$

Where,

$$H = \begin{bmatrix} g(\vec{a}_1, \vec{b}_1, \vec{x}_1) & g(\vec{a}_1, \vec{b}_1, \vec{x}_2)\cdots & g(\vec{a}_1, \vec{b}_1, \vec{x}_N) \\ \vdots & \ddots & \vdots \\ g(\vec{a}_1, \vec{b}_1, \vec{x}_M) & \cdots & g(\vec{a}_1, \vec{b}_1, \vec{x}_N) \end{bmatrix} \tag{8}$$

$$V = \begin{bmatrix} V_1^T \\ . \\ V_n^T \end{bmatrix}_{n\times m}, \quad Y = \begin{bmatrix} Y_1^T \\ . \\ Y_N^T \end{bmatrix}_{N\times m} \tag{9}$$

Where $a_i = [a_{i1}, a_{i2},\ldots, a_{in}]^T$, are the weights connecting the i^th^ input nodes and hidden layer, b_j is the bias of the j^{th} hidden node, and $v_i = [v_{j1}, v_{j2},\ldots,v_{jn}]^T$ are the weights connecting the output layer and the hidden node j. Here, H denotes the output matrix of the neural network. The input weights a_{ij} and the bias of the hidden layer also need to be set; Using a series of liner transformations, the output weights V can be obtained easily. Hence, to obtain the output weights V using ELM can be divided into the following three steps.

Step 1. Select numerical values randomly between 0 and 1 for setting the input weights a_{ij} and similarly the bias of the hidden layer b_j.

Step 2. Calculate H, the output matrix.

Step 3. Calculate V, the output weights:

$$V = H \dagger Y \tag{10}$$

where H† is the generalized inverse matrix of the output matrix H (Ding et al. 2015).

The BBO model structure is completed through migration operation, that is

$$H_i\,(SIV) \leftarrow H_j\,(SIV) \tag{11}$$

Considering the probability of C_k species (habitants) is contained in the k^{th} habitat is P_k, $C_k = 1,2,\ldots S_{max}$ and S_{max} is the maximum number of species, from times t to (t+Δt), the change in P_k is (Huang et al. 2012)

$$P_k\,(t+\Delta t) = P_k\,(t)\,(1-\lambda_k\Delta t - \mu_k\Delta t) + P_{k-1}\lambda_{k-1}\Delta t + P_{k+1}\mu_{k+1}\Delta t \tag{12}$$

where λ_k is the immigration and β_k is the emigration rates. When taking the limit of Eq. (12) as $\Delta t \rightarrow 0$, there are

$$\dot{P}_k = \begin{bmatrix} -\lambda_0 P_0 + \mu_1 P_1 & k = 0 \\ -(\lambda_k + \mu_k)P_k + \lambda_{k-1}P_{k-1} + \mu_{k+1}P_{k+1}, & 1<k< S_{max} - 1 \\ -\mu_k P_k + \lambda_{k-1}P_{k-1} & k< S_{max} \end{bmatrix} \tag{13}$$

where $\dot{P}_k$ is defined as the derivation of P_k, μ_o=0, and $\lambda_{Smax} = S_0$. The linear relationship between λ_k and μ_k is given as follows

$$\lambda_k = I\,(1 - C_k/S_{max}) \tag{14}$$

$$\mu_k = EC_k\,/\,S_{max} \tag{15}$$

The article has incorporated SLFNs (single-hidden layer feed forward neural networks) with RBF kernels-RBF networks. The RBF network output having $\tilde{N}$ kernels for an input vector $x \in R^d$. It is given by

$$f_{\tilde{N}} = \Sigma_{i=1}^{\tilde{N}}\, \beta_i\, \phi_i\,(x) = \phi\,(\mu_i, \sigma_i, x) \tag{16}$$

Where $\beta_i = [\beta_{i1}, \beta_{i2},\ldots,\beta_{im}]^T$ is the weight vector. β_i connects the i^th^ kernel which is generally Gaussian

$$\phi_i\,(x) = \phi\,(\mu_i, \sigma_i, x) = \exp((\|x - \mu_i\|)^2 / \sigma_i) \tag{17}$$

The formulation for OS-ELM can be described as; for N arbitrary distinct samples $(X_j, t_j) \in R^d \times R^1$, SLNFs (single hidden layer feed forward neural networks)

$$\Sigma_{i=1}^n\, \beta_i\, G(a_i, b_i, x_j), \quad j = 1, 2, 3,\ldots N, \tag{18}$$

Where a_i is the weight vector establishing the connection between input nodes and hidden nodes (i^{th}), b_i is the weight vector establishing the connection between output nodes and hidden nodes (i^th^), b_i act as the threshold of the i^{th} hidden node. The function $g_i(X_j) = G(a_i, b_i, X_j)$ stands as the output of the i^{th} hidden node with respect to input. These N samples are approximated with no error through SLFNs operation. This means that there exists (a_i, b_i) and β_i such that

$\Sigma_{i=1}^{n} \beta_i G(a_i, b_i, x_j), \quad j = 1, 2, 3, \ldots N,$

The above N equations can be written compactly as,

$$H\beta = T, \tag{19}$$

Where

$$H = \begin{bmatrix} h_1 \\ \cdot \\ h_n \end{bmatrix} = \begin{bmatrix} G(a_i, b_i, x_j) & \ldots & G(a_i, b_i, x_j) \\ \vdots & & \vdots \\ G(a_i, b_i, x_j) & \ldots & G(a_i, b_i, x_j) \end{bmatrix}_{N \times n} \tag{20}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \cdot \\ \beta_N^T \end{bmatrix}_{n \times 1} \tag{21}$$

$$T = \begin{bmatrix} t_1^T \\ \cdot \\ t_N^T \end{bmatrix}_{n \times 1} \tag{22}$$

The hidden layer output matrix of the network is denoted as H; the output vector of the hidden layer with respect to input x_j is the j^{th} row of H and the i^{th} hidden node's output vector with respect to inputs $x_1, x_2, \ldots x_N$ is the i^{th} column of H.

Furthermore, in order to attain the online learning scenario, ELM is improvised in an online version called as OSELM. In order to facilitate learning from the training samples successively and incrementally, OSELM is used. It functions in two phases namely initialization and sequential learning phase. The learning procedure of OSELM consists of an initialization phase and a following sequential learning phase, and the one by-one OSELM is summarized as follows.

In initialization phase, given an initial training set

$\Omega_{(k-1)} = \{(x_j, t_j)\}, j = 1, 2, \ldots k-1$, according to (Huang et al. 2006).

The expression

$\beta_{(k-1)} = P_{(k-1)} H^T_{(k-1)} T_{(k-1)}$, is used to assign initial output weights,

where,

$$P_{(k-1)} = (H^T_{(k-1)} T_{(k-1)})^{-1}, H_{(k-1)} = [h_1^T, h_2^T \ldots, h^T_{(k-1)}]^T, \text{ and, } T_{(k-1)} = [t_1^T, t_2^T \ldots, t^T_{(k-1)}]^T, \tag{23}$$

The Recursive Least Square algorithm in the sequential learning phase, is used to update the output weights. The partial hidden layer output matrix for another sample received as $(x_k\ t_k)$, is calculated as $h_k = [G(a_1, b_1, x_k) \ldots G(a_n, b_n, x_k)]$, and then equations (24) and (25) can used to calculate the output weights updates given as follows:

$$P_k = P_{(k-1)} - [(P_{(k-1)} h_k^T h_{kP_{(k-1)}})/(1+h_{kP_{(k-1)}} h_k^T)], \tag{24}$$

$$\beta_k = \beta_{(k-)} + P_k h_k^T (t_k - h_k \beta_{(k-)}), \tag{25}$$

(Lu et al. 2016) describes the output weights of OSELM are recursive in nature which are updated based on two results/data namely, the newly arrived data and the intermediate results from the last iteration. In order to reduce the memory requirement and the computational overhead, these can be discarded as soon as they have been learnt. The above one-by-one OSELM algorithm can be easily extended to chunk-by-chunk type.

All the models explained above are implied to attain an accurate, efficient and progressive model (ELM, RBF-ELM, OS-ELM, BBO-ELM). These models were used to optimize WQI prediction by tuning it using trial and error method. The initially selected parameters that were in tune with the models were varied in the trials for achieving the best fit measure. Using the 'Matlab 2018b' toolbox, the hybrid ELM model scripts and other related models have been developed.

In order to attain the objective over the above discussed developed models, training and testing of the data set were done to predict WQI through hybrid models. These models were further tested for robustness also. These simple ELM model the hybrid ELM model codes were developed in 'Matlab 2018b' toolbox.

Moreover, after the computation through performing the model experiments a comparative study is required to be established to find the robustness of models. This research proposes the ELM model for WQI modelling. Moreover, a detailed comparative study with other three hybrid models (BBO-ELM, RBF-ELM and OS-ELM) is performed.

To develop the BBO-ELM, RBF-ELM, OS-ELM, ELM based models, the dataset is divided into two parts.

1) Training dataset: It is the group of randomly chosen data for the model development purpose.
2) Testing dataset: It is the group of randomly chosen data for the model testing purpose of the model developed.

There is no such specific rule of thumb for data partition and it is practiced differently by different researchers. E.g. Sahu et al. (2011), Kumar et al. (2019b), Coulibaly and Baldwin (2005) and Pal (2016) have used 75%, 70%, 90% and 69% of the data as training data, respectively. For the better precision in used models three different data partitions (70 % for training and 30 % for testing, 75 % for training and 25 % for testing, 80 % for training and 20 % for testing) are used.

Model Performance Assessment Metric

The evaluation of hydrological model is based on RMSE (Chai and Draxler, 2014).

The proposed models are investigated with the help of percentage root mean square error (%RSME) and coefficient of correlation (R) to analyze the performance, Ratio of RMSE to the standard deviation of the observation (RSR), mean absolute error (MAE) and coefficient of persistence (Cp) were used as shown in Equations (26-31)

$$r = \frac{\Sigma_{i=1}^{l}(WQI_{E_i} - WQI_{E_i})(WQI_{O_i} - WQI_{O_i})}{\sqrt{\Sigma_{i=1}^{l}(WQI_{E_i} - WQI_{E_i})^2\ \Sigma_{i=1}^{n}(WQI_{O_i} - WQI_{O_i})^2}} \tag{26}$$

RMSE is a function prepared to present the suitability of a model also it successfully fulfills the triangle inequality theorem (Chai and Draxler 2014). The standard deviation of the model prediction error is represented by the RMSE. Smaller the value of RMSE the performance of the model is better.

$$RMSE = ([\Sigma_{i=1}^{l}(WQI_{E_i} - WQI_{O_i})^2] / l) \tag{27}$$

MAE (Mean Absolute Error) is a mathematical formulation for testing the model's goodness of fit. It is the difference between Estimated and Observed.

$$MAE = (1/N)\ \Sigma_{i=1}^{l} \mid Q_{O_i} + Q_{E_i} \mid \tag{28}$$

KGE analysis is done to calculate the relative importance of correlation, bias and variability in WQI modelling. Accurate models have KGE value near to one.

$$KGE = 1 - \sqrt{(s[1] * (r-1)^2 + (s[2] * (\alpha - 1)^2 + (s[3] * (\beta - 1)^2)} \tag{29}$$

s = represents the scaling factors to be used for r

– scaling the criteria space before computing the Eucledian distance where r is the Pearson product moment correlation coefficient

$\beta = \mu_E / \mu_0 \quad ; \quad \alpha = \sigma_E / \sigma_0$

$$d = \frac{1 - \Sigma_{i=1}^{l}(WQI_{O_i} - WQI_{E_i})^2}{\Sigma_{i=1}^{l}(\mid WQI_{E_i} - WQI_{O_i} \mid - \mid WQI_{O_i} - WQI_{O_i} \mid)^2} \tag{30}$$

$$VE = \frac{1 - (\mathrm{sum}(\mathrm{abs}(WQI_{O_i} - WQI_{E_i})))}{\mathrm{sum}(WQI_{O_i})} \tag{31}$$

Where WQI_{E_i} is the i^{th} estimated water quality index using models; WQI_{O_i} is the i^{th} observed water quality index; WQI_{E_i} is the average of the estimated water quality index; WQI_{O_i} is the average of the water quality index data for k number of observations.

RESULTS AND DISCUSSION

Opting the above discussed proposed methodology using a simple model and three hybrid models (ELM, RBF-ELM, OS-ELM, BBO-ELM) to compute the water quality index of the groundwater. The water quality parameters are related to the water quality index in order to map the features. The generalised relationship between the water quality parameters (which are included) and the water quality index (WQI) are represented as shown in the equation (32).

$$WQI = f(Total\ Hardness,\ Chloride,\ Total\ Alkalinity, \ldots) \tag{32}$$

In assessing the quality of water, *WQI* is very much proved significant to understand the acceptability for different purposes and also to know the status of water quality around the region for the remediation point of view.

WQI being highly dynamic and non-linearly driven factor which usually influenced by many environmental factors. Hence, it becomes a challenging task to modelise the WQI considering appropriate input parameters. The previous studies have already adopted a number of physical based models for WQI assessment (Ramakrishnaiah et al. 2009; Tyagi et al. 2017). All these models have governing equations engulfing the different attributes of water and environment. These physical models faced the biggest limitation as they have been developed assuming the idealistic approaches and assumptions which are largely absent in in-situ practicality. Furthermore, the physical models are highly capitalised when it comes to experiment setups and also they usually deal with complex formulations. Lastly, when the numbers of predictors stay limited and forecasting becomes important then these pattern recognitions based tools are handy make the job comparatively much easier.

Table 1 gives the gist of the WQI index of the region. Water quality index of the samples varies from 38.4 to 237.71, out of which 96.15 percent (63.46+32.69) of samples are observed as suitable for drinking water Purpose. 63.46 percent of samples are of good quality; 32.69 percent of samples are excellently fit for drinking water. Only 3.85 percent (3.21 +0.64) of samples are observed unsuitable for drinking purpose.

WQI is plotted using inverse distance weighted (IDW) interpolation in the region of Gaya district as shown below in Fig. 2. The use of the geospatial interpolation technique improves the spatial mapping and level of risk assessments. This kind of identification of regions has a benefit in targeting the local monitoring programs and land use planning.

The input parameters (pH, TDS, total hardness, chloride, Total alkalinity etc.) for models were selected to predict the WQI based on the PCA analysis.

Using statistical approach PCA reduces the number of parameters hence reducing the number of variables. Data here is projected in independent axes called as PCs. From all the PCs generated first Six PCs were considered to be applied as an input to the models used. Table 2 shows the initial 6 PCs account for 88.744 % of the total variance.

Table 1. Water quality index analysis

The WQI range, type of water and percentage of samples in the study region.			
WQI Range	Type of water	Number of Samples	% of samples
<50	Excellent	51	32.69
50 to 100	Good	99	63.46
100 to 200	Poor	5	3.21
200 to 300	Very poor	1	0.64
>300	Not suitable for drinking	Nil	0.00

Table 2. Cumulative variation of principal components

Principal Components (PC)	Variance (%)	Cumulative Variance (%)
PC1	43.130	43.130
PC2	14.530	57.660
PC3	12.250	69.910
PC4	7.606	77.514
PC5	6.244	83.759
PC6	4.985	88.744
PC7	4.594	93.338
PC8	3.287	96.625
PC9	2.002	98.628
PC10	1.144	99.771
PC11	0.229	100
PC12	0	100

Figure 3 and 4 show the performance of training and testing dataset respectively when 70 % of the data are used for training purpose. Figure 5 and 6 show the performance of training and testing dataset respectively when 75 % of the data are used for training purpose. Figure 7 and 8 show the performance of training and testing dataset respectively when 80 % of the data are used for training purpose. Out of them the performance of 75 % partition of data has shown comparatively better results.

The model considers a set of variable input data. The selection of input data in optimum number is an important task during model development and is considered to be tedious. Hence, Partition of data is done based on the suggestion given by Haghabi et.al.(2018). Upon analyzing the results, it was found that the model developed were quite good at performing the prediction of water quality index and shows quite impressive fitness index. Table 3 explains about the comparative performance of models when the 70 % of all the

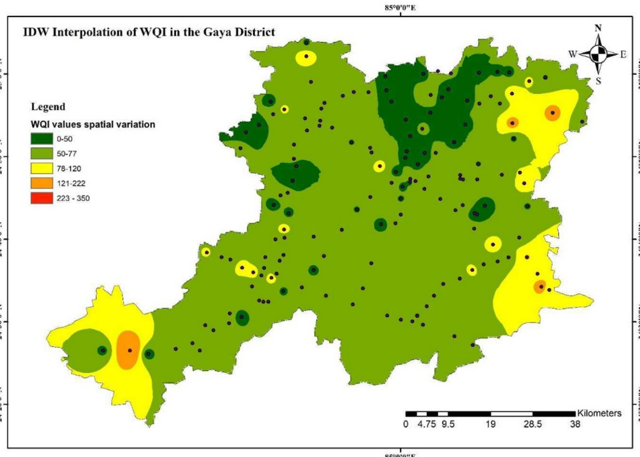

Fig.2. Spatial variation for the WQI values.

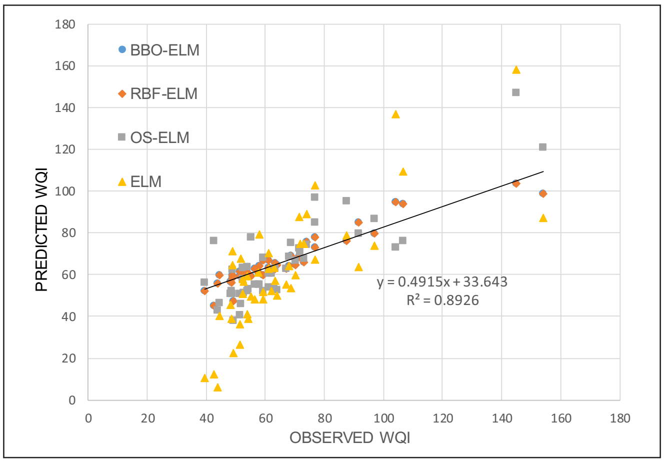

Fig.3. Performance of the used model based on 70% training data set.

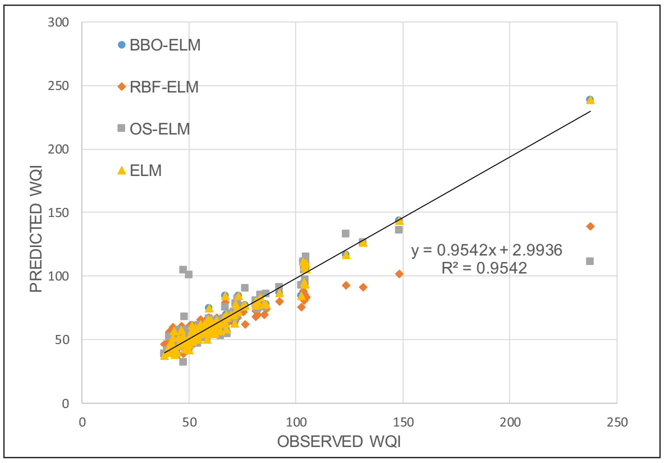

Fig.4. Performance of the used model based on 70 % training data set (predicted)

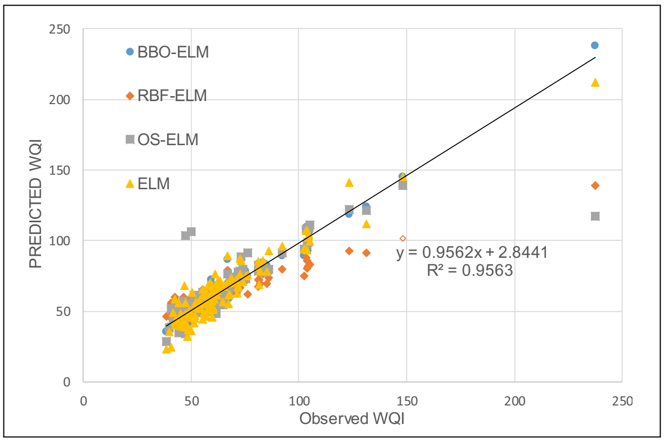

Fig.5. Performance of the used model based on 75% training data set

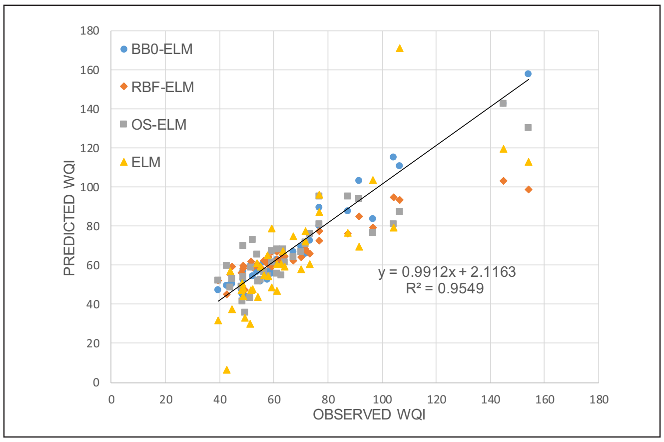

Fig.6. Performance of the used model based on 75% training data set (predicted)

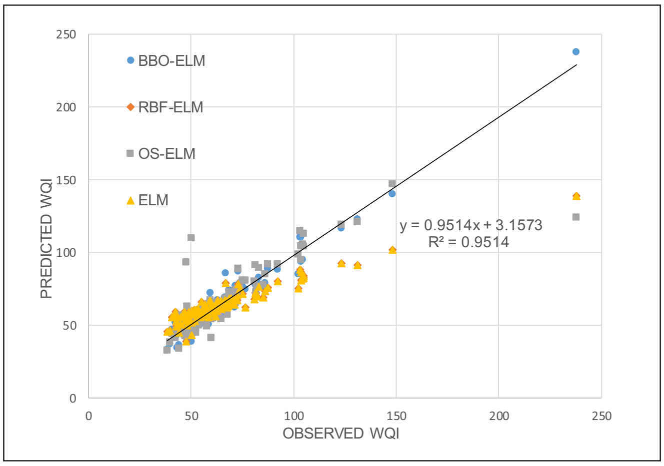

Fig.7. Performance of the used model based on 80% training data set

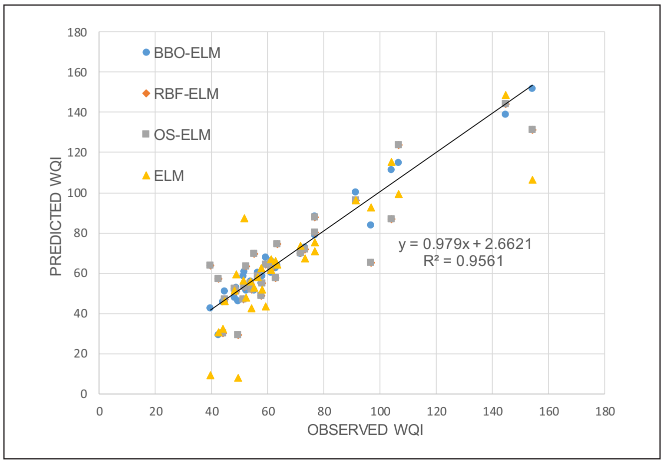

Fig.8. Performance of the used model based on 80% training data set (predicted)

randomly shuffled data are taken for training and the rest is used for testing the fitness of the model. Table 4 explains about the comparative performance of models when the 75 % of all the randomly shuffled data are taken for training and the rest is used for testing the fitness of the model. Table 5 explains about the comparative performance of models when the 80 % of all the randomly shuffled data are taken for training and the rest is used for testing the fitness of the model.

During, the training period BBO-ELM outperformed in the terms of R2 (Coefficient of Determination) followed by RBF-ELM then OS-ELM as explained below. And during the testing period the data is initially separated in a group of three different orientations as explained previously. When the data is randomly separated in ratio of 70:30 the BBO-ELM (training-0.954, testing-0.955) outperformed the RBF-ELM (training-0.891, testing-0.893), OS-ELM (training-0.65, testing-0.725) and ELM (training-0.959, testing-0.593). The performance of

Table 3. Comparison of the performance of models when the 70 % of all the randomly shuffled data are taken

Name of Model	TRAINING						TESTING					
	RMSE	R2	NSE	MAE	d	VE	RMSE	R2	NSE	MAE	d	VE
BBO-ELM	5.517	0.954	0.954	3.969	0.988	0.939	5.838	0.955	0.937	4.037	0.985	0.939
RBF-ELM	14.467	0.891	0.685	8.648	0.855	0.868	12.468	0.893	0.712	8.047	0.873	0.879
OS-ELM	15.235	0.65	0.65	6.274	0.884	0.904	12.201	0.725	0.724	8.255	0.915	0.875
ELM	5.233	0.959	0.959	4.072	0.99	0.938	18.683	0.593	0.354	14.15	0.854	0.786

Table 4. Comparison of the performance of models when the 75 % of all the randomly shuffled data are taken

Name of Model	TRAINING						TESTING					
	RMSE	R2	NSE	MAE	d	VE	RMSE	R2	NSE	MAE	d	VE
BBO-ELM	5.22	0.956	0.956	3.9	0.989	0.94	5.615	0.955	0.95	4.32	0.988	0.936
RBF-ELM	14.01	0.89	0.686	8.256	0.856	0.873	13.40	0.897	0.714	8.636	0.874	0.872
OS-ELM	14.40	0.668	0.668	6.182	0.891	0.905	10.79	0.818	0.815	8.178	0.944	0.879
ELM	8.24	0.894	0.891	6.349	0.972	0.902	17.54	0.637	0.511	12.12	0.881	0.82

Table 5. Comparison of the performance of models when the 80 % of all the randomly shuffled data are taken

Name of Model	TRAINING						TESTING					
	RMSE	R2	NSE	MAE	d	VE	RMSE	R2	NSE	MAE	d	VE
BBO-ELM	5.375	0.951	0.951	3.882	0.987	0.94	5.912	0.956	0.953	4.524	0.988	0.934
RBF-ELM	13.572	0.891	0.69	8.036	0.859	0.876	14.479	0.902	0.721	9.346	0.878	0.863
OS-ELM	13.301	0.702	0.702	5.809	0.905	0.911	11.844	0.815	0.813	8.647	0.943	0.873
ELM	6.846	0.922	0.921	5.25	0.98	0.919	21.337	0.64	0.394	12.411	0.869	0.818

the BBO-ELM (training-0.955, testing-0.956) still outperformed the (training-0.89, testing-0.897), OS-ELM (training-0.668, testing-0.818) and ELM (training-0.894, testing-0.0.637) when the separation of data is done in the ratio of 75:25. The trend kept followed after the separation of data in done in the ratio 80:20, the BBO-ELM (training-0.951, training-0.956) outperforms than RBF-ELM (training-0.891, testing-0.902), OS-ELM (training-0.702, testing-0.815) and ELM (training-0.922, testing-0.64).

The consistency of each developed models is checked using degree of agreement denoted as d. When the data is randomly separated in ratio of 70:30 the BBO-ELM (training-0.988, testing-0.985) outperformed the ELM (training-0.99, testing-0.854), OS-ELM (training-0.855, testing-0.873) and ELM (training-0.884, testing-0.915). The performance of the BBO-ELM (training-0.989, testing-0.988) still outperformed the OS-ELM (training-0.891, testing-0.944), ELM (training-0.972, testing-0.881) and RBF-ELM (training-0.856, testing-0.874) when the separation of data is done in the ratio of 75:25. The trend kept followed after the separation of data in done in the ratio 80:20, the BBO-ELM (training-0.987, testing-0.988) outperformed the OS-ELM (training-0.905, testing-0.948), ELM (training-0.98, testing-0.869) and RBF-ELM (training-0.859, testing-0.878).

Out of all the models discussed above were applied, configured and lastly validated at different spatial locations, the BBO-ELM combination of 75-25 dataset was seen to be outperformed the other models based on different fitness measures (see Table 3,4 and 5). The study shows that the BBO-ELM model can ascertain WQI precisely (configuration- Fig 8; validation - Fig. 9).

From the comparison in the analysis of results, BBO-ELM performed well in terms of all fitness parameters followed by RBF-ELM and OS-ELM as explained in the validation plot comparison chart shown below (Fig. 9).

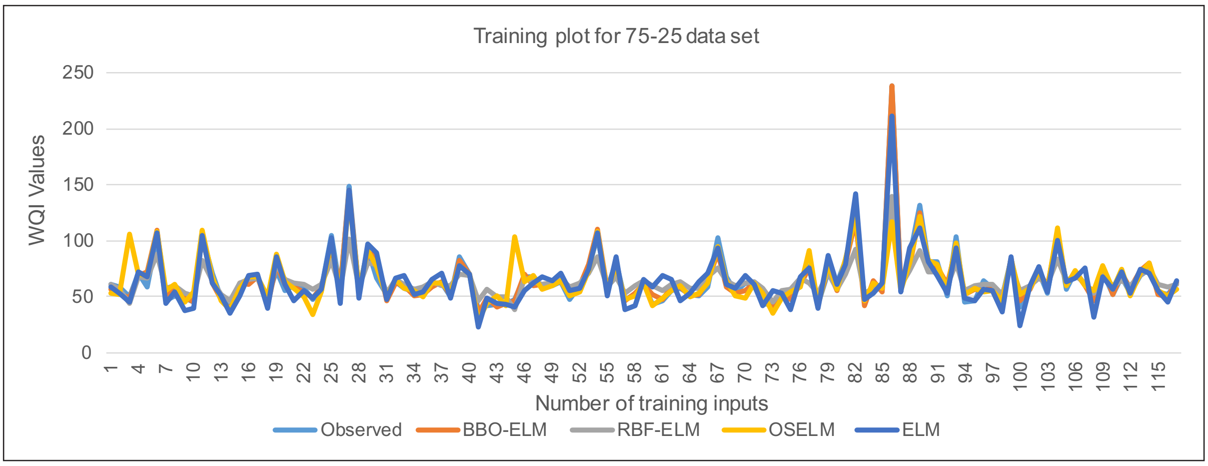

Fig.9 Training plot for different models used

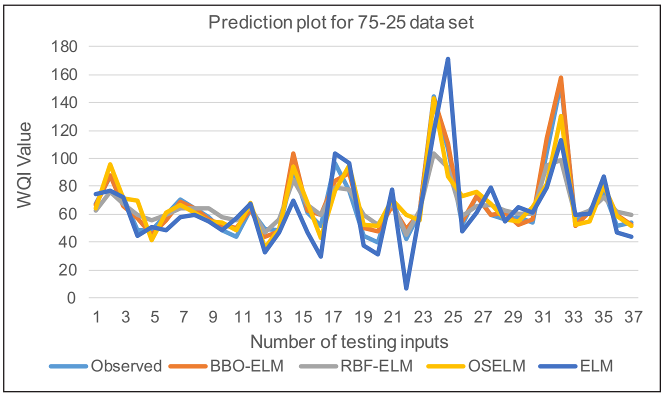

Fig.10. Prediction plot for different models used

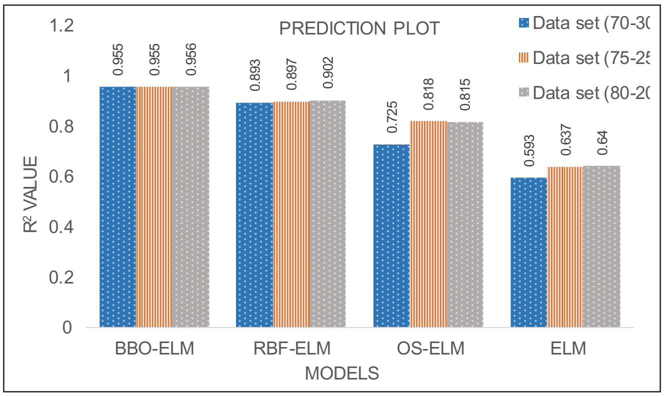

Fig.11. Comparison of coefficient of determination.

CONCLUSIONS

This study precisely predicts the WQI for the groundwater sample as it is essential to analysis and reduce the environmental effects which finally for insures a prosperous health. When we talk about WQI more than 90 % of the samples either are of good quality or very good quality. So, it can be said that the overall quality of water is fit for drinking purpose. This research adopts three hybrids of soft computing technique i.e. ELM, RBF-ELM, OS-ELM, BBO-ELM for predicting WQI. From the analysis of result, BBO-ELM performed well in terms of all fitness parameters followed by RBF-ELM and OS-ELM (with R^2 value of 0.955 in validation stage). The outcomes of this study adopts BBO-ELM as a possible alternative or substitute to assist the water quality assessment for the groundwater and can be readily applied an efficient data-driven methodology. BBO-ELM emerged as the generalized hybrid model for calculating WQI. In future the above discussed models can be applied for different locations for WQI predictions.

Acknowledgement: The financial assistance for data collection of of water quality parameters for this work was provided through Board of Research and Nuclear Sciences (BRNS Project Ref. No.: 36(4)/14/10/2014-BRNS) under Department of Atomic Energy, India.

References

Abba, S.I., Hadi, S.J., Sammen, S.S., Salih, S.Q., Abdulkadir, R.A., Pham, Q.B., Yaseen, Z.M. (2020) Evolutionary computational intelligence algorithm coupled with self-tuning predictive model for water quality index determination. Jour. Hydrol., v.587, pp.124974. DOI: 10.1016/j.jhydrol.2020.124974

Abbasi, T., Abbasi, S.A. (2012) Approaches to WQI Formulation. *In:* Water Quality Indices**.** Internat. Res. Jour. Engg. Tech., pp.3-7. doi:10.1016/C2010 -0-69472 -7

Abdi, H., Williams, L.J. (2010) Principal component analysis. Wiley Interdiscip. Rev. Comput. Stat.

Adimalla, N., Li, P. (2019) Occurrence, health risks, and geochemical mechanisms of fluoride and nitrate in groundwater of the rock-dominant semi-arid region, Telangana State, India. Hum. Ecol. Risk Assess. doi:10.1080/10807039.2018.1480353

Akkoyunlu, A., Akiner, M.E. (2012) Pollution evaluation in streams using water quality indices: A case study from Turkey's Sapanca Lake Basin. Ecol. Indic., v.18, pp.501-511. doi: 10.1016/j.ecolind.2011.12.018

APHA (2005) standard methods for examination of water and wastewater (American Public Health Association Washington), DC, USA

Arumugam K, Elangovan K (2009) Hydrochemical characteristics and groundwater quality assessment in Tirupur region, Coimbatore District, Tamil Nadu, India. Environ. Geol., v.58, pp.1509-1520. doi:10.1007/s00254-008-1652-y

CGWB (2013) Ground water information booklet Gaya district Bihar state. Central ground water board Ministry of Water Resources. Govt. of India

Chai, T., Draxler, R.R. (2014) Root mean square error (RMSE) or mean absolute error (MAE). Geosci. Model Dev. Discuss. doi:10.5194/gmdd-7-1525-2014

Chen, X., Chen, C., Jin, L. (2011) Principal Component Analyses in Anthropological Genetics. Adv. Anthropol., v.1, pp.9-14. doi:10.4236/aa.2011.12002

Chu, K., Liu, W., She, Y., Hua, Z., Tan, M., Liu, X., Gu, L., Jia, Y. (2018) Modified principal component analysis for identifying key environmental indicators and application to a large-scale tidal flat reclamation. Water (Switzerland). doi:10.3390/w10010069

Cotter, A.S., Chaubey, I., Costello, T.A., Soerens, T.S., Nelson, M.A. (2003) Water quality model output uncertainty as affected by spatial resolution of input data. Jour. Amer. Water Resour. Assoc., doi:10.1111/j.1752-1688.2003.tb04420.x

Coulibaly, P., Baldwin, C.K. (2005) Nonstationary hydrological time series forecasting using nonlinear dynamic methods. Jour. Hydrol. doi:10.1016/j.jhydrol.2004.10.008

Ding, S., Zhang, N., Xu, X., Guo, L., Zhang, J. (2015) Deep Extreme Learning Machine and Its Application in EEG Classification. Math Probl. Eng., doi:10.1155/2015/129021

Fisher, R.A., Mackenzie, W.A. (1923) Studies in crop variation: II. The manurial response of different potato varieties. Jour. Agricul. Sci., doi:10.1017/S0021859600003592

Gazzaz, N.M., Yusoff, M.K., Aris, A.Z., Juahir, H., Ramli, M.F. (2012) Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. Mar. Pollut. Bull., v.64, pp.2409-2420. doi:10.1016/j.marpolbul.2012.08.005

Goz, E., Yuceer, M., Karadurmus, E. (2019) Machine Learning Application of Dissolved Oxygen Prediction in River Water Quality. Proc 4th World Congr Civil, Struct. Environ. Eng. 7-8. doi:10.11159/iceptp19.119

Gradilla-Hernández, M.S., de Anda, J., Garcia-Gonzalez, A., Meza-Rodríguez, D., Yebra Montes, C., Perfecto-Avalos, Y. (2020) Multivariate water quality analysis of Lake Cajititlán, Mexico. Environ. Monit. Assess., doi:10.1007/s10661-019-7972-4

Guo, W., Xu, T., Tang, K., Yu, J., Chen, S. (2018) Online Sequential Extreme Learning Machine with Generalized Regularization and Adaptive Forgetting Factor for Time-Varying System Prediction. Math Probl Eng. doi:10.1155/2018/6195387

Haghiabi, A.H., Nasrolahi, A.H., Parsaie, A. (2018) Water quality prediction using machine learning methods. Water Qual. Res. Jour. Canada, v.53, pp.3-13. doi:10.2166/wqrj.2018.025

Heddam, S., Kisi, O. (2017) Extreme learning machines: a new approach for modeling dissolved oxygen (DO) concentration with and without water quality variables as predictors. Environ. Sci. Pollut. Res., v.24, pp.16702-16724. doi:10.1007/s11356-017-9283-z

Horton, R.K. (1965) An Index Number System for Rating Water Quality. Jour. Water Pollut. Control Fed., v.37(3). pp.300-305.

Huang, G. Bin, Slew, C.K. (2004) Extreme learning machine: RBF network case. In: 2004 8th International Conference on Control, Automation, Robotics and Vision (ICARCV)

Huang, G. Bin, Zhou, H., Ding, X., Zhang, R. (2012) Extreme learning machine for regression and multiclass classification. IEEE Trans. Syst. Man., Cybern Part B Cybern. doi:10.1109/TSMCB.2011.2168604

Huang, G. Bin, Zhu, Q.Y., Siew, C.K. (2004) Extreme learning machine: A

new learning scheme of feedforward neural networks. *In:* IEEE International Conference on Neural Networks - Conference Proceedings
Huang, G. Bin, Zhu, Q.Y., Siew, C.K. (2006) Extreme learning machine: Theory and applications. Neurocomputing, v.70, pp.489-501. doi:10.1016/j.neucom.2005.12.126
Huang, G-B, Siew, C-K (2005a) Extreme learning machine with randomly assigned RBF kernels. Internat. Jour. Inform. Tech., v.11, pp.16-24
Huang, G, Huang, G. Bin, Song, S., You, K. (2015) Trends in extreme learning machines: A review. Neural Networks, v.61, pp.32-48.
Hurley, T., Sadiq, R., Mazumder, A. (2012) Adaptation and evaluation of the Canadian Council of Ministers of the Environment Water Quality Index (CCME WQI) for use as an effective tool to characterize drinking source water quality. Water Resour., doi:10.1016/j.watres.2012.03.061
Jain, C.K., Bandyopadhyay, A., Bhadra, A. (2010) Assessment of ground water quality for drinking purpose, District Nainital, Uttarakhand, India. Environ. Monit. Assess., doi:10.1007/s10661-009-1031-5
Khan, R., Jhariya, D.C. (2017) Groundwater quality assessment for drinking purpose in Raipur city, Chhattisgarh using water quality index and geographic information system. Jour. Geol. Soc. India, v.90, pp.69-76. doi:10.1007/s12594-017-0665-0
Kumar, D., Singh, A., Jha, R.K., Sahoo, B.B., Sahoo, S.K., Jha, V. (2019a) Source characterization and human health risk assessment of nitrate in groundwater of middle Gangetic Plain, India. Arab Jour. Geosci., doi:10.1007/s12517-019-4519-5
Kumar, P.J.S., James, E.J. (2013) Development of Water Quality Index (WQI) model for the groundwater in Tirupur district, South India. Chinese Jour. Geochem. doi:10.1007/s11631-013-0631-5
Kumar, D., Singh, A., Jha, R.K., Sahoo, S.K., Jha, V. (2018) Using spatial statistics to identify the uranium hotspot in groundwater in the mid-eastern Gangetic plain, India. Environ. Earth Sci., doi:10.1007/s12665-018-7889-1
Kumar, D., Singh, A., Samui, P., Jha, R.K. (2019b) Forecasting monthly precipitation using sequential modelling. Hydrol. Sci. Jour., v.64, pp.690-700. doi.:10.1080/02626667.2019.1595624
Li, J. andLi, M. (2019) Prediction of ultra-short-term wind power based on BBO-KELM method. Jour. Renew Sustain. Energy, v.11. doi:10.1063/1.5113555
Li, S., Gu, S., Liu, W., Han, H., Zhang, Q. (2008) Water quality in relation to land use and land cover in the upper Han River Basin, China. Catena. doi:10.1016/j.catena.2008.06.005
Lu, X., Li, L.Y., Lei, K., Wang, L., Zhai, Y., Zhai, M. (2010) Water quality assessment of Wei River, China using fuzzy synthetic evaluation. Environ. Earth Sci., doi:10.1007/s12665-009-0303-2
Lu, X.J., Zhou, C., Huang, M.H., Lv, W.B. (2016) Regularized online sequential extreme learning machine with adaptive regulation factor for time-varying nonlinear system. Neurocomputing. doi:10.1016/j.neucom.2015.09.068
Mohebbi, M.R., Saeedi, R., Montazeri, A., Azam, Vaghefi, K., Labbafi, S., Oktaie, S., Abtahi, M., Mohagheghian, A. (2013) Assessment of water quality in groundwater resources of Iran using a modified drinking water quality index (DWQI). Ecol. Indic., v.30, pp.28-34. doi:10.1016/j.ecolind.2013.02.008
Nagaraju, A., Muralidhar, P., Sreedhar, Y. (2016) Hydrogeochemistry and Groundwater Quality Assessment of Rapur Area, Andhra Pradesh, South India. Jour. Geosci. Environ. Prot. doi:10.4236/gep.2016.44012
Pal, M. (2006) Support vector machines-based modelling of seismic liquefaction potential. Internat. Jour. Numer. Anal. Methods Geomech. doi:10.1002/nag.509
Parinet, B., Lhote, A., Legube, B. (2004) Principal component analysis: An appropriate tool for water quality evaluation and management - Application to a tropical lake system. Ecol. Modell., v.178, pp.295-311. doi:10.1016/j.ecolmodel.2004.03.007
Ramakrishnaiah, C.R., Sadashivaiah, C., Ranganna, G. (2009) Assessment of water quality index for the groundwater in Tumkur taluk, Karnataka state, India. E-Journal Chem., v.6, pp.523-530. doi:10.1155/2009/757424
Roy, B., Singh, M.P., Singh, A. (2019) A novel approach for rainfall-runoff modelling using a biogeography-based optimization technique. Internat. Jour. River Basin Managmt., pp.1-14. doi:10.1080/15715124.2019.1628035
Rudwick, M.J.S. (1998) Lyell and the Principles of Geology. Geol. Soc. Spec. Publ. doi:10.1144/GSL.SP.1998.143.01.02
Sahu, M., Mahapatra, S.S., Sahu, H.B., Patel, R.K. (2011) Prediction of Water Quality Index Using Neuro Fuzzy Inference System. Water Qual. Expo. Heal., v.3, pp.175-191. doi:10.1007/s12403-011-0054-7
Sharmin Yousuf Rikta, Md Shiblur Rahaman, Jakia Jerin Mehjabin, Md Khabir Uddin, Mohammad Mahbub Kabir, Shafi Mohammad Tareq (2016) Evaluation of water quality parameters and Humic substance status of Bangshi, Dhaleshwari and Padma Rivers in Bangladesh. Internat. Jour. Environ. Sci., v.6, pp.155-168. doi:10.6088/ijes.6018
Tripathi, M., Singal, S.K. (2019) Use of Principal Component Analysis for parameter selection for development of a novel Water Quality Index: A case study of river Ganga India. Ecol. Indic. doi:10.1016/j.ecolind.2018.09.025
Tyagi, S., Sharma, B. (2014) Water Quality Assessment in Terms of Water Quality Index Water Quality Assessment in Terms of Water Quality Index Water Quality Assessment in Terms of Water Quality Index. Amer. Jour. Water Resour., v.1(3), pp.34-38. doi:10.12691/ajwr-1-3-3
Tyagi, S., Sharma, B., Singh, P., Dobhal, R. (2017) Water quality assessment in terms of Water Quality Index. Interant. Res. Jour. Engg. Tech., v.1, pp.34-38.
Vasanthavigar, M., Srinivasamoorthy, K., Vijayaragavan, K., Rajiv, Ganthi, R., Chidambaram, S., Anandhan, P., Manivannan, R., Vasudevan, S. (2010) Application of water quality index for groundwater quality assessment: Thirumanimuttar sub-basin, Tamilnadu, India. Environ. Monit. Assess. doi:10.1007/s10661-009-1302-1
Water G, Booklet I, District G, State B, Region M Gaya, Bihar, 2009
WHO (2008) Guidelines for drinking water quality, third edition, incorporating first and second edition, Geneva
Yadav, B., Ch, S., Mathur, S., Adamowski, J. (2016) Discharge forecasting using an Online Sequential Extreme Learning Machine (OS-ELM) model: A case study in Neckar River, Germany. Meas. Jour. Int Meas Confed. doi:10.1016/j.measurement.2016.06.042
Yidana, S.M. (2010) Groundwater classification using multivariate statistical methods: Southern Ghana. Jour. African Earth Sci. doi10.1016/j.jafrearsci.2009.12.002

(Received: 18 September 2020; Revised form accepted: 22 February 2021)