Water Practice & Technology

IWA PUBLISHING

Water Practice & Technology Vol 18 No 1, 201 doi: 10.2166/wpt.2022.156

Surface water quality assessment by Random Forest

Pramod Kumar Jena [a,b,*], Sayed Modinur Rahaman[a], Pradeep Kumar Das Mohapatra [a], Durga Prasad Barik [c] and Dikshya Surabhi Patra [d]

[a] Department of Microbiology, Raiganj University, Raiganj 733134, Uttar Dinajpur, West-Bengal, India
[b] Eureka Forbes Institute of Environment and Science, Jadavpur-700032, Kolkata, West-Bengal, India
[c] Department of Botany and Biotechnology, Ravenshaw University, Cuttack – 753 003 Odisha, India
[d] Department of Electronics and Communication, Sambalpur University Institute Of Information Technology (SUIIT), Jyoti Vihar-768019, Burla, Odisha, India
*Corresponding author. E-mail: pramodkumarjena@gmail.com

PKJ, 0000–0002-2200-5553; PKD, 0000-0002-9292-992X; DPB, 0000-0002-0319-968X; DSP, 0000-0001-7184-1067

ABSTRACT

The energetic nature of these important water resources makes them the most vulnerable to contamination from additional waste from multiple sources. Water quality monitoring is critical to water environmental management, and successful monitoring provides direction and confirms the effectiveness of water management. Models based on artificial intelligence are fundamental for anticipating appropriate moderation measures for surface water quality. In any case, it remains a challenge and requires a requirement to improve display accuracy. Faster and cheaper control is required due to the real-world impact of low water quality. With this inspiration, this research examines an array of machine-learning calculations to estimate water quality. The proposed approach uses Random Forest for modeling and is also useful for predicting surface water quality in the Kulik geographic region of West Bengal, India. It is a good tool for assessing the quality and ensuring the safe use of drinking water. Various water quality parameters (iron, fluoride, total coliform, fecal coliform, pH, total dissolved solids, magnesium, alkalinity, chloride, total hardness, nitrate, calcium, and *Escherichia coli*) were measured seasonally (winter, summer, rain) over 10 years (2010–2019). The estimated water quality parameters in this study were total dissolved solids (TDS), pH, and iron.

Key words: artificial intelligence, Kulik River, modeling, prediction water quality, Random Forest, surface water

HIGHLIGHTS

- Most of the north-Bengal people are depend on Kulik River for multiple purposes like settlement, cultivation, irrigation, fishing and various primary activities, so there is a need for water quality monitoring and management of Kulik River.
- Analysis and prediction of 13 parameters will be helpful for society.
- The proposed approach used Random Forest for modeling and assessing the water quality.

1. INTRODUCTION

The water quality includes a coordinated effect on the open well-being and the environment. Water is used for various households such as drinking water, horticulture, and industries. Recently, the advancement of water sports and excitement has done much to attract visitors (Jennings 2007). Among various water delivery providers, rivers have often been further utilized for the development of human societies due to smooth access. Using various water sources, including soil water and seawater, helped with problems at times. For example, the use of groundwater without adequate replenishment leads to subsidence (Motagh *et al.* 2017), and the use of seawater is usually associated with the transfer of pollutants (El-Kowrany *et al.* 2016). Therefore, the use of rivers has attracted attention. Observing water from rivers is not an uncommon job topic in earth science.

The study of the excellence of river processes is considered, together with the measurement of the excellent additions of the water and the definition of the pollutant transfer mechanism (Kashefipour 2002; Kashefefipour & Falconer 2012; Naseri Maleki & Kashefipour 2012; Qishlaqi *et al.* 2016). Among the water quality components, measuring dissolved oxygen (DO), chemical oxygen demand (COD), biochemical oxygen demand (BOD), electrical conductivity (EC), pH, temperature, K, Na, Mg, etc. have been proposed (Şener *et al.* 2017). To this end,

governments have built hydrometric stations along rivers originating from urban regions, agro-commercial tasks, industrial zones, and rivers that are part of reservoirs (Herschy 1993; Kejiang 1993). Water quality assessment is a basic degree for improving agricultural tasks in terms of the devotion to cultivation patterns, the form of irrigation machines, and structures of water purification for industry (Chen *et al.* 2017). To study the mechanism of pollutant transfer, superior numerical techniques including computational hydraulics, photo processing, and GIS techniques were applied in addition to the sector and laboratory experiments (Parsai & Haghiabi 2015, 2017a, 2017b).

By reviewing the time records of prominent water additives, investigators have attempted to estimate fate values. Currently, researchers have tried to adequately study the temporal accumulation of water-soluble additives and their internal relationship by using advanced soft computational strategies in the fields of water and environmental engineering (May *et al.* 2008; Palani *et al.* 2008; Haghiabi 2016a, 2016b; Jaddi & Abdullah 2017). In this regard, Emamgholizadeh *et al.* (2013) have done a study on the prediction of Multilayer Perceptron (MLP), Radial Basis Function (RBF), and an Adoptive Neuro-Fuzzy Inference System (ANFIS) for Water Excellent Additions to the Karoon River. They said that anyone who implemented modes had a reasonable overall performance for predicting water quality additions: however, the MLP modes turned into barely extra correct. Shokoohi *et al.* (2017) did an excellent job of controlling the water using a water dispensing machine. They consider this an optimization problem and use state-of-the-art optimization techniques to solve it. Zhang *et al.* (2010) brought a brand-new method for water allocation.

They consider water to be one of the most important elements of their method. Nikoo & Mahjouri (2013) have developed a PSVM (Probabilistic Support Vector Machines) version related to the GIS method for making plans for the nature and distribution of soil and groundwater in Iran. They said that using these techniques could provide correct statistics for feasibility research of water conservation tasks. Heddam (2016a, 2016b, 2016c, 2016d, 2016e) has applied synthetic neural networks to predict the excellent additives in water in numerous case studies.

He said synthetic intelligence strategies have reasonable overall performance for modeling and predicting the intrinsic relationship between the water additives and modeling their time collection. The review of the literature shows that excellent water assessment and forecasting is an essential matter for growing water conservation tasks, and synthetic intelligence strategies have been proposed for this purpose. Therefore, based on this observation, it was expected that the water additions of the Kulik River, the main river of the city of Raiganj, would be utilized by Random Forest.

2. MATERIALS AND METHODS

2.1. Study area

Uttar Dinajpur district is one of the backward districts of West Bengal, India whose economy is primarily based on agriculture. This district has 761 backward mouzas and the general cultivable land is 2,60,947.00ha. Strong minor irrigation sports have lifted the irrigation repute to an awesome quantity have some stage in the ultimate decades. (http://uttardinajpur.nic.in/waterresources.html). Raiganj is the headquarters of the Uttar Dinajpur district. The Kulik River is a transboundary river that flows through the Indian states of West Bengal, Bihar, and Bangladesh. In the Kulik River basin (Figure 1), the latitude is 25.635841° N and the longitude is 88.1222748° E. The Kulik River has started its journey from a wetland situated in Bangladesh. Kulik enters Uttar Dinajpur district at the side of the north-eastern part of Paharpur village in Hemtabad block. Then it flows through Bahin, Balia, and Raiganj in direction of the northeast to the southwest and finally meets with Nagorat at the place of the West Bengal–Bihar Border.

2.2. Methodology

All the samples were collected from the four hydrological stations (Table 1) by the authors. Over 10 years (2010–2019) (Table 2), the surface water quality of the river basin was assessed through systematic sampling. Similar work was conducted by Roy *et al.* (2022). The samples for the parameters in the used data set are represented numerically. For prediction, either a water quality assessment index can be made, or a regression model can be used to make the prediction. Out of 14 physiological and biological water quality parameters such as chloride, alkalinity, total hardness, magnesium, total iron, calcium, *Escherichia coli*, fecal coliforms, and total coliforms, only three parameters (pH, TDS, and iron) were considered and used for modeling and prediction – Random Forest process is applied.

2.3. Random Forest

For classification and regression problems, many people use supervised machine learning, and Random Forest is one of them. Breiman (2001) proposed the Random Forest algorithm, which was extremely successful as a

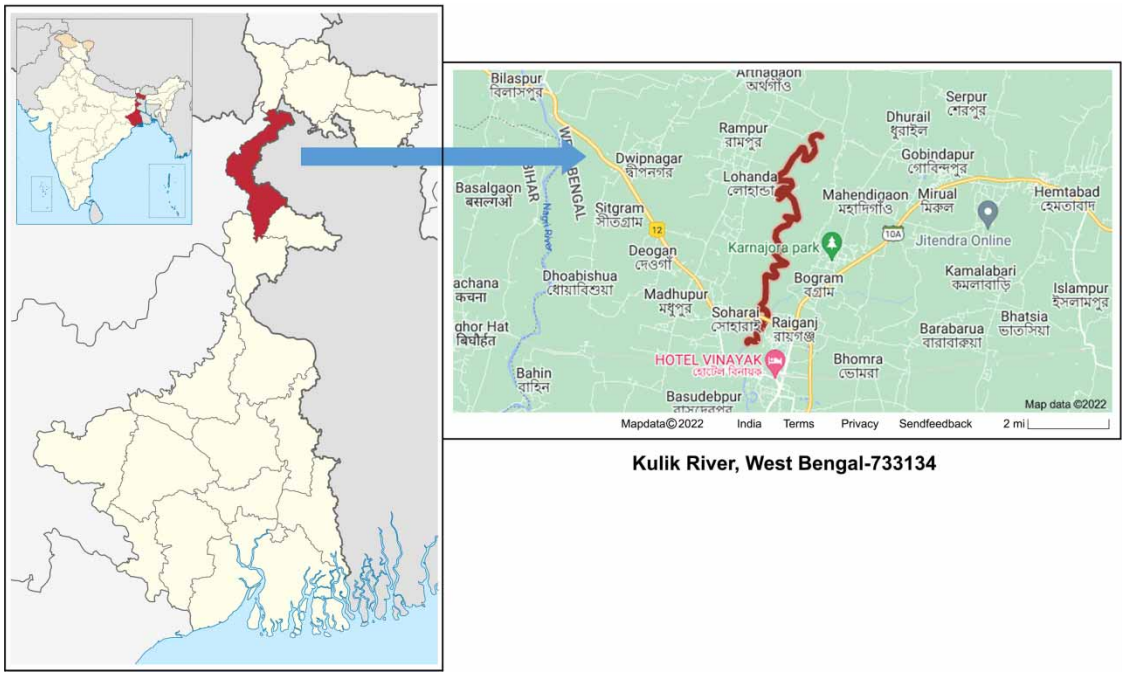

Figure 1 | Kulik River map taken from Wikipedia and Google Map.

Table 1 | Geo:co-ordinates of the four sampling sites of Kulik River

Sl. No.	Sampling sites	Name of the locality	Latitude	Longitude
1	L-1	Kalibari	25° 38′ 10″N	88° 07′ 25″E
2	L-2	Kulik bridge (on NH-12)	25° 38′ 06.7″N	88° 07′ 19.9″E
3	L-3	Abdulghata	25° 38′ 10″N	88° 07′ 25″E
4	L-4	Bamuyaghat	25° 40′ 26″N	88° 09′ 06″E

general-purpose classification and regression technique. The approach, which shuffles numerous randomized selection trees and aggregates their predictions by averaging, has shown an excellent overall performance in settings where the set of variables is much larger than the number of observations. In addition, it is flexible enough to be applied to large-scale problems, easily adaptable to various ad hoc study tasks, and returns measures of different meanings.

The Random Forest regression algorithm was chosen for the following two key reasons:

- Multivariate regression analysis: The target parameter can be dependent on multiple attributes/parameters. This type of many-to-one relationship requires multivariate regression analysis instead of the usual one-to-one linear regression analysis.
- Relatively small dataset: As the total number of samples in the used dataset is less than 5,000, it is considered to be a small dataset. Small datasets are difficult to analyse as sufficient samples are required to train a model as well as for the model testing and validation process

2.4. Software used for Random Forest

Python was used for model building and prediction analysis. From the scikit-learn package, Random Forest Regressor algorithm was imported from the ensemble methods available. The dataset was split into train and test samples in a 7:3 ratio using the train_test_split method from the sklearn package. For visualization, matplotlib and seaborn packages are used. Pandas package was used in the formatting of the dataset, and pre-processing methods.

Table 2 | Summary of descriptive statistics for water quality parameters

Year	TDS (mg/L) AVG	AVG pH	AVG Total Alkalinity (mg/L)	AVG Total Hardness (mg/L)	AVG Calcium as Ca(mg/L)	AVG Magnesium as Mg(mg/L)	AVG Chloride as Cl(mg/L)	AVG Sulfate as SO4(mg/L)	AVG Nitrate as NO3(mg/L)	AVG Total Iron as Fe(mg/L)	AVG Fluoride as F(mg/L)	AVG Total Coliforms (MPN/100 ml)	AVG Fecal Coliforms (MPN/100 ml)
2010	71.84	6.92	36.99	35.56	18.96	6.41	12.91	5.36	1.12	0.94	0.10	34.80	17.22
2011	69.35	6.94	36.85	34.19	18.85	6.34	12.71	5.33	1.12	0.94	0.10	30.45	12.87
2012	72.01	6.99	36.99	34.52	18.88	6.40	12.84	5.33	1.12	1.04	0.10	36.58	25.23
2013	74.35	6.89	36.99	34.62	18.88	6.44	12.91	5.33	1.12	0.74	0.10	31.76	12.64
2014	72.35	6.89	36.99	34.56	18.95	6.44	12.91	5.33	1.12	0.64	0.10	37.76	14.64
2015	73.68	6.89	36.80	34.66	18.90	6.40	12.91	5.32	1.12	0.74	0.10	43.76	18.64
2016	68.01	6.98	37.02	34.86	19.03	6.44	13.01	5.32	1.12	0.72	0.10	33.21	13.27
2017	69.48	6.98	37.02	34.98	18.98	6.40	13.01	5.33	1.12	0.93	0.10	32.57	15.98
2018	72.66	6.94	37.09	38.59	19.05	6.44	13.04	5.33	1.12	1.32	0.10	31.07	12.84
2019	77.01	6.83	37.23	40.62	19.05	6.41	13.07	5.34	1.10	1.40	0.10	33.93	22.01

2.5. Model validation

The consequences derived from the version were assessed using several statistical tests. R^2 is used to assess the relationship between located values and expected values. The equation for the calculation is as follows:

X = data. Drop ('TDS', axis = 1)
y = data['TDS']

reg = Random Forest Regressor (estimators = 100, max_ depth = 3,
max features = '$auto$', $min\ samples\ leaf$ = 4, bootstrap = True,
n_jobs = −1, random state = 0)
reg_fit = reg.fit(X, y)

2.6. Model accuracy

Here, we have taken two key parameters which are used to measure whether the model can predict the target parameter with high accuracy.

- R^2 (coefficient of determination): It is a statistical measure for how well the regression line is able to approximate the actual data.

$$R^2 = 1 - \left[\frac{\sum (y_i - y_i')^2}{\sum (y_i - Y)^2}\right]$$

where y_i is the actual ith sample; y_i' is the predicted ith sample; Y is the mean of the target parameter; R^2 has a range from 0 to 1. Higher the R^2 for a model, better the model can predict for the target parameter.

- RMSE (root mean squared error): It gives the standard deviation of the prediction errors(residuals). It measures how spread out the errors are from the main concentration of actual data points.

$$\text{RMSE} = \left[\frac{\sum (y_i' - y_i)^2}{N}\right]^{1/2}$$

where N is the total number of samples.

Lower the RMSE for a model, the prediction is more precise with less residuals.

2.7. Dataset

The dataset used in this paper is made of parameters that are considered to be vital for a healthy water ecosystem, such as, total iron content, pH, sulfate, and nitrate levels. The breakdown of organic matter is measured in terms of TDS, the total amount of coliforms, and fecal coliform contents. Fluoride content, hardness, and alkalinity of the water are considered for the ergonomic use of the river water.

The data are collected from four different locations on the Kulik River bed. The samples for the parameters are in numerical representation except for the presence of *E. coli* bacteria.

2.8. Random Forest regression algorithm

Random Forest is a type of ensemble learning algorithm as it uses multiple decision trees to estimate a prediction with high accuracy. When multiple attributes are heavily correlated to the target parameter, a decision tree selects the parameter with the highest correlation with the target. From there, it starts the prediction process with a sequence of comparisons with other parameters based on pre-learned threshold values. Starting from the top (parameter with the highest correlation with the target), it works its way to the lowest level nodes (with the least correlation with the target), resulting in a leaf (decision/prediction) at the end of the tree. The comparison is done using MSE (mean squared error) to determine how the data branches from each node. It is given by the MSE equation.

3. RESULTS

Ravindra *et al.* (2022) had done the analysis of the surface water quality of the Amba River. Sipra & Baliarsingh (2017) have done the surface water quality analysis of the Kathojodi River for prediction and modeling. Results of a study of the physicochemical and microbial parameters of the Kulik River were studied and represented. We have studied 10 years of data from 2010 to 2019, out of which 2019 shows the highest TDS value, whereas the lowest was found in 2011. During the study of pH, water sample is slightly acidic pH (6.83) during 2019 while in 2012 pH is neutral (6.99). A similar type of work was carried out by Rabindra *et al.* for Amba River.

3.1. Random Forest model's development

Data for the Kulik flow were collected over a decade (2010–2019) for TDS, pH, total iron content, fluoride content, presence or absence of *E. coli* and its content, chloride, magnesium, calcium, total alkalinity, total hardness, sulfate, and nitrate levels were documented.

A Random Forest model can help accurately predict values for multiple predictors. Regression analysis can help determine which variables most affect the value to be predicted. The data were processed by individual date.

3.1.1. Prediction with Random Forest for TDS

Figures 2 and 3 shows the graphs for TDS after processing and correlating TDS among other parameters data, respectively. From Figure 4, we found that none of them have a cross-correlation of approximately close to 1. Table 3 shows the use of correlation to select the features most influenced by the target parameter, TDS. Except for pH, all other parameters are valued with unit mg/L. From Table 3, it seems that the most influencing features for TDS are pH, Ca, and Mg.

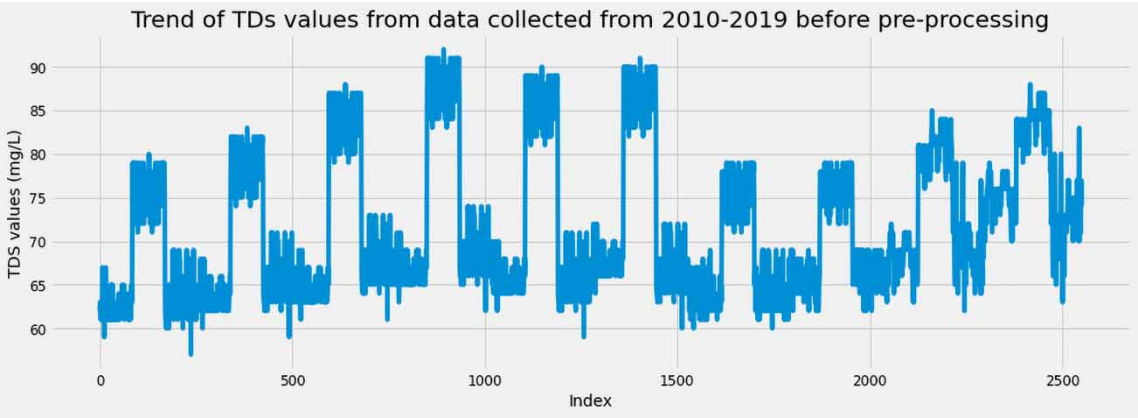

Figure 2 | Plot for TDS.

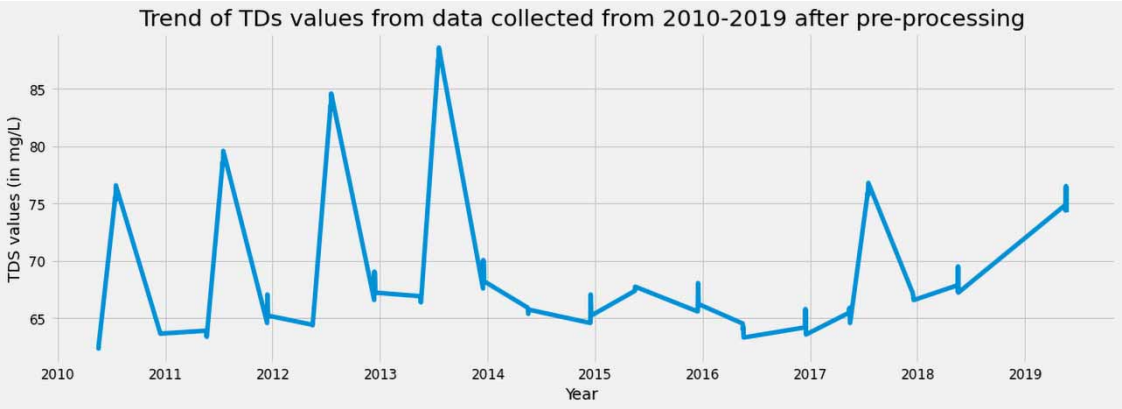

Figure 3 | Plot for TDS after pre-processing.

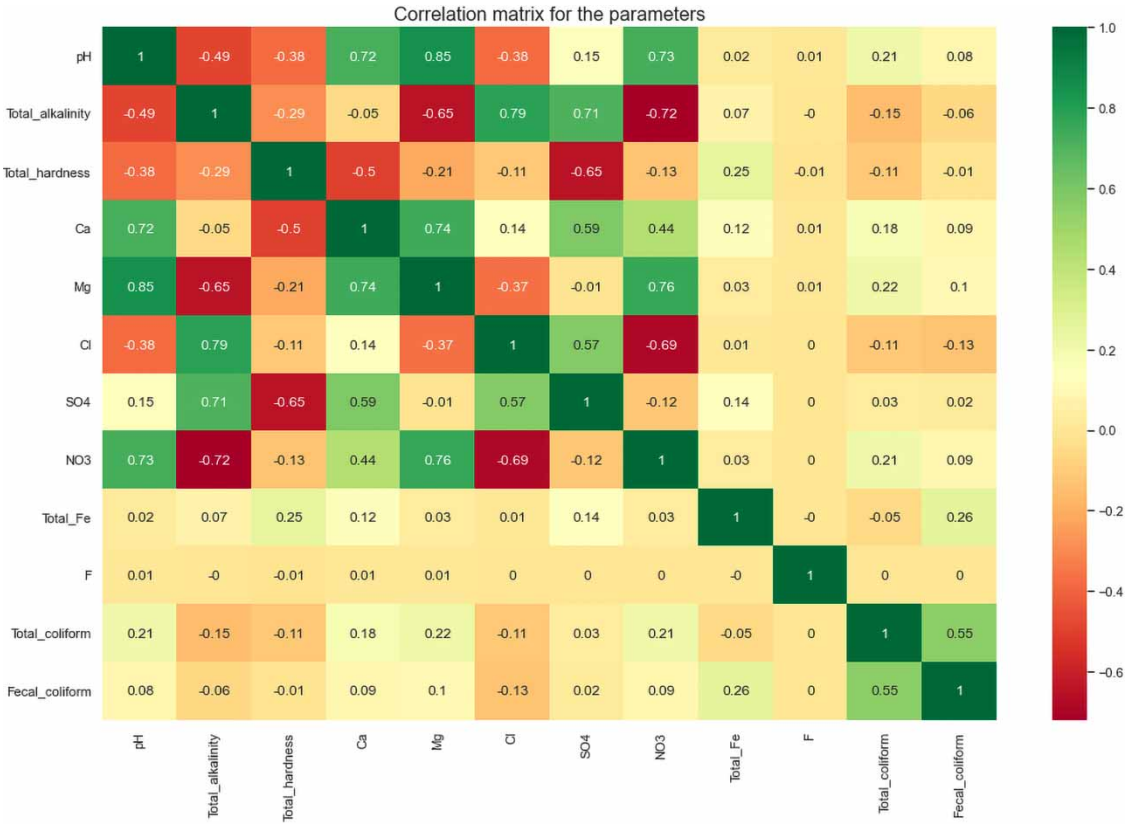

Figure 4 | Plot shows the correlation among other parameters.

Table 3 | Correlation table for the features and target, TDS

Features	Feature importance (correlation with target)
pH	0.71
Total alkalinity	−0.33
Total hardness	−0.17
Calcium	0.765
Magnesium	0.79
Chlorine	−0.18
SO_4	0.31
NO_3	0.66
Total iron	0.25
Fluorine	0.06
Total coliform	0.29
Fecal coliform	0.12

Figure 5 shows the least to most influential characteristics of TDS. Therefore, the important features to consider are pH, Ca, and Mg for building a Random Forest regression model. Figure 6 shows the model evaluation and comparison of predicted and actual values of TDS. Here, the $R^2 = 90.8\%$, RMSE = 2.313, and accuracy is 97.74.

3.1.2. Prediction with Random Forest for pH

Figure 7 and 8 show the graphs for pH after processing and correlating with others respectively. Figure 9 shows the correlation between the features and target (pH). Table 4 shows the use of correlation to select the features

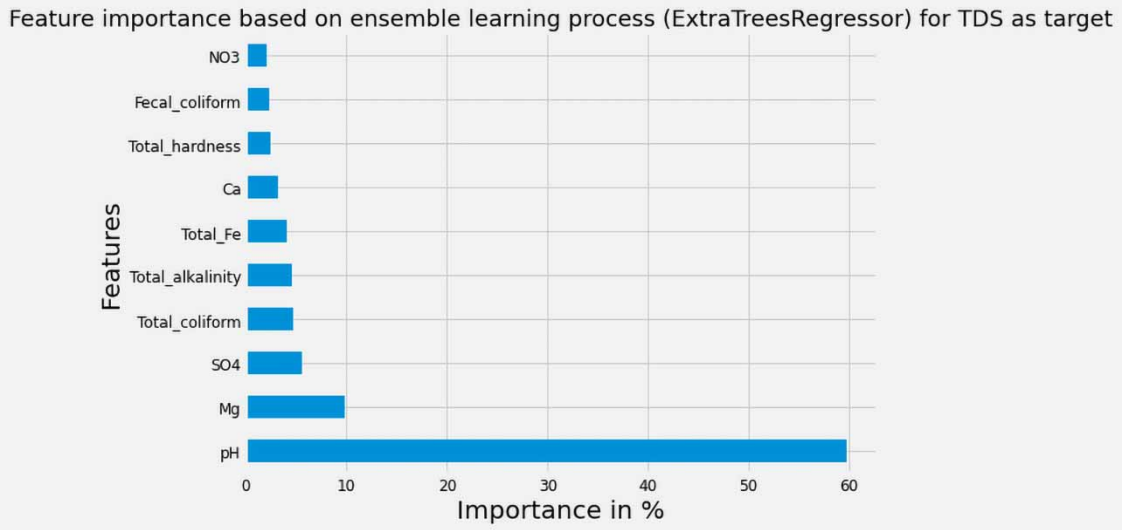

Figure 5 | Extra-tree regressor for the parameter TDS.

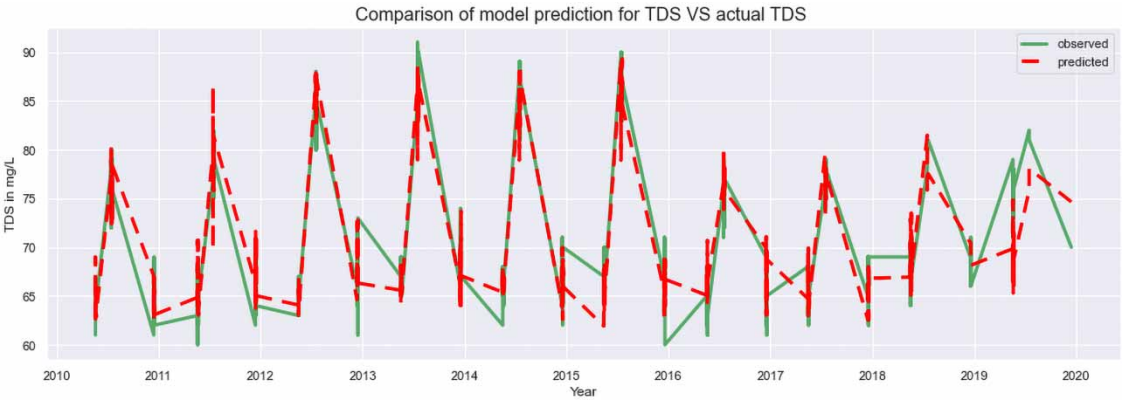

Figure 6 | Comparison of predicted and actual values for TDS.

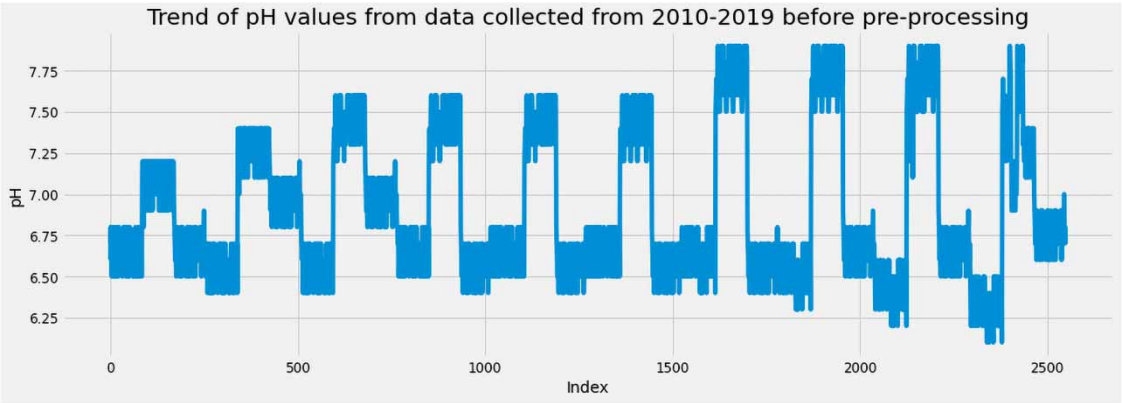

Figure 7 | Plot for pH.

most influential to the target parameter pH. Features that stand out are Mg, NO3, TDS, Ca. Figure 10 shows an additional tree regressor, and Mg, Ca, NO_3, and TDS are the most influential features for pH. Figure 11 shows the comparison of predicted and actual pH values. Here, the model score is as follows: $R^2 = 86.08\%$, RMSE = 0.14, and accuracy = 98.66.

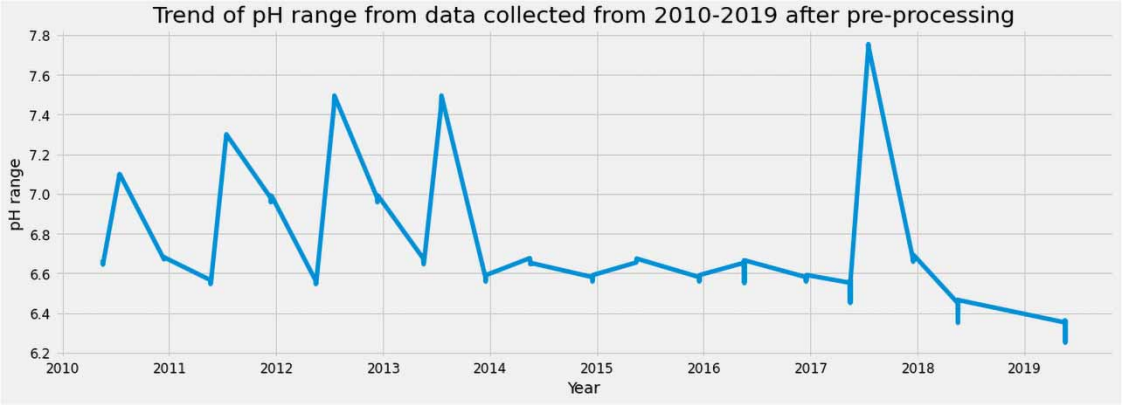

Figure 8 | Plot for pH after pre-processing.

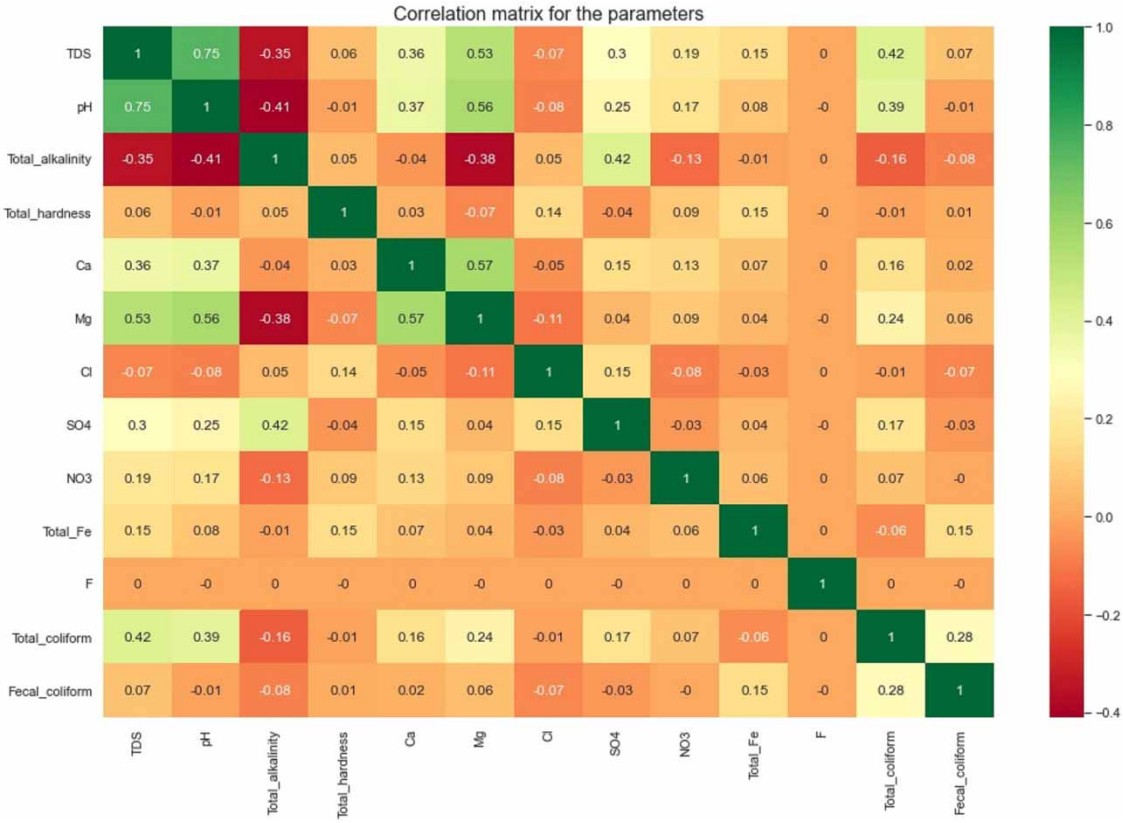

Figure 9 | Plot shows the co-relation among other parameters.

3.1.3. Prediction with Random Forest for iron

Figures 12–14 show the graphs for total iron after processing and correlation among other parameters, respectively, with data. Here, in Table 5, we have used correlation to select the features most influential to the target parameter, iron. Except for pH, all other parameters are valued with unit mg/L. Therefore, the most influential characteristics are pH, total hardness, and TDS.

Figure 15 shows the comparison of the predicted and actual iron values. Here, the model score is as follows: $R^2 = 46.78\%$, RMSE = 0.137, and accuracy = 98.65

Table 4 | Correlation table for the features and target, pH

Features	Feature importance (correlation with target)
TDS	0.71
Total alkalinity	–0.49
Total hardness	–0.37
Calcium	0.71
Magnesium	0.85
Chlorine	–0.38
SO_4	0.14
NO_3	0.72
Total iron	0.02
Fluorine	0.06
Total coliforms	0.21
Fecal coliforms	0.07

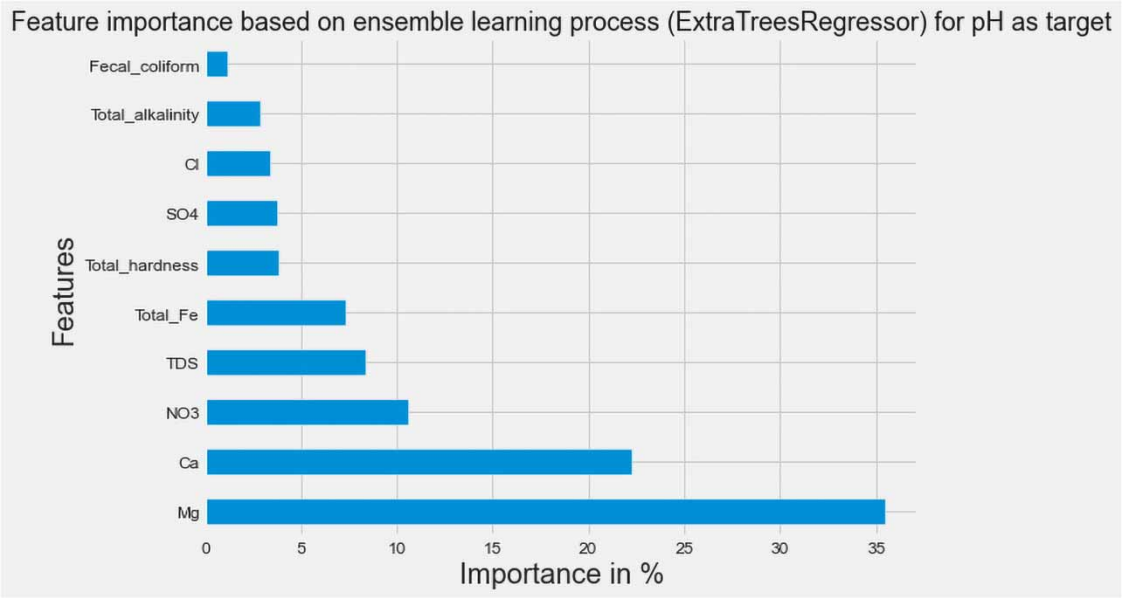

Figure 10 | Extra-tree regressor for the parameter pH.

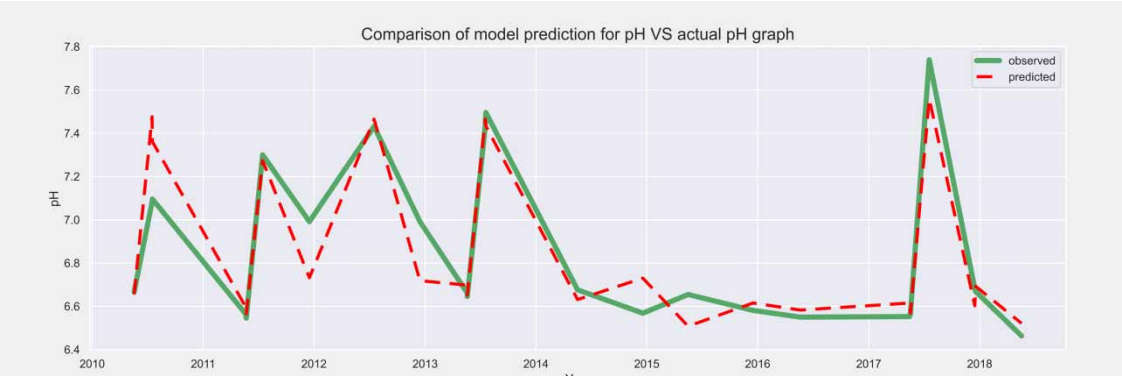

Figure 11 | Comparison of predicted and actual values for pH.

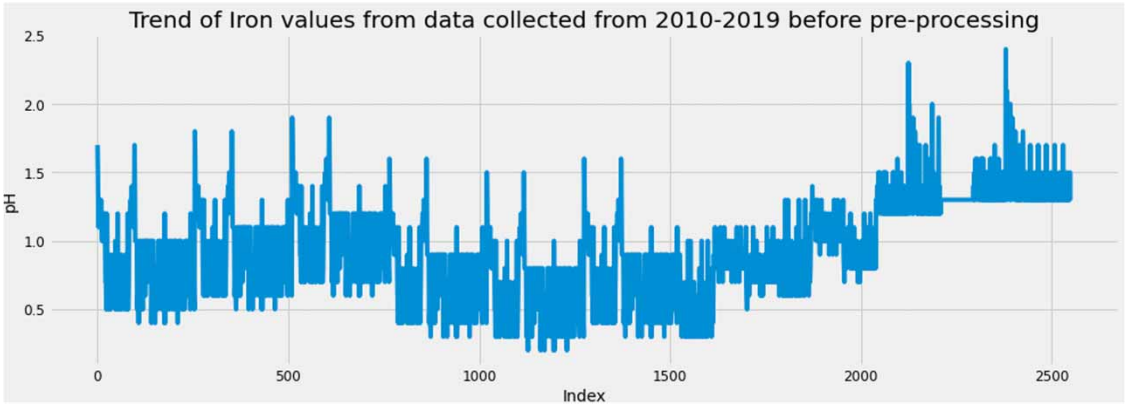

Figure 12 | Plot for total iron.

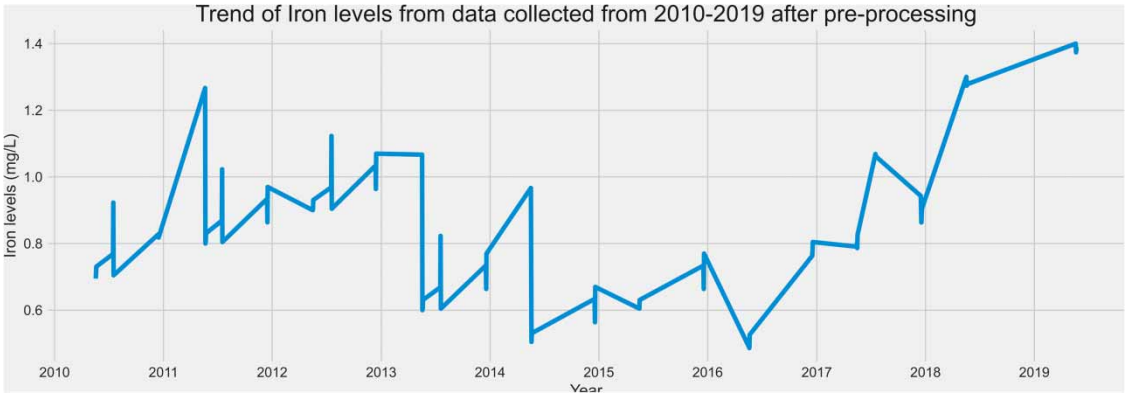

Figure 13 | Plot for total iron after pre-processing.

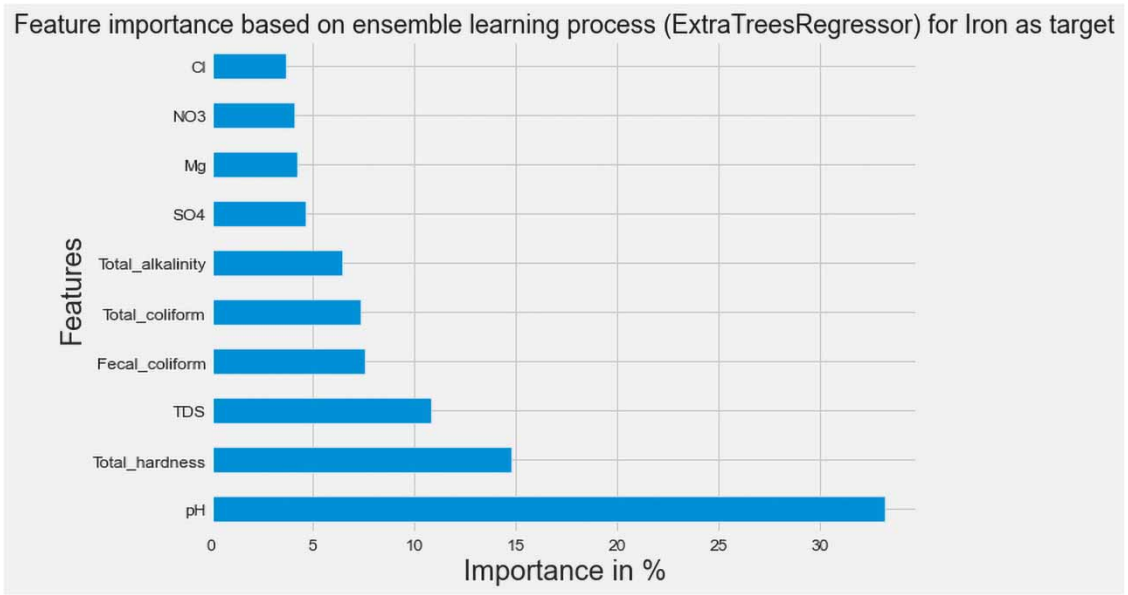

Figure 14 | Extra-tree regressor for the parameter iron.

Table 5 | Correlation table for the features and target: iron

Features	Feature importance (correlation with target)
pH	0.02
Total alkalinity	0.07
Total hardness	0.25
Calcium	0.12
Magnesium	0.03
Chlorine	0.01
SO_4	0.13
NO_3	0.03
TDS	0.25
Fluorine	−0.05
Total coliforms	−0.05
Fecal coliforms	0.26

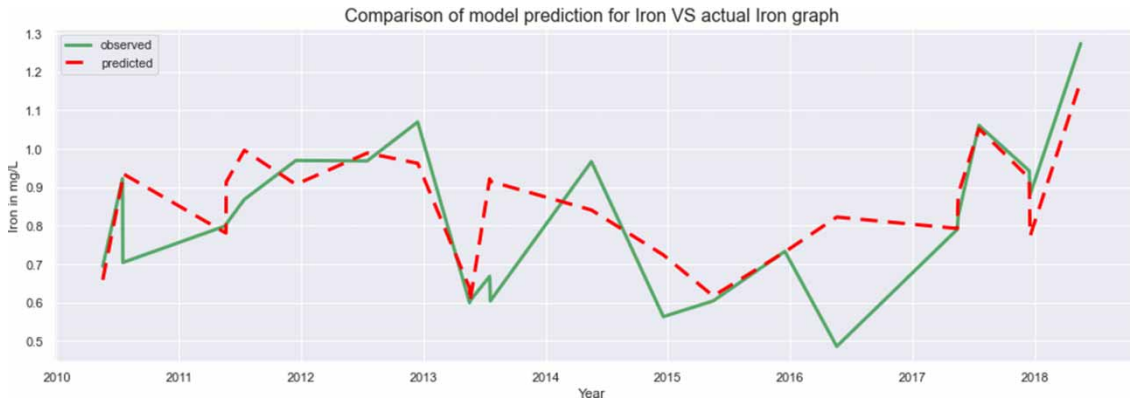

Figure 15 | Comparison of predicted and actual values for iron.

4. DISCUSSION

Predictions of the internal relationships between water quality elements are shown in this section of the paper. Traditional GEP, SVM, RF, ANN, DT, and regression-based models are used in most published works on modeling surface water quality parameters. Many models and forecast water quality indicators use traditional AI algorithms, but the results were not as expected. Consequently, it is extremely important to mix modeling processes and optimization algorithms to achieve powerful and correct modeling results. Few researchers integrate modeling and state search for input optimization in addition to modern work.

AI-Mukhtar & AI-Yaseen (2019) discovered that regression models and ANN outperform regression models and ANN in predicting EC and TDS in a comparison to previous and current studies that used modeling and optimization techniques. In addition, AliKhan *et al.* (2021) reported improved model results for predicting surface water salinity of the Indus using Random Forest.

The results of modern studies have proved that the input optimization system can be used to achieve modeling accuracy, the highest quality structure, reduced computation time, input optimization, and reduced version complexity. In addition, built-in optimization algorithms perform better than standalone ANN, SVM, GEP, RF, and other regression analyses, delivering a powerful version with advanced output.

5. CONCLUSION

The score obtained from this test confirmed TDS, pH, total iron modeling, and prediction. Regardless of the version in the water quality of the river, the version will advance correctly. The overall performance of advanced

models was evaluated through the use of special statistical standards, e.g. modeling accuracy (R^2) and error evaluation standards (RMSE). Input optimization reduced modeling complexity, which is useful for reducing information series and processing overhead. The accuracy of TDS, pH, and total iron was 97.74, 98.66, and 98.65, respectively. The advantage of using the RF version proposed in this document is the accurate assessment of soil water pollutant levels, and furthermore, it allows to avoid lengthy calculations feared with traditional water quality index (WQI).

FUNDING

This research received no external funding.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

AliKhan, M., Izhar Shah, M. & Javed, M. F. 2021 Application of random forest for modelling of surface water salinity. *Ain Shams Eng. J.* **13**(2022), 101635.

AI-Mukhtar, M. & AI-Yaseen, F. 2019 Modeling water quality parameters using data driven models, a case study Abu-Ziriq marsh in south of Iraq. *Hydrology* **6**(1), 24.

Breiman, L. 2001 Random forests. *Mach. Learn.* **45**, 5–32. https://doi.org/10.1023/A:1010933404324.

Chen, X., Chen, Y., Shimizu, T., Niu, J., Nakagami, K. i., Qian, X., Jia, B., Nakajima, J., Han, J. & Li, J. 2017 Water resources management in the urban agglomeration of the Lake Biwa region, Japan: an ecosystem services-based sustainability assessment. *Sci. Total Environ.* **586**(Suppl. C), 174–187.

El-Kowrany, S. I., El- Zamarany, E. A., El-Nouby, K. A., El-Mehy, D. A., Abo Ali, E. A., Othman, A. A., Salah, W. & El-Ebiary, A. A. 2016 Water pollution in the Middle Nile Delta, Egypt: an environmental study. *J. Adv. Res.* **7**(5), 781–794.

Emamgholizadeh, S., Kashi, H., Marofpoor, I. & Zalaghi, E. 2013 *Prediction of Water Quality Parameters of Karoon River (Iran)*.

Haghiabi, A. H. 2016a Modeling river mixing mechanism using data driven model. *Water Resour. Manage.* **31**(3), 811–824.

Haghiabi, A. H. 2016b Prediction of longitudinal dispersion coefficient using multivariate adaptive regression splines. *J. Earth Syst. Sci.* **125**(5), 985–995.

Heddam, S. 2016a Generalized regression neural network based approach as a new tool for predicting total dissolved gas (TDG) downstream of spillways of dams: a case study of Columbia River Basin Dams, USA. *Environ. Process.* **4**(1), 235–253.

Heddam, S. 2016b Multilayer perceptron neural network-based approach for modeling phycocyanin pigment concentrations: case study from lower Charles River buoy, USA. *Environ. Sci. Pollut. Res.* **23**(17), 17210–17225.

Heddam, S. 2016c New modelling strategy based on radial basis function neural network (RBFNN) for predicting dissolved oxygen concentration using the components of the Gregorian calendar as inputs: case study of Clackamas River, Oregon, USA. *Model. Earth Syst. Environ.* **2**(4), 162–167.

Heddam, S. 2016d Secchi disk depth estimation from water quality parameters: artificial neural network versus multiple linear regression models? *Environ. Process.* **3**(2), 525–536.

Heddam, S. 2016e Simultaneous modelling and forecasting of hourly dissolved oxygen concentration (DO) using radial basis function neural network (RBFNN) based approach: a case study from the Klamath River, Oregon, USA. *Model. Earth Syst. Environ.* **2**(3), 117–135.

Herschy, R. 1993 National and international standards in streamflow measurement. *Flow Meas. Instrum.* **4**(1), 53–55.

Jaddi, N. S. & Abdullah, S. 2017 A cooperative-competitive masterslave global-best harmony search for ANN optimization and water-quality prediction. *Appl. Soft Comput.* **51**, 209–224.

Jennings, G. 2007 *Water-based Tourism, Sport, Leisure, and Recreation Experiences*. Elsevier, Oxford.

Kashefipour, S. M. 2002 *Modelling Flow, Water Quality and Sediment Transport Processes in Reverine Basins*. PhD thesis, Cardiff University, Cardiff.

Kashefipour, S. M. & Falconer, R. A. 2002 Longitudinal dispersion coefficients in natural channels. *Water Res.* **36**(6), 1596–1608.

Kejiang, C. 1993 Flow measurement in large rivers in China. *Flow Meas. Instrum.* **4**(1), 47–50.

May, R. J., Dandy, G. C., Maier, H. R. & Nixon, J. B. 2008 Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems. *Environ. Model. Softw.* **23**(10–11), 1289–1299.

Motagh, M., Shamshiri, R., Haghshenas Haghighi, M., Wetzel, H.-U., Akbari, B., Nahavandchi, H., Roessner, S. & Arabi, S. 2017 Quantifying groundwater exploitation induced subsidence in the Rafsanjan plain, southeastern Iran, using InSAR time-series and in situ measurements. *Eng. Geol.* **218**, 134–151.

Naseri Maleki, M. & Kashefipour, S. M. 2012 Application of numerical modeling for solution of flow equations and estimation of water quality pollutants in rivers (Case study: Karkheh River). *Civil Environ. Eng.* **42.3**(68), 51–60.

Nikoo, M. R. & Mahjouri, N. 2013 Water quality zoning using probabilistic support vector machines and self-organizing maps. *Water Resour. Manage.* **27**(7), 2577–2594.

Palani, S., Liong, S.-Y. & Tkalich, P. 2008 An ANN application for water quality forecasting. *Mar. Pollut. Bull.* **56**(9), 1586–1597.

Parsaie, A. & Haghiabi, A. 2015 The effect of predicting discharge coefficient by neural network on increasing the numerical modeling accuracy of flow over side weir. *Water Resour. Manage.* **29**(4), 973–985.

Parsaie, A. & Haghiabi, A. H. 2017a Computational modeling of pollution transmission in rivers. *Appl. Water Sci.* **7**(3), 1213–1222.

Parsaie, A. & Haghiabi, A. H. 2017b Numerical routing of tracer concentrations in rivers with stagnant zones. *Water Sci. Technol. Water Supply* **17**(3), 825–834.

Qishlaqi, A., Kordian, S. & Parsaie, A. 2016 Hydrochemical evaluation of river water quality – a case study. *Appl. Water Sci.* **7**(5), 2337–2342.

Ravindra, J., Pramod, N., Manik, A., Dipen, P., Uday, K., Jayesh, J., Utkarsh, M. & Pramod, K. 2022 Analysis of seasonal variation in surface water quality and water quality index (WQI) of Amba River from Dolvi Region, Maharastra India. *Arabian J. Geo-Sci.* **15**, 1261. https://doi.org/10.1007/s12517-022-10542-3.

Roy, V., Saha, B. K., Saha, J. & Pal, A. 2022 Assessment of water quality of Kulik River of Raiganj with reference to physicochemical characteristics and potability. *Curr. World Environ.* **17**(2). http://dx.doi.org/10.12944/CWE.17.2.19

Şener, Ş., Şener, E. & Davraz, A. 2017 Evaluation of water quality using water quality index (WQI) method and GIS in Aksu River (SW-Turkey). *Sci. Total Environ.* **584–585**, 131–144.

Shokoohi, M., Tabesh, M., Nazif, S. & Dini, M. 2017 Water quality based multi-objective optimal design of water distribution systems. *Water Resour. Manage.* **31**(1), 93–108.

Sipra, M. & Baliarsingh, F. 2017 Surface water quality assessment and prediction modelling of Kathojodi River. *Int. J. Emerg. Res. Manage. Technol.* **6**(8), 447.

Zhang, W., Wang, Y., Peng, H., Li, Y., Tang, J. & Wu, K. B. 2010 A coupled water quantity–quality model for water allocation analysis. *Water Resour. Manage.* **24**(3), 485–511.

First received 25 August 2022; accepted in revised form 28 November 2022. Available online 9 December 2022