



Water quality forecasting using a Novel Hybrid DNN-MBGD optimization and WAWQI technique for assessment of surface water quality index in 10 districts of Uttar Pradesh

SUBHA SINHA

Department of Civil Engineering, Bakhtiyapur College of Engineering, Bakhtiyapur, Department of Science and Technology, Government of Bihar, India.

e-mail: starsubha@gmail.com

MS received 1 November 2022; revised 10 March 2023; accepted 19 March 2023

Chemicals from farm fields, salts from the road, pollutants from the atmospheric circulation, garbage from domestic and industrial wastes, hazardous wastes and animal wastes can lead to water pollution. Chemicals, bacteria and viruses from the septic system can also infect water contaminated with synthetic materials such as gasoline. This polluting water affects human health and is unsuitable for drinking, so water quality is important. This study estimates the water quality index by the Weighted Arithmetic Water Quality Index Method (WAWQI) using 20 water quality parameters in Uttar Pradesh for 10 districts in one year (January 2019–January 2020). Also, a novel hybrid Deep learning Neural Network-Mini Batch gradient descent optimization (DNN-MBGD) is used for water quality prediction. Based on the results, water from the experimental sites is unsuitable for drinking and other purposes like domestic and irrigation. The WQI value for these 10 sites is much greater than 100. After WQI calculation, the relationship between two water quality parameters are determined by a correlation matrix. Based on the feature importance score, input features were selected, and the performance of K was determined. Then the performance of K was predicted by the hybrid DNN-MBGD model and compared with the ANN (artificial neural network), SVM (support vector machine) and GMDH (group method of data handling). From the performance comparison and error analysis hybrid DNN-MBGD model result gives better performance than other models.

Keywords. Water pollution; water quality; DNN; MBGD; WAWQI; water resource management; ANN; water quality parameters; Uttar Pradesh.

1. Introduction

In the world above, 97% of the water is salt water occupied by oceans. The remaining 3% of the water is fresh water. Over 50% of the freshwater is used for drinking purposes and is also used for

agriculture and irrigation purposes. Water source is always available; streams and rivers are constantly dry and surface water is polluted in various ways. Horvat *et al.* (2021) investigated the water quality parameters in the fourth sector of Lake Palic in Serbia. This paper evaluated the water

quality parameters from 2011 to 2019 for 9 years. The data were subjected to principal component analysis (PCA) and machine learning (ML) classification algorithms that identified seasonality related to lake water quality. WQI was determined using two approaches to provide a general view of the overall quality of a lake. From 2011 to 2019, they monitored 13 water quality parameters. The final result compares the WA WQI with the CCME WQI (Canadian Council of Ministers of the Environment Water Quality Index) methods. The determination of the water quality index categorized the quality of the lake as good or poor. The overall temporal tendency was not shown in the result. The water quality parameters were examined by using the method of ML.

By using WQI and analysis of statistics, Ustaoglu *et al.* (2020) analyzed the water quality of the stream and river systems. From February 2017 to January 2018, the biology, water quality, oxygenation, and nutrient parameters were examined. Standard methods were used to investigate the performance of parameters. Pearson correlation index (PCI), PCA and clustering analysis (CA) are the statistical methods used to analyze the results. Hazard index (HI), stream water quality, and hazard quotient (HQ) were examined by WQI. This was used to solve the risk to human health. The results showed that the Turnasuyu has very good water quality characteristics and that the identified trace elements are not within the dangerous limit for public health. Dey and Vijay (2021) monitored the water quality parameters using geospatial techniques. Laboratory and fieldwork were reduced by using geospatial techniques. The water quality was assessed using the bio-optical method and the spectral ration method. The remote sensing analysis observed temperature, total suspended solids (TSS), and salinity. The results indicated the estimation of the non-optical parameters and proved the remote sensing technique for water quality measurement.

Valentini *et al.* (2021) adapted a WQI technique for calculating the parameters in Brazil. This study was designed to monitor water quality parameters through WQI correlation analysis. Using statistical methods to monitor water quality, the highest coefficient was indicated by ML4 point. Using ML models based on big data, Chen *et al.* (2020) identified the water quality parameters and studied the surface water quality performance. From 2012 to 2018 in China, the execution of water quality

parameters for lakes and rivers was compared using ten ML models and predicted the water quality parameters by recall, precision, weighted FI score and FI score. Deep Cascade Forest (DCF), Random Forest (RF) and Decision Tree (DT) identified the water parameters such as $\text{NH}_3\text{-N}$, DO, COD, etc.

Shalumon *et al.* (2021) estimated the water quality parameters for five districts of Kerala by using the WQI method. During monsoon season, 11 water quality parameters and uranium were analyzed for 30 samples. It was clearly described that the water quality requirement was satisfied, the uranium concentration for the study area was low, and the water was also acceptable for drinking purposes. Based on the Python framework and Raspberry Pi3 board development, Khatri *et al.* (2020) presented a real-time monitoring system for water quality. The validation result of water quality parameters such as PH, electrical conductivity, temperature, and dissolved oxygen were compared with benchmark equipment. A developed setup was used for real-time environment analysis, processing and logging of the data. Detection of contamination, network distribution monitoring, water quality for rivers and lakes and environmental monitoring are the important applications of real-time monitoring.

Kothari *et al.* (2021) investigated the water quality parameters for five districts of Uttarakhand in India based on the WQI calculation, biological and hydro-geochemical parameters. This study found suitable water quality parameters based on BIS standards for drinking purposes. The combination of WQI with multiple parameters was used to calculate the correlation coefficient by statistical analysis. The result showed that the correlation index describes the total coli form, concentration of total iron and faecal coli form. The concentration of chloride ions, sulphate and conductivity of correlation identified nitrate in the water. At the location of the Awash River in Ethiopia, Shishaye and Asfaw (2020) determined and analyzed the various water quality parameters for 10 years. Also, this study examined the composition of the geochemical and origin of the river water. From the result analysis, the 10 years samples and total dissolved solids (TDS), EC and PH were measured. Ray *et al.* (2021) presented a two-seasonal study of water quality parameters at 66 locations in Kerala. This study compared chemical water parameters with ecological water parameters with the help of

PCA, and the study result was that fast-growing algal species were identified.

The water quality parameters were determined by Najafzadeh and Niazmardi (2021). The algorithm of Multiple-Kernel Support Vector Regression (MKSVR), also known as Support Vector Regression (SVR), was the novel method used to evaluate the optimization problem. Using WQI, Biochemical Oxygen Demand (BOD) and Chemical Oxygen Demand (COD) were determined by MKSVR. Comparison of random forest regression (RFR) with observed SVR for the final result showed corresponding values of root mean squared error (RMSE), RFR and SVR. Xu *et al.* (2022) analyzed the easy and simple water quality parameters, which in turn, predicted the appearance of trihalomethane (THMs). To improve the THMs models, the radial basis function (RBF) adopted artificial neural network (ANN) and linear/log-linear regression models (LRM) and obtained 64 results from samples used to model development. Limited ability prediction and LRM parameters were shown in the results. For best results, Gaussian and a maximum number of neuron (MN) functions are used to enhance the different types of LRMs.

Pandey *et al.* (2020) determined the chemical and physical parameters of pre- and post-monsoon groundwater collected at the Koradi Thermal Power Plant Site in the Nagpur region. The result analysis has been determined during the pre- and post-monsoon. It was analyzed by statistical methods, which showed the electrical conductivity, total hardness and dissolved solids. GW-3, GW-5 and GW-9 were also known as groundwater resources used to determine the parameters. Chatterjee and Lataye (2020) investigated surface water quality parameters in Futala Lake, Nagpur. Eleven water quality parameters were determined between September 2017 and October 2017 using a geographic information system (GIS). The parameters were determined by kriging interpolation across the lake. The percentage difference of parameters was calculated, and the result indicated the water quality. Parameters were compared for immersion and pre-immersion, and affected areas were treated during immersion.

SBHS (Asadollah *et al.* 2021) analyzed a prediction of WQI for rivers using extra tree regression (ETR), also known as the ML model. The performance of ETR was compared with classic stand-alone models, decision tree regression (DTR)

and SVR. The result shows that the ETR model produces more accurate WQI predictions for the training and testing phases. Although all 10 input variables achieved the highest prediction performance, the combination of input parameters, including only BOD, phosphate and turbidity concentration, provided the second-highest prediction accuracy.

Based on the sensitivity analysis, Khan *et al.* (2021) proposed an ANN pattern obtained from the SLA algorithm. To estimate the model's performance, MATLAB software was used to calculate the determinant of the coefficient and MSE. The result showed the best performance of EC, pH, TDS, and turbidity. The errors were reduced by monitoring the bacteria in real-time. An irrigation WQI was developed by Jahin *et al.* (2020) using MA for surface water in Egypt. Based on the National Sanitation Foundation Water Quality Index (NSFWQI), parameter weights were adopted using FAO 29 guidelines. Twenty-two parameters from the summer, winter, autumn and spring seasons of 2018 were collected and analyzed from 45 geo-references. FA, Ca, and PCA were used to analyze the data. The tested water quality was good, which was shown in the results. The performances of the seven parameters were better, and the weights of water quality were estimated. Based on the 3D-EEM (Excitation–Emission Matrix) technique and hyper spectra, Zhang *et al.* (2020) determined the parameters for surface water quality in China. Forty-eight samples were derived from hyper spectra and analyzed 12 water quality parameters by WQI. This study improved the Back Propagation-Artificial Neural Network (BPANN) method and determined the WQP. The study's outcome revealed that water quality estimation by WQP and surface water quality monitoring have improved.

The existing studies determined various water quality parameters by various analyzing techniques in various places. Some existing studies (Khatoun *et al.* 2013; Maji and Chaudhary 2019; Singh *et al.* 2020; Isaac and Siddiqui 2022) conducted water quality assessments in the Kali River, Yamuna River and Ganga River in Uttar Pradesh state. Some previous studies (Banerjee *et al.* 2021; Chaurasia *et al.* 2021; Ram *et al.* 2021) estimated water quality in one district. This study used a novel hybrid DNN-MBGD for water quality prediction and established the water quality index by

WAWQI using 20 water quality parameters in Uttar Pradesh. The hybrid DNN-MBGD model predicts the best performance of the selected parameter.

The objective of the study is to check the water quality, whether it is suitable for drinking purposes and to predict the water quality parameters by hybrid DNN-MBGD model. The proposed methodology is given in section 2, the result and discussion are explained in section 3, and section 4 consists of the conclusion of this study.

2. Methodology

The data on the water quality parameters are collected from the surface water quality station-wise report (Jan 2019–Jan 2020). Solving the WQI equation using Excel 2007 software determined the WQI value, using these 20 water quality parameters. There are three steps involved in WQI determination. The first step is determining the weightage level (w_i) for each water quality parameter. The second step is to determine the relative weight (W_i) for each water quality parameter and then determine the quality rating value (Q_i). The third and last step is to find the values of WQI for quality parameters. WQI was used to determine the water quality of experimental sites. After that correlation between WQI and 20 water quality parameters was determined. Input water quality parameters are considered by important feature analysis, and output values are predicted using a novel hybrid DNN-MBGD technique. In this study, the DNN model is combined with MBGD to update the weights and develop the accuracy for better prediction. The performance of hybrid DNN-MBGD compared with ANN, SVM and GMDH models. Finally, error analyses were determined for better accuracy. Figure 1 shows the proposed research methodology for this study.

2.1 Study area

Uttar Pradesh (27°40'N; 80°00'E) is the study area of this study and was executed from January 2019 to January 2020. In this study, 10 districts of Uttar Pradesh were selected for water quality prediction and WQI estimation because this state is an industrialized and the highest populous state in India. Minerals, textiles, and electronics are the

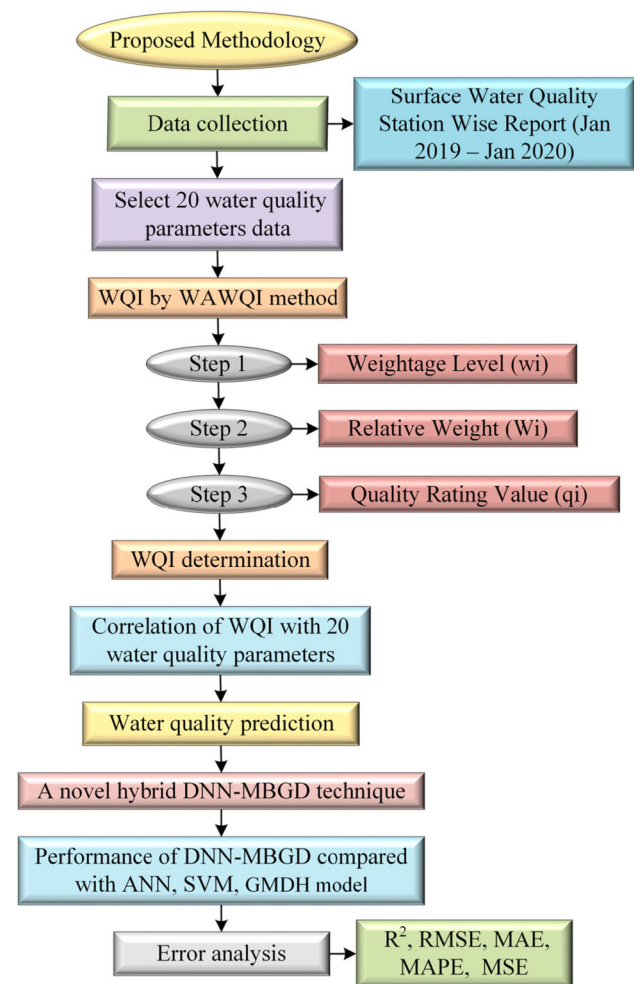


Figure 1. Flowchart for proposed research methodology.

main industries in Uttar Pradesh. In this state, surface water can easily get polluted by discharging toxic wastes from these industries and extracting chemical wastes. So author analyzed the water quality in 10 districts (Jaunpur, Deoria, Fatehpur, Prayagraj, Ambedkar Nagar, Gorakhpur, Lucknow, Amroha, Hardoi and Meerut) of Uttar Pradesh to check whether the water is suitable for drinking. Maighat site is selected for the Jaunpur district.

The following sites are selected for water quality prediction. The sites are, Deoria, Fatehpur, Prayagraj, Ambedkar Nagar, Gorakhpur, Lucknow, Amroha, Hardoi and Meerut, Turtipar, Kora-old, Allahabad, Akbarpur, Birdghat, Garhmukteshwar, Ankinghat and Galeta. Figure 2 shows the location of Uttar Pradesh in India and its experimental districts, respectively.

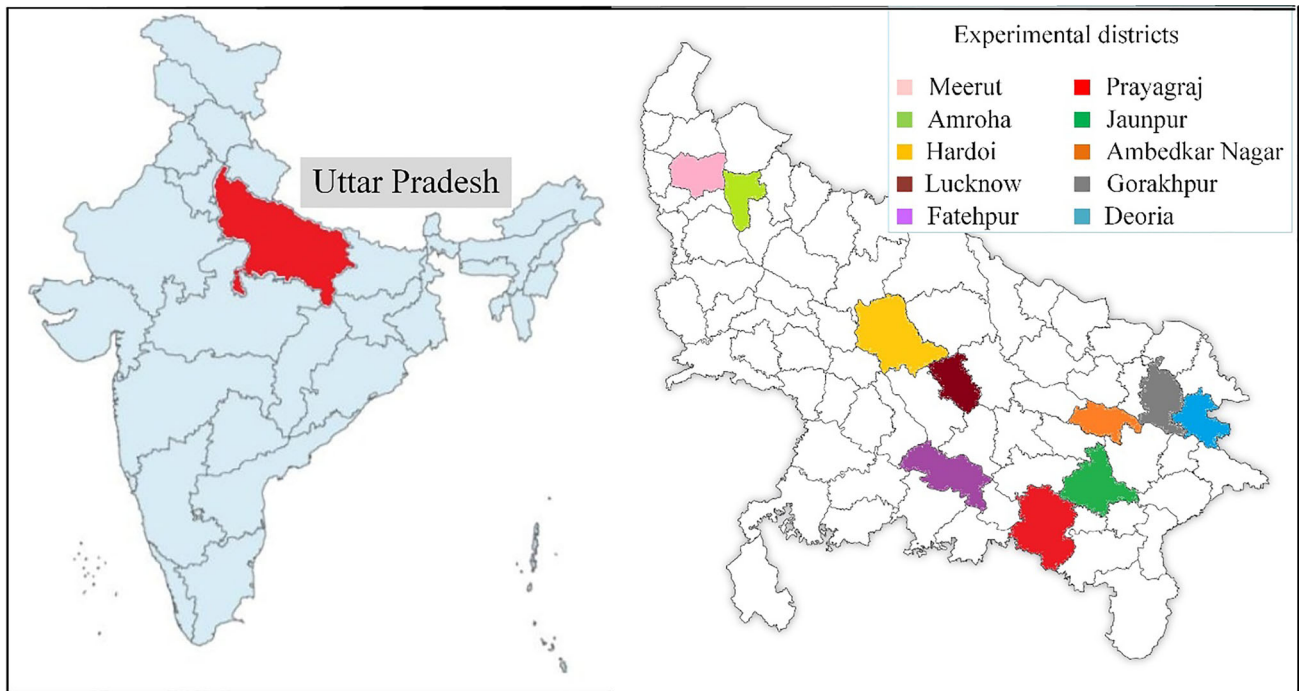


Figure 2. Uttar Pradesh’s location in India and its experimental districts.

2.2 Dataset collection

India’s most industrialized state is Uttar Pradesh. Uttar Pradesh has several industries centred on textiles, minerals, and electronics. Surface water in this state is readily contaminated by chemical waste extraction and discharge from these companies. There is a chance that water will get contaminated. So in this study, 10 districts of Uttar Pradesh were selected for water quality prediction and WQI estimation. Calculation of the WQI value used to determine if the water is fit for human consumption. Between January 2019 and January 2020, the data for 10 sites of 20 water quality parameters were gathered weekly. In total 10400 historical data were collected for 20 water quality parameters from 10 sites of Uttar Pradesh. For each site, 1040 water quality parameters were collected from January 2019 to January 2020. For water quality prediction, the entire data was considered. In this study, the average values of water quality parameters data for 10 sites from January 2019 and January 2020 are given in table 3.

2.3 Methods

The study aims to forecast water quality using a novel hybrid DNN-MBGD optimization and

WAWQI technique to assess WQI in 10 districts of Uttar Pradesh.

2.3.1 Weighted Arithmetic Water Quality Index Method (WAWQI)

WQI determines water quality, and the index number indicates the WQI. WQI is determined by various water quality parameters such as chemical, biological, and physical. Compared to other methods like NSF-WQI, CCME WQI and OWQI, WAWQI gives accurate results, so this study used the WAWQI method for WQI calculation. Many studies (Gautam *et al.* 2021; Menberu *et al.* 2021; Rojamadhuvanthi *et al.* 2021; Hagage *et al.* 2022; Lien *et al.* 2022; Nandi *et al.* 2022; Ojukwu *et al.* 2022; Prathibha *et al.* 2022; Rahul *et al.* 2022; Zait *et al.* 2022) was used the WAWQI method to calculate WQI. The ecological status of water is performed by WAWQI methods, which transform the multiple WQ parameters into a mathematical equation. By solving the constant value K and S_i , finding out the w_i value, which is equation (1). The summation of the unit weights for 20 parameters is used to determine the relative weight, shown in equation (2). The ratio of observed values of each parameter to the standard values of each

WQ parameter is known as the Quality Index, represented in equation (3). Multiplying Q_i and W_i gives the subindex value shown in equation (4). Summation of the subindex for 20 parameters used to determine the WQI, shown in equation (5).

The following steps calculate the WQI of water quality parameters. The first step is to determine the unit weight (w_i) of each water quality parameter.

$$w_i = \frac{K}{S_i} \tag{1}$$

where w_i = unit weight of each WQ parameter, K = constant value.

$$K = \frac{1}{\frac{1}{S_1} + \frac{1}{S_2} + \dots + \frac{1}{S_{20}}} = \frac{1}{\sum_{i=1}^{20} S_i}.$$

Here, S_i represents the standard values of each WQ parameter and i indicates the total number of parameters used, here $i = 20$.

The second step is to determine the relative weightage (W_i), and the third step is to determine the quality index (Q_i) of each WQ parameter.

$$W_i = \sum_{i=1}^{20} w_i \tag{2}$$

$$Q_i = \frac{C_i}{S_i} \tag{3}$$

where W_i = relative weightage of each parameter, Q_i = quality index of each parameter, C_i = observed values of each parameter.

After finding Q_i , the sub-index SI is determined. The WQI is calculated from the SI value.

$$SI_i = W_i \times Q_i, \tag{4}$$

$$WQI = \sum SI. \tag{5}$$

The WAWQI method divides the WQI value into five categories. The WQI value is greater than

100, which means the water is not suitable for drinking. The water condition is very poor if the WQI value is between 76 and 100. The water condition is poor if the WQI value is between 51 and 75. The water condition is good if the WQI value is between 26 and 50. The WQI value is less than 20, so the water is suitable for drinking. Table 1 explains the five categories of WQI values.

2.3.2 Deep learning neural network (DNN)

Deep learning neural networks (DNN) are referred to as feed-forward neural networks (FFNNs). DNN's data does not run backwards; it runs only in the forward direction. This network divides multiple data and the massive amount of data trained by DNN. There are three layers involved in DNN: input, hidden and output. These networks handled unlabeled data, non-structured data and non-linear data. The signal was passed from one neuron to another based on the input data. The output data has been neglected if the threshold value is smaller than the signal value. A function is done in each neuron, called the Function of activation, and various neurons are involved in the layers. The output layer was obtained by an input layer influenced by weight. DNN gives a better prediction by optimizing the weights with MBGD optimization. A mathematical approach was used to apply the DNN. The process of DNN was performed by transferring one neuron to another neuron. The structure of DNN is given in figure 3. In figure 3, pink colour circles indicate the input layer, hidden layers are represented by blue colour, and green colour represents the output layers. Each layer is fully connected with other layers.

To determine the precision of DNN-MBGD model, the coefficient of determination (R^2), RMSE, MAE and MSE were used. These are the important step in the machine learning model. This study examines the coefficient of determination (R^2) and RMSE values for 20 water quality parameters.

The linear regression technique is used to describe the proportion of variance in the dependent variable is known as the coefficient of determination (R^2). The values of R^2 should be less than 1. Equation (6) illustrates the determination of the coefficient of determination.

$$R^2 = 1 - \frac{\sum (x_i - x_j)^2}{\sum (x_i - x_k)^2}. \tag{6}$$

Table 1. Classification of WQI by WAWQI method.

Water quality conditions	Values of WQI
Not suitable for drinking purposes	WQI > 100
Very poor	WQI = 76–100
Poor	WQI = 51–75
Good	WQI = 26–50
Suitable for drinking purposes	WQI < 20

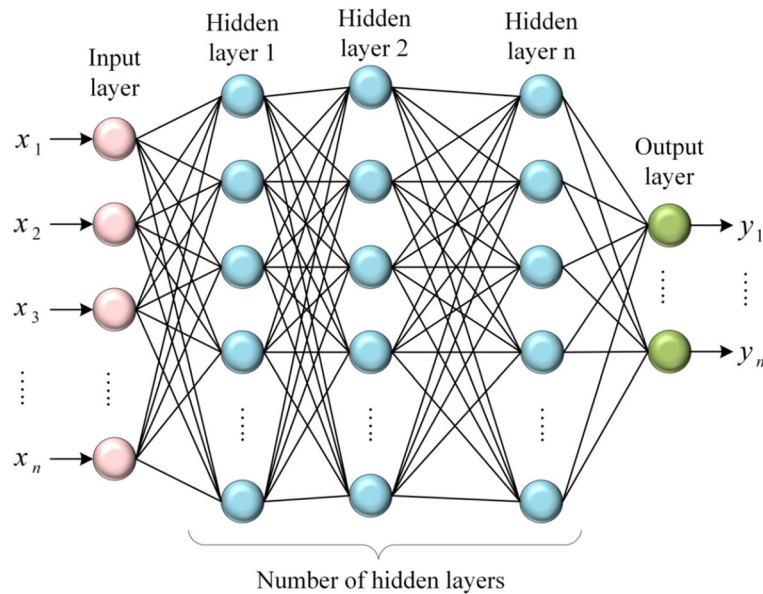


Figure 3. Structure of deep neural network.

The square root of MSE is also called RMSE. RMSE reflects the standard deviation. Equation (7) specifies the RMSE determination.

$$\sqrt{\text{MSE}} = \text{RMSE} = \sqrt{\frac{1}{K} \sum_{i=1}^K (x_i - x_j)^2} \quad (7)$$

Equation (8) is used to predict the output of DNN-MBGD. In this equation, X indicates the prediction of output data, C indicates the weights of hidden layer inputs, Y represents the input data and D indicates the bias.

$$X = CY + D. \quad (8)$$

R^2 and RMSE values are used to determine the best input parameter performance. The predicted values are compared using the ANN technique after predicting the output data.

2.4 Water quality parameters

The following 20 parameters are determined from the water samples, namely boron (B), Biochemical Oxygen Demand (BOD), Carbon trioxide (CO_3), Calcium (Ca), Chemical Oxygen Demand (COD), Chlorine (Cl), Electrical Conductivity (EC), Bicarbonate (HCO_3), Potassium (K), Magnesium (Mg), Ammonia (NH_3), Sodium (Na), Phosphorous (P), Sulphate (SO_4), Silicon dioxide (SiO_2), Total Alkalinity (TA), Temperature (T), Total Coliform (TC), Total Dissolved Solids (TDS) and pH. The

WQI value is determined using these 20 water quality parameters. The data on water quality parameters for the above sites were collected from the Surface Water Quality Station Wise Report (Jan 2019–Jan 2020).

Boron: It is a chemical parameter denoted by B, and the unit is mg/l. The permissible boron limit of the BIS for drinking water is 1 mg/l. Maximum B values occur at site 7, and minimum B values occur at site 1. The B values for each site do not exceed the standard value.

Biochemical oxygen demand: It is a chemical parameter denoted by BOD, and the unit is mg/l, and the BIS permissible BOD limit for drinking water is 2 mg/l. The highest BOD value is at site 10, the lowest BOD value is at site 3. The BOD value exceeds the standard limit with the exception of site 3.

Carbon trioxide: It is denoted by CO_3 ; the unit is mg/l, BIS permissible limit for CO_3 is 75 mg/l. The CO_3 values for site 1, site 3, and site 4 were 12.43, 12.60 and 5.28, respectively. At the other sites, the CO_3 value was zero.

Calcium: It is denoted by Ca, the permissible BIS limit for Ca is 75 mg/l, and the unit is mg/l. The highest amount of Ca occurs at site 10, and the lowest Ca presents at site 8; at all 10 sites, Ca does not exceed the standard limit.

Chemical oxygen demand: It is denoted by COD, the unit is mg/l, and CPCB permissible limit for COD is 250 mg/l. Site 3 has a minimum COD of 5.55, and site 10 has a maximum COD of 131. At

all 10 sites, the COD does not exceed the standard limit.

Chlorine: It is denoted by Cl, the unit is mg/l, BIS permissible limit for Cl is 250 mg/l. The lowest Cl is 10.84 at site 3 and the highest Cl is 88.85 at site 10. At all 10 sites, Cl does not exceed the standard limit.

Electrical conductivity: It is denoted by EC, the unit is mg/l; the limit value for EC permitted by the WHO is 400 mg/l. The minimum EC is 291 at site 3 and the maximum EC is 1209 at site 10. Sites 1, 5, 7, and 10 exceed the standard limit.

Bicarbonate: It is denoted by HCO_3 , the unit is mg/l, and the BIS permissible limit for HCO_3 is 300 mg/l. The lowest HCO_3 is 124 at site 3, and the highest HCO_3 is 391 at site 10. Site 10 exceeds the standard limit.

Potassium: It is denoted by K, the unit is mg/l, and the WHO permissible K limit is 10 mg/l. Minimum K is 3.71 at site 2, and maximum K is 49.23 at site 10. Site 10 exceeds the standard limit.

Magnesium: It is denoted by Mg, the unit is mg/l, and the BIS permissible limit of Mg is 30 mg/l.

Table 2. Drinking water standards for 20 water quality parameters.

Sl. no.	Water quality parameters	Standard values for drinking water
1	B	1 mg/l (BIS)
2	BOD	2 mg/l (BIS)
3	CO_3	75 mg/l (BIS)
4	Ca	75 mg/l (BIS)
5	COD	250 mg/l (CPCB)
6	Cl	250 mg/l (BIS)
7	EC	400 mg/l (WHO)
8	HCO_3	300 mg/l (BIS)
9	K	10 mg/l (WHO)
10	Mg	30 mg/l (BIS)
11	NH_3	0.5 mg/l (BIS)
12	Na	200 mg/l (BIS)
13	P	1 mg/l (BIS)
14	SO_4	200 mg/l (BIS)
15	SiO_2	9.2 mg/l (WHO)
16	TA	200 mg/l (BIS)
17	T	25°C (WHO)
18	TC	10000 (BIS)
19	TDS	500 mg/l (BIS)
20	pH	6.5–8.5 (BIS)

Table 3. Average values of water quality parameters from Jan 2019 to Jan 2020.

Water quality parameters	Values of WQ parameters in 10 districts (Jan 2019–Jan 2020)									
	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10
B	0.04	0.14	0.08	0.05	0.05	0.17	0.33	0.15	0.19	0.23
BOD	3.49	2.33	0.66	3.57	3.09	2.92	10.15	2.37	3.12	51.65
CO_3	12.43	0	12.60	5.28	0	0	0	0	0	0
Ca	35.85	32.76	29.14	32.36	40.32	36.26	33.65	23.77	38.33	69.33
COD	32.43	5.55	8.89	31.68	31.28	8.09	20.36	6.18	7.18	131
Cl	31.86	20.16	10.84	33.70	26	17.64	28.84	11.82	28.48	88.85
EC	405	307	291	363	418	322	501	199	361	1209
HCO_3	178	160	124	155	216	164	191	126	181	391
K	5.48	3.71	4.59	5.45	5.74	3.85	6.58	3.74	6.32	49.23
Mg	19.92	17.22	8.36	13.65	18.82	19.05	25.46	15.29	24.68	32.41
NH_3	0.13	0.26	0.05	0.09	0.09	0.23	0.48	0.26	0.24	16.45
Na	29.17	13.81	21.28	26.54	21.02	9.69	16.61	7.15	16.1	68.78
P	0.09	0.29	0.08	0.09	0.16	0.34	0.47	0.32	0.36	0.63
SO_4	25.93	23.41	13.42	26.39	18.31	21.77	30.3	15.14	34.41	65.03
SiO_2	10.18	6.55	5.55	8.47	14.29	8.12	8.97	7.7	7.15	13.72
TA	164	131	123	136	177	134	156	103	149	352
T	51	50	42	48	43	48	52	43	44	41
TC	36300	1409	3177	33308	18388	2218	50700	2182	3036	8119000
TDS	212	179	85	196	219	192	300	119	211	753
pH	8.71	7.74	8.71	8.05	7.77	7.9	8.55	7.66	7.12	7.57

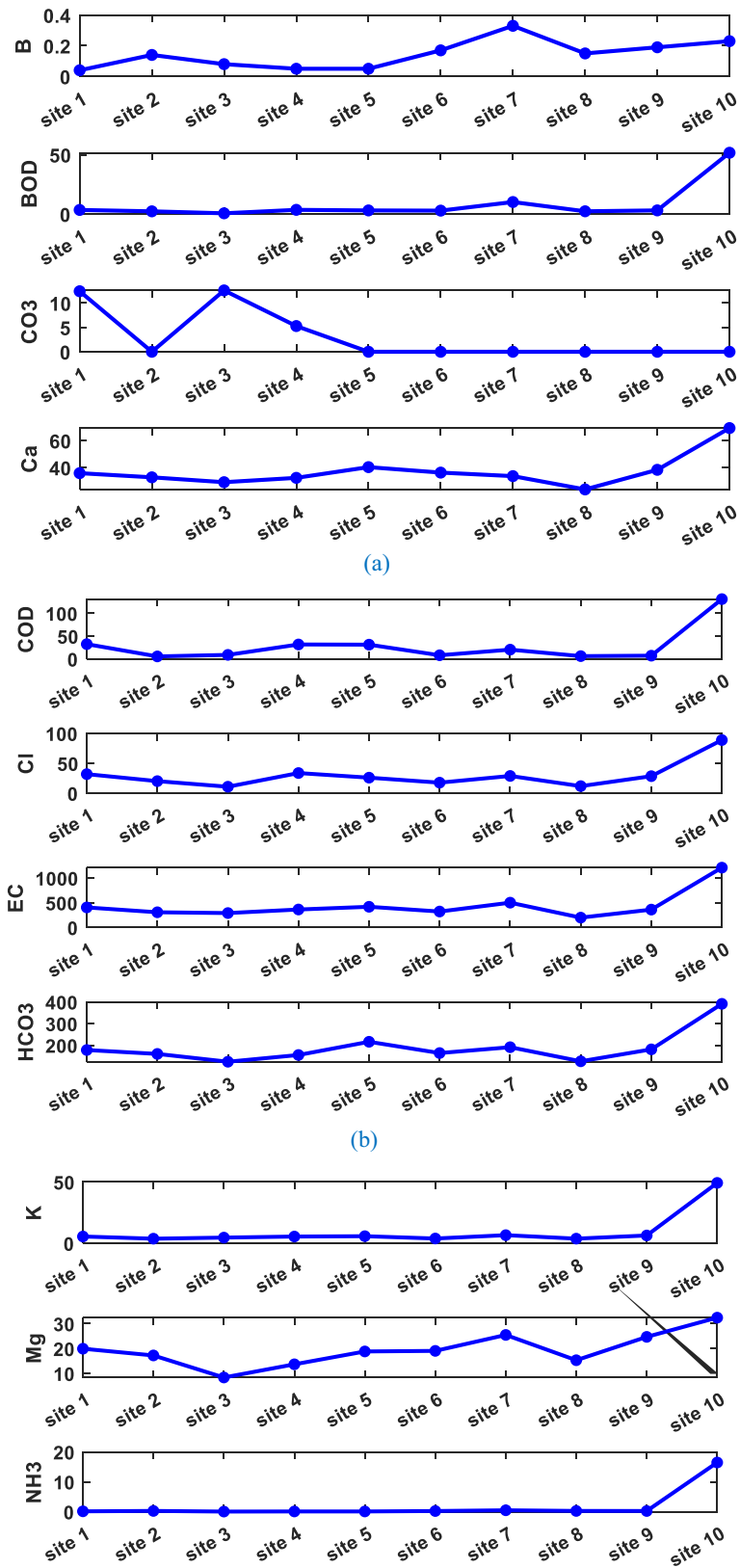


Figure 4. Variation of water quality parameters in 10 sites.

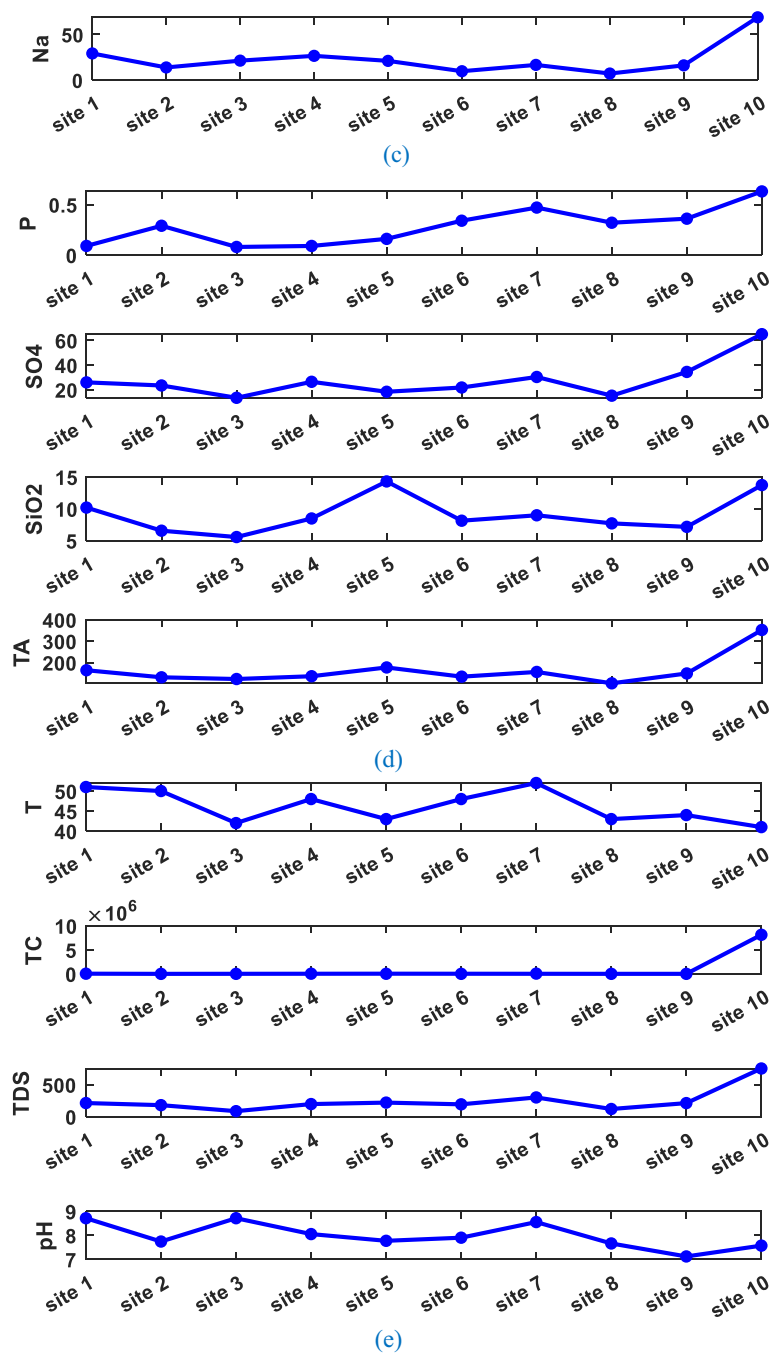


Figure 4. (Continued.)

Minimum Mg is 8.36 at site 3, and maximum Mg is 32.41 at site 10. Site 10 exceeds the standard limit.

Ammonia: It is denoted by NH_3 ; the unit is mg/l, and the BIS permissible limit of NH_3 is 0.5 mg/l. Minimum NH_3 is 0.05 at site 3, and maximum NH_3 is 16.45 at site 10. Site 10 exceeds the standard limit.

Sodium: It is denoted by Na, the unit is mg/l, and the BIS permissible limit of Na is 200 mg/l.

Minimum Na is 7.15 at site 8, and maximum Na is 68.78 at site 10. In all 10 sites, Na does not exceed the standard limit.

Phosphorous: It is denoted by P, the unit is mg/l, and the BIS permissible limit of Phosphorous for drinking water is 1 mg/l. Minimum P is 0.08 at site 3, and maximum P is 0.63 at site 10. In all 10 sites, P does not exceed the standard limit.

Table 4. Examined unit weight and relative weightage of each parameter.

Sl. no.	Water quality parameters	Unit weight of each parameter (w_i)
1	B	0.201723687
2	BOD	0.100861843
3	CO ₃	0.002689649
4	Ca	0.002689649
5	COD	0.000806895
6	Cl	0.000806895
7	EC	0.000504309
8	HCO ₃	0.000672412
9	K	0.020172369
10	Mg	0.006724123
11	NH ₃	0.403447374
12	Na	0.001008618
13	P	0.201723687
14	SO ₄	0.001008618
15	SiO ₂	0.021926488
16	TA	0.001008618
17	T	0.008068947
18	TC	2.01724E-05
19	TDS	0.000403447
20	pH	0.023732198

$$W_i = \sum_{i=1}^{20} w_i = 1$$

Sulphate: It is denoted by SO₄, the unit is mg/l, and the BIS permissible limit of SO₄ is 200 mg/l. Minimum SO₄ is 13.42 at site 3, and maximum SO₄ is 65.03 at site 10. In all 10 sites, SO₄ does not exceed the standard limit.

Silicon dioxide: It is denoted by SiO₂, the unit is mg/l, and the WHO permissible SiO₂ limit is 9.2 mg/l. Minimum SiO₂ is 5.55 at site 3, and maximum SiO₂ is 14.29 at site 5. Site 1, site 4, site 5 and site 10 exceed the standard limit.

Total alkalinity: It is denoted by TA, the unit is mg/l, and the BIS permissible limit of TA is 200 mg/l. Minimum TA is 103 at site 8, and maximum TA is 352 at site 10. Site 10 exceeds the standard limit.

Temperature: It is denoted by T, the unit is °C, and the WHO permissible temperature limit is 25°C. Minimum T is 41°C at site 10 and maximum T is 52°C at site 7. All sites exceed the standard limit.

Total coliform: It is denoted by TC, and the BIS permissible limit of TC is 10000. Minimum TC is 1409 at site 2 and maximum TC is

Table 5. Examined quality index (Q_i) for 10 sites.

Sl. no.	Water quality parameters	Quality index (Q_i)									
		Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10
1	B	4	14	8	5	5	17	33	15	19	23
2	BOD	175	117	33	179	155	146	508	119	156	2583
3	CO ₃	17	0	17	7	0	0	0	0	0	0
4	Ca	48	44	39	43	54	48	45	32	51	92
5	COD	13	2.22	3.56	13	13	3.24	8.14	2.47	2.87	53
6	Cl	13	8.06	4.34	13	10	7.06	12	4.73	11	36
7	EC	101	77	73	91	105	81	125	50	90	302
8	HCO ₃	59	53	41	52	72	55	64	42	60	130
9	K	55	37	46	55	57	39	66	37	63	492
10	Mg	66	57	28	46	63	64	85	51	82	108
11	NH ₃	26	52	10	18	18	46	96	52	48	3290
12	Na	15	6.91	11	13	11	4.85	8.30	3.58	8.05	34
13	P	9	29	8	9	16	34	47	32	36	63
14	SO ₄	13	12	6.71	13	9.15	11	15	7.57	17	32.51
15	SiO ₂	111	71	60	92	155	88	98	84	78	149
16	TA	82	65.5	61.5	68	88.5	67	78	51.5	74.5	176
17	T	204	200	168	192	172	192	208	172	176	164
18	TC	363	14.0	31.8	333	184	22.2	507	21.8	30.3	81190
19	TDS	42.4	35.8	17	39.2	43.8	38.4	60	23.8	42.2	150.6
20	pH	121	24	121	55	27	40	105	16	-38	7

Table 6. WQI for 10 sites in Uttar Pradesh.

Sl. no.	Experimental sites	WQI
1	Site 1	1536 > 100 (Unsuitable for drinking purposes)
2	Site 2	919 > 100 (Unsuitable for drinking purposes)
3	Site 3	787 > 100 (Unsuitable for drinking purposes)
4	Site 4	1335 > 100 (Unsuitable for drinking purposes)
5	Site 5	1257 > 100 (Unsuitable for drinking purposes)
6	Site 6	1002 > 100 (Unsuitable for drinking purposes)
7	Site 7	2167 > 100 (Unsuitable for drinking purposes)
8	Site 8	816 > 100 (Unsuitable for drinking purposes)
9	Site 9	1008 > 100 (Unsuitable for drinking purposes)
10	Site 10	89075 >>> 100 (Unsuitable for drinking purposes)

8119000 at site 10. Except for sites 2, 3, 6, 8 and 9 remaining sites exceed the standard limit.

Total dissolved solids: It is denoted by TDS, the unit is mg/l and the BIS permissible limit of TDS is 500 mg/l. Minimum TDS is 85 at site 3, and maximum TDS is 753 at site 10. Site 10 exceeds the standard limit of TDS.

pH: Basic/acidic water calculated by pH has no unit, and the BIS permissible pH limit is 6.5–8.5. Minimum pH is 7.12 at site 9, and maximum pH is 8.71 at site 1. Sites 1, 3 and 7 exceed the standard limit.

Site 1 represented Maighat site in the Jaunpur district. Site 2 represented Turtipar site in the Deoria district. Site 3 represented Kora-old site in the Fatehpur district. Site 4 represented Allahabad site in the Prayagraj district. Site 5 represented Akbarpur site in the Ambedkar Nagar district. Site 6 represented Maighat site in the Birdghat district. Site 7 represented Lucknow site in the Lucknow district. Site 8 represented Garhmukteshwar site in the Amroha district. Site 9 represented Ankinghat site in the Hardoi district, and Site 10 represented Galeta site in the Meerut district. The standard value of 20 parameters is given in table 2. The values of water quality parameters for the corresponding sites are shown in table 3. Figure 4 shows the variation of water quality parameters in 10 sites.

3. Results and discussion

The result of the proposed methodology describes the prediction of water quality parameters using a hybrid DNN-MBGD technique. It evaluates the WQI values

for 10 sites of UP from January 2019 to January 2020. The WQI values are determined by determining unit weight, relative weight and quality index.

3.1 WQI calculation

WQI determines the water quality and indicates the WQI in terms of an index number. This study uses the WAWQI method to calculate the WQI values. Compared to other methods such as the National Sanitation Foundation-WQI (NSF-WQI), the Canadian Council of Ministers of the Environment Index (CCME WQI) and the Oregon-WQI (OWQI), WAWQI gives accurate results. So WAWQI method was selected for WQI calculation. The first step of WQI calculation is to determine the unit weight of each parameter. After determining the unit weight, the second step determines the relative weightage and the third step determines the quality index of each WQ parameter. Excel 2007 software was used for the WQI calculations.

3.1.1 Calculation of w_i and W_i

w_i and W_i of each parameter is determined by equations (1 and 2). By solving the constant value, K and S_i find out the w_i value. The summation of the unit weights for 20 parameters is used to determine the relative weights shown in equation (2). Table 4 shows the w_i and W_i values for each parameter.

3.1.2 Determination of quality index (Q_i)

The ratio of the observed values of each parameter to the standard values of each WQ parameter is known as the quality index, shown in

Table 7. Correlation matrix of 20 water quality parameters.

WQI																				1																		
pH																			1	-0.26																		
TDS																		1	-0.25	0.95																		
TC																		1	0.96	-0.27	0.99																	
T																		1	-0.45	-0.25	0.44	-0.44																
TA																		1	-0.4	0.95	0.97	-0.2	0.96															
SiO ₂																		1	0.73	-0.2	0.57	0.66	-0.2	0.57														
SO ₄																		1	0.51	0.92	-0.2	0.9	0.96	-0.3	0.9													
P																		1	0.74	0.27	0.62	-0.1	0.67	0.74	-0.5	0.67												
Na																		1	0.4	0.86	0.62	0.94	-0.3	0.92	0.9	-0.1	0.92											
NH ₃																		1	0.91	0.69	0.90	0.56	0.95	-0.44	0.99	0.96	-0.27	0.99										
Mg																		1	0.68	0.58	0.83	0.86	0.57	0.77	0	0.67	0.83	-0.4	0.68									
K																		1	0.70	0.99	0.93	0.67	0.91	0.59	0.97	-0.44	0.99	0.96	-0.26	0.99								
HCO ₃																		1	0.95	0.82	0.93	0.89	0.68	0.92	0.76	0.99	-0.32	0.93	0.98	-0.3	0.93							
EC																		1	0.98	0.97	0.78	0.96	0.94	0.67	0.94	0.67	0.99	-0.3	0.96	0.99	-0.2	0.96						
Cl																		1	0.97	0.96	0.95	0.77	0.93	0.95	0.6	0.96	0.66	0.97	-0.2	0.93	0.97	-0.2	0.94					
COD																		1	0.97	0.97	0.95	0.96	0.66	0.95	0.97	0.53	0.88	0.72	0.97	-0.4	0.96	0.95	-0.1	0.96				
Ca																		1	0.93	0.95	0.96	0.98	0.94	0.78	0.92	0.9	0.62	0.92	0.7	0.98	-0.4	0.92	0.95	-0.3	0.93			
CO ₃																		1	-0.25	-0.1	-0.18	-0.2	-0.31	-0.2	-0.52	-0.22	0.088	-0.69	-0.29	-0.25	-0.19	0.085	-0.2	-0.31	0.747	-0.2		
BOD																		1	-0.2	0.92	0.96	0.95	0.98	0.95	0.99	0.75	0.99	0.91	0.73	0.93	0.59	0.96	-0.4	0.99	0.98	-0.2	0.99	
B																		1	0.44	-0.6	0.28	0.19	0.29	0.39	0.35	0.35	0.66	0.35	0.09	0.87	0.48	0	0.29	0.13	0.33	0.46	-0.2	0.34
	B	BOD	CO ₃	Ca	COD	Cl	EC	HCO ₃	K	Mg	NH ₃	Na	P	SO ₄	SiO ₃	TA	T	TC	TDS	pH	WQI																	

equation (3). Table 5 indicates the Q_i values for each site.

3.1.3 WQI determination

Multiplying Q_i and W_i , we get a sub-index value shown in equation (4). Summation of the sub-index for 20 parameters used to determine the WQI shown in equation (5). If the WQI value exceeds 100, the water is unsuitable for drinking. The water condition is very poor if the WQI value is between 76 and 100. The water condition is poor if the WQI value is between 51 and 75. The water condition is good if the WQI value is between 26 and 50. WQI

value is less than 20, the water is suitable for drinking.

The WQI values of 10 districts of Uttar Pradesh are shown in table 6. Table 6 indicates that the water from these sites is unsuitable for drinking, domestic and irrigation purposes. The WQI value for site 10 is greater than 100. The water has to be treated at 10 sites in Uttar Pradesh. A water purifying system must also be installed. As the water purification system reduced excess physical, chemical and biological parameters, the water quality and harmful containment were also minimized. The water at these 10 sites is unsuitable for drinking purposes. But comparatively, site 3 has a

minimum WQI value. So, the water quality parameters are predicted from site 3 using hybrid DNN-MBGD optimization.

3.2 Correlation matrix of 20 parameters and WQI

Determining the relationship between two water quality parameters is known as correlation. The relationship between the two directions of water quality parameters reflects the correlation. Correlation values can be zero values, negative values and positive values. The direction of two variables transformed in the same direction (1) is known as a positive correlation. The direction of two variables transforming in the opposite direction (-1) is known as a negative correlation, and if there is no relationship between two variables, it is called zero correlation (0). Some existing studies also determine the correlation between water quality parameters (Patil *et al.* 2020; Geng *et al.* 2021; Sayess *et al.* 2021; Srivastava *et al.* 2022). In this study, the 20 water quality parameters are correlated with WQI to determine the relationship between two water quality parameters. A correlation matrix of 20 parameters is shown in table 7. In table 7, the correlation of the same variable shows the value 1, indicating a positive correlation and two zero correlation values are shown in the yellow box. Positive values indicate a positive correlation, meaning that two variables are transforming in the same direction. Negative values indicate a negative correlation, meaning that two variables are transforming in opposite directions, and there is a 2-zero correlation which means there is no relationship between the two variables.

3.3 Proposed hybrid deep learning neural network–mini-batch gradient descent optimization (DNN-MBGD)

MBGD, also known as GD variation, is used to determine the coefficients of an updated model and calculate the model's errors. Training data is divided into small batches by MBGD. Gradient variations are minimized by MBGD and used for implementation by deep learning. In this study, the MBGD optimization is chosen because it gives the highest update frequency than the GD model. Compared to SGD, it provides more efficiency. In this study, the DNN model is hybridized with MBGD to update the weights and to improve the accuracy for better prediction. Each layer consists of multiple neurons and is transmitted by an activation function. In the second

hidden layer, the weights of the water quality parameters are updated for better prediction.

In this study, 80% of the dataset was trained, while the remaining 20% was tested. The hybrid DNN-MBGD was developed using these two sets of datasets. Hence the results showed that the hybrid DNN-MBGD optimization models could predict the various elements of water quality with adequate accuracy. Following dataset division, one hidden layer with several neurons corresponding to the input features is considered in the first stage. The effectiveness of the various transfer functions is now evaluated, and the best one is selected, namely K. The network size is changed to increase the accuracy of the created models in the subsequent phase. This would increase the number of neurons or hidden layers. The weights are optimized by MBGD optimization in the second hidden layer.

After selecting the best performance of the water quality parameter (K), K is predicted based on other parameters. The input feature is more relevant or crucial in predicting the feature importance that the analysis will report the water quality. Based on the feature importance score, the less important feature is eliminated using a backward elimination process. In a range from 0 to 1, the feature importance can be calculated; 0 means the feature is unimportant and 1 means the feature is essential. Negative values for feature significance indicate a feature that is affecting the performance of the model.

The random forest classifier is one of the most commonly used methods for determining a feature's relevance. This study uses Random Forest Classifier to determine the feature importance of water quality parameters.

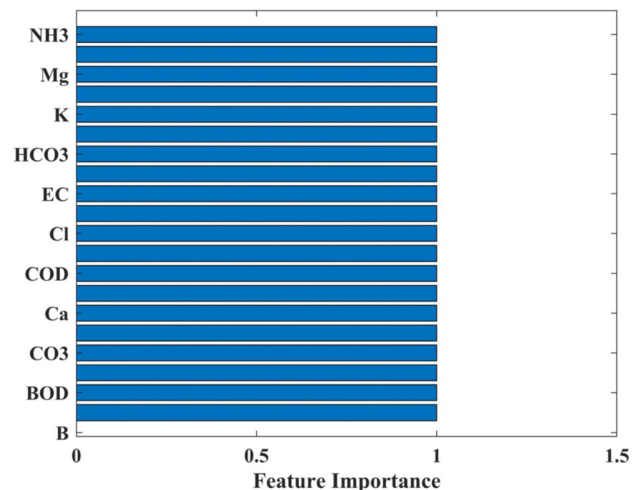


Figure 5. Feature importance score of input parameters.

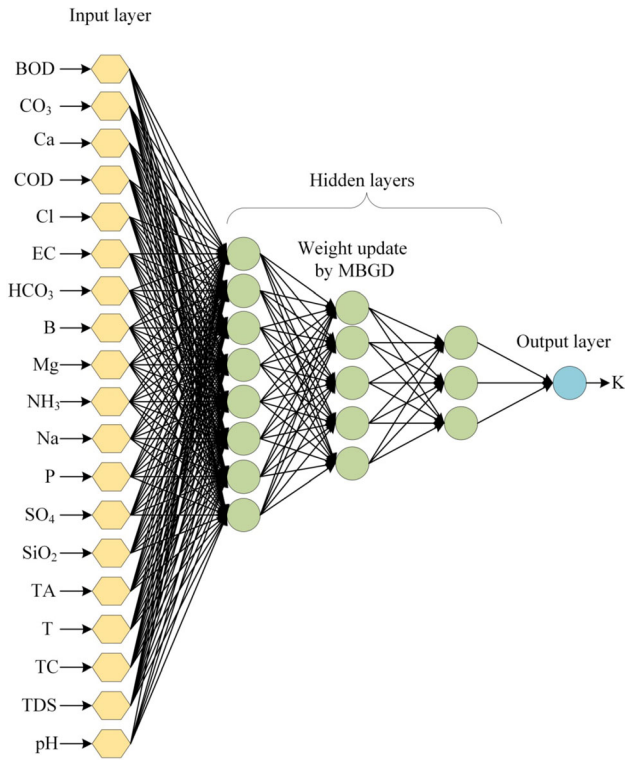


Figure 6. DNN model structure for predicting a component of water quality (K).

Figure 5 shows the feature importance score of input parameters. Based on figure 5, the feature importance is 1 for all input parameters, this means that the input parameters are extremely important for predicting water quality. So, other water quality components, including B, BOD, CO₃, Ca, COD, Cl, EC, HCO₃, Mg, NH₃, Na, P, SO₄, SiO₂, TA, T, TC, TDS and pH were considered as inputs to the prediction of K, as shown in figure 6.

3.4 Prediction of K

Based on the feature importance score, the water quality component K is predicted by the proposed hybrid DNN-MBGD optimization model. Eighty per cent of the dataset was used for training in this study, while 20% was used for testing. These two sets of datasets were employed during the developing of the hybrid DNN-MBGD model. Using LSM, the prediction of potassium (K) is made by equation (8).

$$K = \text{weight} \times B + \text{bias} \quad (9)$$

$$K = \text{weight} \times \text{BOD} + \text{bias} \quad (10)$$

$$K = \text{weight} \times \text{CO}_3 + \text{bias} \quad (11)$$

$$K = \text{weight} \times \text{Ca} + \text{bias} \quad (12)$$

$$K = \text{weight} \times \text{COD} + \text{bias} \quad (13)$$

$$K = \text{weight} \times \text{Cl} + \text{bias} \quad (14)$$

$$K = \text{weight} \times \text{EC} + \text{bias} \quad (15)$$

$$K = \text{weight} \times \text{HCO}_3 + \text{bias} \quad (16)$$

$$K = \text{weight} \times \text{Mg} + \text{bias} \quad (17)$$

$$K = \text{weight} \times \text{NH}_3 + \text{bias} \quad (18)$$

$$K = \text{weight} \times \text{Na} + \text{bias} \quad (19)$$

$$K = \text{weight} \times \text{P} + \text{bias} \quad (20)$$

$$K = \text{weight} \times \text{SO}_4 + \text{bias} \quad (21)$$

$$K = \text{weight} \times \text{SiO}_2 + \text{bias} \quad (22)$$

$$K = \text{weight} \times \text{TA} + \text{bias} \quad (23)$$

$$K = \text{weight} \times \text{T} + \text{bias} \quad (24)$$

$$K = \text{weight} \times \text{TC} + \text{bias} \quad (25)$$

$$K = \text{weight} \times \text{TDS} + \text{bias} \quad (26)$$

$$K = \text{weight} \times \text{pH} + \text{bias} \quad (27)$$

Figure 7 shows the prediction result of the water quality parameter K and the prediction performance of the hybrid DNN-MBGD model compared to machine learning models (Haghiabi *et al.* 2018). As per figure 7, the blue line with the star indicates the observed K values and the green line star demonstrates the predicted performance of the hybrid DNN-MBGD model. Similarly, the red, orange, and violet lines with a star indicate the predicted performances of ANN, SVM and GMDH models. The outcome demonstrated that the observed and predicted K values of the hybrid DNN-MBGD correspond closely. It shows that the hybrid DNN-MBGD-based prediction model is more accurate than other models such as ANN, SVM and GMDH. Because the weight optimization by MBGD in DNN leads to an improvement in model accuracy. Accurate water quality prediction is crucial to improve the pollution control and water management.

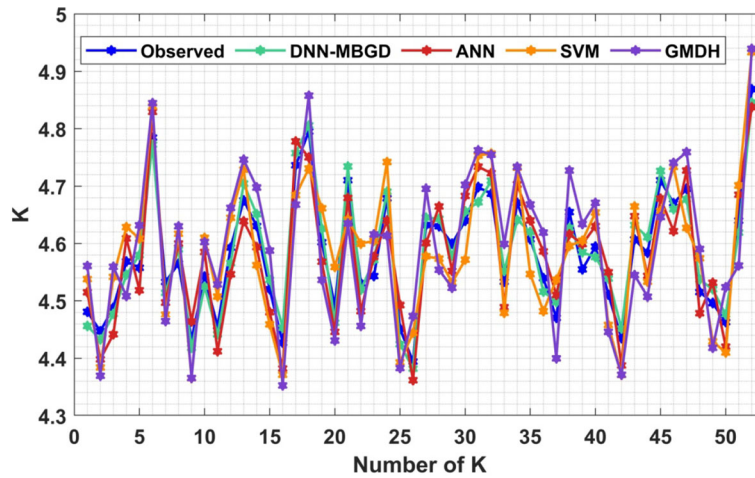


Figure 7. Comparison of prediction performance result with ANN, SVM, and GMDH model.

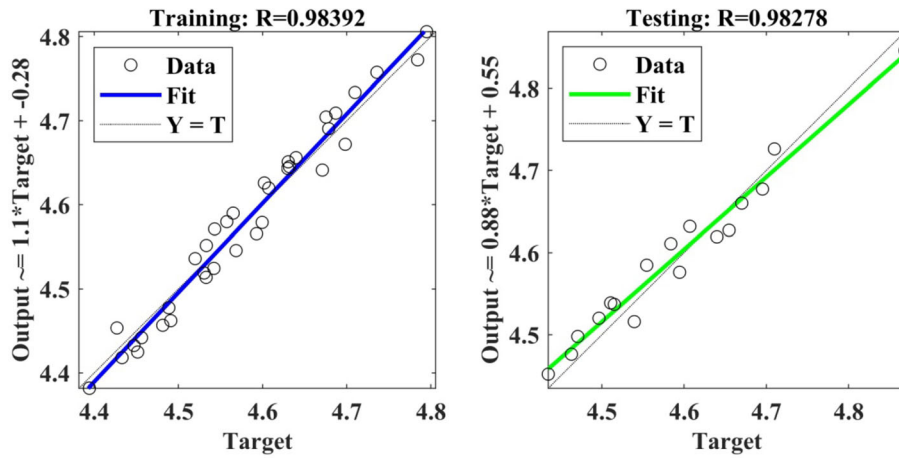


Figure 8. Regression plots by predicting K.

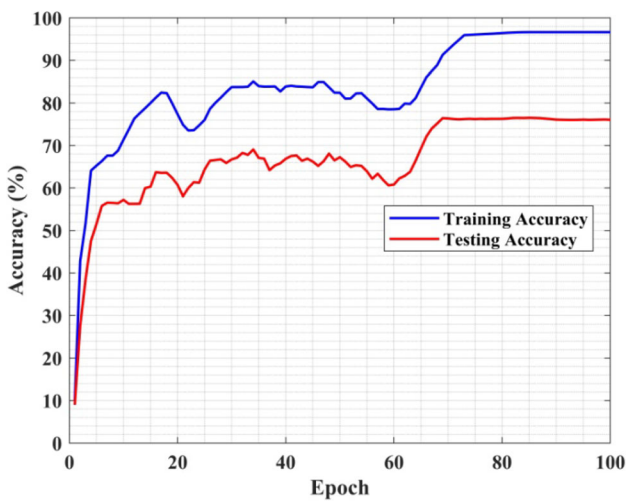


Figure 9. Training and testing accuracy.

Figure 8 shows the regression plots of a proposed model observed using the prediction

performance of K. These regression plots are used to examine and understand complex data relationships. Figure 8 shows that the model's output values are closer to the real target values. Figure 9 shows a proposed model's training and testing accuracy, which is used to establish the model's accuracy. Superior precisions are provided through training and testing accuracy. The data is divided into train and test sets to avoid fitting and overfitting issues and improve the model performance. In this study, the proposed model gives better accuracy. Table 8 shows the model's performance based on R^2 , RMSE, MAE, MAPE, and MSE. From this table 8, the least error obtained in the proposed model shows that the proposed model performance is better than the ANN, SVM and GMDH models. Because of the hybridization of MBGD, optimization with DNN gives more accuracy than non-hybrid models.

Table 8. Performance of the model.

Models	Performance metrics				
	R^2	RMSE	MAE	MAPE	MSE
Proposed DNN-MBGD	0.983	0.0239	0.0204	0.0045	0.0005
ANN	0.938	0.0472	0.0394	0.0086	0.0016
SVM	0.880	0.0189	0.0594	0.0129	0.0036
GMDH	0.865	0.1157	0.0693	0.0151	0.0049

4. Conclusion

This study was used to determine the WQI and predict the water quality parameters using a hybrid approach of DNN-MBGD optimization in 10 districts of Uttar Pradesh. The estimation of the WQI value is used to check whether the water quality is suitable for drinking. After WQI calculation, the relationship between two water quality parameters is determined. A novel hybrid DNN technique predicts water quality parameters and the performance is compared with the ANN, SVM and GMDH models.

The WAWQI is used to calculate the WQI in 10 districts of Uttar Pradesh. The result indicated, the water from 10 sites is unsuitable for drinking and for other purposes such as domestic and irrigation. Because WQI values for 10 sites in Uttar Pradesh are greater than 100, the water from 10 experimental sites in Uttar Pradesh must be treated and is also important for installing a water purifying system. As excess physical, chemical and biological parameters have been reduced by the water purification system which increased water quality and harmful containment is minimized. The correlation matrix of 20 parameters with WQI determines the relationship between each component. In this study, the performance of DNN-MBGD was determined and compared with ANN, SVM and GMDH models. Twenty water quality parameters predicted by DNN-MBGD optimization at ten sites in Uttar Pradesh yield better results. Then the best performance of the parameter K is selected and predicted using a mathematical equation with the other concerned parameters. In addition, comparing the DNN-MBGD to the ANN, SVM and GMDH model and comparing a hybrid DNN-MBGD optimization gives better performance than other models.

Author statement

Subha Sinha: Concept design, data collection, data analysis and interpretation of results, draft manuscript preparation and proofreading.

References

- Asadollah S B H S, Sharafati A, Motta D and Yaseen Z M 2021 River water quality index prediction and uncertainty analysis: A comparative study of machine learning models; *J. Environ. Chem. Eng.* **9**(1) 104599.
- Banerjee K, Kumar S, Tilak L N and Vashistha S 2021 Analysis of groundwater quality using GIS-based water quality index in Noida, Gautam Buddha Nagar, Uttar Pradesh (UP), India; In: *Applications of Artificial Intelligence and Machine Learning*, Springer, Singapore, pp. 171–187.
- Chatterjee R and Lataye D H 2020 Analysis of water quality parameters and their variation for surface water using GIS-based tools; In: *Applications of Geomatics in Civil Engineering*, Springer, Singapore, pp. 289–302.
- Chaurasia A K, Pandey H K, Tiwari S K, Pandey P and Ram A 2021 Groundwater vulnerability assessment using water quality index (WQI) under geographic information system (GIS) framework in parts of Uttar Pradesh, India; *Sustain. Water Resour. Manag.* **7**(3) 1–15.
- Chen K, Chen H, Zhou C, Huang Y, Qi X, Shen R, Liu F *et al.* 2020 Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data; *Water Res.* **171** 115454.
- Dey J and Vijay R 2021 critical and intensive review on assessment of water quality parameters through geospatial techniques; *Environ. Sci. Pollut. Res.* **28**(31) 41,612–41,626.
- Gautam A, Rai S C and Rai S P 2021 Assessment of groundwater quality beneath agriculturally advanced region of Northern Alluvial Plain, India; *Sustainability* **13**(13) 7053.
- Geng M, Wang K, Yang N, Li F, Zou Y, Chen X, Deng Z and Xie Y 2021 Evaluation and variation trends analysis of water quality in response to water regime changes in a typical river-connected lake (Dongting Lake), China; *Environ. Pollut.* **268** 115761.
- Hagage M, Madani A A and Elbeih S F 2022 Quaternary groundwater aquifer suitability for drinking in Akhmim, Upper Egypt: An assessment using water quality index and GIS techniques; *Arab. J. Geosci.* **15**(2) 1–14.
- Haghiabi A H, Nasrolahi A H and Parsaie A 2018 Water quality prediction using machine learning methods; *Water Quality Res. J.* **53**(1) 3–13.
- Horvat M, Horvat Z and Pastor K 2021 Multivariate analysis of water quality parameters in Lake Palic, Serbia; *Environ. Monit. Assess.* **193**(7) 1–18.
- Isaac R and Siddiqui S 2022 Application of water quality index and multivariate statistical techniques for assessment of

- water quality around Yamuna River in Agra Region, Uttar Pradesh, India; *Water Supply* **22**(3) 3399–3418.
- Jahin H S, Abuzaid A S and Abdellatif A D 2020 Using multivariate analysis to develop irrigation water quality index for surface water in Kafr El-Sheikh Governorate, Egypt; *Environ. Technol. Innov.* **17** 100532.
- Khan F M, Gupta R and Sekhri S 2021 Superposition learning-based model for prediction of *E. coli* in groundwater using physico-chemical water quality parameters; *Groundw. Sustain. Dev.* **13** 100580.
- Khatoun N, Khan A H, Rehman M and Pathak V 2013 Correlation study for the assessment of water quality and its parameters of Ganga River, Kanpur, Uttar Pradesh, India; *IOSR J. Appl. Chem.* **5**(3) 80–90.
- Khatri P, Gupta K K and Gupta R K 2020 Assessment of water quality parameters in real-time environment; *SN Comput. Sci.* **1**(6) 1–9.
- Kothari V, Vij S, Sharma S K and Gupta N 2021 Correlation of various water quality parameters and water quality index of districts of Uttarakhnad; *Environ. Sustain. Indic.* **9** 100093.
- Lien B T, Thi Thanh Ngan N, Kumar P, Dang T T T, Hong T T K, Ty T V, Avtar R and Minh H V T 2022 Assessing the impacts of dike systems on water quality in natural reserves of the Vietnamese Mekong Delta; *Urban Sci.* **6**(1) 21.
- Maji K J and Chaudhary R 2019 Principal component analysis for water quality assessment of the Ganga River in Uttar Pradesh, India; *Water Resourc.* **46**(5) 789–806.
- Menberu Z, Mogesse B and Reddythota D 2021 Evaluation of water quality and eutrophication status of Hawassa Lake based on different water quality indices; *Appl. Water Sci.* **11**(3) 1–10.
- Najafzadeh M and Niazmardi S 2021 A novel multiple-kernel support vector regression algorithm for estimation of water quality parameters; *Nat. Resourc. Res.* **30**(5) 3761–3775.
- Nandi I, Pokharia C, Chaturvedi V and Shah K 2022 Developing a robust tool for quality health assessment of rivers using optimized weighted arithmetic water quality index method: A study on River Ganges at Varanasi in India: Health assessment of River Ganga using WAWQI method; *Int. J. Res. Appl. Sci. Biotechnol.* **9**(3) 7–20.
- Ojukwu C K, Chukwu-Okeah G O, Ezekwe C I and Prince C Mmom 2022 Assessment of major water sources in Onicha Local Government Area of Ebonyi State Nigeria using Water Quality Index.
- Pandey V, Ray M and Kumar V 2020 Assessment of water-quality parameters of groundwater contaminated by fly ash leachate near Koradi Thermal Power Plant, Nagpur; *Environ. Sci. Pollut. Res.* **27**(22) 27,422–27,434.
- Patil B, Bhimanagouda V, Pinto S M, Govindaraju T, Hebbalu V S, Bhat V and Kannanur L N 2020 Multivariate statistics and water quality index (WQI) approach for geochemical assessment of groundwater quality – a case study of Kanavi Halla Sub-Basin, Belagavi, India; *Environ. Geochem. Health* **42**(9) 2667–2684.
- Prathibha K S, Kumar R K N, Joseph R S and Subramani S 2022 Predicting the parameters of water quality and calculating the Water Quality Index of Ulsoor Lake, Bangalore, India using Deep Learning Techniques; In: *2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), IEEE*, pp. 1–11.
- Rahul T S, Karunya J B and Wessley G J J 2022 A novel remote sensing based approach to estimate the water quality index using Sentinel-2 multispectral data; In: *2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN), IEEE*, pp. 1–4.
- Ram A, Tiwari S K, Pandey H K, Chaurasia A K, Singh S and Singh Y V 2021 Groundwater quality assessment using water quality index (WQI) under GIS framework; *Appl. Water Sci.* **11**(2) 1–20.
- Ray J G, Santhakumaran P S and Kookal S K 2021 Phytoplankton communities of eutrophic freshwater bodies (Kerala, India) in relation to the physicochemical water quality parameters; *Environ. Dev. Sustain.* **23**(1) 259–290.
- Rojamadhuvanthi C, Harsita M and Harathi P B 2021 Assessment of physico-chemical parameters and water quality index of Bhavanisagar reservoir, Tamil Nadu; *Ind. J. Ecol.* **48**(4) 1120–1123.
- Sayess R, Eyring A M and Reckhow D A 2021 Source and drinking water organic and total iodine and correlation with water quality parameters; *Water Res.* **190** 116686.
- Shalumon C S, Sanu K S, Thomas J R, Aravind U K, Radhakrishnan S, Sahoo S K, Jha S K and Aravindakumar C T 2021 Analysis of uranium and other water quality parameters in drinking water sources of 5 districts of Kerala in southern India and potability estimation using water quality indexing method; *Hydro Res.* **4** 38–46.
- Shishaye H A and Asfaw A T 2020 Analysis and evaluation of the spatial and temporal variabilities of river water quality parameters; *Appl. Water Sci.* **10**(6) 1–20.
- Singh G, Patel N, Jindal T, Srivastava P and Bhowmik A 2020 Assessment of spatial and temporal variations in water quality by the application of multivariate statistical methods in the Kali River, Uttar Pradesh, India; *Environ. Monit. Assess.* **192**(6) 1–26.
- Srivastava M, Srivastava P K, Kumar D and Kumar A 2022 A systematic study of uranium in groundwater and its correlation with other water quality parameters; *Water Supply* **22**(3) 2478–2492.
- Ustaoglu F, Tepe Y and Taş B 2020 Assessment of stream quality and health risk in a subtropical Turkey river system: A combined approach using statistical analysis and water quality index; *Ecol. Indic.* **113** 105815.
- Valentini M H K, Santos G B, Duarte V H, Franz H S, Guedes H A S, Romani R F and Vieira B M 2021 Analysis of the influence of water quality parameters in the final WQI result through statistical correlation methods: Mirim Lagoon, RS, Brazil, case study; *Water Air Soil Pollut.* **232**(9) 1–10.
- Xu Z, Shen J, Qu Y, Chen H, Zhou X, Hong H, Sun H, Lin H, Deng W and Wu F 2022 Using simple and easy water quality parameters to predict trihalomethane occurrence in tap water; *Chemosphere* **286** 131586.
- Zait R, Sluser B, Fighir D, Plavan O and Teodosiu C 2022 Priority pollutants monitoring and water quality assessment in the Siret River Basin Romania; *Water* **14**(1) 129.
- Zhang F, Wang X, Chen Y and Airiken M 2020 Estimation of surface water quality parameters based on hyper-spectral and 3D-EEM fluorescence technologies in the Ebinur Lake Watershed, China; *Phys. Chem. Earth, Parts A/B/C* **118** 102895.