Environmental Science and Pollution Research (2024) 31:30370–30398
https://doi.org/10.1007/s11356-024-33350-6

REVIEW ARTICLE

Prediction of potentially toxic elements in water resources using MLP-NN, RBF-NN, and ANFIS: a comprehensive review

Johnson C. Agbasi[1] · **Johnbosco C. Egbueri**[1,2]

Received: 12 March 2024 / Accepted: 12 April 2024 / Published online: 20 April 2024

Abstract

Water resources are constantly threatened by pollution of potentially toxic elements (PTEs). In efforts to monitor and mitigate PTEs pollution in water resources, machine learning (ML) algorithms have been utilized to predict them. However, review studies have not paid attention to the suitability of input variables utilized for PTE prediction. Therefore, the present review analyzed studies that employed three ML algorithms: MLP-NN (multilayer perceptron neural network), RBF-NN (radial basis function neural network), and ANFIS (adaptive neuro-fuzzy inference system) to predict PTEs in water. A total of 139 models were analyzed to ascertain the input variables utilized, the suitability of the input variables, the trends of the ML model applications, and the comparison of their performances. The present study identified seven groups of input variables commonly used to predict PTEs in water. Group 1 comprised of physical parameters (P), chemical parameters (C), and metals (M). Group 2 contains only P and C; Group 3 contains only P and M; Group 4 contains only C and M; Group 5 contains only P; Group 6 contains only C; and Group 7 contains only M. Studies that employed the three algorithms proved that Groups 1, 2, 3, 5, and 7 parameters are suitable input variables for forecasting PTEs in water. The parameters of Groups 4 and 6 also proved to be suitable for the MLP-NN algorithm. However, their suitability with respect to the RBF-NN and ANFIS algorithms could not be ascertained. The most commonly predicted PTEs using the MLP-NN algorithm were Fe, Zn, and As. For the RBF-NN algorithm, they were NO_3, Zn, and Pb, and for the ANFIS, they were NO_3, Fe, and Mn. Based on correlation and determination coefficients (R, R^2), the overall order of performance of the three ML algorithms was ANFIS > RBF-NN > MLP-NN, even though MLP-NN was the most commonly used algorithm.

Keywords Artificial neural networks · Input variables · PTE prediction modeling · Soft computing models · Input variable selection

Introduction

Water is one of the most important natural resources on earth. One of the fundamental human rights and a requirement for a healthy existence is having access to clean drinking water. Urbanization, population expansion, industrialization, and changes in consumption habits have all led to an increase in the demand for freshwater resources globally (Abu et al. 2024; Bhatt et al. 2024). Like other natural resources, proper management of water resources requires effective and efficient monitoring and assessment strategies which would ensure availability of clean water for present and future generations (Abba et al. 2024; Egbueri 2019). According to Ighalo et al. (2021), water quality monitoring and assessment help to determine its purity and safety. Since a lot of chemicals are utilized in our daily lives and could end up in water resources, monitoring water resources is

Highlights

- Physicochemical parameters and metals were commonly used input variables for predicting PTEs.
- Suitability of the input variables used for predicting PTEs in water was identified.
- Order of algorithms' efficacy was ANFIS > RBF-NN > MLP-NN, though MLP-NN was used most.

Responsible Editor: Marcus Schulz

✉ Johnbosco C. Egbueri
jc.egbueri@coou.edu.ng; johnboscoegbueri@gmail.com

[1] Department of Geology, Chukwuemeka Odumegwu Ojukwu University, Uli, Nigeria

[2] Research Management Office (RMO), Chukwuemeka Odumegwu Ojukwu University, Anambra State, Nigeria

becoming more and more difficult in the twenty-first century (Mhlongo et al. 2018). Prior to the adoption of data science, traditional methods were predominantly used for monitoring and assessment of natural and human resources. Traditional methods employed in the study of water resources generate sufficient data, which is now explored with the help of data science. Data science, popularly known as the "oil of the twenty-first century" (Shah et al. 2021), is the process of extracting clean information from raw data to create insights that can be put into practice (Yasmin 2019; Kumar 2015). Application of data science in the study of water resources has aided the determination of the quality of water resources (Karmakar et al. 2021; Chojnacki et al. 2017), discovery of complicated patterns and the causes of water pollution (Omeka and Egbueri 2022; Unigwe et al. 2022), and prediction of possible future status of water resources (Egbueri 2022). Data science has also offered solutions in a variety of other fields, such as engineering (Brunton and Kutz 2022), business (Gunjal 2022), marketing (Shi 2022), finance (Hasan and Alam 2021), and meteorology (Sarker 2021), among others. To build models and make predictions using algorithms and other methods, data scientists significantly rely on artificial intelligence (Usman et al. 2023; Yassin et al. 2023, 2022; Leonard et al. 2021), notably its subfields of machine learning (ML) (Egger 2022; El Mrabet et al. 2021) and deep learning (Franzen et al. 2021). Due to its strong nonlinear mapping and learning capabilities, high fault tolerance, and improved generalization capabilities, deep learning has recently become one of the most widely used methods for research on hydrological time series prediction (Jiang et al. 2021; Lu et al. 2019). The five stages of the typical data science life cycle are collecting, maintaining, processing, communicating, and predictive analysis (Bellatreche et al. 2022; Han and Trimi 2022; Sabharwal and Miah 2021). With the use of the data mining approach known as predictive modeling, different factors are examined for their potential impact on a future result (Aryadoust and Goh 2014). Regression, multivariate adaptive regression splines, classification and regression trees, neural networks, and their expansions known as ANFIS are a few important prediction models (Zhang et al. 2021, 2020; Chebrolu et al. 2005).

Artificial neural networks (ANNs), also referred to as neural networks (NNs), are mathematical nonparametric models that consist of a network of "neurons," which are flexible and trainable processing units that store empirical information (Abba et al. 2023, 2020; Aryadoust and Goh 2014). Similar to the human brain, ANNs are made up of linked units or neurons that are capable of learning, pattern recognition, categorization, and prediction (Geetha et al. 2022; Karaca and Baleanu 2022; Aryadoust and Goh 2014). ANN has become one of the most popular ML tools, finding use in the domains of quality management of different water sources (Khan et al. 2020). The use of ANN as a tool does not necessitate prior understanding of the mathematical structures underlying interactions between inputs and associated outputs (Ewuzie et al. 2022; Nourani et al. 2011; Shahin et al. 2001). Due to their considerable adaptability, ANNs do not place any restrictions on the connections between dependent and independent variables, including those of normality, linearity, homogeneity of variance, and error independence (Aryadoust and Goh 2014). However, the main difficulty with using ANN is its complicated design problems, which might limit its capacity for data processing (Ewuzie et al. 2022). ANN can perform poorly when trained with less data (Ewuzie et al. 2022), since it needs a large number of parameters to get a good outcome (Khalil et al. 2005). Nevertheless, using a lot of weight might lead to overfitting (Farmaki et al. 2010; Zur et al. 2009). Despite these drawbacks, ANN has proven to be reliable in effective monitoring (Azrour et al. 2022; Nicklow et al. 2010; El-Shafie et al. 2009) and assessment (Kouadri et al. 2022; Than et al. 2021) of water resources.

Several studies have successfully predicted physical parameters (Saleh et al. 2022; Egbueri and Agbasi 2022a; Singh et al. 2009), chemical parameters (Egbueri and Agbasi 2022a; El-Safa et al. 2022), and metals (Alizamir and Sobhanardakani 2016; Egbueri 2021) in water with the aid of ANN. To predict using ANN, input/predictor variables and output/predicted variables are selected. Several factors impact the selection of input variables, including parameters that can be measured easily or cheaply, knowledge of the water source, knowledge from literature, theoretical understanding of the parameter(s), and available data from water authorities and monitoring stations (Ewuzie et al. 2022), among others. There are several kinds of neural networks, and they vary in terms of their structure, data flow, number of neurons employed, density, depth of activation filters, and others (Team 2020). They include multilayer perceptron, radial basis function, recurrent neural networks, convolutional neural networks, feedforward neural networks, and others (Praveena and Vivekanandan 2021; Team 2020; Burse et al. 2010). Additionally, a variety of optimizers are available for neural networks, such as gradient descent, stochastic gradient descent, mini-batch gradient descent, Nesterov accelerated gradient, and AdaGrad, among others (Haji and Abdulazeez 2021; Doshi 2019; Dogo et al. 2018; Ruder 2016). The learning rate and weights of NNs are two examples of the variables that these optimizers can adjust in order to change the NNs behavior and decrease losses (Doshi 2019). By selecting the appropriate optimization method, training time may be drastically decreased (Doshi 2019).

PTEs are elements that have the capacity to cause an upset when introduced into a medium (air, water, or soil). Studies have found arsenic (As), cadmium (Cd), chromium (Cr), copper (Cu), iron (Fe), manganese (Mn), nickel (Ni), nitrate (Ni), lead (Pb), zinc (Zn), and other PTEs in polluted levels within natural water bodies (Agbasi et al. 2024; Ayejoto et al. 2022; Fural et al. 2022). These elements can infiltrate water resources through natural processes (Abugu et al. 2024; Ayejoto et al. 2022; Egbueri and Agbasi 2022b) and human-induced processes (Abba et al. 2024; Fural et al. 2022; Jianfei et al. 2020). PTEs have been linked to a wide range of issues, which include anemia (Fural et al. 2022), anomalies in fetal development in pregnant women (Egbueri et al. 2022a), antisocial behaviors observed in children (Emenike et al. 2019), asthma (Rashid et al. 2022), cancer (Zhang et al. 2022), cardiovascular illness (Genchi et al. 2020), diarrhea (Ukah et al. 2019), mental retardation (Mgbenu and Egbueri 2019), respiratory disorders (Wan et al. 2016), sensory disorder (Egbueri 2020), and others. Despite the risks associated with these elements, research has also shown that some PTEs, when found within their allowable limits, have some health benefits (Alipour et al. 2021; Bini and Wahsha 2014). For instance, iron (Fe), when present in water within permissible limits, is advantageous to organisms as it helps circulate oxygen in the blood (Agbasi and Egbueri 2022; IDPH 2010). Globally, PTEs have impacted water resources significantly (Egbueri et al. 2022b; Wagh et al. 2017, 2018). In response, multiple studies have been conducted to understand the impact of PTEs on water resources (Ayejoto et al. 2022; Moghanm et al. 2020), identify their sources (Ricolfi et al. 2020), eradicate (Abbas 2021), and predict the chances of future occurrences (Egbueri and Agbasi 2022a). Several methods have been employed to forecast PTEs in water resources, which include ANN, ANFIS, M5P, multiple linear regression, and others (Egbueri 2021; Eid et al. 2021; Singha et al. 2021). However, most studies have employed ANN in predicting PTEs in water compared to other algorithms.

Studies have revealed that the choice of input variables influences the performance of ANN prediction (Egbueri and Agbasi 2022c; Lee et al. 2021; May et al. 2008). Several studies that used ANNs to predict various elements in water, including PTEs, have been reviewed (Ewuzie et al. 2022; Zounemat-Kermani et al. 2021; Rajaee et al. 2020; Sit et al. 2020). However, attention has not been given to analyzing the suitability of the input parameters utilized for the prediction of PTEs. It is important that the right input variables are selected for effective prediction of PTEs. There are speculative opinions that it is inappropriate to predict PTEs using physicochemical parameters as input variables. Moreover, trends in the applicability of MLP-NN, RBF-NN, and ANFIS algorithms for PTEs' predictions and the performances of these three modeling techniques have not been analyzed in literature. Therefore, the present review examines the state of art and analyzes the input variables utilized for PTE predictions, the suitability of the input variables, the application trends of the three models, and the comparison of their performances. The specific objectives are to (1) identify the predominantly predicted PTEs in water resources; (2) identify the commonly used input variables for predicting PTEs in water; (3) analyze the suitability of input variables used for predicting PTEs in water; (4) identify the most commonly used ANNs; and (5) compare the performances of MLP-NN, RBF-NN, and ANFIS algorithms in forecasting PTEs in water. To the best of the authors' knowledge, this is the first review conducted to analyze the suitability of input variables utilized in the prediction of PTEs in water, with a focus on the MLP-NN, RBF-NN, and ANFIS algorithms. This study is also novel, as it is the first to categorically analyze other aforementioned objectives. It is hoped that the findings of this study will aid researchers, water managers, and policymakers in selecting the right parameters for efficient prediction of PTEs in water resources. The current review paper is also expected to contribute economically, scholarly, and timely to the knowledge bank and understanding of PTEs' predictions in water resources.

Brief history of ANNs

ANNs

ANNs have been used in numerous engineering and scientific applications since the 1940s (Jorjani et al. 2008). The neurophysiologist Warren McCulloch and logician Walter Pits developed the first artificial neuron model in 1943 (McCulloch and Pitts 1943). Simple electrical circuits were used to simulate a neural network by Warren McCulloch and Walter Pits (Zaqoot et al. 2017). The initial attempt to mimic a neural network was led by IBM researcher Nathanial Rochester (McCarthy et al. 1955). It failed at first, but further attempts were successful (Zaqoot et al. 2017). Different architectures, including at least three layers, are found in ANNs (input, hidden, and output layers). There may be one or more hidden layers, and they are situated between the input and output layers (Bayatzadeh Fard et al. 2017). Several neurons make up each layer depending on the layer's location (Bayatzadeh Fard et al. 2017). Input layer neurons correspond to the number of input variables utilized for prediction, while output layer neurons correspond to the number of variables to be predicted (Bayatzadeh Fard et al. 2017). The neurons have learning, categorization, pattern recognition, and prediction abilities (Aryadoust and Goh 2014). By employing input or independent variables, mathematical functions like the multilayer perceptron (MLP) and radial basis function (RBF) are used to predict output or dependent variables in ANNs with the least amount of error (Aryadoust and Goh 2014). The designer chooses the training algorithm, learning rule, network topology, performance function, and criterion to end the training phase in artificial neural networks, but the system undoubtedly modifies the parameters (Adeoti and Osanaiye 2013). The feedforward topology and the recurrent topology are two significant designs that are frequently used to visualize an ANN (Zaqoot et al. 2017). Due to its link with the backpropagation learning algorithm, a dominant and very reliable learning technique, the feedforward topology is extremely well-liked (Zaqoot et al. 2017). Among the networks using the feedforward topology are the MLP network and the RBF network.

MLP

A perceptron model was first presented in 1958 by Cornell neurobiologist Frank Rosenblatt (Olazaran 1996). The hardware included the perceptron, which he discovered via his scientific efforts (Zaqoot et al. 2017). The MLP is a feedforward neural network ANN with input, hidden, and output layers. It always sends signals in the direction of the output layer (Bayatzadeh Fard et al. 2017). Each layer contains an activation function, which expresses the quantity of output based on the input data mathematically (Aryadoust and Goh 2014). Hyperbolic tangent and logistic functions are examples of mathematical activation functions in the neurons of MLP networks (Aryadoust and Goh 2014). The most commonly used mathematical function in ANN prediction is MLP (Maier and Dandy 2000).

RBF

According to Suen and Eheart (2003), radial basis function neural networks were created about the same time by Powell (1987) and Broomhead and Lowe (1988). Similar to the MLP, the RBF network also consists of an input layer, a hidden layer, and an output layer, with neurons present in each layer. In contrast to MLP, RBF only comprises weights between the hidden layer and the output layer. The hidden layer is where the most significant differences between RBF and MLP can be found. These differences may be divided into structural and functional differences (Ucun Ozel et al. 2020). The number of neurons and the training technique are responsible for the structural difference, but the hidden layer neurons' inclusion of radial functions is responsible for the functional difference (Ucun Ozel et al. 2020). Because the RBF function depends on the distance from the origin to gather the input layer neurons, variations in it are depending on radial distance (Alizamir and Sobhanardakani 2017). The Gaussian function is primarily used as the activation function in the neurons of RBF networks (Asgharnia et al. 2019; Han et al. 2019).

ANFIS

An extension of artificial neural networks (ANNs) and adaptive neuro-fuzzy inference systems (ANFIS) merge ANNs with fuzzy (Landín et al. 2009). The fuzzy inference system (FIS) examines human thinking by incorporating uncertainty into if-then rules and human knowledge (Mousavi and Amiri 2012). The fuzzy rule base, membership functions, which specify the fuzzy sets of fuzzy rules, and a reasoning process are the three fundamental components of ANFIS (Ucun Ozel et al. 2020). In order to determine the membership function parameters, ANFIS employs gradient descent-based optimization techniques (Ucun Ozel et al. 2020). There are three different methodologies employed in developing the ANFIS model, namely the fuzzy C-means (FCM) technique, grid partitioning (GP), and the subtractive clustering method (SCM) (Bayatzadeh Fard et al. 2017). The SCM and FCM methods may be used for multi-output ANFIS, in contrast to the GP approach, which produces a single-output Sugeno-type ANFIS on the data (Bayatzadeh Fard et al. 2017). In general, an input-output data set is needed in order to use the ANFIS approach.

Selection of articles for the review study

The primary focus of this review study is on the adequacy of input variables utilized by previous authors for three ANN techniques (MLP, RBF, and ANFIS) to forecast PTEs in water resources. The relevant papers were found by conducting a keyword search of articles on Google, Goggle scholar, and Research Gate that had been published over the years on the subject, using terms like "ANN," 'predicting," "forecasting," "modelling," "heavy metals," "water," and "potentially toxic elements," alongside with the names of the modeling approaches, like "MLP," "RBF," and "ANFIS." Then, among the search engine results displayed, the most pertinent articles were chosen after careful examinations of their contents.

Classification schemes of model performance metrics

Correlation coefficient (*R*) and coefficient of determination (R^2)

The correlation coefficient (R) is a statistical measure used to assess the degree of association between two quantitative variables recorded in each individual member of a group (Eq. 1; Nazar et al. 2023; Aggarwal and Ranganathan 2016). The coefficient of determination, also known as the square of the correlation coefficient (R^2), is the percentage of variation in one variable that is accounted for by variation in the other variable (Eq. 2; Egbueri and Agbasi 2022c; Aggarwal and Ranganathan 2016). The values of R and R^2 can range from −1.0 to +1.0. The qualitative description of R values according to Egbueri (2021) is as follows: strong correlation (0.75–1.00), moderate correlation (0.50–0.75), and weak correlation ($r < 0.5$).

$$R = \frac{\sum_{i=1}^{n} \left(a_i - \overline{a}_i\right)\left(p_i - \overline{p}_i\right)}{\sqrt{\sum_{i=1}^{n} \left(a_i - \overline{a}_i\right)^2 \sum_{i=1}^{n} \left(p_i - \overline{p}_i\right)^2}} \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(X_{\text{predicted value}} - X_{\text{measured value}}\right)^2}{\sum_{i=1}^{n} \left(X_{\text{predicted value}} - X_{\text{average value}}\right)^2} \quad (2)$$

where p_i and a_i represent the ith predicted and observed scores, respectively. $\overline{p}_i$ and $\overline{a}_i$ denote the mean predicted and observed scores across n total observations.

Mean square error (MSE)

A statistical technique for calculating the difference between an estimator and an estimated outcome is the MSE, often known as the mean squared deviation (MSD) of an estimator (Sara et al. 2019). The MSD or MSE calculates the average of the square of the errors between an estimator and an estimated outcome (Eq. 3; Pande et al. 2024; Sara et al. 2019). The key benefit of utilizing MSE is that it squares the error, which penalizes or prominently highlights large errors (Allwright 2022). Therefore, it is helpful when working on models where it is necessary to minimize sporadic significant errors (Allwright 2022). Models with MSE values nearer to zero are more accurate. Nevertheless, a “good” value for MSE does not exist (Allwright 2022). This is because MSE is an absolute metric specific to each use case; results may only be compared to other MSE values computed for the same dataset (Allwright 2022).

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} \left(y\_i - \hat{y}\_i\right)^2 \quad (3)$$

where n is the total count of data points. For the ith sample, y_i refers to the true, observed value of the target variable, while $\hat{y}\_i$ shows the value forecasted by the model for that same sample.

Root mean square error (RMSE)

The square root of the MSE is what generates the RMSE (Eq. 4; Pande et al. 2024; Sara et al. 2019). According to Draper et al. (2013), the RMSE is also referred to as the root mean square deviation (RMSD). The difference between an estimator’s forecasted value and the actual value is often measured using the RMSE (Sara et al. 2019). It calculates the differences in predicting errors from the various estimators for a certain variable and analyzes the error size, making it the ideal accuracy measure (Sara et al. 2019). Like the MSE, models with RMSE values closer to zero are more accurate.

$$\text{RMSE} = \sqrt{MSE} \quad (4)$$

Predominantly predicted PTEs in water resources

MLP

The predicted PTEs in water resources using the MLP-NN algorithm by the studies reviewed (24 articles) include Al, As, Cd, Cr, Cu, Fe, Hg, Mn, Ni, NO_3, Pb, Sb, Sr, Ti, and Zn. The degree to which they were predicted by the studies is as follows: Pb > Zn > As > Fe > Mn > Cu > NO_3 > Ni > Cd > Al = Cr = Hg = Sb = Sr = Ti. A graphical representation of their distribution can be found in Fig. 1a.

RBF

The predicted PTEs in water resources using the RBF-NN algorithm by the studies reviewed (8 articles) include As, Cd, Cu, Fe, Hg, Mn, Ni, NO_3, Pb, and Zn. The degree to which they were predicted by the studies is as follows: NO_3 > Zn = Pb > As = Cd, Cu = Fe = Hg = Mn = Ni. A graphical representation of their distribution can be found in Fig. 1b.

ANFIS

The predicted PTEs in water resources using the ANFIS algorithm by the studies reviewed (10 articles) include As, Cd, Cu, Fe, Mn, Ni, NO_3, Pb, and Zn. The degree to which they were predicted by the studies is as follows: NO_3 = Fe = Mn = Zn > As = Cd = Cu = Pb > Ni. A graphical representation of their distribution can be found in Fig. 1c.

Commonly used input parameters for predicting PTEs in water resources

From the articles reviewed (42 articles), seven groups of input variables were found to be commonly used for predicting PTEs in water resources. These groups of input variables include Group 1, a combination of physical parameters (P), chemical parameters (C), and metals (M); Group 2, a combination of P and C only; Group 3, a combination of P and M

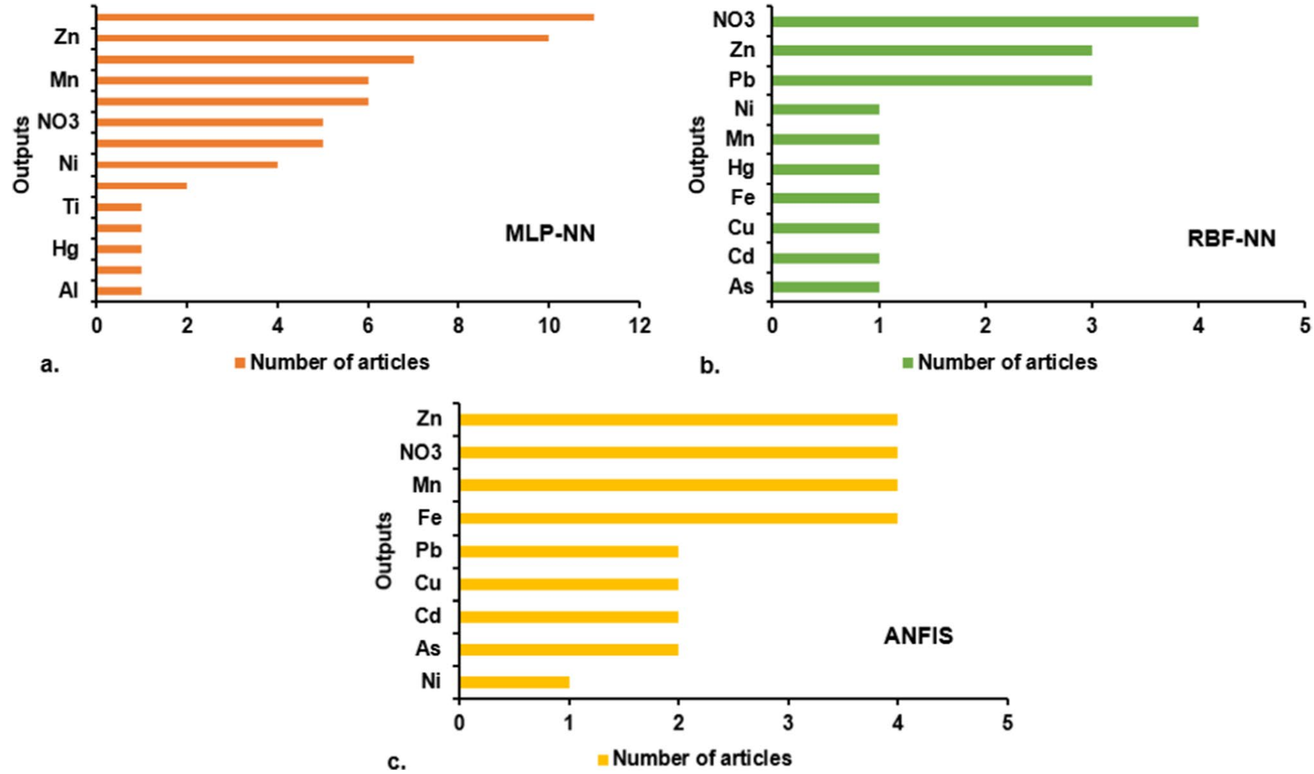

Fig. 1 Bar charts showing PTEs and number of studies that predicted them using **a** MLP-NN, **b** RBF-NN, **c** ANFIS

only; Group 4, a combination of C and M only; Group 5, P only; Group 6, C only; and Group 7, M only. Physical parameters include pH, temperature, TDS, SS, BOD, and COD; chemical parameters include HCO_3, SO_4, NO_3, and Cl; and metals include As, Cd, Cr, Pb, and Zn. The full details of the input parameters utilized by 42 articles for forecasting PTEs in water resources can be found in Tables 1, 2, and 3.

MLP

The pictorial representation of the distribution of input variables used to predict PTEs in water resources with the aid of the MLP-NN algorithm can be visualized in Fig. 2a. It was observed that most of the studies used Group 1 input variables for their prediction. The degree to which the reviewed studies utilized input variables when the MLP-NN algorithm was employed is as follows: Group 1 > Group 5 > Group 7 > Group 2 = Group 4 > Group 6 > Group 3.

RBF

The pictorial representation of the distribution of input variables used to predict PTEs in water resources with the aid of the RBF-NN algorithm can be visualized in Fig. 2b. It was observed that most of the studies used Groups 5 and 7 input variables for their prediction. The degree to which the reviewed studies utilized input variables when the RBF-NN algorithm was employed is as follows: Group 5 = Group 7 > Group 1 = Group 2 = Group 3 > Group 4 = Group 6.

ANFIS

The pictorial representation of the distribution of input variables used to predict PTEs in water resources with the aid of the ANFIS algorithm can be visualized in Fig. 2c. It was observed that most of the studies used Group 1 input variables for their prediction. The degree to which the reviewed studies utilized input variables is as follows: Group 1 > Group 2 = Group 5 > Group 7 > Group 3 = Group 4 = Group 6.

Suitability of input variables in the prediction of PTEs in water resources

To ascertain the suitability of input variables in the prediction of PTEs in water resources using MLP-NN, RBF-NN, and ANFIS algorithms, data from 148 models (MLP-NN algorithm, 86 models, RBF-NN algorithm, 29 models, and ANFIS algorithm, 33 models) was extracted from the 42 articles reviewed. However, 9 models (MLP-NN algorithm-1 model, RBF-NN algorithm-2 models,

Table 1 Synopsis of MLP-NN algorithm used for forecasting PTEs in water resources

Article no.	References	Model no.	Input variables	Output variable	R^2	R	MAE	MSE	RMSE/SEE
1	Kanj et al. 2022	Model 1	pH, Electrical conductivity (EC), TDS (Total dissolved solids), Total Oxidized Nitrogen (TON), calcium (Ca) and magnesium (Mg)	Mercury (Hg)	-	0.903	0.383	-	0.452
		Model 2	pH, EC, TDS, Mg, and TON	Hg	-	0.842	0.461	-	0.709
		Model 3	pH, EC, TDS and TON	Hg	-	0.882	0.400	-	0.501
		Model 4	pH, EC and TON	Hg	-	0.915	0.328	-	0.407
		Model 5	pH and EC	Hg	-	0.831	0.468	-	0.616
		Model 6	EC	Hg	-	0.500	1.084	-	1.732
2	Chaal and Aboutafail 2021	Model 7	Temperature, pH, dissolved oxygen (DO), conductivity (Cond), TDS, bicarbonate (HCO_3), total alkalinity (as $CaCO_3$), Mg, sodium (Na), potassium (K), chlorides (Cl), Ca, sulfate (SO_4), nitrate (NO_3), phosphorus (P), and ammoniacal nitrogen (NH_4)	Copper (Cu)	0.99	-	-	-	-
		Model 8	Temperature, pH, DO, Cond, TDS, HCO_3, $CaCO_3$, Mg, Na, K, Cl, Ca, SO_4, NO_3, P, and NH_4	Manganese (Mn)	0.99	-	-	-	-
		Model 9	Temperature, pH, DO, Cond, TDS, HCO_3, $CaCO_3$, Mg, Na, K, Cl, Ca, SO_4, NO_3, P, and NH_4	Zinc (Zn)	0.99	-	-	-	-
3	Egbueri 2021	Model 10	pH, Fe, Zn, Ni, Cr, and Pb	NO_3	0.794	-	-	-	-
		Model 11	pH, Zn, and HCO_3	Nickel (Ni)	0.678	-	-	-	-
		Model 12	pH, Zn, and HCO_3	Lead (Pb)	0.828	-	-	-	-
4	Boudaghpour and Malekmohammadi 2020	Model 13	Flow rate, hydraulic gradient, pH, electrical conductivity, lifetime, water level, and abstraction	Pb	0.88	-	-	-	-
5	Ucun Ozel et al. 2020	Model 14	Temperature, pH, EC, chemical oxygen demand (COD), biological oxygen demand (BOD), and suspended matter (SS)	Cu	0.954	-	0.0014	-	0.0015
		Model 15	Temperature, pH, EC, COD, BOD, and SS	Fe	0.842	-	0.1029	-	0.1316
		Model 16	Temperature, pH, EC, and COD	Zn	0.966	-	0.0143	-	0.0165
		Model 17	Temperature, pH, and EC	Mn	0.770	-	0.0100	-	0.0110
		Model 18	Temperature, pH, and EC	Ni	0.941	-	0.0018	-	0.0024
		Model 19	Temperature, EC, and COD	Pb	0.805	-	0.0006	-	0.0008
6	Lu et al. 2019	Model 20	water temperature (WT), pH, SS, turbidity (TUB), total nitrogen (TN), ammonia nitrogen (NH_3-N), nitrate nitrogen (NO_3-N), total phosphorous (TP), orthophosphate (PO_4-P), and permanganate index (COD_{Mn})	Ti (titanium)	-	0.97	-	251.14	-
		Model 21	pH, SS, COD_{Mn}, TP and WT	Ti	-	0.99	-	197.43	-
		Model 22	WT, pH, SS, TUB, TN, NH_3-N, NO_3-N, TP, PO_4-P and COD_{Mn}	Cr	-	0.87	-	4.79	-
		Model 23	pH, SS, COD_{Mn}, TP and WT	Cr	-	0.79	-	2.77	-

Table 1 (continued)

Article no.	References	Model no.	Input variables	Output variable	R^2	R	MAE	MSE	RMSE/SEE
		Model 24	WT, pH, SS, TUB, TN, NH_3-N, NO_3-N, TP, PO_4-P and COD_{Mn}	Mn	-	0.81	-	465.34	-
		Model 25	pH, SS, COD_{Mn}, TP and WT	Mn	-	0.67	-	1070.77	-
		Model 26	WT, pH, SS, TUB, TN, NH_3-N, NO_3-N, TP, PO_4-P and COD_{Mn}	Ni	-	0.88	-	103.33	-
		Model 27	pH, SS, COD_{Mn}, TP and WT	Ni	-	0.93	-	57.36	-
		Model 28	WT, pH, SS, TUB, TN, NH_3-N, NO_3-N, TP, PO_4-P and COD_{Mn}	As	-	0.86	-	46.76	-
		Model 29	pH, SS, COD_{Mn}, TP and WT	As	-	0.84	-	58.06	-
		Model 30	WT, pH, SS, TUB, TN, NH_3-N, NO_3-N, TP, PO_4-P and COD_{Mn}	Cd	-	0.82	-	0.39	-
		Model 31	pH, SS, COD_{Mn}, TP and WT	Cd	-	0.83	-	0.34	-
		Model 32	WT, pH, SS, TUB, TN, NH_3-N, NO_3-N, TP, PO_4-P and COD_{Mn}	Sb (antimony)	-	0.86	-	0.40	-
		Model 33	pH, SS, COD_{Mn}, TP and WT	Sb	-	0.58	-	0.99	-
		Model 34	WT, pH, SS, TUB, TN, NH_3-N, NO_3-N, TP, PO_4-P and COD_{Mn}	Pb	-	0.76	-	58.01	-
		Model 35	pH, SS, COD_{Mn}, TP, and WT	Pb	-	0.83	-	49.03	-
7	Alizamir et al. 2019	Model 36	As, Zn, and Pb	As	0.672	0.819	-	-	1.310
		Model 37	As, Zn, and Pb	Zn	0.672	0.819	-	-	13.516
		Model 38	As, Zn, and Pb	Pb	0.708	0.841	-	-	1.017
8	Alayat et al. 2018	Model 39	Ca, Mg, Na, K, Cl, SO_4, HCO_3, and NO_3	Fe	-	0.79	-	-	-
9	Alizamir and Sobhanardakani 2017	Model 40	As, Zn, and Pb	As	0.9302	-	-	-	0.2592
		Model 41	As, Zn, and Pb	Zn	0.9774	-	-	-	0.4797
		Model 42	As, Zn, and Pb	Pb	0.9626	-	-	-	0.2880
10	Zaqoot et al. 2017	Model 43	Pressure, temperature, pH, and EC	NO_3	-	0.98972	3.4713	27.7861	5.2713
11	Alizamir et al. 2017	Model 44	Cd, Pb, and Zn	Cd	0.9465	-	-	-	0.2601
		Model 45	Cd, Pb, and Zn	Pb	0.9821	-	-	-	0.7627
		Model 46	Cd, Pb, and Zn	Zn	0.9373	-	-	-	5.0030
12	Venkatramanan et al. 2017	Model 47	Si, Al, Fe, K, Ca, Na, Mg, Cl, Mn, Sr, and Br	Al	0.8651	-	-	-	-
		Model 48	Si, Al, Fe, K, Ca, Na, Mg, Cl, Mn, Sr, and Br	Fe	0.7145	-	-	-	-
		Model 49	Si, Al, Fe, K, Ca, Na, Mg, Cl, Mn, Sr, and Br	Mn	0.3199	-	-	-	-
		Model 50	Si, Al, Fe, K, Ca, Na, Mg, Cl, Mn, Sr, and Br	Sr (strontium)	0.0847	-	-	-	-
13	Alizamir and Sobhanardakani 2016	Model 51	As, Pb, and Zn	As	0.9288	-	-	-	4.0303
		Model 52	As, Pb, and Zn	Pb	0.9823	-	-	-	0.7638
		Model 53	As, Pb, and Zn	Zn	0.8876	-	-	-	7.2230
14	Bayatzadeh Fard et al. 2017	Model 54	SO_4, Cl, and TDS	Fe	0.54	-	-	0.01	-

Table 1 (continued)

Article no.	References	Model no.	Input variables	Output variable	R^2	R	MAE	MSE	RMSE/SEE
		Model 55	SO_4, Cl, and TDS	Mn	0.59	-	-	0.07	-
		Model 56	SO_4, Cl, and TDS	Pb	0.68	-	-	0.02	-
		Model 57	SO_4, Cl, and TDS	Zn	0.52	-	-	0.08	-
15	Ghadimi 2015	Model 58	HCO_3 and SO_4	Pb	0.71	-	-	0.158	-
		Model 59	HCO_3 and SO_4	Zn	0.71	-	-	0.124	-
		Model 60	HCO_3 and SO_4	Cu	0.70	-	-	0.094	-
16	Shakeri et al. 2013	Model 61	EC, TDS, temperature, and pH	Cu	-	0.9184	-	0.00010	-
17	Zare et al. 2011	Model 62	Na, Mg, temperature, Ca, HCO_3, SO_4, Cl, TH, TDS, EC, and pH	NO_3	-	0.84	9.84	-	14.78
		Model 63	Na, Mg, Ca, HCO_3, Cl, TH, TDS, EC, and pH	NO_3	-	0.84	8.02	-	12.68
		Model 64	Na, Mg, Ca, HCO_3, Cl, TH, EC, and pH	NO_3	-	0.87	7.77	-	10.46
		Model 65	Na, Mg, Ca, HCO_3, Cl, EC, and pH	NO_3	-	0.81	8.68	-	13.51
		Model 66	Na, Mg, Ca, HCO_3, Cl, and EC	NO_3	-	0.79	9.28	-	13.56
		Model 67	Na, Mg, Ca, HCO_3, and EC	NO_3	-	0.80	10.47	-	14.25
18	Cho et al. 2011	Model 68	pH, EC, TDS, temperature, and redox potential (Eh)	As	-	-	88.93	-	-
19	Rooki et al. 2011	Model 69	pH, SO_4, and Mg	Cu	-	0.921	-	-	-
		Model 70	pH, SO_4, and Mg	Fe	-	0.22	-	-	-
		Model 71	pH, SO_4, and Mg	Mn	-	0.915	-	-	-
		Model 72	pH, SO_4, and Mg	Zn	-	0.92	-	-	-
20	Gholami et al. 2011	Model 73	pH, SO4, HCO_3, TDS, EC, Mg, and Ca	Fe	-	0.88	-	-	0.3
		Model 74	pH, SO4, HCO_3, TDS, EC, Mg, and Ca	Ni	-	0.901	-	-	0.4
21	Li et al. 2020	Model 75	Fe, River flow, pH, WT, and DO	As	0.550	-	-	-	0.383
		Model 76	Fe, pH, and DO	As	0.415	-	-	-	0.376
		Model 77	Fe	As	0.442	-	-	-	0.163
		Model 78	Fe, River flow, pH, NO_3–N, EC	Pb	0.703	-	-	-	1.290
		Model 79	Fe and River flow	Pb	0.666	-	-	-	0.807
		Model 80	Fe	Pb	0.632	-	-	-	0.794
		Model 81	Fe, River flow, pH, NO_3–N, and EC	Zn	0.780	-	-	-	6.702
		Model 82	Fe and NO_3–N	Zn	0.714	-	-	-	5.033
		Model 83	Fe	Zn	0.632	-	-	-	3.499
22	Purkait et al. 2008	Model 84	pH, specific conductivity (EC_w), TDS, salinity, DO, Eh, and depth of tube well water	As	0.6672	-	114.29	-	184.11
23	Yesilnacar et al. 2007	Model 85	pH, temperature, EC_w, and groundwater level	NO_3	-	0.92	-	0.0794	-
24	Diamantopoulou et al. 2005	Model 86	Discharge, WT, EC_w, DO, Na, Ca, SO_4, Cl, and HCO_3	NO_3	-	0.9278	0.962	-	1.333

Table 2 Synopsis of RBF-NN algorithm used for forecasting PTEs in water resources

Article no.	References	Model no.	Input variables	Output variable	R^2	R	MAE	MSE	RMSE/SEE
25	Kanj et al. 2022	Model 87	pH, EC, TDS, TON, Ca, and Mg	Hg	-	0.870	0.518	-	0.656
		Model 88	pH, EC, TDS, Mg, and TON	Hg	-	0.849	0.484	-	0.705
		Model 89	pH, EC, TDS, and TON	Hg	-	0.864	0.418	-	0.658
		Model 90	pH, EC, and TON	Hg	-	0.870	0.491	-	0.582
		Model 91	pH and EC	Hg	-	0.740	0.600	-	1.384
		Model 92	EC	Hg	-	0.600	0.474	-	0.867
26	Ucun Ozel et al. 2020	Model 93	Temperature, pH, EC, COD, BOD, and SS	Cu	0.920	-	0.0011		0.0013
		Model 94	Temperature, pH, EC, COD, and BOD	Fe	0.892	-	0.0962	-	0.1281
		Model 95	Temperature, pH, EC, COD, BOD, and SS	Zn	0.989	-	0.0105	-	0.0115
		Model 96	Temperature, pH, EC, COD, BOD, and SS	Mn	0.831	-	0.0077	-	0.0083
		Model 97	Temperature, pH, EC, and COD	Ni	0.773	-	0.0044	-	0.0049
		Model 98	Temperature, pH, EC, COD, BOD, and SS	Pb	0.829	-	0.0007	-	0.0008
27	Zaqoot et al. 2018	Model 99	pH, EC, TDS, TH, Ca, Mg, and abstraction rate (Abs)	NO_3	-	0.785	55.906	-	103.26
		Model 100	pH, EC, TDS, Ca, Mg, and Abs	NO_3	-	0.638	58.500	-	105.32
		Model 101	pH, EC, TDS, Mg, and Abs	NO_3	-	0.907	42.202	-	70.815
		Model 102	pH, EC, TDS, and Abs	NO_3	-	0.725	54.232	-	86.655
		Model 103	pH, EC, and Abs	NO_3	-	0.735	53.194	-	82.402
		Model 104	pH and EC	NO_3	-	0.612	56.357	-	94.095
		Model 105	EC	NO_3	-	0.733	55.066	-	71.423
28	Alizamir and Sobhanardakani 2017	Model 106	As, Zn, and Pb	As	0.9199	-	-	-	0.3666
		Model 107	As, Zn, and Pb	Zn	0.8953	-	-	-	0.5064
		Model 108	As, Zn, and Pb	Pb	0.9531	-	-	-	0.3867
29	Zaqoot et al. 2017	Model 109	Pressure, temperature, pH, and EC	NO_3	-	0.99511	2.7143	17.6708	4.2037
30	Alizamir et al. 2017	Model 110	Cd, Pb, and Zn	Cd	0.9402	-	-	-	0.2904
		Model 111	Cd, Pb, and Zn	Pb	0.9790	-	-	-	0.8237
		Model 112	Cd, Pb, and Zn	Zn	0.8742	-	-	-	7.6509
31	Ehteshami et al. 2016	Model 113	NO_3, soil organic matter content, soil nitrogen content, and pH	NO_3	-	0.7	-	0.69	-
32	Suen and Eheart 2003	Model 114	Precipitation, streamflow, and temperature	NO_3	-	-	-	-	2.946
		Model 115	Precipitation, streamflow, and temperature	NO_3	-	-	-	-	2.567

and ANFIS algorithm-6 models) did not report R^2 or R values and thus were not considered (Tables 4, 5, and 6). Thus, a total of 139 models (MLP-NN algorithm, 85 models; RBF-NN algorithm, 27 models; and ANFIS algorithm, 27 models) were used to ascertain the suitability of input variables in forecasting PTEs in water resources using MLP-NN, RBF-NN, and ANFIS algorithms. The models were evaluated based on the seven groups of input variables mentioned in the previous section. The seven groups of input variables utilized by the three ML

Table 3 Synopsis of ANFIS algorithm used for forecasting PTEs in water resources

Article no.	References	Model no.	Input variables	Output variable	R^2	R	MAE	MSE	RMSE/SEE
33	Abd El-Mageed et al. 2022	Model 116	Locations, year, season, COD, BOD, NH_4, and NO_3	NO_3	-	-	0.564	-	0.575
34	Agah and Soleimanpour-moghadam 2020	Model 117	pH, SO_4, and Mg	Cu	-	-	-	0.3079	0.0055
		Model 118	pH, SO_4, and Mg	Fe	-	-	-	0.1020	0.0010
		Model 119	pH, SO_4, and Mg	Mn	-	-	-	0.4031	0.0063
		Model 120	pH, SO_4, and Mg	Zn	-	-	-	0.1639	0.0040
35	Elzain 2020	Model 121	Depth to water, net recharge, aquifer media, soil media, topographic slope, impact to vadose zone, hydraulic conductivity, and land-use layers	NO_3	-	0.60	0.580	-	0.836
36	Ucun Ozel et al. 2020	Model 122	Temperature, pH, EC, COD, BOD, and SS	Cu	0.879	-	0.0017	-	0.0020
		Model 123	Temperature, pH, and COD	Fe	0.813	-	0.0342	-	0.0425
		Model 124	Temperature, pH, EC, COD, BOD, and SS	Zn	0.896	-	0.0204	-	0.0235
		Model 125	Temperature, pH, EC, COD, and SS	Mn	0.826	-	0.0103	-	0.0115
		Model 126	Temperature, pH, COD, and BOD	Ni	0.807	-	0.0044	-	0.0049
		Model 127	Temperature, pH, EC, COD, BOD, and SS	Pb	0.834	-	0.0006	-	0.0007
37	Jebastina and Prince Arulraj 2018	Model 128	Ca and EC	NO_3	0.6400	0.8000	-	-	0.1631
		Model 129	Ca, EC, and hardness	NO_3	0.7200	0.8500	-	-	0.1401
		Model 130	Ca, EC, Na, and hardness	NO_3	0.7570	0.8700	-	-	0.1212
		Model 131	Ca, EC, Na, K, and hardness	NO_3	0.8870	0.9420	-	-	0.1000
		Model 132	Ca, EC, Na, Cl, K, and hardness	NO_3	0.9000	0.9500	-	-	0.0934
38	Sonmez et al. 2018	Model 133	Fe, Cu, Mn, Zn, Ni, Cr	Cd	0.9130	-	-	0.0000	0.0004
39	Bayatzadeh Fard et al. 2017	Model 134	SO_4, Cl, and TDS	Fe	0.60	-	-	0.05	-
		Model 135	SO_4, Cl, and TDS	Mn	0.92	-	-	0.01	-
		Model 136	SO_4, Cl, and TDS	Pb	0.84	-	-	0.01	-
		Model 137	SO_4, Cl, and TDS	Zn	0.78	-	-	0.69	-
40	Chang et al. 2014	Model 138	Antecedent rainfall, temperature, Cd, EC, Hg, nitrite nitrogen, pH, DO, Zn, Pb, nitrate nitrogen, Cr, Cl, and Cu	As	-	-	-	-	0.011
41	Valente et al. 2013	Model 139	pH, EC, and SO_4	Fe	0.9695	-	-	-	-
		Model 140	pH, EC, and SO_4	As	0.9890	-	-	-	-
		Model 141	pH, EC, and SO_4	Cd	0.9929	-	-	-	-
		Model 142	pH, EC, and SO_4	Zn	0.9634	-	-	-	-
		Model 143	pH, EC, and SO_4	Cu	0.9762	-	-	-	-
		Model 144	pH, EC, and SO_4	Mn	0.9799	-	-	-	-
42	Mousavi and Amiri 2012	Model 145	EC, HCO_3, Ca, and hardness	NO_3	0.9300	-	-	-	1.17

Table 3 (continued)

Article no.	References	Model no.	Input variables	Output variable	R^2	R	MAE	MSE	RMSE/SEE
		Model 146	EC, HCO_3, Ca, Mg, and hardness	NO_3	0.9100	-	-	-	1.9
		Model 147	EC, HCO_3, and hardness	NO_3	0.8800	-	-	-	2.3
		Model 148	EC and hardness	NO_3	0.6800	-	-	-	2.94

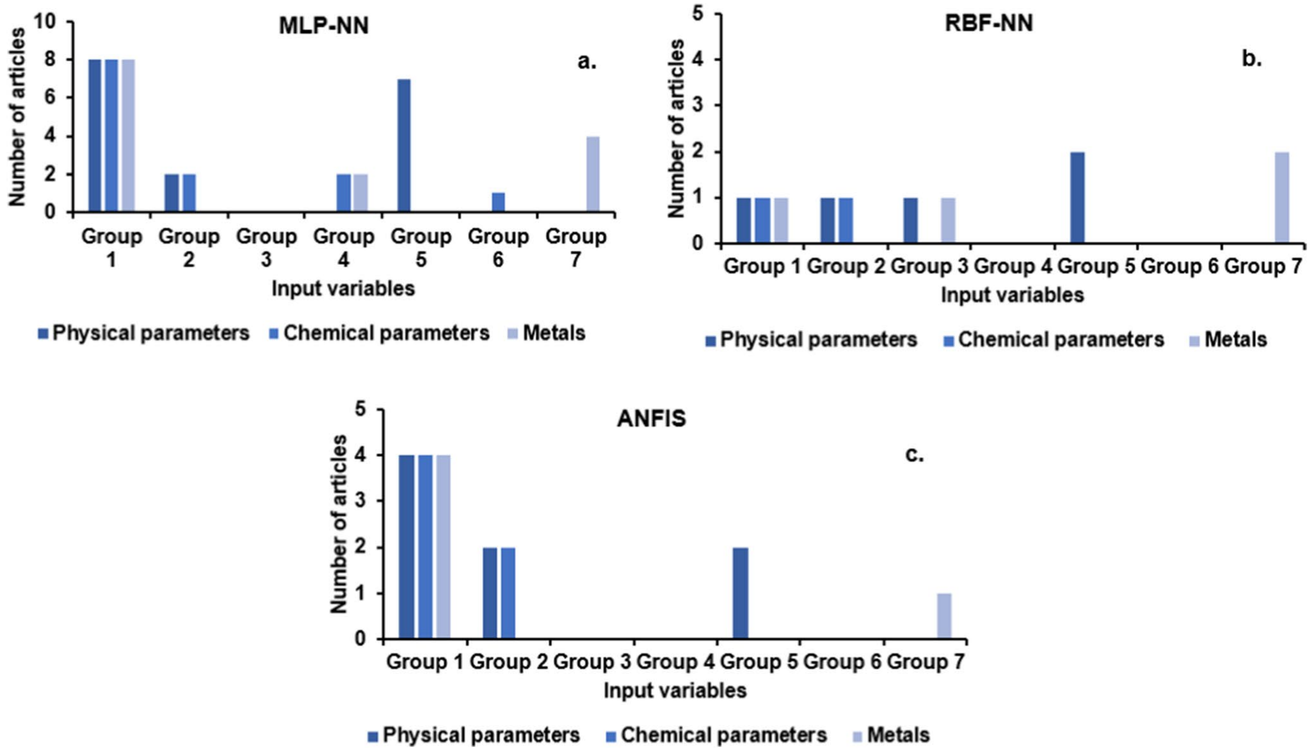

Fig. 2 Bar charts showing input variables utilized for prediction of PTEs employing **a** MLP-NN, **b** RBF-NN, **c** ANFIS

algorithms were compared with the model's R^2 or R. In scenarios where R^2 and R were given, the outcomes of R^2 were taken. This is because R^2 has a more straightforward interpretation and is more widely used than R for ANN modeling. Moreover, it aligns more with the study objectives, as it represents the proportion of variance in the dependent variable that is explained by the independent variable(s) in the model. Based on their R^2 or R values, the performances of the models were classified into weak, moderate, and strong correlation.

MLP

Physical parameters, chemical parameters, and metals combined as input variables

A total of 22 models combined Group 1 parameters (P, C, and M) as input variables for prediction of PTEs in water resources using the MLP-NN algorithm. Based on the performance metrics used (R^2 or R values), 1/22, 2/22, and 19/22 of the models can be classified as having weak, moderate, and strong correlations, respectively (Table 7). The distribution of the model performance can be visualized in Fig. 3a. The distribution shows that 4.50%, 9.10%, and 86.40% of models that used Group 1 parameters as input variables for predicting PTEs in water resources had weak, moderate, and strong correlations, respectively. In summary, 95.5% of the models evaluated had an acceptable model performance (moderate-strong correlation). Thus, it can be concluded that a combination of only physical parameters, chemical parameters, and metals as input variables is suitable for predicting PTEs in water resources employing the MLP-NN algorithm.

Table 4 Models of PTEs in water resources produced using MLP-NN algorithm

Article no.	References	Model no.	Input variables type	Output variable	Performance
1	Kanj et al. 2022	Model 1	Group 1	Hg	Strong correlation
		Model 2	Group 1	Hg	Strong correlation
		Model 3	Group 2	Hg	Strong correlation
		Model 4	Group 2	Hg	Strong correlation
		Model 5	Group 5	Hg	Strong correlation
		Model 6	Group 5	Hg	Moderate correlation
2	Chaal and Aboutafail 2021	Model 7	Group 1	Cu	Strong correlation
		Model 8	Group 1	Mn	Strong correlation
		Model 9	Group 1	Zn	Strong correlation
3	Egbueri 2021	Model 10	Group 3	NO_3	Strong correlation
		Model 11	Group 1	Ni	Moderate correlation
		Model 12	Group 1	Pb	Strong correlation
4	Boudaghpour and Malekmohammadi 2020	Model 13	Group 5	Pb	Strong correlation
5	Ucun Ozel et al. 2020	Model 14	Group 5	Cu	Strong correlation
		Model 15	Group 5	Fe	Strong correlation
		Model 16	Group 5	Zn	Strong correlation
		Model 17	Group 5	Mn	Strong correlation
		Model 18	Group 5	Ni	Strong correlation
		Model 19	Group 5	Pb	Strong correlation
6	Lu et al. 2019	Model 20	Group 2	Ti	Strong correlation
		Model 21	Group 2	Ti	Strong correlation
		Model 22	Group 2	Cr	Strong correlation
		Model 23	Group 2	Cr	Strong correlation
		Model 24	Group 2	Mn	Strong correlation
		Model 25	Group 2	Mn	Moderate correlation
		Model 26	Group 2	Ni	Strong correlation
		Model 27	Group 2	Ni	Strong correlation
		Model 28	Group 2	As	Strong correlation
		Model 29	Group 2	As	Strong correlation
		Model 30	Group 2	Cd	Strong correlation
		Model 31	Group 2	Cd	Strong correlation
		Model 32	Group 2	Sb	Strong correlation
		Model 33	Group 2	Sb	Moderate correlation
		Model 34	Group 2	Pb	Strong correlation
		Model 35	Group 2	Pb	Strong correlation
7	Alizamir et al. 2019	Model 36	Group 7	As	Moderate correlation
		Model 37	Group 7	Zn	Moderate correlation
		Model 38	Group 7	Pb	Moderate correlation
8	Alayat et al. 2018	Model 39	Group 4	Fe	Strong correlation
9	Alizamir and Sobhanardakani 2017	Model 40	Group 7	As	Strong correlation
		Model 41	Group 7	Zn	Strong correlation
		Model 42	Group 7	Pb	Strong correlation
10	Zaqoot et al. 2017	Model 43	Group 5	NO_3	Strong correlation
11	Alizamir et al. 2017	Model 44	Group 7	Cd	Strong correlation
		Model 45	Group 7	Pb	Strong correlation
		Model 46	Group 7	Zn	Strong correlation
12	Venkatramanan et al. 2017	Model 47	Group 4	Al	Strong correlation
		Model 48	Group 4	Fe	Moderate correlation
		Model 49	Group 4	Mn	Weak correlation
		Model 50	Group 4	Sr	Weak correlation

Table 4 (continued)

Article no.	References	Model no.	Input variables type	Output variable	Performance
13	Alizamir and Sobhanardakani 2016	Model 51	Group 7	As	Strong correlation
		Model 52	Group 7	Pb	Strong correlation
		Model 53	Group 7	Zn	Strong correlation
14	Bayatzadeh Fard et al. 2017	Model 54	Group 2	Fe	Moderate correlation
		Model 55	Group 2	Mn	Moderate correlation
		Model 56	Group 2	Pb	Moderate correlation
		Model 57	Group 2	Zn	Moderate correlation
15	Ghadimi 2015	Model 58	Group 6	Pb	Moderate correlation
		Model 59	Group 6	Zn	Moderate correlation
		Model 60	Group 6	Cu	Moderate correlation
16	Shakeri et al. 2013	Model 61	Group 5	Cu	Strong correlation
17	Zare et al. 2011	Model 62	Group 1	NO_3	Strong correlation
		Model 63	Group 1	NO_3	Strong correlation
		Model 64	Group 1	NO_3	Strong correlation
		Model 65	Group 1	NO_3	Strong correlation
		Model 66	Group 1	NO_3	Strong correlation
		Model 67	Group 1	NO_3	Strong correlation
18*	Cho et al. 2011	Model 68	Group 5	As	-
19	Rooki et al. 2011	Model 69	Group 1	Cu	Strong correlation
		Model 70	Group 1	Fe	Weak correlation
		Model 71	Group 1	Mn	Strong correlation
		Model 72	Group 1	Zn	Strong correlation
20	Gholami et al. 2011	Model 73	Group 1	Fe	Strong correlation
		Model 74	Group 1	Ni	Strong correlation
21	Li et al. 2020	Model 75	Group 3	As	Moderate correlation
		Model 76	Group 3	As	Weak correlation
		Model 77	Group 7	As	Weak correlation
		Model 78	Group 1	Pb	Moderate correlation
		Model 79	Group 3	Pb	Moderate correlation
		Model 80	Group 7	Pb	Moderate correlation
		Model 81	Group 1	Zn	Strong correlation
		Model 82	Group 4	Zn	Moderate correlation
		Model 83	Group 7	Zn	Moderate correlation
22	Purkait et al. 2008	Model 84	Group 5	As	Moderate correlation
23	Yesilnacar et al. 2007	Model 85	Group 5	NO_3	Strong correlation
24	Diamantopoulou et al. 2005	Model 86	Group 1	NO_3	Strong correlation

* indicates that the model was not considered

Physical and chemical parameters only combined as input variables

A total of 22 models combined Group 2 parameters (P and C only) as input variables for prediction of PTEs in water resources using the MLP-NN algorithm. Based on the performance metrics used, 0/22, 6/22, and 16/22 of the models can be classified as having weak, moderate, and strong correlations, respectively (Table 7). The distribution of the model performance can be visualized in Fig. 3b. The distribution shows that 0%, 27.27%, and 72.73% of models that used Group 2 parameters as input variables for predicting PTEs in water resources had weak, moderate, and strong correlations, respectively. In summary, 100% of the models evaluated had acceptable model performance. Thus, it can be concluded that a combination of only physical and chemical parameters as input variables is suitable for predicting PTEs in water resources employing the MLP-NN algorithm.

Table 5 Models of PTEs in water resources produced using RBF-NN algorithm

Article no.	References	Model no.	Input variables type	Output variable	Performance
25	Kanj et al. 2022	Model 87	Group 1	Hg	Strong correlation
		Model 88	Group 1	Hg	Strong correlation
		Model 89	Group 2	Hg	Strong correlation
		Model 90	Group 2	Hg	Strong correlation
		Model 91	Group 5	Hg	Moderate correlation
		Model 92	Group 5	Hg	Moderate correlation
26	Ucun Ozel et al. 2020	Model 93	Group 5	Cu	Strong correlation
		Model 94	Group 5	Fe	Strong correlation
		Model 95	Group 5	Zn	Strong correlation
		Model 96	Group 5	Mn	Strong correlation
		Model 97	Group 5	Ni	Strong correlation
		Model 98	Group 5	Pb	Strong correlation
27	Zaqoot et al. 2018	Model 99	Group 3	NO_3	Strong correlation
		Model 100	Group 3	NO_3	Moderate correlation
		Model 101	Group 3	NO_3	Strong correlation
		Model 102	Group 5	NO_3	Moderate correlation
		Model 103	Group 5	NO_3	Moderate correlation
		Model 104	Group 5	NO_3	Moderate correlation
		Model 105	Group 5	NO_3	Moderate correlation
28	Alizamir and Sobhanardakani 2017	Model 106	Group 7	As	Strong correlation
		Model 107	Group 7	Zn	Strong correlation
		Model 108	Group 7	Pb	Strong correlation
29	Zaqoot et al. 2017	Model 109	Group 5	NO_3	Strong correlation
30	Alizamir et al. 2017	Model 110	Group 7	Cd	Strong correlation
		Model 111	Group 7	Pb	Strong correlation
		Model 112	Group 7	Zn	Strong correlation
31	Ehteshami et al. 2016	Model 113	Group 2	NO_3	Moderate correlation
32*	Suen and Eheart 2003	Model 114	Group 5	NO_3	-
*		Model 115	Group 5	NO_3	-

* indicates that the model was not considered

Physical parameters and metals only combined as input variables

A total of 4 models combined Group 3 parameters (P and M only) as input variables for prediction of PTEs in water resources using the MLP-NN algorithm. Based on the performance metrics used, 1/4, 2/4, and 1/4 of the models can be classified as having weak, moderate, and strong correlations, respectively (Table 7). The distribution of the model performance can be visualized in Fig. 3c. The distribution shows that 25%, 50%, and 25% of models that used Group 3 parameters as input variables for predicting PTEs in water resources had weak, moderate, and strong correlations, respectively. In summary, 75% of models that used Group 3 parameters as input variables had an acceptable model performance. Thus, it can be concluded that a combination of physical parameters and metals as input variables is suitable for predicting PTEs in water resources employing the MLP-NN algorithm.

Chemical parameters and metals only combined as input variables

A total of 6 models combined Group 4 parameters (C and M only) as input variables for prediction of PTEs in water resources using the MLP-NN algorithm. Based on the performance metrics used, 2/6, 2/6, and 2/6 of the models can be classified as having weak, moderate, and strong correlations, respectively (Table 7). The distribution of the model performance can be visualized in Fig. 3d. The distribution shows that 33.33%, 33.33%, and 33.33% of models that used Group 4 parameters as input variables for predicting PTEs in water resources had weak, moderate, and strong correlations, respectively. In summary, 66.66% of the models

Table 6 Models of PTEs in water resources produced using ANFIS algorithm

Article no.	References	Model no.	Input variables type	Output variable	Performance
33*	Abd El-Mageed et al. 2022	Model 116	Group 2	NO_3	-
34*	Agah and Soleimanpourmoghadam 2020	Model 117	Group 1	Cu	-
*		Model 118	Group 1	Fe	-
*		Model 119	Group 1	Mn	-
*		Model 120	Group 1	Zn	-
35	Elzain 2020	Model 121	Group 5	NO_3	Moderate correlation
36	Ucun Ozel et al. 2020	Model 122	Group 5	Cu	Strong correlation
		Model 123	Group 5	Fe	Strong correlation
		Model 124	Group 5	Zn	Strong correlation
		Model 125	Group 5	Mn	Strong correlation
		Model 126	Group 5	Ni	Strong correlation
		Model 127	Group 5	Pb	Strong correlation
37	Jebastina and Prince Arulraj 2018	Model 128	Group 3	NO_3	Moderate correlation
		Model 129	Group 3	NO_3	Moderate correlation
		Model 130	Group 3	NO_3	Moderate correlation
		Model 131	Group 3	NO_3	Strong correlation
		Model 132	Group 1	NO_3	Strong correlation
38	Sonmez et al. 2018	Model 133	Group 7	Cd	Strong correlation
39	Bayatzadeh Fard et al. 2017	Model 134	Group 2	Fe	Moderate correlation
		Model 135	Group 2	Mn	Strong correlation
		Model 136	Group 2	Pb	Strong correlation
		Model 137	Group 2	Zn	Strong correlation
40*	Chang et al. 2014	Model 138	Group 1	As	-
41	Valente et al. 2013	Model 139	Group 2	Fe	Strong correlation
		Model 140	Group 2	As	Strong correlation
		Model 141	Group 2	Cd	Strong correlation
		Model 142	Group 2	Zn	Strong correlation
		Model 143	Group 2	Cu	Strong correlation
		Model 144	Group 2	Mn	Strong correlation
42	Mousavi and Amiri 2012	Model 145	Group 1	NO_3	Strong correlation
		Model 146	Group 1	NO_3	Strong correlation
		Model 147	Group 2	NO_3	Strong correlation
		Model 148	Group 5	NO_3	Moderate correlation

* indicates that the model was not considered

Table 7 Groups of input variables and model performances in forecasting PTEs in water using the MLP-NN algorithm

Groups/model performances	Weak correlation	Moderate correlation	Strong correlation
Group 1	70	11, 78	1, 2, 7, 8, 9, 12, 62, 63, 64, 65, 66, 67, 69, 71, 72, 73, 74, 81, 86
Group 2	Nil	25, 33, 54, 55, 56, 57	3, 4, 20, 21, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32, 34, 35
Group 3	76	75, 79	10
Group 4	49, 50	48, 82	39, 47
Group 5	Nil	6, 84	5, 13, 14, 15, 16, 17, 18, 19, 43, 61, 85
Group 6	Nil	58, 59, 60	Nil
Group 7	77	80, 83	36, 37, 38, 40, 41, 42, 44, 45, 46, 51, 52, 53

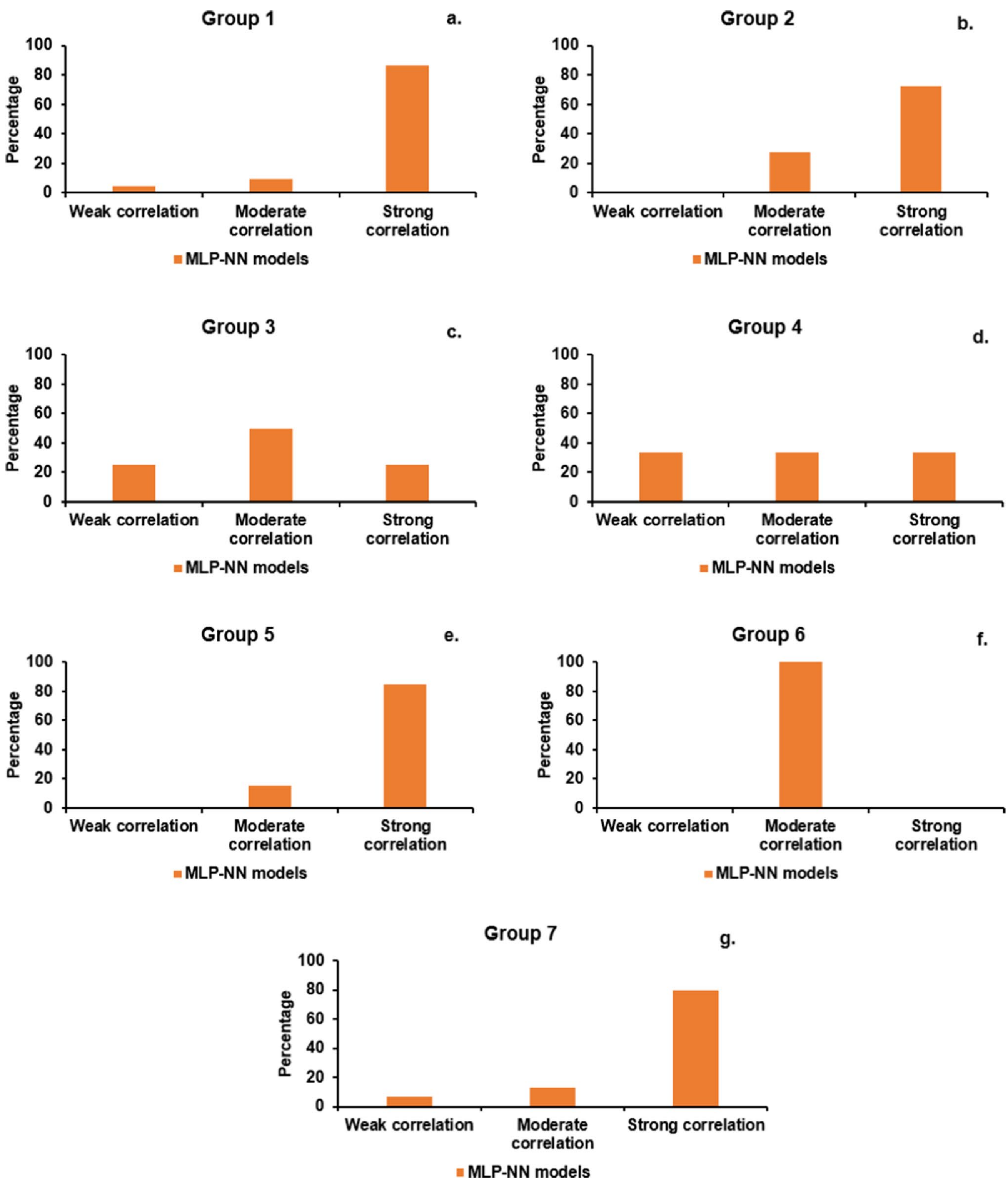

Fig. 3 Bar charts showing performance of MLP-NN models that predicted PTEs in water using the following as input variables: **a** Group 1, **b** Group 2, **c** Group 3, **d** Group 4, **e** Group 5, **f** Group 6, and **g** Group 7

evaluated had acceptable model performance. Thus, it can be concluded that a combination of chemical parameters and metals as input variables is suitable for predicting PTEs in water resources employing the MLP-NN algorithm.

Physical parameters only as input variables

A total of 13 models combined Group 5 parameters (P only) as input variables for prediction of PTEs in water resources using the MLP-NN algorithm. Based on the performance metrics used, 0/13, 2/13, and 11/13 of the models can be classified as having weak, moderate, and strong correlations, respectively (Table 7). The distribution of the model performance can be visualized in Fig. 3e. The distribution shows that 0%, 15.38%, and 84.62% of models that used Group 5 parameters as input variables for predicting PTEs in water resources had weak, moderate, and strong correlations, respectively. In summary, 100% of the models evaluated had acceptable model performance. Thus, it can be concluded that utilizing only physical parameters as input variables is suitable for predicting PTEs in water resources employing the MLP-NN algorithm.

Chemical parameters only as input variables

A total of 3 models combined Group 6 parameters (C only) as input variables for prediction of PTEs in water resources using the MLP-NN algorithm. Based on the performance metrics used, 0/3, 3/3, and 0/3 of the models can be classified as having weak, moderate, and strong correlations, respectively (Table 7). The distribution of the model performance can be visualized in Fig. 3f. The distribution shows that 0%, 100%, and 0% of models that used Group 6 parameters as input variables for predicting PTEs in water resources had weak, moderate, and strong correlations, respectively. In summary, 100% of the models evaluated had acceptable model performance. Thus, it can be concluded that utilizing only chemical parameters as input variables is suitable for predicting PTEs in water resources employing the MLP-NN algorithm.

Metals only as input variables

A total of 15 models combined Group 7 parameters (M only) as input variables for prediction of PTEs in water resources using the MLP-NN algorithm. Based on the performance metrics used, 1/15, 2/15, and 12/15 of the models can be classified as having weak, moderate, and strong correlations, respectively (Table 7). The distribution of the model performance can be visualized in Fig. 3g. The distribution shows that 6.67%, 13.33%, and 80% of models that used Group 7 parameters as input variables for predicting PTEs in water resources had weak, moderate, and strong correlations, respectively. In summary, 93.33% of the models evaluated had acceptable model performance. Thus, it can be concluded that utilizing only metals as input variables is suitable for predicting PTEs in water resources employing the MLP-NN algorithm.

RBF

Physical parameters, chemical parameters, and metals combined as input variables

A total of 2 models combined only physical parameters (P), chemical parameters (C), and metals as input variables for prediction of PTEs in water resources using the RBF-NN algorithm. Based on the performance metrics used, 0/2, 0/2, and 2/2 of the models can be classified as having weak, moderate, and strong correlations, respectively (Table 8). The distribution of the model performance can be visualized in Fig. 4a. The distribution shows that 0%, 0%, and 100% of models that combined P, C, and M parameters as input variables for predicting PTEs in water resources had weak, moderate, and strong correlations, respectively. In summary, 100% of the models evaluated had acceptable model performance. Thus, it can be concluded that a combination of only physical parameters, chemical parameters, and metals as input variables is suitable for predicting PTEs in water resources employing the RBF-NN algorithm.

Table 8 Groups of input variables and model performances in forecasting PTEs in water using the RBF-NN

Groups/model performance	Weak correlation	Moderate correlation	Strong correlation
Group 1	Nil	Nil	87, 88
Group 2	Nil	113	89, 90,
Group 3	Nil	100	99, 101
Group 4	-	-	-
Group 5	Nil	91, 92, 102, 103, 104, 105	93, 94, 95, 96, 97, 98, 109
Group 6	-	-	-
Group 7	Nil	Nil	106, 107, 108, 110, 111, 112

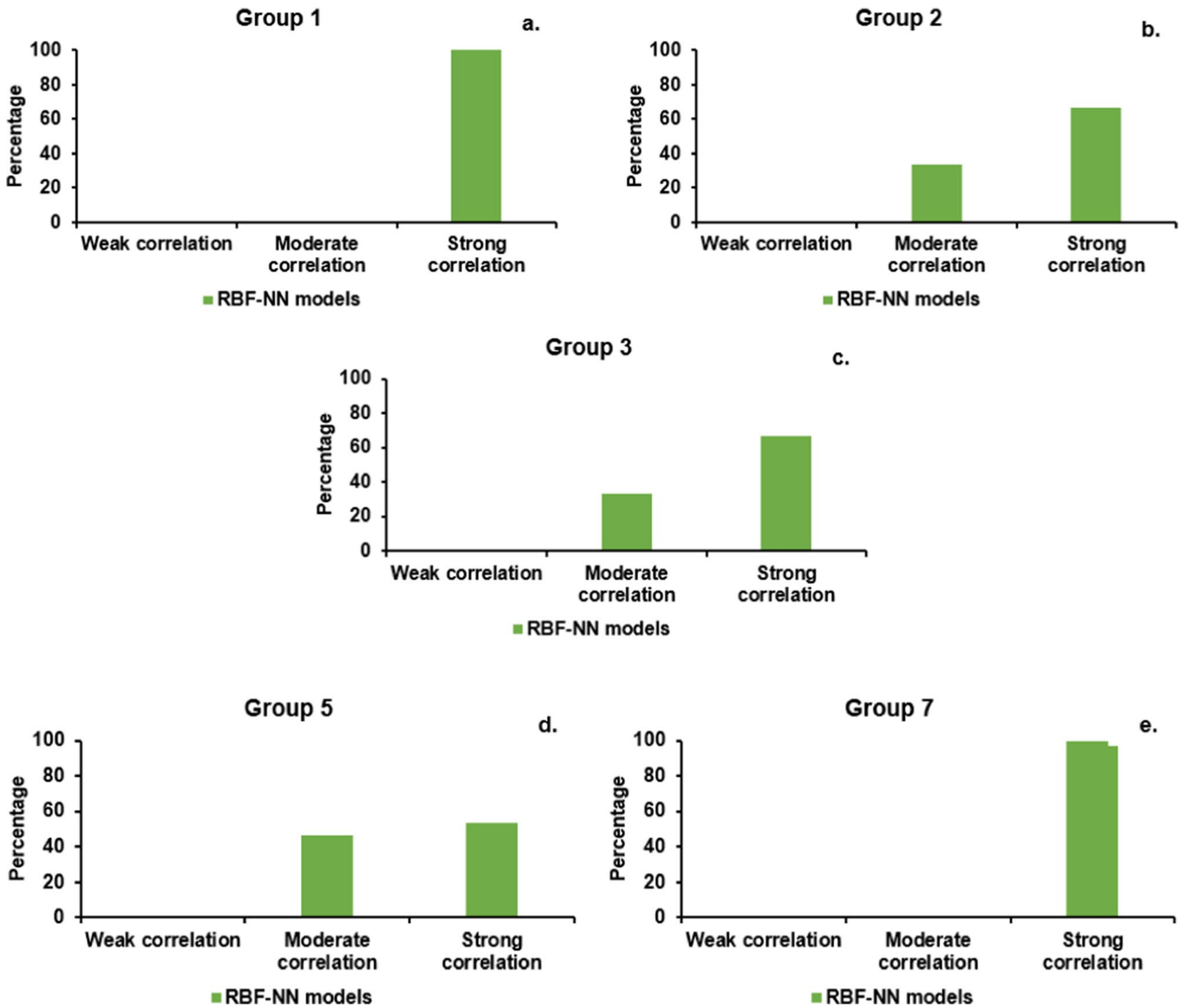

Fig. 4 Bar charts showing performance of RBF-NN models that predicted PTEs in water using the following as input variables: **a** Group 1, **b** Group 2, **c** Group 3, **d** Group 5, and **e** Group 7

Physical and chemical parameters only combined as input variables

A total of 3 models combined only physical and chemical parameters as input variables for the prediction of PTEs in water resources using the RBF-NN algorithm. Based on the performance metrics used, 0/3, 1/3, and 2/3 of the models can be classified as having weak, moderate, and strong correlations, respectively (Table 8). The distribution of the model performance can be visualized in Fig. 4b. The distribution shows that 0%, 33.33%, and 66.67% of models that combined only P and C parameters as input variables for predicting PTEs in water resources had weak, moderate, and strong correlations, respectively. In summary, 100% of the models evaluated had acceptable model performance. Thus, it can be concluded that a combination of only physical and chemical parameters as input variables is suitable for predicting PTEs in water resources employing the RBF-NN algorithm.

Physical parameters and metals only combined as input variables

A total of 3 models combined only physical parameters and metals as input variables for prediction of PTEs in water resources using the RBF-NN algorithm. Based on the performance metrics used, 0/3, 1/3, and 2/3 of the models can be classified as having weak, moderate, and strong correlations, respectively (Table 8). The distribution of the model performance can be visualized in Fig. 4c.

The distribution shows that 0%, 33.33%, and 66.67% of models that combined only P and M parameters as input variables for predicting PTEs in water resources had weak, moderate, and strong correlations, respectively. In summary, 100% of the models evaluated had acceptable model performance. Thus, it can be concluded that a combination of physical parameters and metals as input variables is suitable for predicting PTEs in water resources employing the RBF-NN algorithm.

Chemical parameters and metals only combined as input variables

Out of the 139 models considered in the review, none combined only chemical parameters and metals as input variables for prediction of PTEs in water resources using the RBF-NN algorithm (Table 8). Therefore, there is no conclusion with regard to the suitability of combining only chemical parameters and metals as input variables for predicting PTEs in water resources employing the RBF-NN algorithm.

Physical parameters only as input variables

A total of 13 models combined only physical parameters as input variables for the prediction of PTEs in water resources using the MLP-NN algorithm. Based on the performance metrics used, 0/13, 6/13, and 7/13 of the models can be classified as having weak, moderate, and strong correlations, respectively (Table 8). The distribution of the model performance can be visualized in Fig. 4d. The distribution shows that 0%, 46.15%, and 53.85% of models that combined only P parameters as input variables for predicting PTEs in water resources had weak, moderate, and strong correlations, respectively. In summary, 100% of the models evaluated had acceptable model performance. Thus, it can be concluded that utilizing only physical parameters as input variables is suitable for predicting PTEs in water resources employing the RBF-NN algorithm.

Chemical parameters only as input variables

Out of the 139 models considered in the review, none combined only chemical parameters as input variables for prediction of PTEs in water resources using the RBF-NN algorithm (Table 8). Therefore, there is no conclusion with regard to the suitability of combining only chemical parameters as input variables for predicting PTEs in water resources employing the RBF-NN algorithm.

Metals only as input variables

A total of 6 models combined only metals as input variables for prediction of PTEs in water resources using the MLP-NN algorithm. Based on the performance metrics used, 0/6, 0/6, and 6/6 of the models can be classified as having weak, moderate, and strong correlation, respectively (Table 8). The distribution of the model performance can be visualized in Fig. 4e. The distribution shows that 0%, 0%, and 100% of models that combined only M parameters as input variables for predicting PTEs in water resources had weak, moderate, and strong correlations, respectively. In summary, 100% of the models evaluated had acceptable model performance. Thus, it can be concluded that utilizing only metals as input variables is suitable for predicting PTEs in water resources employing the RBF-NN algorithm.

ANFIS

Physical parameters, chemical parameters, and metals combined as input variables

A total of 3 models combined only physical parameters (P), chemical parameters (C), and metals as input variables for prediction of PTEs in water resources using the ANFIS algorithm. Based on the performance metrics used, 0/3, 0/3, and 3/3 of the models can be classified as having

Table 9 Groups of input variables and model performances in forecasting PTEs in water using the ANFIS

Groups/model performance	Weak correlation	Moderate correlation	Strong correlation
Group 1	Nil	Nil	132, 145, 146
Group 2	Nil	134	135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 147
Group 3	Nil	128, 129, 130	131
Group 4	-	-	-
Group 5	Nil	121, 148	122, 123, 124, 125, 126, 127
Group 6	-	-	-
Group 7	Nil	Nil	133

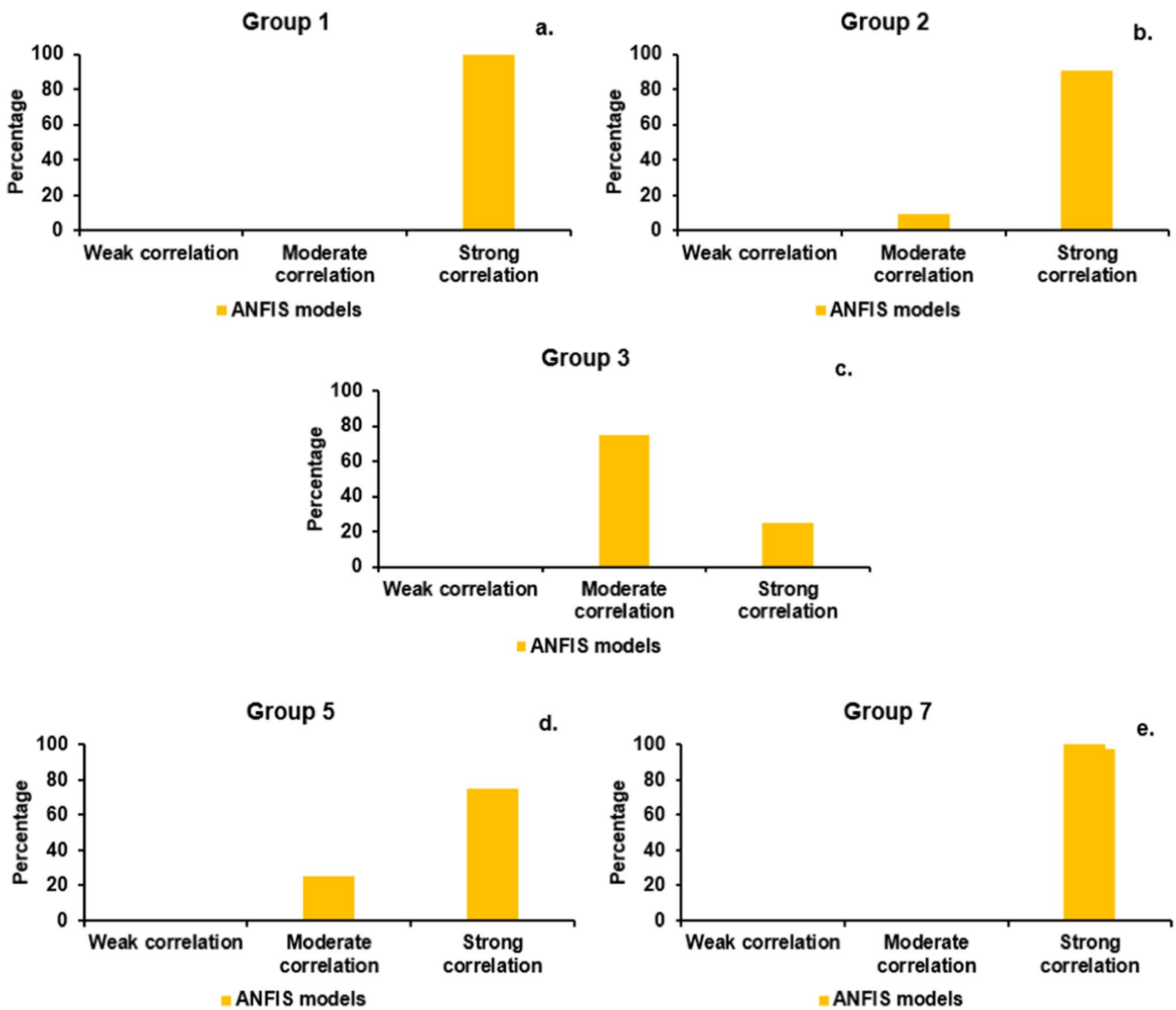

Fig. 5 Bar charts showing performance of ANFIS models that predicted PTEs in water using the following as input variables: **a** Group 1, **b** Group 2, **c** Group 3, **d** Group 5, and **e** Group 7

weak, moderate, and strong correlations, respectively (Table 9). The distribution of the model performance can be visualized in Fig. 5a. The distribution shows that 0%, 0%, and 100% of models that combined P, C, and M parameters as input variables for predicting PTEs in water resources had weak, moderate, and strong correlations, respectively. In summary, 100% of the models evaluated had acceptable model performance. Thus, it can be concluded that a combination of only physical parameters, chemical parameters, and metals as input variables is suitable for predicting PTEs in water resources employing the ANFIS algorithm.

Physical and chemical parameters only combined as input variables

A total of 11 models combined only physical and chemical parameters as input variables for the prediction of PTEs in water resources using the ANFIS algorithm. Based on the performance metrics used, 0/11, 1/11, and 10/11 of the models can be classified as having weak, moderate, and strong correlations, respectively (Table 9). The distribution of the model performance can be visualized in Fig. 5b. The distribution shows that 0%, 9.09%, and 90.91% of models that combined only P and C parameters as input variables for predicting PTEs in water resources had weak, moderate, and strong correlations, respectively. In summary, 100% of the models evaluated had acceptable model performance. Thus, it can be concluded that a combination of only physical and chemical parameters as input variables is suitable for predicting PTEs in water resources employing the ANFIS algorithm.

Physical parameters and metals only combined as input variables

A total of 4 models combined only physical parameters and metals as input variables for prediction of PTEs in water resources using the ANFIS algorithm. Based on the performance metrics used, 0/4, 3/4, and 1/4 of the models can be classified as having weak, moderate, and strong correlations, respectively (Table 9). The distribution of the model performance can be visualized in Fig. 5c. The distribution shows that 0%, 75%, and 25% of models that combined only P and M parameters as input variables for predicting PTEs in water resources had weak, moderate, and strong correlations, respectively. In summary, 100% of the models evaluated had acceptable model performance. Thus, it can be concluded that a combination of physical parameters and metals as input variables is suitable for predicting PTEs in water resources employing the ANFIS algorithm.

Chemical parameters and metals only combined as input variables

Out of the 139 models considered in the review, none combined only chemical parameters and metals as input variables for prediction of PTEs in water resources using the ANFIS algorithm (Table 9). Therefore, there is no conclusion with regard to the suitability of combining only chemical parameters and metals as input variables for predicting PTEs in water resources employing the ANFIS algorithm.

Physical parameters only as input variables

A total of 8 models combined only physical parameters as input variables for the prediction of PTEs in water resources using the ANFIS algorithm. Based on the performance metrics used, 0/8, 2/8, and 6/8 of the models can be classified as having weak, moderate, and strong correlations, respectively (Table 9). The distribution of the model performance can be visualized in Fig. 5d. The distribution shows that 0%, 25%, and 75% of models that combined only P parameters as input variables for predicting PTEs in water resources had weak, moderate, and strong correlations, respectively. In summary, 100% of the models evaluated had acceptable model performance. Thus, it can be concluded that utilizing only physical parameters as input variables is suitable for predicting PTEs in water resources employing the ANFIS algorithm.

Chemical parameters only as input variables

Out of the 139 models considered in the review, none combined only chemical parameters as input variables for prediction of PTEs in water resources using the ANFIS algorithm (Table 9). Therefore, there is no conclusion with regard to the suitability of combining only chemical parameters as input variables for predicting PTEs in water resources employing the ANFIS algorithm.

Metals only as input variables

One model combined only metals as input variables for prediction of PTEs in water resources using the ANFIS algorithm. Based on the performance metrics used, the model can be classified as having a strong correlation (Table 9). The distribution of the model performance can be visualized in Fig. 5e. The distribution shows that 100% of models that combined only M parameters as input variables for predicting PTEs in water resources had a strong correlation. Therefore, 100% of the models evaluated had acceptable model performance. Thus, it can be concluded that utilizing only metals as input variables is suitable for predicting PTEs in water resources employing the ANFIS algorithm.

Table 10 Overall performance of the MLP-NN models in forecasting PTEs in water

Groups/model performance	Weak correlation	Moderate correlation	Strong correlation
All groups	70, 76, 49, 50, 77	11, 78, 25, 33, 54, 55, 56, 57,75, 79, 48, 82, 6, 84, 58, 59, 60, 80, 83	1, 2, 7, 8, 9, 12, 62, 63, 64, 65, 66, 67, 69, 71, 72, 73, 74, 81, 86, 3, 4, 20, 21, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32, 34, 35, 10, 39, 47, 5, 13, 14, 15, 16, 17, 18, 19, 43, 61, 85, 36, 37, 38, 40, 41, 42, 44, 45, 46, 51, 52, 53
Synopsis	5/85	19/85	61/85
Percentage (%)	5.88%	21.35%	71.77%

Performances of MLP-NN, RBF-NN, and ANFIS in predicting PTEs in water

To understand the overall performance of the three algorithms (MLP-NN, RBF-NN, and ANFIS) in forecasting PTEs in water resources, we aggregated their performances across the seven groups of input variables. For the MLP-NN algorithm, 85 models were evaluated and 5.88%, 22.35%, and 71.77% of the models were classified into weak, moderate, and strong correlation, respectively (Table 10, Fig. 6a). Thus, 94.12% of the models that used the MLP-NN in forecasting PTEs in water resources had acceptable model performance (moderate-strong correlation). For the RBF-NN algorithm, 27 models were evaluated and 0%, 29.63%, and 70.37% of the models were classified into weak, moderate, and strong correlation, respectively (Table 11, Fig. 6b).

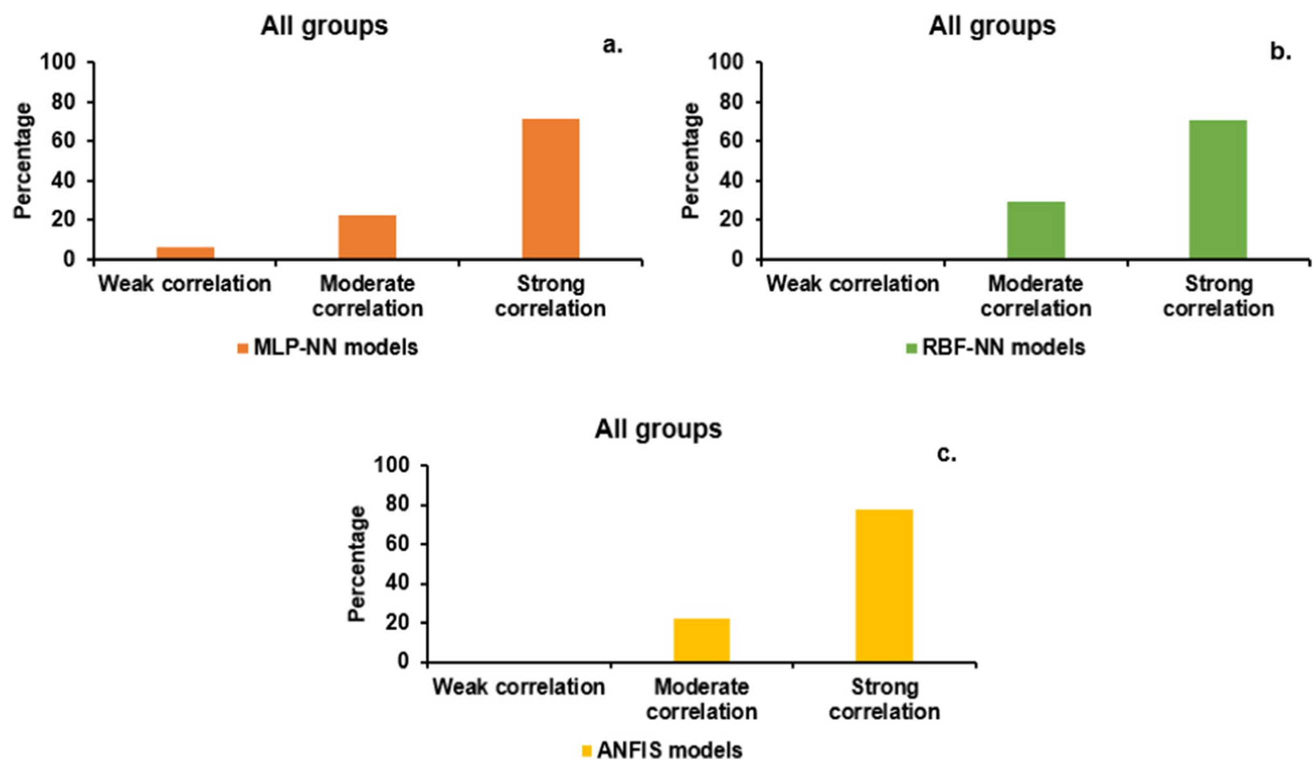

Fig. 6 Bar charts showing overall performance of models that predicted PTEs in water using **a** MLP-NN, **b** RBF-NN, and **c** ANFIS

Table 11 Overall performance of the RBF-NN models in forecasting PTEs in water

Groups/model performance	Weak correlation	Moderate correlation	Strong correlation
All groups	Nil	113, 100, 91, 92, 102, 103, 104, 105	87, 88, 89, 90, 99, 101, 93, 94, 95, 96, 97, 98, 109, 106, 107, 108, 110, 111, 112
Synopsis	0/27	8/27	19/27
Percentage (%)	0%	29.63%	70.37%

Table 12 Overall performance of the ANFIS models in forecasting PTEs in water

Groups/model performance	Weak correlation	Moderate correlation	Strong correlation
All groups	Nil	134, 128, 129, 130 121, 148	132, 145, 146, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 147, 131, 122, 123, 124, 125, 126, 127, 133
Synopsis	0/27	6/27	21/27
Percentage (%)	0%	22.22%	77.78%

Thus, 100% of the models that used the RBF-NN in forecasting PTEs in water resources had acceptable model performance. For the ANFIS algorithm, 27 models were evaluated, and 0%, 22.22%, and 77.78% of the models were classified into weak, moderate, and strong correlation, respectively (Table 12, Fig. 6c). Thus, 100% of the models that used ANFIS in forecasting PTEs in water resources had acceptable model performance. Based on the percentage of acceptable models of PTEs produced by the three algorithms, the RBF-NN and ANFIS clearly outperformed the MLP-NN. Nevertheless, the ANFIS algorithm had a higher percentage of strongly correlated models (77.78%), compared to 70.37% ascertained by the RBF-NN algorithm (Tables 10 and 12). Therefore, the overall performance of the three algorithms in forecasting PTEs in water resources can be rated as follows: ANFIS > RBF-NN > MLP-NN.

Conclusions

It is anticipated that the findings of the review study would ultimately help in protecting the lives of a wide range of human populations, especially those who are the most vulnerable to water pollution by PTEs. In addition, insights drawn from this study will aid cost-effective and efficient water quality monitoring, assessment, prediction, management, and sustainability. The current review study identified the most commonly used ANN algorithm, the most commonly predicted PTEs, the most commonly used input variables, the most suitable input variables, and the most efficient ANN algorithm for predicting PTEs in water. These were achieved after a careful evaluation of 139 models from 42 articles that employed the three ML algorithms (MLP-NN, RBF-NN, and ANFIS) to predict PTEs in water resources. Based on the findings, the following conclusions were made:

- The MLP-NN was the most commonly employed algorithm among the three analyzed ML algorithms.
- The most commonly predicted PTEs using the MLP-NN algorithm were Fe, Zn, and As. For the RBF-NN algorithm, they were NO_3, Zn, and Pb, and for the ANFIS, they were NO_3, Fe, and Mn.
- The input variables utilized by the three ML algorithms to predict PTEs in water resources can be grouped into seven: Group 1 consists of physical parameters (P), chemical parameters (C), and metals (M). Group 2 contains only P and C; Group 3 contains only P and M; Group 4 contains only C and M; Group 5 contains only P; Group 6 contains only C; and Group 7 contains only M.
- The MLP-NN algorithm used parameters in Group 1 most and Group 3 least as input variables for prediction of PTEs in water resources. The RBF-NN algorithm used parameters in Groups 5 and 7 most and least used those in Groups 4 and 6 as input variables for prediction of PTEs in water resources. The ANFIS algorithm used parameters in Group 1 most, while using Groups 3, 4, and 6 least as input variables for prediction of PTEs in water resources.
- For MLP-NN, RBF-NN, and ANFIS algorithms, the use of Groups 1, 2, 3, 5, and 7 parameters proved to be suitable input variables for forecasting PTEs in water resources. Thus, we encourage future research to make use of mentioned input variables to predict PTEs. However, the suitability of Groups 4 and 6 parameters using the RBF-NN and ANFIS algorithms could not be ascertained due to non-selection of the mentioned groups as input variables. Nevertheless, models that utilized Groups 4 and 6 parameters as input variables using the MLP-NN algorithm showed that they were suitable for the predictions.
- The overall order of performance of the three algorithms in predicting PTEs in water resources is ANFIS > RBF-NN > MLP-NN.
- The findings of the review agree with previous studies which suggest that MLP-NN, RBF-NN, and ANFIS are reliable algorithms for predicting PTES in water resources.

While key input variables have been identified for predicting PTEs in water bodies, it is crucial to recognize that regional peculiarities can significantly influence model performance. Therefore, in addition to the identified variables, dominant regional water quality influencers should be accounted for to enhance the robustness and applicability of the predictive models across diverse geographical contexts.

Recommendations and perspectives for future research

Based on the findings of the literature review and analysis, the following concerns and recommendations are provided to be addressed in future studies:

- The key challenge faced during the course of this study was extracting data from literature. Future studies should ensure model performances are summarized in a table for easier access.
- Future studies could consider the suitability of Group 4 (chemical and physical parameters only) and Group 6 (chemical parameters only) as input variables in modeling of PTEs in water resources using RBF-NN and ANFIS algorithms.

- Future studies could analyze the suitability and sensitivity of the specific (individual) input variables in the prediction of PTEs in water resources.
- Future studies could analyze the suitability of groups of input variables in the prediction of specific (individual) PTEs (i.e., the outputs) in water resources.
- Future studies could investigate new groups of input variables that can be used individually or combined to forecast PTEs in water resources using the ML algorithms.
- For regions where existing data is available, principal component analysis and correlation techniques can facilitate the selection of the most relevant input variables from the identified high-performing groups of predictors.
- Future studies could employ more advanced ML algorithms in forecasting PTEs in water resources and compare them with these neural network algorithms.

Availability of data and materials Not applicable

Author contribution Johnson C. Agbasi: Conceptualization, manuscript design, data computation, analysis and interpretation, manuscript writing, editing, review and revision. Johnbosco C. Egbueri: Conceptualization, manuscript design, supervision, manuscript editing, review and revision. Both authors read, commented on, and approved the final version.

Declarations

Ethical approval Not applicable

Consent to participate Not applicable

Consent for publication Not applicable

Competing interests The authors declare no competing interests.

References

Abba SI, Benaafi M, Usman AG, Aljundi IH (2023) Sandstone groundwater salinization modelling using physicochemical variables in Southern Saudi Arabia: application of novel data intelligent algorithms. Ain Shams Eng J 14(3):101894. https://doi.org/10.1016/j.asej.2022.101894

Abba SI, Usman AG, Danmaraya YA, Usman AG, Abdullahi HU (2020) Modeling of water treatment plant performance using artificial neural network: case study Tamburawa Kano-Nigeria. Dutse J Pure Appl Sci (DUJOPAS) 6(3):135–114

Abba SI, Yassin MA, Shah SMH, Egbueri JC, Elzain HE, Agbasi JC, Saini G, Usaman J, Khan NA, Aljundi IH (2024) Trace element pollution tracking in the complex multi-aquifer groundwater system of Al-Hassa oasis (Saudi Arabia) using spatial, chemometric and index-based techniques. Environ Res:118320. https://doi.org/10.1016/j.envres.2024.118320

Abbas H (2021) Removal of potentially toxic elements (PTEs) from contaminated water using microcrystalline cellulose extracted from rice Straw. Environ Biodiver Soil Security 5(2021):31–40

Abd El-Mageed AM, Enany TA, Goher ME, Hassouna ME (2022) Forecasting water quality parameters in Wadi El Rayan Upper Lake, Fayoum, Egypt using adaptive neuro-fuzzy inference system. The Egypt J Aqua Res 48(1):13–19. https://doi.org/10.1016/j.ejar.2021.10.001

Abu M, Akurugu BA, Egbueri JC (2024) Understanding groundwater mineralization controls and the implications on its quality (Southwestern Ghana): insights from hydrochemistry, multivariate statistics, and multi-linear regression models. Acta Geophys. https://doi.org/10.1007/s11600-023-01271-6

Abugu HO, Egbueri JC, Agbasi JC, Ezugwu AL, Omeka ME, Ucheana IA, Aralu CC (2024) Hydrochemical characterization of ground and surface water for irrigation application in nigeria: a review of progress. Chem Afr. https://doi.org/10.1007/s42250-024-00917-2

Adeoti OA, Osanaiye PA (2013) Effect of training algorithms on the performance of ANN for pattern recognition of bivariate process. Int J Comput Appl 69(2)

Agah A, Soleimanpourmoghadam N (2020) Design and implementation of heavy metal prediction in acid mine drainage using multi-output adaptive neuro-fuzzy inference systems (ANFIS)-a case study. Int J Mining Geo-Eng 54(1):59–64

Agbasi JC, Egbueri JC (2022) Assessment of PTEs in water resources by integrating HHRISK code, water quality indices, multivariate statistics, and ANNs. Geocarto Int. https://doi.org/10.1080/10106049.2022.2034990

Agbasi JC, Ezugwu AL, Omeka ME, Ucheana IA, Aralu CC, Abugu HO, Egbueri JC (2024) More about making profits or providing safe drinking water? A state-of-the-art review on sachet water contamination in Nigeria. J Environ Sci Health C:1–43. https://doi.org/10.1080/26896583.2024.2319009

Aggarwal R, Ranganathan P (2016) Common pitfalls in statistical analysis: the use of correlation techniques. Perspect Clin Res 7(4):187. https://doi.org/10.4103/2229-3485.192046

Alayat H, El Badaoui H, Abdallaoui A, Abrid D, El Hmaidi A (2018) Development of mathematical models for predicting the iron concentrations of lake Oubeira waters (ne Algerian). J Fundam Appl Sci 10(1):83–96. https://doi.org/10.4314/jfas.v10i1.6

Alipour M, Sarafraz M, Chavoshi H, Bay A, Nematollahi A, Sadani M, Fakhri Y, Vasseghian Y, Khaneghah AM (2021) The concentration and probabilistic risk assessment of potentially toxic elements in fillets of silver pomfret (Pampus argenteus): a global systematic review and meta-analysis. J Environ Sci 100:167–180

Alizamir M, Sobhanardakani S (2016) Forecasting of heavy metals concentration in groundwater resources of Asadabad plain using artificial neural network approach. J Adv Environ Health Res 4(2):68–77

Alizamir M, Sobhanardakani S (2017) A comparison of performance of artificial neural networks for prediction of heavy metals concentration in groundwater resources of toyserkan plain. Avicenna J Environ Health Eng 4(1):11792–11792. https://doi.org/10.5812/ajehe.11792

Alizamir M, Sobhanardakani S, Shahrabadi AH (2019) Prediction of heavy metals concentration in the groundwater resources in Razan Plain: extreme learning machine vs. artificial neural network and multivariate adaptive regression spline. Ann Military Health Sci Res 17(4). https://doi.org/10.5812/amh.98554

Alizamir M, Sobhanardakani S, Taghavi L (2017) Modeling of groundwater resources heavy metals concentration using soft computing methods: application of different types of artificial neural networks. J Chem Health Risks 7(3)

Allwright, S. (2022). What is a good MSE value? (Simply explained). https://stephenallwright.com/good-mse-value/#. Retrieved September 7, 2022.

Aryadoust V, Goh CC (2014) Predicting listening item difficulty with language complexity measures: a comparative data mining study. CaMLA Working Papers, 2014-01

Asgharnia A, Jamali A, Shahnazi R, Maheri A (2019) Load mitigation of a class of 5-MW wind türbine with RBF neural network based fractional-order PID controller. ISA Trans 96:272–286. https://doi.org/10.1016/j.isatra.2019.07.006

Ayejoto DA, Agbasi JC, Egbueri JC, Echefu K (2022) Assessment of oral and dermal health risk exposures associated with contaminated water resources: an update in Ojoto area, southeast Nigeria. Int J Environ Anal Chem. https://doi.org/10.1080/03067319.2021.2023515

Azrour M, Mabrouki J, Fattah G, Guezzaz A, Aziz F (2022) Machine learning algorithms for efficient water quality prediction. Model Earth Syst Environ 8(2):2793–2801

Bayatzadeh Fard Z, Ghadimi F, Fattahi H (2017) Use of artificial intelligence techniques to predict distribution of heavy metals in groundwater of Lakan lead-zinc mine in Iran. J Mining Environ 8(1):35–48. https://doi.org/10.22044/jme.2016.592

Bellatreche L, Ordonez C, Méry D, Golfarelli M (2022) The central role of data repositories and data models in data science and advanced analytics. Futur Gener Comput Syst 129:13–17

Bhatt S, Mishra AP, Chandra N, Sahu H, Chaurasia SK, Pande CB, Agbasi JC, Khan MYA, Abba SI, Egbueri JC, Hunt J (2024) Characterizing seasonal, environmental and human-induced factors influencing the dynamics of Rispana River's water quality: implications for sustainable river management. Res Eng:102007. https://doi.org/10.1016/j.rineng.2024.102007

Bini C, Wahsha M (2014) Potentially harmful elements and human health. In: PHEs, environment and human health. Springer, Dordrecht, pp 401–463

Boudaghpour S, Malekmohammadi S (2020) Modeling prediction of dispersal of heavy metals in plain using neural network. J Appl Water Eng Res 8(1):28–43. https://doi.org/10.1080/23249676.2020.1719219

Broomhead DS, Lowe D (1988) Multivariable functional interpolation and adaptive networks. Complex Syst 2:321–355

Brunton SL, Kutz JN (2022) Data-driven science and engineering: machine learning, dynamical systems, and control. Cambridge University Press

Burse K, Yadav RN, Shrivastava SC (2010) Channel equalization using neural networks: a review. IEEE Trans Syst Man Cybern Part C Appl Rev 40(3):352–357

Chaal RE, Aboutafail MO (2021) Development of stochastic mathematical models for the prediction of heavy metal content in surface waters using artificial neural network and multiple linear regression. In: E3S Web of Conferences, vol 314. EDP Sciences, p 02001. https://doi.org/10.1051/e3sconf/202131402001

Chang FJ, Chung CH, Chen PA, Liu CW, Coynel A, Vachaud G (2014) Assessment of arsenic concentration in stream water using neuro fuzzy networks with factor analysis. Sci Total Environ 494:202–210. https://doi.org/10.1016/j.scitotenv.2014.06.133

Chebrolu S, Abraham A, Thomas JP (2005) Feature deduction and ensemble design of intrusion detection systems. Comput Secur 24(4):295–307

Cho KH, Sthiannopkao S, Pachepsky YA, Kim KW, Kim JH (2011) Prediction of contamination potential of groundwater arsenic in Cambodia, Laos, and Thailand using artificial neural network. Water Res 45(17):5535–5544. https://doi.org/10.1016/j.watres.2011.08.010

Chojnacki A, Dai C, Farahi A, Shi G, Webb J, Zhang DT, Abernethy J, Schwartz E (2017) A data science approach to understanding residential water contamination in flint. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 1407–1416

Diamantopoulou MJ, Papamichail DM, Antonopoulos VZ (2005) The use of a Neural Network technique for the prediction of water quality parameters. Oper Res 5(1):115–125. https://doi.org/10.1007/BF02944165

Dogo EM, Afolabi OJ, Nwulu NI, Twala B, Aigbavboa CO (2018) A comparative analysis of gradient descent-based optimization algorithms on convolutional neural networks. In: In 2018 international conference on computational techniques, electronics and mechanical systems (CTEMS). IEEE, pp 92–99

Doshi S (2019) Various optimization algorithms for training neural network. Towards data science, p 13

Draper C, Reichle R, de Jeu R, Naeimi V, Parinussa R, Wagner W (2013) Estimating root mean square errors in remotely sensed soil moisture over continental scale domains. Remote Sens Environ 137:288–298. https://doi.org/10.1016/j.rse.2013.06.013

Egbueri JC (2019) Groundwater quality assessment using pollution index of groundwater (PIG), ecological risk index (ERI) and hierarchical cluster analysis (HCA): a case study. Groundw Sustain Dev. https://doi.org/10.1016/j.gsd.2019.100292

Egbueri JC (2020) Heavy metals pollution source identification and probabilistic health risk assessment of shallow groundwater in Onitsha, Nigeria. Anal Lett 53(10):1620–1638. https://doi.org/10.1080/00032719.2020.1712606

Egbueri JC (2021) Prediction modeling of potentially toxic elements' hydrogeopollution using an integrated Q–mode HCs and ANNs machine learning approach in SE Nigeria. Environ Sci Pollut Res 28(30):40938–40956. https://doi.org/10.1007/s11356-021-13678-z

Egbueri JC (2022) Predicting and analysing the quality of water resources for industrial purposes using integrated data-intelligent algorithms. Groundw Sustain Dev 18:100794. https://doi.org/10.1016/j.gsd.2022.100794

Egbueri JC, Agbasi JC (2022a) Combining data-intelligent algorithms for the assessment and predictive modeling of groundwater resources quality in parts of southeastern Nigeria. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-022-19818-3

Egbueri JC, Agbasi JC (2022b) Data-driven soft computing modeling of groundwater quality parameters in southeast Nigeria: comparing the performances of different algorithms. Environ Sci Pollut Res 29(25):38346–38373. https://doi.org/10.1007/s11356-022-18520-8

Egbueri JC, Agbasi JC (2022c) Performances of MLR, RBF-NN, and MLP-NN in the evaluation and prediction of water resources quality for irrigation purposes under two modeling scenarios. Geocarto Int. https://doi.org/10.1080/10106049.2022.2087758

Egbueri JC, Ayejoto DA, Agbasi JC (2022b) Pollution assessment and estimation of the percentages of toxic elements to be removed to make polluted drinking water safe: a case from Nigeria. Toxin Rev. https://doi.org/10.1080/15569543.2021.2025401

Egbueri JC, Unigwe CO, Agbasi JC, Nwazelibe VE (2022a) Indexical and artificial neural network modeling of the quality, corrosiveness, and encrustation potential of groundwater resources in industrialized metropolises, Southeast Nigeria. Environ Dev Sustain. https://doi.org/10.1007/s10668-022-02687-8

Egger R (2022) Machine learning in tourism: a brief overview. In: Applied data science in tourism: Interdisciplinary approaches, methodologies, and applications. Springer, pp 85–107

Ehteshami M, Farahani ND, Tavassoli S (2016) Simulation of nitrate contamination in groundwater using artificial neural networks. Mod Earth Sys Environ 2:28. https://doi.org/10.1007/s40808-016-0080-3

Eid EM, Dakhil MA, Hassan LM, Salama SG, Galal TM (2021) Uptake prediction of eight potentially toxic elements by Pistia stratiotes L. grown in the Al-Sero Drain (South Nile Delta, Egypt): A biomonitoring approach. Sustainability 13(9):5276

El Mrabet MA, El Makkaoui K, Faize A (2021) Supervised machine learning: a survey. In: In 2021 4th International Conference on Advanced Communication Technologies and Networking (CommNet). IEEE, pp 1–10

El-Safa MMA, Elsayed S, Elsherbiny O, Elmetwalli AH, Gad M, Moghanm FS, Eid EM, Taher MA, El-Morsy MH, Osman HE, Saleh AH (2022) Environmental assessment of potentially toxic elements using pollution indices and data-driven modeling in surface sediment of the Littoral Shelf of the Mediterranean Sea Coast and Gamasa Estuary, Egypt. J Marine Sci Eng 10(6):816

El-Shafie A, Abdin AE, Noureldin A, Taha MR (2009) Enhancing inflow forecasting model at Aswan high dam utilizing radial basis neural network and upstream monitoring stations measurements. Water Resour Manag 23(11):2289–2315

Elzain HEEO (2020) Improvement of groundwater contamination vulnerability assessment using the adaptive neuro-fuzzy inference system with metaheuristic optimization algorithms. Ph.D. thesis, Pukyong Nat'l Univ., Korea, p 120

Emenike PC, Nnaji CC, Tenebe IT, Agunwamba JC (2019) Hydrogeochemical imprints and spatio-temporal health risk assessment of lead in drinking water sources of Abeokuta, south-western Nigeria. Int J Environ Sci Technol. https://doi.org/10.1007/s13762-019-02506-0

Ewuzie U, Bolade OP, Egbedina AO (2022) Application of deep learning and machine learning methods in water quality modeling and prediction: a review. In: Current Trends and Advances in Computer-Aided Intelligent Environmental Data Engineering. Elsevier, pp 185–218

Farmaki EG, Thomaidis NS, Efstathiou CE (2010) Artificial neural networks in water analysis: theory and applications. Int J Environ Anal Chem 90(2):85–105. https://doi.org/10.1080/03067310903094511

Franzen M, Kloetzer L, Ponti M, Trojan J, Vicens J (2021) Machine learning in citizen science: promises and implications. The Sci Citizen Sci 183

Fural Ş, Kükrer S, Cürebal İ, Aykır D (2022) Ecological degradation and non-carcinogenic health risks of potential toxic elements: a GIS-based spatial analysis for Doğancı Dam (Turkey). Environ Monit Assess 194(4):1–18

Geetha R, Balasubramanian M, Devi KR (2022) COVIDetection: deep convolutional neural networks-based automatic detection of COVID-19 with chest x-ray images. Res Biomed Eng 38(3):955–964

Genchi G, Carocci A, Lauria G, Sinicropi MS, Catalano A (2020) Nickel: human health and environmental toxicology. Int J Environ Res Public Health 17:679

Ghadimi F (2015) Prediction of heavy metals contamination in the groundwater of Arak region using artificial neural network and multiple linear regression. J Tethys 3(3):203–215

Gholami R, Kamkar-Rouhani A, Doulati Ardejani F, Maleki S (2011) Prediction of toxic metals concentration using artificial intelligence techniques. Appl Water Sci 1(3):125–134

Gunjal BL (2022) Real-world applications of data science. Data Science: Techniques and Intelligent Applications, p 33

Haji SH, Abdulazeez AM (2021) Comparison of optimization techniques based on gradient descent algorithm: a review. PalArch's J Archaeol Egypt/Egyptol 18(4):2715–2743

Han S, Wang H, Tian Y, Christov N (2019) Time-delay estimation based computed torque control with robust adaptive RBF neural network compensator for a rehabilitation exoskeleton. ISA Trans 97:171–181. https://doi.org/10.1016/j.isatra.2019.07.030

Han H, Trimi S (2022) Towards a data science platform for improving SME collaboration through Industry 4.0 technologies. Technol Forecast Soc Chang 174:121242

Hasan N, Alam M (2021) Applications of big data analytics in supply-chain management. In: Big Data Analytics. Auerbach Publications, pp 173–199

Ighalo JO, Adeniyi AG, Adeniran JA, Ogunniyi S (2021) A systematic literature analysis of the nature and regional distribution of water pollution sources in Nigeria. J Clean Prod 283:124566

Illinois Department of Public Health (IDPH). 2010. Iron in drinking water. http://www.idph.state.il.us/envhealth/factsheets/ironFS.htm. Accessed 11th July 2022.

Jebastina N, Prince Arulraj G (2018) Spatial prediction of nitrate concentration using GIS and ANFIS modelling in groundwater. Bull Environ Contam Toxicol 101(3):403–409. https://doi.org/10.1007/s00128-018-2406-5

Jianfei C, Chunfang L, Lixia Z, Quanyuan W, Jianshu L (2020) Source apportionment of potentially toxic elements in soils using APCS/MLR, PMF and geostatistics in a typical industrial and mining city in Eastern China. PLoS One 15(9). https://doi.org/10.1371/journal.pone.0238513

Jiang Y, Li C, Sun L, Guo D, Zhang Y, Wang W (2021) A deep learning algorithm for multi-source data fusion to predict water quality of urban sewer networks. J Clean Prod 318:128533

Jorjani E, Chehreh Chelgani S, Mesroghli SH (2008) Application of artificial neural networks to predict chemical desulfurization of Tabas coal. Fuel 87:2727–2734

Kanj F, Sawaya R, Halwani J, Nehmeh N (2022) Mercury prediction in groundwater of Naameh landfill using an artificial neural network (ANN) model. Green Technol Resil Sustain 2(1):1–14

Karaca Y, Baleanu D (2022) Artificial neural network modeling of systems biology datasets fit based on Mittag-Leffler functions with heavy-tailed distributions for diagnostic and predictive precision medicine. In: Multi-Chaos, Fractal and Multi-fractional Artificial Intelligence of Different Complex Systems. Academic Press, pp 133–148

Karmakar B, Singh MK, Choudhary BK, Singh SK, Egbueri JC, Gautam S, Rawat SK (2021) Investigation of the hydrogeochemistry, groundwater quality and associated health risks in industrialized regions of Tripura, northeast In-dia. Environ Forensic. https://doi.org/10.1080/15275922.2021.2006363

Khalil A, Almasri MN, McKee M, Kaluarachchi JJ (2005) Applicability of statistical learning algorithms in groundwater quality modeling. Water Resour Res 41(5)

Khan U, Cook FJ, Laugesen R, Hasan MM, Plastow K, Amirthanathan GE, Bari MA, Tuteja NK (2020) Development of catchment water quality models within a realtime status and forecast system for the Great Barrier Reef. Environ Model Softw 132:104790. https://doi.org/10.1016/j.envsoft.2020.104790

Kouadri S, Pande CB, Panneerselvam B, Moharir KN, Elbeltagi A (2022) Prediction of irrigation groundwater quality parameters using ANN, LSTM, and MLR models. Environ Sci Pollut Res 29(14):21067–21091

Kumar B (2015) An encyclopedic overview of 'big data'analytics. Int J Appl Eng Res 10(3):5681–5705

Landín M, Rowe RC, York P (2009) Advantages of neurofuzzy logic against conventional experimental design and statistical analysis in studying and developing direct compression formulations. Eur J Pharm Sci 38(4):325–331

Lee C, Jung DE, Lee D, Kim KH, Do SL (2021) Prediction performance analysis of artificial neural network model by input variable combination for residential heating loads. Energies 14(3):756

Leonard KC, Hasan F, Sneddon HF, You F (2021) Can artificial intelligence and machine learning be used to accelerate sustainable chemistry and engineering? ACS Sustain Chem Eng 9(18):6126–6129

Li P, Hua P, Gui D, Niu J, Pei P, Zhang J, Krebs P (2020) A comparative analysis of artificial neural networks and wavelet hybrid approaches to long-term toxic heavy metal prediction. Sci Rep 10(1):1–15

Lu H, Li H, Liu T, Fan Y, Yuan Y, Xie M, Qian X (2019) Simulating heavy metal concentrations in an aquatic environment using artificial intelligence models and physicochemical indexes. Sci Total Environ 694:133591. https://doi.org/10.1016/j.scitotenv.2019.133591

Maier HR, Dandy GC (2000) Neural Networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications. Environ Model Softw 15:101–124

May RJ, Dandy GC, Maier HR, Nixon JB (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

McCarthy J, Minsky ML, Rochester N (1955) A proposal for the Dartmouth Summer Research Project on artificial intelligence. IBM Corporation, C.E, Shannon, Bell telephone Lab

McCulloch W, Pitts W (1943) A logical calculus of ideas immanent in nervous activity. Bull Math Biophys 5:115–133

Mgbenu CN, Egbueri JC (2019) The hydrogeochemical signatures, quality indices and health risk assessment of water resources in Umunya district, southeast Nigeria. Appl Water Sci 9(1):22. https://doi.org/10.1007/s13201-019-0900-5

Mhlongo S, Mativenga PT, Marnewick A (2018) Water quality in a mining and water-stressed region. J Clean Prod 171:446–456. https://doi.org/10.1016/j.jclepro.2017.10.030

Moghanm FS, El-Banna A, El-Esawi MA, Abdel-Daim MM, Mosa A, Abdelaal KA (2020) Genotoxic and anatomical deteriorations associated with potentially toxic elements accumulation in water hyacinth grown in drainage water resources. Sustainability 12(5):2147

Mousavi SF, Amiri MJ (2012) Modelling nitrate concentration of groundwater using adaptive neural-based fuzzy inference system. Soil Water Res 7(2):73–83

Nazar S, Yang J, Amin MN, Khan K, Ashraf M, Aslam F, Javed MF, Eldin SM (2023) Machine learning interpretable-prediction models to evaluate the slump and strength of fly ash-based geopolymer. J Mater Res Technol 24:100–124. https://doi.org/10.1016/j.jmrt.2023.02.180

Nicklow J, Reed P, Savic D, Dessalegne T, Harrell L, Chan-Hilton A, Karamouz M, Minsker B, Ostfeld A, Singh A, Zechman E (2010) State of the art for genetic algorithms and beyond in water resources planning and management. J Water Resour Plan Manag 136(4):412–432

Nourani V, Ejlali RG, Alami MT (2011) Spatiotemporal groundwater level forecasting in coastal aquifers by hybrid artificial neural network-geostatistics model: a case study. Environ Eng Sci 28(3):217–228

Olazaran M (1996) A sociological study of the official history of the perceptron's controversy. Soc Stud Sci 26(3):611–659

Omeka ME, Egbueri JC (2022) Hydrogeochemical assessment and health related risks due to toxic element ingestion and dermal contact within the Nnewi-Awka urban areas, Nigeria. Environ Geochem Health. https://doi.org/10.1007/s10653-022-01332-7

Pande CB, Egbueri JC, Costache R, Sidek LM, Wang Q, Alshehri F, Din NM, Gautam VK, Pal SC (2024) Predictive modeling of land surface temperature (LST) based on Landsat-8 satellite data and machine learning models for sustainable development. J Clean Prod 444:141035. https://doi.org/10.1016/j.jclepro.2024.141035

Powell MJD (1987) Radial basis functions for multivariable interpolation: a review. In: Mason JC, Cox MG (eds) Proc., IMA Conf. on Algorithms for Approximation. Oxford University Press, New York, pp 143–167

Praveena N, Vivekanandan K (2021) A review on deep neural network design and their applications. In: In 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), vol 1. IEEE, pp 1495–1501

Purkait B, Kadam SS, Das SK (2008) Application of artificial neural network model to study arsenic contamination in groundwater of Malda District, Eastern India. J Environ Inf 12(2)

Rajaee T, Khani S, Ravansalar M (2020) Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: a review. Chemom Intell Lab Syst 200:103978

Rashid A, Ayub M, Khan S, Ullah Z, Ali L, Gao X, Li C, El-Serehy H, Kaushik P, Rasool A (2022) Hydrogeochemical assessment of carcinogenic and non-carcinogenic health risks of potentially toxic elements in aquifers of the Hindukush ranges, Pakistan: insights from groundwater pollution indexing, GIS-based, and multivariate statistical approaches. Environ Sci Pollut Res 29(50):75744–75768

Ricolfi L, Barbieri M, Muteto PV, Nigro A, Sappa G, Vitale S (2020) Potential toxic elements in groundwater and their health risk assessment in drinking water of Limpopo National Park, Gaza Province, Southern Mozambique. Environ Geochem Health 42(9):2733–2745

Rooki R, Doulati Ardejani F, Aryafar A, Bani Asadi A (2011) Prediction of heavy metals in acid mine drainage using artificial neural network from the Shur River of the Sarcheshmeh porphyry copper mine, Southeast Iran. Environ Earth Sci 64(5):1303–1316

Ruder, S. (2016). An overview of gradient descent optimization algorithms. https://arxiv.org/abs/1609.04747.

Sabharwal R, Miah SJ (2021) A new theoretical understanding of big data analytics capabilities in organizations: a thematic analysis. J Big Data 8(1):1–17

Saleh AH, Elsayed S, Gad M, Elmetwalli AH, Elsherbiny O, Hussein H, Moghanm FS, Qazaq AS, Eid EM, El-Kholy AS, Taher MA (2022) Utilization of pollution indices, hyperspectral reflectance indices, and data-driven multivariate modelling to assess the bottom sediment quality of Lake Qaroun, Egypt. Water 14(6):890

Sara U, Akter M, Uddin MS (2019) Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study. J Comput Commun 7(3):8–18. https://doi.org/10.4236/jcc.2019.73002

Sarker IH (2021) Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective. SN Comput Sci 2(5):1–22

Shah SIH, Peristeras V, Magnisalis I (2021) Government big data ecosystem: definitions, types of data, actors, and roles and the impact in public administrations. ACM J Data Inform Qual 13(2):1–25

Shahin MA, Jaksa MB, Maier HR (2001) Artificial neural network applications in geotechnical engineering. Aust Geomech 36(1):49–62

Shakeri AA, Gholamalizadeh AA, Soltani J (2013) Artificial neural network (ANN) approach for predicting Cu concentration in drinking water of Chahnimeh1 reservoir in Sistan-Balochistan. Iran Health Scope 2(1):31–38

Shi Y (2022) Advances in big data analytics: theory, algorithms and practices. Springer Nature

Singh KP, Basant A, Malik A, Jain G (2009) Artificial neural network modeling of the river water quality—a case study. Ecol Model 220(6):888–895

Singha S, Pasupuleti S, Singha SS, Singh R, Kumar S (2021) Prediction of groundwater quality using efficient machine learning technique. Chemosphere 276:130265

Sit M, Demiray BZ, Xiang Z, Ewing GJ, Sermet Y, Demir I (2020) A comprehensive review of deep learning applications in hydrology and water resources. Water Sci Technol 82(12):2635–2670

Sonmez AY, Kale S, Ozdemir RC, Kadak AE (2018) An adaptive neuro-fuzzy inference system (ANFIS) to predict of cadmium (Cd) concentrations in the Filyos River, Turkey. Turk J Fish Aquat Sci 18(12):1333–1343. https://doi.org/10.4194/1303-2712-v18_12_01

Suen JP, Eheart JW (2003) Evaluation of neural networks for modeling nitrate concentrations in rivers. J Water Resour Plan Manag 129(6):505–510

Team, G. L. (2020). Types of neural networks and definition of neural network. https://www.mygreatlearning.com/blog/types-of-neural-networks/. Accessed 11 July 2022.

Than NH, Ly CD, Van Tat P (2021) The performance of classification and forecasting Dong Nai River water quality for sustainable water resources management using neural network techniques. J Hydrol 596:126099

Ucun Ozel H, Gemici BT, Gemici E, Ozel HB, Cetin M, Sevik H (2020) Application of artificial neural networks to predict the heavy metal contamination in the Bartin River. Environ Sci Pollut Res 27(34):42495–42512. https://doi.org/10.1007/s11356-020-10156-w

Ukah BU, Egbueri JC, Unigwe CO, Ubido OE (2019) Extent of heavy metals pollution and health risk assessment of groundwater in a densely populated industrial area, Lagos, Nigeria. Int J Energy Water Resour 3(4):291–303. https://doi.org/10.1007/s42108-019-00039-3

Unigwe CO, Egbueri JC, Omeka ME (2022) Geospatial and statistical approaches to nitrate health risk and groundwater quality assessment of an alluvial aquifer in SE Nigeria for drinking and irrigation purposes. J Indian Chem Soc 99:100479. https://doi.org/10.1016/j.jics.2022.100479

Usman J, Salami BA, Gbadamosi A, Adamu H, Usman AG, Benaafi M, Abba SI, Othman MHD, I. H., & Aljundi, I. H. (2023) Intelligent optimization for modelling superhydrophobic ceramic membrane oil flux and oil-water separation efficiency: evidence from wastewater treatment and experimental laboratory. Chemosphere 331:138726. https://doi.org/10.1016/j.chemosphere.2023.138726

Valente T, Ferreira MJ, Grande JA, de la Torre ML, Borrego J (2013) pH, electric conductivity and sulfate as base parameters to estimate the concentration of metals in AMD using a fuzzy inference system. J Geochem Explor 124:22–28. https://doi.org/10.1016/j.gexplo.2012.07.013

Venkatramanan S, Chung SY, Selvam S, Son JH, Kim YJ (2017) Interrelationship between geochemical elements of sediment and groundwater at Samrak Park Delta of Nakdong River Basin in Korea: multivariate statistical analyses and artificial neural network approaches. Environ Earth Sci 76(13):1–12. https://doi.org/10.1007/S12665-017-6795-2

Wagh V, Panaskar D, Muley A, Mukate S, Gaikwad S (2018) Neural network modelling for nitrate concentration in groundwater of Kadava River basin, Nashik, Maharashtra, India. Groundw Sustain Dev 7:436–445. https://doi.org/10.1016/j.gsd.2017.12.012

Wagh VM, Panaskar DB, Muley AA (2017) Estimation of nitrate concentration in groundwater of Kadava river basin-Nashik district, Maharashtra, India by using artificial neural network model. Model Earth Syst Environ 3:36. https://doi.org/10.1007/s40808-017-0290-3

Wan D, Zhan C, Yang G, Liu X, Yang J (2016) Preliminary assessment of health risks of potentially toxic elements in settled dust over Beijing urban area. Int J Environ Res Public Health 13(5):491

Yasmin M (2019) Big data analytics capabilities and firm performance: an MCDM approach. (Master's thesis, İbn Haldun Üniversitesi, Lisansüstü Eğitim Enstitüsü)

Yassin MA, Tawabini B, Al-Shaibani A, Adetoro JA, Benaafi M, Al-Areeq AM, Usman AG, Abba SI (2022) Geochemical and spatial distribution of topsoil HMs coupled with modeling of Cr using chemometrics intelligent techniques: case study from Dammam Area, Saudi Arabia. Molecules 27(13):4220. https://doi.org/10.3390/molecules27134220

Yassin MA, Usman AG, Abba SI, Ozsahin DU, Aljundi IH (2023) Intelligent learning algorithms integrated with feature engineering for sustainable groundwater salinization modelling: Eastern Province of Saudi Arabia. Res Eng 20:101434. https://doi.org/10.1016/j.rineng.2023.101434

Yesilnacar MI, Sahinkaya E, Naz M, Ozkaya B (2007) Neural network prediction of nitrate in groundwater of Harran Plain, Turkey. Environ Geol 56(1):19–25. https://doi.org/10.1007/S00254-007-1136-5

Zaqoot HA, Hamada M, Miqdad S (2018) A comparative study of Ann for predicting nitrate concentration in groundwater wells in the southern area of Gaza Strip. Appl Artif Intell 32(7-8):727–744. https://doi.org/10.1080/08839514.2018.1506970

Zaqoot H, Aish A, Abdeljawad S (2017) Application of artificial neural networks for predicting water quality. LAMBERT Academic Publishing

Zare A, Bayat V, Daneshkare A (2011) Forecasting nitrate concentration in groundwater using artificial neural network and linear regression models. Int Agrophys 25(2)

Zhang L, Yang Z, Peng M, Cheng X (2022) Contamination levels and the ecological and human health risks of potentially toxic elements (PTEs) in soil of Baoshan Area, Southwest China. Appl Sci 12(3):1693

Zhang W, Wu C, Li Y, Wang L, Samui P (2021) Assessment of pile drivability using random forest regression and multivariate adaptive regression splines. Georisk 15(1):27–40

Zhang W, Zhang R, Wu C, Goh ATC, Lacasse S, Liu Z, Liu H (2020) State-of-the-art review of soft computing applications in underground excavations. Geosci Front 11(4):1095–1106

Zounemat-Kermani M, Batelaan O, Fadaee M, Hinkelmann R (2021) Ensemble machine learning paradigms in hydrology: a review. J Hydrol 598:126266

Zur RM, Jiang Y, Pesce LL, Drukker K (2009) Noise injection for training artificial neural networks: a comparison with weight decay and early stopping. Med Phys 36(10):4810–4818