Environmental Science and Pollution Research (2022) 29:38346–38373
https://doi.org/10.1007/s11356-022-18520-8

RESEARCH ARTICLE

Data-driven soft computing modeling of groundwater quality parameters in southeast Nigeria: comparing the performances of different algorithms

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

Received: 19 October 2021 / Accepted: 1 January 2022 / Published online: 25 January 2022

Abstract
In recent decades, the simulation and modeling of water quality parameters have been useful for monitoring and assessment of the quality of water resources. Moreover, the use of multiple modeling techniques, rather than a standalone model, tends to provide more robust and reliable insights. In this present paper, several soft computing techniques were integrated and compared for the modeling of groundwater quality parameters (pH, electrical conductivity (EC), total dissolved solids (TDS), total hardness (TH), modified heavy metal index (MHMI), pollution load index (PLI), and synthetic pollution index (SPI)) in Ojoto area, SE Nigeria. Standard methods were employed in the physicochemical analysis of the groundwater resources. It was found that anthropogenic and non-anthropogenic activities influenced the concentrations of the water quality parameters. The PLI, MHMI, and SPI revealed that about 20–25% of the groundwater samples are unsuitable for drinking. Simple linear regression indicated that strong agreements exist between the results of the water quality indices. Principal component and Varimax-rotated factor analyses showed that Pb, Ni, and Zn influenced the judgment of the water quality indices most. Q-mode hierarchical and K-means clustering algorithms grouped the water samples based on their pH, EC, TDS, TH, MHMI, PLI, and SPI values. Multiple linear regression (MLR) and artificial neural network (ANN) algorithms were used for the simulation and prediction of the pH, EC, TDS, TH, PLI, MHMI, and SPI. The MLR performed better than the ANN model in predicting EC, TH, and TDS. Nevertheless, the ANN model predicted the pH better than the MLR model. Meanwhile, both MLR and ANN performed equally in the prediction of PLI, MHMI, and SPI.

Keywords Artificial neural network · Factor analysis · Groundwater quality modeling · K-means clustering · Multiple linear regression · Q-mode hierarchical modeling

Introduction

Water is a resource that living things cannot do without, just like air. Over 70% of the earth's surface is covered with water; however, only a little portion of the earth's water resources is easily accessible to humans. Surface water and groundwater are the primary sources of water for living things. The quality of water resources is a crucial influencer of environmental and human health and the total well-being of ecosystems. Also, the sustainable development of any nation substantially relies on the quality of groundwater supply. Groundwater quality is defined by its physical, chemical, and biological properties (Orouji et al. 2013; Egbueri 2018). In recent times, numerous studies from different regions of the world have revealed that groundwater resources are rapidly being exposed to several point and nonpoint pollution sources (Ghobadi et al. 2021; Alizamir et al. 2019; Alizamir and Sobhanardakani 2017a). Such scenario leads to the depreciation of groundwater quality and quantity, globally. Potentially toxic elements (PTEs) have been identified as a threat to good water quality and thus have attracted the interest of many environmental scientists (Papazotos 2021; Pourret et al. 2021; Pourret and Hursthouse 2019). PTEs have also been linked to the cause of some detrimental health challenges to living things (Sobhanardakani 2019; Sobhanardakani et al. 2018; Başyiğit and Tekin-Özan 2013; Sobhanardakani 2017; Rezaei Raja

Responsible Editor: Xianliang Yi

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

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

et al. 2016). Additionally, this suggests that there is a continuous increase in environmental degradation across many regions. Thus, it is important to continuously assess the quality of this essential resource.

Due to the fact that groundwater is a primary resource that largely supports the existence of mankind and other living things, its preservation for future generations (Selvaraj et al. 2020) requires collaborative efforts of researchers, local inhabitants, policymakers, governments, and nongovernmental organizations. This would ensure that long-term environmental and socioeconomic sustainability is provided (El-Sayed et al. 2017). The development of preservation strategies for water resources often starts with adequate monitoring and assessment of water quality in an area. There are several physicochemical indicators usually employed for the determination of groundwater quality. They include pH, electrical conductivity (EC), total dissolved solids (TDS), total hardness (TH), chemical ions (the cations and anions), and potentially toxic elements. The concentration of these parameters is analyzed in water since their excessive quantity could affect water quality and deteriorate the environment and anthropogenic norms (Bai et al. 2011; Saravanakumar and Kumar 2011; Barzegar et al. 2017; Rupakheti et al. 2017; Kükrer and Mutlu 2019; Egbueri et al. 2021b, c; Gad et al. 2021). Interestingly, several water quality indices (such as pollution index of groundwater (PIG), modified heavy metal index (MHMI), water quality index (WQI), pollution load index (PLI), and synthetic pollution index (SPI)) have been utilized to summarily present and interpret water quality (Subba Rao 2012; Egbueri and Unigwe 2019; Solangi et al. 2019; Egbueri et al. 2020, 2021a, b, c). It is fundamental to mention that water quality indices have been utilized across the globe to simulate analyzed physicochemical parameters and generate single whole numbers that are representatives of the actual water quality. Hence, these water quality indices are a function of the measured physicochemical quality variables considered in the evaluation (Roy et al. 2017). Moreover, the various water quality indices have their defined scales or classification schemes for clear interpretation of water quality (Tyagi et al. 2013; Mohammadpour et al. 2016; Roy et al. 2017; Setshedi et al. 2021; Egbueri et al. 2021c).

Furthermore, statistical and machine learning techniques (such as correlation analysis, principal component analysis (PCA), factor analysis (FA), hierarchical cluster analysis (HCA), K-means clustering (KMC), multiple linear regression (MLR), and artificial neural networks (ANNs)) have also been globally used for simulating and predicting water quality parameters and the possible sources of pollution. Over time, these soft computing modeling techniques have been shown to be efficient and reliable means of simulating and predicting the spatial and temporal variations of water quality indicators. As a matter of fact, many environmental and water quality problems have been solved using these statistical and machine learning techniques (Goyal et al. 2015; Swathi and Lokeshappa 2015; Chatterjee et al. 2017; Hong et al. 2018; Roy and Majumder 2018a, b; Shamshirband et al. 2019; Kargar et al. 2020; Chen et al. 2020; Pham et al. 2020; Egbueri 2021; Senapati et al. 2021; Alizamir and Sobhanardakani 2016). Furthermore, these data-driven intelligent computing techniques are considerably less expensive, less tedious, and time-saving more than the conventional water quality monitoring and assessment programs (Maier et al. 2004; Orouji et al. 2013; Maroufpoor et al. 2020; Egbueri 2021). Often, water quality monitoring and assessment require large numbers of physicochemical parameters (Belkhiri et al. 2018), and that usually attract high financial costs and human labor. While the sampling size is influenced by funding, the choice of physicochemical parameters to be analyzed is usually influenced by both funding and research expert discretion. However, where large-scale water sampling and analysis are not feasible, due to high cost and inadequate facilities (especially in underdeveloped and developing countries), the soft computational algorithms become reliable alternatives for water quality monitoring, assessment, simulation, and prediction (Luo et al. 2003; Hameed et al. 2016; Mahmoudi et al. 2016; Yaseen et al. 2018; Najafzadeh and Ghaemi 2019; Maroufpoor et al. 2020; Egbueri 2021; Shah et al. 2021).

Numerous applications of statistical and machine learning predictive modeling of water quality parameters have been reported. For example, the prediction of total hardness of water resources reported by Roy and Majumder (2018a) could help managers in decision-making regarding the desired level of water hardness to maintain during industrial processing. Their research could also help the industrial managers in protecting industrial water heaters from scale formation and minimizing the risk of bridging process flow. Moreover, such predictions could also save the managers the cost of installing pricy hardness monitoring devices (Roy and Majumder 2018a). Furthermore, Pan et al. (2019) utilized integrated machine learning techniques in predicting the TDS and other parameters of a groundwater system in Canada. Excess TDS in water leads to esthetic problems (Sun et al. 2021). TDS in water often correlates well with water salinity, EC, and TH (Roy and Majumder 2018a, b). High salinity has been reported to pose negative effects on the use of water for drinking and irrigation purposes (Egbueri et al. 2021c). Machine learning prediction of TDS and salinity could assist policymakers and water managers in planning and management of water quality for drinking and irrigation purposes (Asadollahfardi et al. 2012; Pan et al. 2019), especially in areas exposed to multiple pollutants (Sun et al. 2021). Several other studies have estimated and predicted some chemical species (including anions, cations, and potentially toxic metals) and water quality indices using

machine learning approaches (Kadam et al. 2019; Ozel et al. 2020; Egbueri 2021; Alizamir et al. 2017; Alizamir and Sobhanardakani 2016, 2017b, 2018). Although statistical and machine learning approaches are known to be reliable, they also have some drawbacks: (1) they require expert knowledge, (2) they can be time-consuming, (3) they require careful handling, as inputting of a single wrong dataset could be costly, and lead to misinformation.

Unfortunately, in Nigeria, not many research works have been carried out to estimate and predict water quality using machine learning methods. Due to unplanned and poorly regulated socioeconomic activities, available water resources are exposed to contamination and pollution. Although this scenario seems to have led to a decline in water quality, there also seems to be inconsistent and unsystematic measurements of the rate and extent of pollution. However, multivariate statistical tools, such as correlation analysis, PCA, FA, and HCA, have commonly been used across Nigeria for pollution source identification. Meanwhile, the use of MLR and ANN (or other advanced methods) has not gotten wide application across the country. Nevertheless, Egbueri (2021) recently estimated and predicted some anions (including SO_4, NO_3, and Cl) and potentially toxic elements (including Zn, Fe, Cr, Ni, and Pb) in Ojoto region (southeastern Nigeria) using only ANNs. Although these predictions have been made in this region, more works are encouraged, as such predictive modeling studies are scarce in Nigeria. Importantly, there is a need to expand the prediction studies of water quality parameters in Nigeria using soft computing techniques.

Therefore, this paper targets to develop, integrate, and compare the efficacies of several machine learning models (algorithms) for predicting groundwater quality parameters (such as pH, EC, TDS, and TH) of Ojoto and its surroundings. Most previous studies in other parts of the world, which predicted these quality parameters, were conducted on surface waters, with lesser attention on groundwater. Additionally, water quality indices (namely, PLI, MHMI, and SPI) were simulated and predicted using machine learning algorithms too. Thus, in this paper, several physicochemical variables were used to predict the listed groundwater quality indicators. The specific research contributions of this paper are the following: (1) assessment of the physicochemical characteristics of the Ojoto groundwater using robust approach; (2) evaluation of the extent of pollution and drinking suitability of the groundwater resources using multiple numerical models (PLI, MHMI, and SPI); (3) categorization of the groundwater quality based on K-means partitional and Q-mode hierarchical clustering algorithms; (4) investigation of possible contamination sources and influencers using PCA and FA; (5) computation and prediction of the groundwater pH, EC, TDS, TH, PLI, MHMI, and SPI using MLR and ANN algorithms; and (6) comparative analysis of the performances of the soft computing models used herein. The use of multiple assessment and modeling approaches in this paper was to minimize the possible biases that arise from the use of a single, standalone model (Roy and Majumder 2019; Egbueri et al. 2020). This paves way for a more comprehensive and realistic assessment and modeling of groundwater quality. To the best of the authors' search and knowledge, this paper is the first attempt to predict these groundwater quality parameters in Nigeria using these machine learning methods. Moreover, it is the first to simulate and predict MHMI, PLI, and SPI globally. So, it is believed that this work would fill significant knowledge gaps in water research literature regarding soft computing simulations of the abovementioned parameters. Its findings could provide baseline information for future water quality simulation and prediction, locally and internationally, based on the studied parameters.

Materials and methods

Case study description

The study is conducted in a tropical region and is located within longitudes 06°50′E–07°00′E and latitudes 06°00′N–06°05′N. The present study area includes Ojoto and its environs in the southeastern part of Nigeria (Fig. 1). The study area is a suburb where commercial, industrial, agricultural, and residential activities that require water resources thrive. As a matter of fact, due to the increasing socioeconomic activities going on in this area, the human population and demand for high-quality surface and groundwater supplies by the inhabitants are also on the rise. Unfortunately, the study area currently suffers high-volume waste generation per capita; unregulated waste disposal; irregularities in commercial, agricultural, and industrial operations; and poor water resources planning and management strategies. These anthropogenic activities and shortcomings seem to heighten the environmental challenges that expose the available natural water resources to contamination and pollution threats (Egbueri et al. 2019). Due to the continual rise in human population in this area, the exploitation of the available water resources could lead to problems such as over-abstraction consequences and pollution effects (Avci et al. 2018; Chen et al. 2018).

Being predominantly characterized by tropical climate, the area experiences two main atmospheric seasons, annually—the rainy season and the dry season. While the former is noticed to span for a longer timeframe (April–November), the latter usually lasts for a shorter timeframe (December–March). The annual rainfall (precipitation) in this region

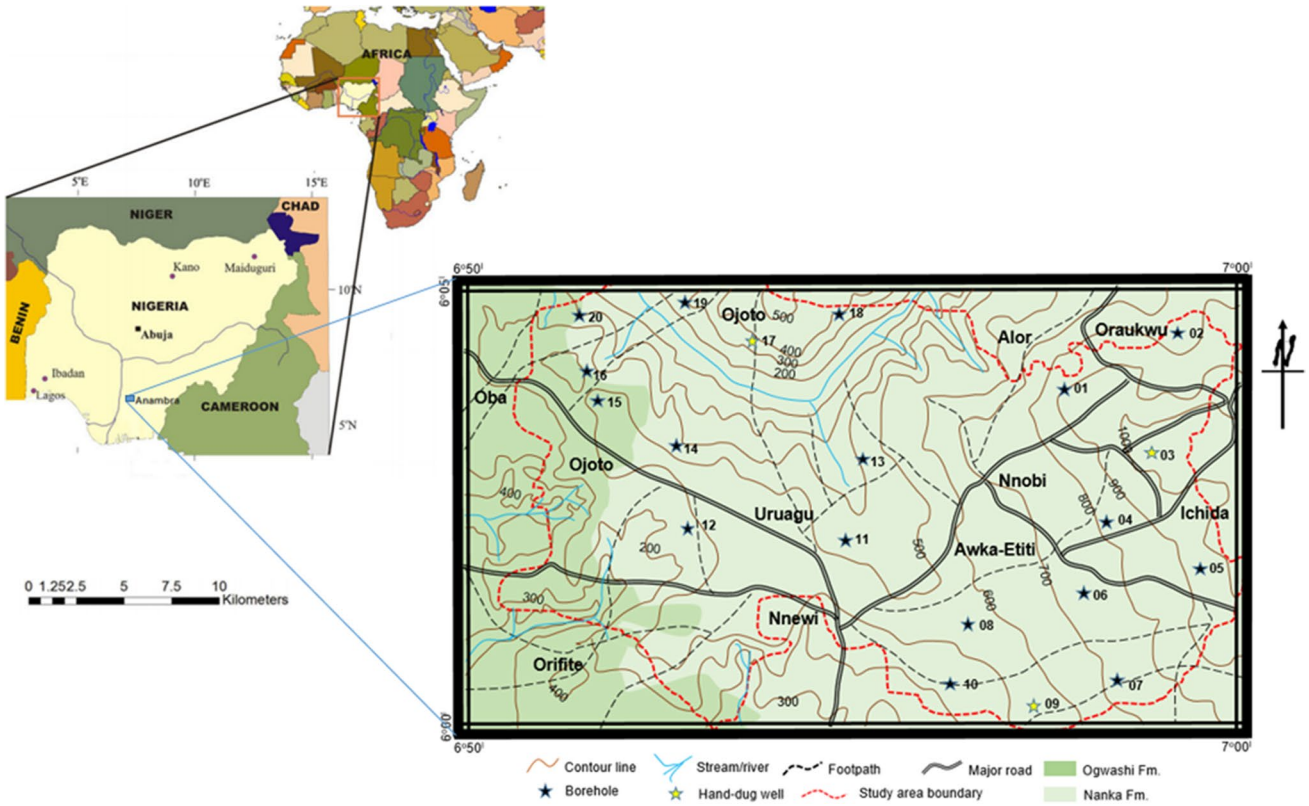

Fig. 1 Map of the study area showing the location and the geologic formations

has been estimated to be around 1500–2000 mm, whereas the average temperature is about 32 °C (Nwajide 2013). Ojoto and its environs are mostly drained by the Idemili River, which has a number of tributaries and distributaries. In turn, the Idemili River empties into the large Niger River, located some few kilometers westward of the study area. In other words, the Idemili River is a tributary of the Niger River. The surface water bodies in the area are believed to be ready aquifer recharge systems. However, the hygiene of the streams and rivers seems to be poor due to anthropogenic impacts, and these could, in turn, be influencing the quality of the water environment.

With undulating and uneven topographical configuration, the area is known to be geologically underlain by mostly sandstones, with few deposits of mudrocks, lignite, limestones, ironstones, and siltstones (Nwajide 2013; Egbueri et al. 2019). The main lithostratigraphic units underlying the study area are the Nanka Sands (Eocene) and Ogwashi Formation (Oligocene–Miocene) (Fig. 1). However, the Nanka Formation underlies most parts of the study area (Fig. 1). The compressive movements of the Campanian–Eocene that resulted in the formation of the Abakaliki Anticlinorium are believed to be responsible for the formation of depositional spaces for the retrograde deposition of the present Nanka Formation (Nwachukwu 1972; Nwajide 2013). Towards the end of the Eocene's tectonic activity, the Ogwashi Formation was deposited. Thus, based on chronology, the Ogwashi Formation is younger and overlies the Nanka Formation.

While the Nanka Sand is lithologically composed of mostly friable fine-coarse sands, sandy shales, shales, claystones, and bands of ironstones and limestones, the Ogwashi Formation is lithologically composed of mostly medium-coarse sands, light-colored mudrocks, and lignite seams (Reyment 1965; Kogbe 1976; Arua 1986; Nwajide 2013). Some hydrogeological and hydrogeophysical investigations conducted by previous researchers provided details of the aquifer characteristics of the Nanka and Ogwashi formations (Nfor et al. 2007; Okoro et al. 2010a, b; Akpoborie et al. 2011). Generally, their studies revealed that both geologic formations have prolific aquifer systems at varying depths. However, more investigations are encouraged to provide updated information on the climatic and aquifer characteristics of the study area.

Groundwater sampling and physicochemical testing

Random sampling was employed for the collection of a total of 20 test groundwater samples across the case study locality. Both boreholes ($n = 17$) and hand-dug wells ($n = 3$) were sampled. The distribution of the sample stations

is also shown in Fig. 1. The samples were collected from the boreholes after few minutes of pumping. Water samples were collected using 1-L polyethylene containers prewashed with diluted HCl. Ice-chested coolers were also used to preserve the samples prior to laboratory analyses. After collection, each water sample was acidified using 1 ml of concentrated HNO_3 to prevent cationic precipitation. Moreover, the samples were also filtered prior to analysis using cellulose acetate filter (size = 0.45 µm milli pore). The groundwater samples were analyzed for several physicochemical variables, including pH, total dissolved solids (TDS), electrical conductivity (EC), total suspended solids (TSS), total hardness (TH); bicarbonate (HCO_3^-), nitrate (NO_3^-), chloride (Cl^-), sulfate (SO_4^{2-}), potassium (K^+), sodium (Na^+), magnesium (Mg^{2+}), calcium (Ca^{2+}), lead (Pb), nickel (Ni), chromium (Cr), iron (Fe), and zinc (Zn). While the pH, TDS, EC, and TSS were measured in situ, the TH, HCO_3^-, NO_3^-, Cl^-, SO_4^{2-}, K^+, Na^+, Mg^{2+}, Ca^{2+}, Pb, Ni, Cr, Fe, and Zn were measured in the laboratory. Details of the procedures and equipment used for the physicochemical analyses are presented in Table 1. These physicochemical tests were conducted following the procedures and guidelines recommended by the American Public Health Association (APHA 2005). The ion balance was estimated to validate the efficiency of the physiochemical test results (Eq. 1). An ion balance of less than 5% was obtained; this shows that the results of the analysis are reliable.

$$\text{Ion balance error} = \frac{\sum \text{cations} - \sum \text{anions}}{\sum \text{cations} + \sum \text{anions}} \times 100 \quad (1)$$

Indexical methods for groundwater pollution and quality assessment

Pollution load index

The PLI model was utilized in this study to determine the extent of toxic metal pollution in the groundwater samples. Prior to the computation of the PLI, the contamination factor (CF) of the PTEs was calculated using Eq. 2 (Håkanson 1980). The PTEs considered in this paper are Zn, Fe, Ni, Pb, and Cr. The PLI was then calculated using Eq. 3 (Tomlinson et al. 1980).

$$\text{CF} = \frac{\text{Cn}}{\text{BV}} \quad (2)$$

$$\text{PLI} = \sqrt[n]{\text{CF} \times \text{CF} \times \text{CF} \times \ldots \text{CF}_n} \quad (3)$$

where Cn equals the measured PTE concentration in the groundwater, BV equals the background value, *n* represents the number of analyzed PTEs, and CF is the contamination factor of the PTEs. The BV in this paper corresponds to the SON (2015) standard limits of the elements.

Modified heavy metal index

To further analyze the extent of pollution in the groundwater resources, the MHMI model was also computed. In the MHMI calculations, weights are assigned to the PTEs on a scale of 1–5 (Egbueri et al. 2020). Temporary weights and relative weights assigned to the groundwater quality parameters are shown in Table 2. The temporary weights were assigned based on the significance of the parameters in water quality assessment and their potential impact on human health whereas the relative weights of the parameters were calculated using Eq. 4 (Egbueri et al. 2020). The final MHMI scores of the groundwater samples were obtained using Eq. 5.

$$R_\text{w} = \frac{wi}{\sum_{i=1}^{n} wi} \quad (4)$$

$$\text{MHMI} = \sum_{i=1}^{n} \left[R_\text{w} \times \frac{M_i}{S_i} \right] \quad (5)$$

Table 2 Weightage of parameters for the calculation of MHMI

Parameter	Fe	Zn	Ni	Cr	Pb
Temporary weight (w_i)	4	2	5	5	5
Relative weight (R_w)	0.1905	0.0952	0.2381	0.2381	0.2381

Table 1 Procedures and equipment for physicochemical analysis

Parameter	Procedure/method	Point of analysis
pH, EC, TDS, and TSS	Testr-2, conductivity/TDS/meter; HM Digital COM-100	In situ (on site)
SO_4^{2-} and NO_3	Spectrophotometry technique	Laboratory
Cl^-	$AgNO_3$ titration	Laboratory
K^+ and Na^+	Flame photometer (model: Systronics Flame Photometer 128)	Laboratory
HCO_3^-	Water titration using H_2SO_4	Laboratory
Mg^{2+} and Ca^{2+}	Volumetric technique (0.05 N EDTA and 0.01 N)	Laboratory
Fe, Zn, Ni, Cr, and Pb	Atomic absorption spectrophotometric method (model: Bulk Scientific 210 VGP)	Laboratory

where R_w represents the relative weight of each parameter, w_i represents the temporary weight assigned to each quality parameter, n represents the total number of parameters considered in the MHMI computation, M_i represents the metal concentration in the water sample, and S_i represents the WHO (2017) standard limit for each PTE (Egbueri et al. 2020). The R_w of all the parameters usually sums up to 1 or ≈ 1.

Synthetic pollution index

The drinking suitability of the groundwater resources was assessed using the SPI model. The SPI model is more elaborate than the PLI and MHMI models, in the sense that all the analyzed water quality parameters were considered in the SPI computation. Equations 6–8 were used for the calculation of the SPI (Solangi et al. 2019; Egbueri and Unigwe 2019; Egbueri et al. 2021a). As can be seen, the SPI model does not require weight assignment to the analyzed parameters.

$$K = \frac{1}{\left(\sum_{i=1}^{n} \frac{1}{\mathrm{Vs}}\right)} \tag{6}$$

$$\mathrm{Wi} = \frac{K}{\mathrm{Vs}} \tag{7}$$

$$\mathrm{SPI} = \sum_{i}^{n} \frac{\mathrm{Vn}}{\mathrm{Vs}} \times \mathrm{Wi} \tag{8}$$

where n represents the number of measured water quality parameters, Vn represents the concentrations of each parameter in the groundwater sample, Vs represents the WHO (2017) standard limit of each parameter, Wi represents the weight coefficient of the parameter, and K represents the proportionality constant.

Linear regression of indexical models

Simple linear regression is often used to show interrelationships between parameters of interest. Results of the MHMI, PLI, and SPI models were subjected to linear regression analysis to determine the agreements, relationships, and consistencies between them in identifying and classifying the level of pollution in the groundwater samples. The linear regression analysis was performed using Microsoft Excel (v. 2016). Three regression models were produced: (i) SPI vs MHMI, (ii) PLI vs MHMI, and (iii) SPI vs PLI.

Soft computing methods for groundwater quality classification

Q-mode hierarchical clustering

In this paper, the groundwater quality was first clustered using Q-mode hierarchical clustering algorithm. This allows for spatial understanding and classification of the groundwater quality association between the analyzed samples. The hierarchical clustering in this study was performed using the Ward (1963) linkage method, which has been justified by previous researchers to provide a more effective clustering (Yidana 2010; Egbueri 2020, 2021; Ozel et al. 2020; Egbueri et al. 2020). The clusters were normalized using Z-score standardization algorithm to minimize bias. The similarities and dissimilarities of the water samples were determined based on squared Euclidean distances. Seven dendrograms were produced to classify the water quality based on the pH, EC, TDS, TH, MHMI, PLI, and SPI. The dendrograms for MHMI, PLI, and SPI could also provide useful information for confirming the relationships and consistencies between the three indexical models in classifying the groundwater quality. On the other hand, the dendrograms for pH, EC, TDS, and TH could provide insights regarding the influence of these physical parameters on the quality of the groundwater system.

K-means partitional clustering

The K-means partitional clustering algorithm was also utilized in this paper to test the assumptions and the findings of the Q-mode agglomerative hierarchical clustering algorithm. Several water quality assessment projects have utilized K-means clustering for the identification and classification of quality peculiarities of water samples (Zubaidah et al. 2018; Zou et al. 2015). Similarities and dissimilarities of the clusters were determined based on Euclidean distances. Although the K-means clustering seems to be simpler than the hierarchical clustering, the latter seems to be more widely employed by researchers than the former. In this study, the relationships between both clustering algorithms were also determined. Similar to the hierarchical clustering of the groundwater samples, seven cluster groups were also generated for the K-means. Accordingly, samples were clustered based on the pH, EC, TDS, TH, MHMI, PLI, and SPI. Both the K-means clustering and the Q-mode hierarchical clustering were performed using IBM SPSS (v. 22). Both algorithms and their results would provide useful information that would greatly enhance local planning and decision-making regarding groundwater quality monitoring and assessment and pollution mitigation programs in the area of study.

Soft computing methods for groundwater contamination source apportionment

Principal component analysis and Varimax-rotated factor analysis

Datasets, such as groundwater quality dataset, are often complex and multidimensional in nature. Hence, it is often difficult to directly interpret such datasets. However, the use of dimensionality reduction machine learning techniques has often proven useful for clearer analysis and interpretation of water quality. Both unrotated principal component analysis and Varimax-rotated factor analysis were used for the dimension reduction and interpretation of the groundwater quality datasets. Both were carried out in IBM SPSS (v. 22). It is pertinent to mention that the scores obtained from these analyses are, similar to simple correlation analysis, classified into three categories. While component loadings < 0.50 are considered to be weak and insignificant, those in the range of 0.50 < loadings < 0.75 are said to be moderate and significant. However, strong (high) loadings are found in ranges > 0.75. Traditionally, the strong loadings are said to be very significant and explain more details about a given dataset.

Soft computing methods for groundwater quality modeling

Multiple linear regression modeling

More advanced than simple linear regression, multiple linear regression is a data-driven algorithm for the prediction of the linear relationships between input (independent or predictor) variables and the output (dependent, target, outcome, or predicted) variable(s). The MLR is an algorithm that adopts the simple least-squares rule in its judgment. This predictive modeling approach has been used in predicting water quality parameters and indices in different parts of the world (Chen and Liu 2015; Kadam et al. 2019; Gaya et al. 2020). Mathematically, MLR is simply expressed as shown in Eq. 9:

$$y = b_0 + b_1 x_2 + b_2 x_2 + \cdots + b_i x_i + \varepsilon \quad (9)$$

where y = predicted target, b_0 = regression constant, b_i = regression coefficient of the ith predictor, and x_i = value of the ith predictor, whereas ε = residual or error for individual i.

The MLR was applied in the present study to simulate and predict pH, EC, TDS, TH, MHMI, PLI, and SPI. For the predictions of pH, EC, TDS, and TH, the predicted is considered as the output parameter and all other physicochemical variables used as input. Alternatively stated, seventeen input variables were utilized for the predictions of pH, EC, TDS, and TH. On the other hand, the predictions of MHMI, PLI, and SPI utilized all the eighteen physicochemical parameters analyzed on the groundwater samples as input variables.

The IBM SPSS (v. 22) was also used for the MLR modeling. The performances of the MLR models were evaluated using four different statistical methods, including multiple correlation coefficient (R), standard error of estimate (SEE), coefficient of determination (R^2), and adjusted R^2. The integration of the multiple statistical methods for validating the accuracies of the models seems to better guarantee the reliability of such models, as a single measure of validity might lead to bias.

Artificial neural network modeling

Unlike the MLR that only considers the multiple linearity between predictor and the predicted variables, ANN is a more advanced soft computing simulation approach for predicting linear, nonlinear, and complex relationships. Since the relationships between water quality parameters are usually complex, it is important to also include ANN in the present modeling study. Similar thought has been expressed by Senapati et al. (2021). Interestingly, numerous water researchers have adopted ANN modeling as a reliable approach to water quality prediction, even though there are newer and more advanced soft computing approaches that are now usable for the same course. It is sufficed to say that the reliability of ANN modeling in predictive study of water quality parameters could be one of the main reasons it is still adopted, regardless of newer prediction algorithms. According to Chen and Liu (2015), ANN modeling is usually adopted in water quality prediction due to the fact that there are several challenges and difficulties often associated with the computation of water quality indices and the simulation of the conditions that influence water quality based on hydrodynamics.

In this paper, scaled conjugate gradient (SCG) optimization algorithm was utilized in developing seven ANNs that predicted the seven variables of interest—pH, EC, TDS, TH, MHMI, PLI, and SPI. Thus, seven SCG-ANNs are the models also utilized, alongside the seven MLR models, to simulate and predict these water quality parameters. Similar to the MLR, seventeen input parameters were used for the predictions of the pH, EC, TDS, and TH, whereas eighteen input variables were utilized for the predictions of the MHMI, PLI, and SPI. The ANN modeling methodology for the present paper is summarized in Table 3. All the ANN models were produced in IBM SPSS (v. 22). The performances and accuracies of the ANN models were evaluated in SPSS using different statistical methods such as R^2, relative error (RE), sum of square errors (SOSE), and residual error plots (REPs).

Table 3 ANN modeling requirements and instructions for the current predictive study

Model parameter	Instruction/activity report
Input variables	The predictor variables are pH, EC, TDS, TSS, TH, Na, K, Ca, Mg, Cl, SO_4, HCO_3, NO_3, Fe, Zn, Ni, Cr, and Pb. All these were used as input for the predictions of MHMI, PLI, and SPI. However, for the predictions of pH, EC, TDS, and TH, the predicted is made the dependent variable and all others used as predictors.
Output variable	The predicted parameters are pH, EC, TDS, TH, MHMI, PLI, and SPI.
Input layer activation function	Hyperbolic tangent
ANN type	Multilayer perceptron (MLP)
Rescaling of covariates	Normalized
Partitioning of dataset	Randomly assigned cases based on relative number of cases: pH: Training (75%), testing (25%), and validity of cases (100%) EC: Training (80%), testing (20%), and validity of cases (100%) TDS: Training (90%), testing (10%), and validity of cases (100%) TH: Training (80%), testing (20%), and validity of cases (100%) MHMI: Training (80%), testing (20%), and validity of cases (100%) PLI: Training (75%), testing (25%), and validity of cases (100%) SPI: Training (70%), testing (30%), and validity of cases (100%)
Number of hidden layers	One (1)
Hidden layer activation function	Hyperbolic tangent
Number of units	Automatically computed
Rescaling of scale dependent variables	Adjusted normalized (correction = 0.02)
Type of training	Batch
Optimization algorithm	Scaled conjugate gradient (SCG)

Results and discussion

Overview of groundwater physicochemical properties and quality

Summary of the physicochemical testing is presented in Table 4. It has been reported that most chemical reactions in both surface and groundwater systems are greatly influenced by the water pH (Idriss et al. 2020). Thus, pH is regarded as an important water quality parameter. The pH values, which were found to range between 4.0 and 6.4, revealed that the water resources are mostly acidic in nature. The pH range that has been stipulated by WHO (2017) and SON (2015) as acceptable benchmark is 6.5–8.5. Also, pH values in the range of 6.22 to 7.43 have been noted to be the best fit for the existence and survival of aquatic life (Garg et al. 2010). The measured pH scores indicate that the majority of the groundwater resources are unsuitable for man's consumption, aquatic life, and food production.

The acidic nature of the analyzed groundwater could be corrosive to metallic domestic, irrigational, and industrial wares (Egbueri et al. 2021a, c) and could also affect human skin and eyes negatively (Li et al. 2017). A scenario whereby acidic water deteriorates metallic wares and influences public health could amount to socioeconomic losses. The groundwater acidity could be attributed to acidic rain and other geochemical processes or reactions. Also, the release of nitrogen oxides and sulphur gases by fossil fuel burning could trigger precipitation of acid rain (Egbueri et al. 2021c). The weathering and leaching of minerals rich in sulphur and chloride and subsequent oxidation/hydrogenation into water could heighten its acidity.

Total dissolved solids (TDS) and electrical conductivity (EC) are also essential physicochemical properties of water (Weber-Scannell and Duffy 2007). The TDS and EC of water are part of the key indicators used for estimating the level of water quality deterioration. Thus, variations in the level of water TDS and EC seem to provide useful insights on the extent and possible predominant sources of water contamination and pollution. Water resources with elevated TDS and EC values have a higher chance of possessing contaminants and pollutants. Since they can be used as reliable indicators, they can be used for managing pollution in water environment (Shah et al. 2021; Sun et al. 2021). Moreover, both can be used to infer the salinity of water. In this study, TDS ranged from 8.00 to 76.00, whereas EC was found to range between 8.00 and 102.00 mg/L (Table 4). These values were found to be within their standard permissible limits of 600–1000 mg/L and 1000 µS/cm, respectively (WHO 2017; SON 2015). Thus, with respect to the TDS and EC, it is anticipated that all the water samples have low level of contaminants.

Water samples with higher TDS values could have esthetic problems with respect to staining, taste, and scaling

Table 4 Physicochemical register of the groundwater samples

Sample no.	Source	pH	EC	TDS	TSS	TH	Na^+	K^+	Ca^{2+}	Mg^{2+}	Cl^-	SO_4^{2-}	HCO_3^-	NO_3^-	Fe	Zn	Ni	Cr	Pb
1	BH	5.800	16.000	35.000	0.000	10.000	15.000	5.000	2.000	0.200	2.000	70.000	2.000	0.900	0.400	0.200	0.020	0.001	2.000
2	BH	4.600	22.000	13.000	0.000	13.000	13.000	4.000	4.000	0.300	4.000	10.000	<DL	<DL	0.400	0.030	0.000	0.000	0.021
3	HW	4.800	14.000	15.000	1.000	16.000	17.000	8.000	8.000	0.600	8.000	115.000	3.200	0.020	0.400	0.110	0.000	0.000	<DL
4	BH	4.700	73.000	62.000	2.000	45.000	10.000	3.000	20.000	0.700	4.000	48.000	<DL	0.030	0.500	0.320	0.000	0.015	<DL
5	BH	4.800	33.000	28.000	0.000	16.000	8.000	7.000	8.000	0.500	2.000	13.000	<DL	1.900	0.200	0.040	0.000	0.000	0.012
6	BH	4.800	16.000	11.000	0.000	18.000	10.000	4.000	3.000	0.600	21.000	24.000	0.400	0.030	0.300	<DL	0.000	0.000	<DL
7	BH	5.900	102.000	76.000	1.000	52.000	27.000	7.000	25.000	16.000	20.000	67.000	3.500	<DL	0.100	0.100	0.000	0.000	<DL
8	BH	4.300	14.000	12.000	0.000	8.000	15.000	3.000	6.000	0.400	12.000	8.000	3.200	0.200	0.100	0.060	0.000	0.000	0.061
9	HW	5.100	20.000	13.000	0.000	8.000	15.000	<DL	4.000	0.200	3.000	40.000	3.200	<DL	2.400	<DL	0.000	0.000	<DL
10	BH	6.000	42.000	19.000	1.000	24.000	16.000	4.000	11.000	0.600	5.000	25.000	2.000	0.300	0.400	0.010	0.120	0.000	0.001
11	BH	4.000	53.000	32.000	0.000	33.000	15.000	12.000	10.000	0.000	3.000	33.000	0.400	18.480	0.400	<DL	0.000	0.000	<DL
12	BH	5.000	10.000	8.000	0.000	6.000	17.000	<DL	2.000	0.500	3.000	18.000	<DL	8.400	0.300	<DL	0.000	0.000	<DL
13	BH	6.400	43.000	33.000	1.000	32.000	12.000	5.000	8.000	0.600	4.000	26.000	3.000	0.060	0.400	0.010	0.120	0.000	0.001
14	BH	4.900	28.000	12.000	3.000	8.000	38.000	9.000	4.000	0.400	33.000	72.000	5.000	4.200	0.100	0.400	0.340	0.004	1.980
15	BH	5.300	32.000	21.000	4.000	12.000	12.000	8.000	3.000	0.300	25.000	8.000	0.400	1.680	0.300	0.120	0.000	0.000	<DL
16	BH	4.600	19.000	31.000	1.000	14.000	14.000	3.000	3.000	0.300	23.000	25.000	0.600	0.030	0.200	0.000	0.000	0.000	<DL
17	HW	4.100	15.000	11.000	2.000	16.000	20.000	8.000	7.000	0.300	9.000	130.000	0.600	0.030	0.000	0.110	0.000	0.000	<DL
18	BH	4.400	8.000	10.000	0.000	8.000	11.000	3.000	5.000	0.000	10.000	64.000	<DL	1.800	0.400	0.000	0.000	0.000	0.011
19	BH	4.800	12.000	11.000	2.000	10.000	12.000	11.000	7.000	1.000	21.000	15.000	<DL	14.400	0.200	0.340	0.230	0.012	1.430
20	BH	5.400	21.000	23.000	0.000	10.000	14.000	8.000	2.000	0.300	3.000	70.000	1.500	1.240	0.300	0.200	0.020	0.001	2.000
	DL	0.2	-	-	-	-	0.01	0.01	0.01	0.01	0.01	0.01	0.05	0.001	0.900	0.900	-	-	0.900

BH, borehole; *HW*, hand-dug well; *DL*, detection limit. All parameters measured in mg/L, except pH and EC (measured in µS/cm). The DL of Fe, Zn, and Pb are measured in nM

or precipitation (Sibanda et al. 2014; Sun et al. 2021). On the other hand, water with high EC could cause certain undesired reactions in the human body. Both the TDS and EC are greatly influenced by the TDS (inorganic salts and organic matter) present in the water (McNeil and Cox 2000; Subramani et al. 2005; Miranda and Krishnakumar 2015; Wagh et al. 2020; Sun et al. 2021; Shah et al. 2021). The dissolved minerals which primarily account for TDS may be related to the release of salts from surface soil and aquifer materials and some human-induced activities (Mukate et al. 2019). The electronic conductivity of the water may be related to evaporation processes in the phreatic zone, extensive rock-water interactions, and man-made contamination exercises (Wagh et al. 2020). Furthermore, it is suspected that the low TDS and EC observed in this study could be due to dilution by rain infiltrating through the aquifer system.

The groundwater samples were also tested for total suspended solids (TSS) and total hardness (TH). The TSS has no WHO (2017) and SON (2015) guidelines. However, the TSS values obtained in this study were seen to be generally low (Table 4). On the other hand, the TH values ranged from 6 to 52 mg/L (Table 4), which is below the 150-mg/L threshold value (SON 2015). TH has been recognized as a crucial parameter for assessing the suitability of water resources for domestic and industrial uses (Lower 2007). Water quality may be classified into four categories on the basis of TH as follows: soft water (TH < 60 mg/L), moderately hard water (TH range of 60–120 mg/L), hard water (TH range of 120–180 mg/L), and very hard water (TH > 180 mg/L) (McGowan 2000; Sawyer and McCarthy 1967). Based on this classification, all the groundwater samples are adjudged to be soft. This implies that they would readily allow the formation of lather; thus, would not cause wastage of soap in domestic laundry and other sanitary activities.

Salt deposition due to rock-water interaction might be responsible for elevated TH values in water (Kadam et al. 2021). TH concentration above the standard limit could cause health hazards (Kožíšek 2003) such as eczema (Arnedo-Pena et al. 2007; Miyake et al. 2004; McNally et al. 1998) and cardiovascular issues (Marque et al. 2003; Rubenowitz et al. 1999; Pocock et al. 1981). Moreover, scale formation on water boilers and clogging of water distribution pipes are some of the problems posed by hard water in homes and industries (Lower 2007; Roy and Majumder 2018a; Egbueri et al. 2021a, c). However, for the present study, the low TH values could be attributed to low concentrations of Ca and Mg salts (WHO 2011).

The cations analyzed in the groundwater samples are Na^+, K^+, Ca^{2+}, and Mg^{2+}, whereas the anions include Cl^-, SO_4^{2-}, HCO_3^-, and NO_3^-. Their results are also presented in Table 4. Overall, all the chemical ions were noticed to be mostly within their permissible limits of 200 mg/L (Na^+), 12 mg/L (K^+), 75 mg/L (Ca^{2+}), 0.20–50 mg/L (Mg^{2+}), 250 mg/L (Cl^-), 100–250 mg/L (SO_4^-), 250 mg/L (HCO_3^-), and 50 mg/L (NO_3^-) (SON 2015; WHO 2017). This implies that the cations and anions mostly occurred at contamination levels, not polluted levels. Low concentration of Na^+ might be due to the absence of confirmed rock salt deposits in the area (Egbueri et al. 2021a). Dissolved potassium seems to be low due to its high resistance in minerals, especially in clay structures (Srinivas et al. 2017). Due to this fact, studies have suggested that K^+ in surface water or groundwater is normally found in few parts per million to about 0.1 ppm in rainwater (Matthess 1982).

Majority of the groundwater sample exceeded the permissible limits of Mg^{2+} in water; thus, they are adjudged to be polluted with Mg^{2+}. This suggests that the people who consume the water resources are exposed to health risks due to the ingestion of excess Mg. Excessive concentration of Mg^{2+} in water could be a result of irrigation return flow and dissolution of evaporites and ferromagnesian minerals (Haritash et al. 2008; Egbueri et al. 2020). Moreover, carbonate species present in an area might also have obvious influence on the Mg^{2+} content of water. However, the presence of carbonates is mostly favored by alkaline conditions of water, unlike the acidic nature observed in the study area. The acidic nature of the waters could be one of the reasons for the low HCO_3^- measured in the samples.

Dissolved chlorides and SO_4^{2-} seem to be the most predominant anions measured in the groundwater samples, as their concentrations were generally seen to be higher than those of the HCO_3^- and NO_3^- (Table 4). Human-induced activities such as the use of fertilizer in farms and detergents in homes could influence the concentrations of SO_4 in water (Kadam et al. 2021). However, chloride in water could be influenced by anthropogenic exercises such as improper waste disposal in dumpsites, sewage, and agrarian flows (Egbueri 2018; Mukate et al. 2017; Kumar et al. 2008). Both SO_4^{2-} and Cl^- can also be influenced by geogenic processes such as rock weathering, mineral dissolution, and decomposition of sulfide and carbon-based substances (Egbueri 2019a, b; Kadam et al. 2021).

Inorganic nitrogen in soil is very essential for the growth and development of plants and crops (Pisciotta et al. 2015). From the soil, nitrate can be added to water. In recent times, nitrate pollution of water systems has attracted the attention of many researchers on a global scale (Rahman et al. 2021). In this paper, NO_3 ranged from 0.00 to 18.48 mg/L in the groundwater system (Table 4). Nitrate, when occurring excessively in water, could lead to several health hazards (including methemoglobinemia in children and stomach cancer in adults) over a long time of ingestion. Agricultural return flows (Vetrimurugan et al. 2013) as well as population explosion, rapid

industrialization and rapid urbanization, human and animal wastes, improper sewage systems, and excessive use of fertilizers (Egbueri 2019a, b; Pisciotta et al. 2015; Egbueri et al. 2020; Idriss et al. 2020; Kadam et al. 2021; Rahman et al. 2021) are examples of the commonest human exercises that contribute to nitrate water pollution.

Potentially toxic elements (PTEs) were also analyzed in the groundwater resources. Their results are presented in Table 4. These are elements which often have higher toxicity than the chemical ions, even in little concentrations. They are nonbiodegradable and could be more persistent in different environmental compartments (Ukah et al. 2020). Thus, they have the potential to bioaccumulate in human systems. It is important to mention that they can naturally occur in the water environment. However, elevated values of PTEs are highly toxic to human beings and their concentrations could be greatly influenced by anthropogenically induced activities. Fe, Zn, Ni, Cr, and Pb were found to range from 0.00 to 2.40 mg/L, 0.00 to 0.40 mg/L, 0.00 to 0.34 mg/L, 0.00 to 0.015 mg/L, and 0.00 to 2.00 mg/L, respectively. Based on their concentration levels, their decreasing order is Fe > Pb > Zn > Ni > Cr.

Amongst the analyzed metals, Fe and Zn are required in the human system for some biochemical functions. Similarly, Cr has been recognized as an essential nutrient for insulin action (Proshad et al. 2021). Hence, these elements can be considered essential elements. Nevertheless, there is a broad range of health issues associated with cumulative consumption of water resources polluted with PTEs. Several ailments, such as gangrene, cancer, peripheral vascular disease, fatal cardiac arrest, hyperkeratosis, liver disease, nausea and vomiting, restrictive lung diseases, hypertension, kidney failure, abdominal pain, headache, gliomas, nerve damage, loose stools, and fatal organ failure, are attributed to the ingestion of high concentration of PTEs into the human system (Alsubih et al. 2021; Enyigwe et al. 2021; Proshad et al. 2021; Ukah et al. 2020). Studies have also shown that there is an increment in the rate of dissolution and absorption of toxic metals in humans when the system is acidic (Egbueri et al. 2021c). Since the water resources are mostly acidic in nature, the bioaccumulation and bioavailability of these metals in humans would be enhanced.

Not only are the PTEs injurious to humans, they can also affect plants and aquatic life. For instance, Proshad et al. (2021) reported that Pb in excess of 0.5 µg/L can inhibit enzymatic functions required for photosynthesis in algae. This portends danger for the use of the water for agricultural purpose. Likewise, fishes, which are more affected by Pb than algae, can develop gill diseases due to high content of Pb in water (Proshad et al. 2021). This, therefore, portends danger for the use of the Pb-contaminated groundwater in fishponds and fisheries.

Indexical methods for groundwater pollution and quality assessment

Pollution load index

The extent of PTEs loading on the groundwater system was first examined using PLI. The obtained PLI results are presented in Table 5. Prior to the PLI final computation, the contamination factor of each of the metals was computed and the result is also shown in Table 5. With respect to contamination factor (CF), pollution level can be classified into four categories: low contamination is indicated by $CF < 1$, moderate contamination is represented by CF range of $1 \leq CF < 3$, considerable level of contamination is signified by CF range of $3 \leq CF < 6$, and very high contamination level is indicated by $CF > 6$ (Håkanson 1980; Egbueri et al. 2021c). Based on the CF results shown in Table 5, it was learnt that all the samples had low contamination with respect to Zn and Cr. Although majority of the samples also showed low contaminations of Fe, Ni, and Pb, some recorded moderate to very high levels of contamination. Furthermore, the final PLI values were seen to range between 0.000 and 1.4491 (Table 5). The PLI classifies water quality into four groups: no pollution is signified by $PLI < 1$, moderate pollution level is indicated by $1 < PLI < 2$, heavy pollution level is signified by $2 < PLI < 3$, and $PLI \geq 3$ represents extreme heavy pollution (Tomlinson et al. 1980; Egbueri et al. 2021c). With respect to this classification scheme, 90% of the total samples were identified as having no pollution, whereas 10% had moderate pollution. The samples with moderate pollution are adjudged to have been exposed to more contaminants, linked to anthropogenic genesis, which elevated the concentrations of the PTEs more than in the other samples.

Modified heavy metal index

The computed MHMI results are shown in Table 5. MHMI classifies water quality into five categories: excellent water quality is represented by $MHMI < 50$, good water quality is indicated by $50 \leq MHMI < 100$, poor water quality is signified by $100 \leq MHMI < 200$, $200 \leq MHMI < 300$ indicates very poor water quality, and $MHMI \geq 300$ marks unsuitable water quality for human consumption (Egbueri et al. 2020). In the present study, it was realized that the MHMI results of the analyzed water samples ranged between 0.0026 and 48.3924 (Table 5). Overall, this result suggests that all the groundwater samples are suitable for human consumption. Nevertheless, it was also observed that the majority (about 80%) of the samples had $MHMI < 2$. But samples 1, 14, 19, and 20 had higher MHMI scores in the range of 35.0229–48.3924. Based on the MHMI values of the water samples mentioned above (Table 5) and the concentration of PTEs in them (Table 4), it was noticed that these four

Table 5 Pollution load index (PLI) and modified heavy metal index (MHMI) of the groundwater samples

		a. Results of PLI computation						b. Results of MHMI computation					
Sample no.	Source	CF (Fe)	CF (Zn)	CF (Ni)	CF (Cr)	CF (Pb)	Final PLI	MHMI (Fe)	MHMI (Zn)	MHMI (Ni)	MHMI (Cr)	MHMI (Pb)	Final MHMI
1	BH	1.3333	0.0500	0.2857	0.0200	200.0000	0.5976	0.2540	0.0048	0.0680	0.0048	47.6200	47.9516
2	BH	1.3333	0.0075	0.0000	0.0000	2.1000	0.0000	0.2540	0.0007	0.0000	0.0000	0.5000	0.7547
3	HW	1.3333	0.0275	0.0000	0.0000	0.0000	0.0000	0.2540	0.0026	0.0000	0.0000	0.0000	0.2566
4	BH	1.6667	0.0800	0.0000	0.3000	0.0000	0.0000	0.3175	0.0076	0.0000	0.0714	0.0000	0.3965
5	BH	0.6667	0.0100	0.0000	0.0000	1.2000	0.0000	0.1270	0.0010	0.0000	0.0000	0.2857	0.4137
6	BH	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1905	0.0000	0.0000	0.0000	0.0000	0.1905
7	BH	0.3333	0.0250	0.0000	0.0000	0.0000	0.0000	0.0635	0.0024	0.0000	0.0000	0.0000	0.0659
8	BH	0.3333	0.0150	0.0000	0.0000	6.1000	0.0000	0.0635	0.0014	0.0000	0.0000	1.4524	1.5173
9	HW	8.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.5240	0.0000	0.0000	0.0000	0.0000	1.5240
10	BH	1.3333	0.0025	1.7143	0.0000	0.1000	0.0000	0.2540	0.0002	0.4082	0.0000	0.0238	0.6862
11	BH	1.3333	0.0000	0.0000	0.0000	0.0000	0.0000	0.2540	0.0000	0.0000	0.0000	0.0000	0.2540
12	BH	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1905	0.0000	0.0000	0.0000	0.0000	0.1905
13	BH	1.3333	0.0025	1.7143	0.0000	0.1000	0.0000	0.2540	0.0002	0.4082	0.0000	0.0238	0.6862
14	BH	0.3333	0.1000	4.8571	0.0800	198.0000	1.2073	0.0635	0.0095	1.1565	0.0190	47.1438	48.3924
15	BH	1.0000	0.0300	0.0000	0.0000	0.0000	0.0000	0.1905	0.0029	0.0000	0.0000	0.0000	0.1934
16	BH	0.6667	0.0000	0.0000	0.0000	0.0000	0.0000	0.1270	0.0000	0.0000	0.0000	0.0000	0.1270
17	HW	0.0000	0.0275	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0000	0.0000	0.0000	0.0026
18	BH	1.3333	0.0000	0.0000	0.0000	1.1000	0.0000	0.2540	0.0000	0.0000	0.0000	0.2619	0.5159
19	BH	0.6667	0.0850	3.2857	0.2400	143.0000	1.4491	0.1270	0.0081	0.7823	0.0571	34.0483	35.0229
20	BH	1.0000	0.0500	0.2857	0.0200	200.0000	0.5641	0.1905	0.0048	0.0680	0.0048	47.6200	47.8881

samples had the highest Pb pollution. They were also identified to have the highest PLI scores amongst the water samples. Although the MHMI has classed them as suitable, prior treatment before human consumption is advised. This recommendation is based on the WHO (2017) and SON (2015) acceptable limits for Pb in water.

Synthetic pollution index

The SPI model is more elaborate and comprehensive as it considered all the analyzed parameters. Thus, it gives a clearer picture of the drinking water quality. The computed SPI for the water samples is given in Table 6, and their results ranged between 0.001 and 144.805. The individual contributions of the water quality parameters to the final SPI scores of the water samples are also presented in Table 6. The SPI classifies drinking water quality into five different categories: excellent and suitable drinking water (SPI < 0.2); slightly polluted drinking water (SPI 0.2–0.5); moderately polluted drinking water (SPI 0.5–1.0); highly polluted drinking water (SPI 1.0 and 3.0); and extremely polluted and unsuitable drinking water source (SPI > 3.0) (Solangi et al. 2019; Egbueri and Unigwe 2019). The present study reveals that 50% of the total samples are excellent and suitable drinking waters, whereas the other water samples fall into the following classification: slightly polluted (10%), moderately polluted (10%), highly polluted (5%), or extremely polluted (25%) drinking water resources. Therefore, the SPI results confirmed that 50% of the groundwater samples are of excellent quality and suitable for drinking, domestic, industrial, and irrigation purposes. However, the other 50% are confirmed to have questionable quality. Thus, they would require adequate treatment before human consumption, as they may constitute health risks to the people who drink from the sources.

Linear regression of indexical models

Graphical and statistical methods are usually useful for clear representation of complex datasets and in the recognition of patterns. Thus, in this paper, the interrelationships between indexical methods of water quality assessment were established in this paper using simple trend graph, linear regression models, and parity plots. This was considered necessary as the different index methods have varied scores and meanings. Prior to the linear regression modeling, simple line graph (shown in Fig. 2a) was used to depict the trend of the indexical models. Figure 2a portrays that MHMI results agreed best with the results of PLI and SPI. Alternatively stated, the groundwater samples that were identified to have elevated pollution were observed to have similar trend as per the MHMI and SPI model. Furthermore, the simple regression models shown in Fig. 2b–d seem to confirm the pollution trend/pattern shown in Fig. 2a. It was observed that the strength of agreement between MHMI and SPI (R^2 = 0.9996; Fig. 2b) was stronger than that between MHMI and PLI (R^2 = 0.7415; Fig. 2c) and PLI and SPI (R^2 = 0.7346; Fig. 2d).

It is recalled that while the MHMI model required weights to be assigned to the water quality parameters by the assessors, the SPI model did not require weightage assignment by the assessors. Traditionally, the assignment of weights to water quality parameters is usually based on the assessors' discretion regarding the possible significance and health impacts of the considered variables. Unfortunately, it has been reported by several authors that weightage assignment by water quality assessors often introduces some levels of bias in water quality assessment (Li et al. 2010; Amiri et al. 2014; Egbueri et al. 2020; Ukah et al. 2020). However, the very strong agreement between the MHMI and SPI models suggests that the weights assigned to the parameters in the present MHMI analysis seem to be the best fits and typical representatives of the actual water quality scenario. Although the PLI did not also require weight assignment to the water quality variables, its judgments seem to be less correlative to those of the MHMI and SPI. In this study, combining indexical, statistical (artificial intelligence), and human intelligence interestingly seems to have provided more robust and reliable assessment.

Soft computing methods for groundwater quality classification

Q-mode hierarchical clustering

The grouping of the groundwater samples based on their quality was first performed with the hierarchical cluster method. Dendrograms produced for this analysis are presented in Fig. 3. Figure 3a represents the water quality classification based on pH. Three main clusters were identified. The first cluster is composed of samples 13, 1, 10, and 7; the second cluster has samples 18, 8, 17, and 11 as members; the third cluster is made up of groundwater samples 9, 14, 12, 20, 15, 4, 16, 2, 5, 3, 19, and 6 (Fig. 3a). Based on the pH scores presented in Table 4, the second cluster represents the water samples with the highest acidity, as their pH values ranged from 4.000 to 4.400. Following a similar trend, the third cluster also has acidic water samples with pH range of 4.600–5.400. However, the first cluster represents water sample class with their pH values (5.800–6.400) tending towards alkalinity. The order in which corrosion and associated health risks of the acidic waters are anticipated is cluster 2 > cluster 3 > cluster 1. Thus, cluster 2 contains samples with high-risk exposure due to acidic pH, and cluster 1 has samples with the lower risk exposure.

Table 6 Synthetic pollution index (SPI) of the groundwater samples

Sample no.	Source	SPI (pH)	SPI (EC)	SPI (TDS)	SPI (TH)	SPI (Na)	SPI (K)	SPI (Ca)	SPI (Mg)	SPI (Cl)	SPI (SO4)	SPI (HCO3)	SPI (NO3)	SPI (Fe)	SPI (Zn)	SPI (Ni)	SPI (Cr)	SPI (Pb)	Final SPI
1	BH	0.0008567	0.0000001	0.0000007	0.0000072	0.0000027	0.0002513	0.0000026	0.0000006	0.0000004	0.0000081	0.0000002	0.0000026	0.0321640	0.0000905	0.0295386	0.0028948	144.7396000	144.805
2	BH	0.0006795	0.0000002	0.0000003	0.0000094	0.0000023	0.0002010	0.0000051	0.0000009	0.0000007	0.0000012	0.0000000	0.0000000	0.0321640	0.0000136	0.0000000	0.0000000	1.5197658	1.553
3	HW	0.0007090	0.0000001	0.0000003	0.0000115	0.0000031	0.0004020	0.0000102	0.0000017	0.0000014	0.0000133	0.0000004	0.0000001	0.0321640	0.0000497	0.0000000	0.0000000	0.0000000	0.033
4	BH	0.0006943	0.0000005	0.0000012	0.0000324	0.0000018	0.0001508	0.0000256	0.0000020	0.0000007	0.0000056	0.0000000	0.0000001	0.0402050	0.0001447	0.0000000	0.0434220	0.0000000	0.085
5	BH	0.0007090	0.0000002	0.0000006	0.0000115	0.0000014	0.0003518	0.0000102	0.0000015	0.0000004	0.0000015	0.0000000	0.0000055	0.0160820	0.0000181	0.0000000	0.0000000	0.8684376	0.886
6	BH	0.0007090	0.0000001	0.0000002	0.0000130	0.0000018	0.0002010	0.0000038	0.0000017	0.0000038	0.0000028	0.0000000	0.0000001	0.0241230	0.0000000	0.0000000	0.0000000	0.0000000	0.025
7	BH	0.0008715	0.0000007	0.0000015	0.0000374	0.0000049	0.0003518	0.0000320	0.0000464	0.0000036	0.0000078	0.0000004	0.0000000	0.0080410	0.0000452	0.0000000	0.0000000	0.0000000	0.009
8	BH	0.0006352	0.0000001	0.0000002	0.0000058	0.0000027	0.0001508	0.0000077	0.0000012	0.0000022	0.0000009	0.0000004	0.0000006	0.0080410	0.0000271	0.0000000	0.0000000	4.4145578	4.423
9	HW	0.0007533	0.0000001	0.0000003	0.0000058	0.0000027	0.0000000	0.0000051	0.0000006	0.0000005	0.0000046	0.0000004	0.0000000	0.1929840	0.0000000	0.0000000	0.0000000	0.0000000	0.194
10	BH	0.0008863	0.0000003	0.0000004	0.0000173	0.0000029	0.0002010	0.0000141	0.0000017	0.0000009	0.0000029	0.0000002	0.0000009	0.0321640	0.0000045	0.1772314	0.0000000	0.0723698	0.283
11	BH	0.0005909	0.0000004	0.0000006	0.0000238	0.0000027	0.0006030	0.0000128	0.0000000	0.0000005	0.0000038	0.0000000	0.0000536	0.0321640	0.0000000	0.0000000	0.0000000	0.0000000	0.033
12	BH	0.0007386	0.0000001	0.0000002	0.0000043	0.0000031	0.0000000	0.0000026	0.0000015	0.0000005	0.0000021	0.0000000	0.0000244	0.0241230	0.0000000	0.0000000	0.0000000	0.0000000	0.025
13	BH	0.0009454	0.0000003	0.0000007	0.0000230	0.0000022	0.0002513	0.0000102	0.0000017	0.0000007	0.0000030	0.0000003	0.0000002	0.0321640	0.0000045	0.1772314	0.0000000	0.0723698	0.283
14	BH	0.0007238	0.0000002	0.0000002	0.0000058	0.0000068	0.0004523	0.0000051	0.0000012	0.0000059	0.0000084	0.0000006	0.0000122	0.0080410	0.0001809	0.5021557	0.0115792	143.2922040	143.815
15	BH	0.0007829	0.0000002	0.0000004	0.0000086	0.0000022	0.0004020	0.0000038	0.0000009	0.0000045	0.0000009	0.0000000	0.0000049	0.0241230	0.0000543	0.0000000	0.0000000	0.0000000	0.025
16	BH	0.0006795	0.0000001	0.0000006	0.0000101	0.0000025	0.0001508	0.0000038	0.0000009	0.0000041	0.0000029	0.0000001	0.0000001	0.0160820	0.0000000	0.0000000	0.0000000	0.0000000	0.017
17	HW	0.0006056	0.0000001	0.0000002	0.0000115	0.0000036	0.0004020	0.0000090	0.0000009	0.0000016	0.0000151	0.0000001	0.0000001	0.0000000	0.0000497	0.0000000	0.0000000	0.0000000	0.001
18	BH	0.0006499	0.0000001	0.0000002	0.0000058	0.0000020	0.0001508	0.0000064	0.0000000	0.0000018	0.0000074	0.0000000	0.0000052	0.0321640	0.0000000	0.0000000	0.0000000	0.7960678	0.829
19	BH	0.0007090	0.0000001	0.0000002	0.0000072	0.0000022	0.0005528	0.0000090	0.0000029	0.0000038	0.0000017	0.0000000	0.0000418	0.0160820	0.0001538	0.3396936	0.0347376	103.4888140	103.881
20	BH	0.0007977	0.0000001	0.0000005	0.0000072	0.0000025	0.0004020	0.0000026	0.0000009	0.0000005	0.0000081	0.0000002	0.0000036	0.0241230	0.0000905	0.0295386	0.0028948	144.7396000	144.797

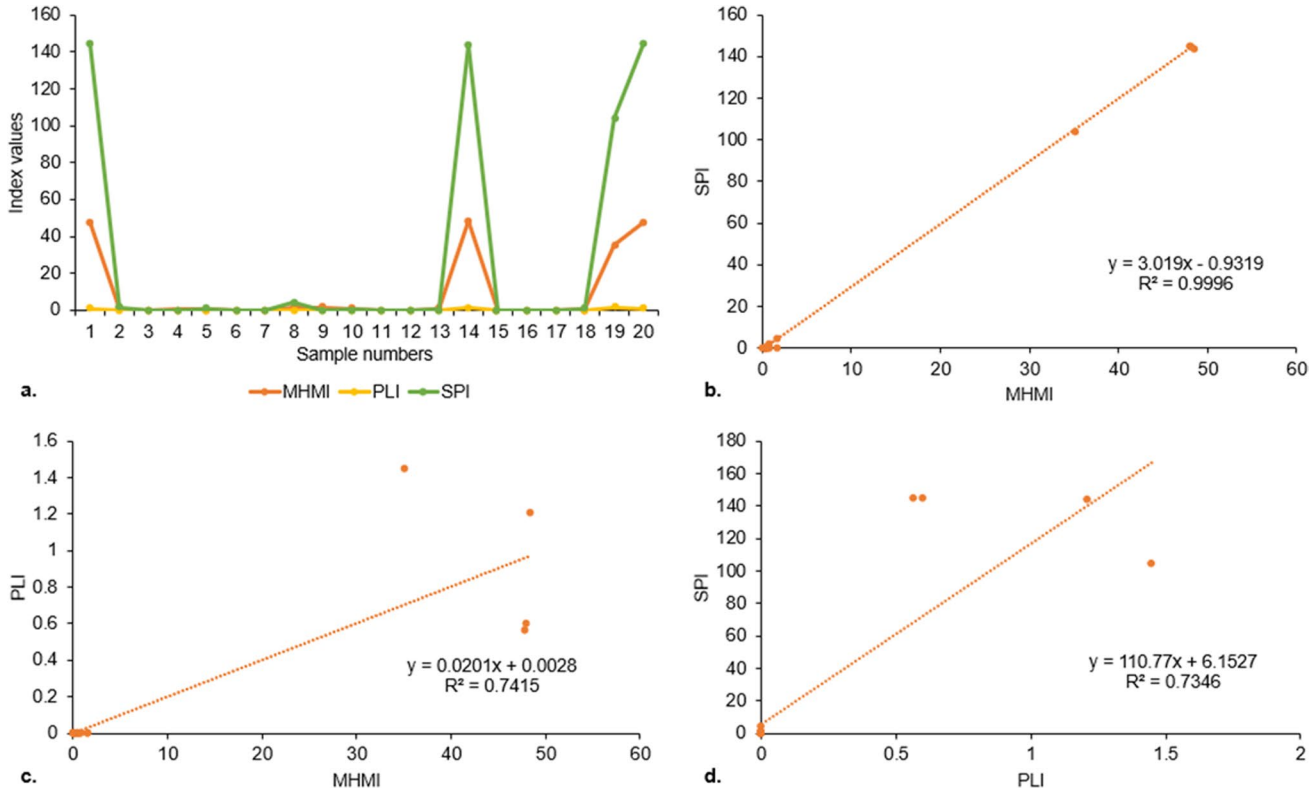

Fig. 2 Graphical comparison of the agreement between MHMI, PLI, and SPI (**a**), MHMI and SPI (**b**), MHMI and PLI (**c**), and PLI and SPI (**d**)

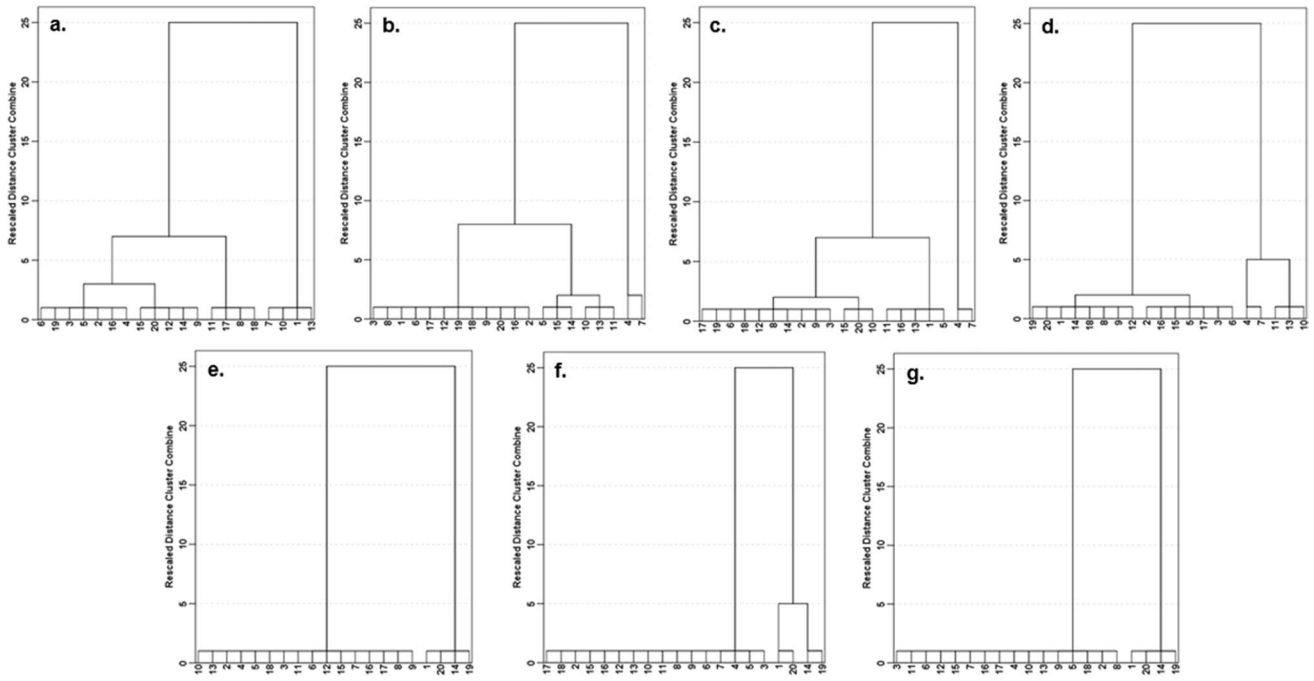

Fig. 3 Q-mode hierarchical clustering of water quality based on pH (**a**), EC (**b**), TDS (**c**), TH (**d**), MHMI (**e**), PLI (**f**), and SPI (**g**)

The groupings of the water samples on the basis of EC and TDS are presented in Fig. 3 b and c, respectively. In Fig. 3b, water samples 4 and 7 formed the first cluster, whereas the second cluster is constituted by water samples 11, 13, 10, 14, 15, 6, 2, 16, 20, 9, 18, 19, 12, 17, 6, 1, 8, and 3. While cluster 1 has the samples with the highest EC values (ranging between 73 and 102 µS/cm), cluster 2 has samples with the lowest EC scores (< 73 µS/cm) (Fig. 3b; Table 4). The negative effects of water EC would be much pronounced around the water sites that formed the first cluster than around those in cluster 2. However, the TDS dendrogram has three main clusters. Cluster 1 of this dendrogram also has samples 4 and 7 as members (Fig. 3c). However, cluster 2 is constituted by water samples 5, 1, 13, 16, and 11, while the cluster 3 has samples 10, 20, 16, 3, 9, 2, 14, 8, 12, 18, 6, 19, and 17 (Fig. 3c). It was observed that the TDS decreased from cluster 1 to cluster 3. While cluster 1 has a TDS range of 62–76 mg/L, cluster 2 has a TDS range of 28–35 mg/L, and Cluster 3 has TDS < 28 mg/L (Fig. 3c; Table 4). Accordingly, the risk of TDS in water would more likely decrease from cluster 1 to cluster 3.

Composed of two main water quality clusters, the dendrogram that classified the groundwater samples on the basis of TH is shown in Fig. 3d. While cluster 1 has five water samples as members (10, 13, 11, 4, and 7), cluster 2 has fifteen water samples (6, 3, 17, 5, 15, 16, 2, 12, 9, 8, 18, 14, 1, 20, and 19) (Fig. 3d). The first cluster represents water sites that have the higher TH concentrations (in the range of 24–52 mg/L) amongst the samples. On the other hand, the second cluster represents water sites with lower TH values < 24 mg/L (Fig. 3d; Table 4). Scale formation in water distribution networks, which is often linked to TH, would be more anticipated in the water sites that formed cluster 1. In other words, cluster 2 would pose little or no scaling (encrustation) risk to well-screen, domestic, and industrial pipes and wares. This order seems to hold true for the use of the waters for laundry. Ease of lather formation would be higher in cluster 2 than in cluster 1.

The results of the computed pollution and water quality indices were also utilized in classifying the samples. The dendrograms for the MHMI, PLI, and SPI are shown in Fig. 3 e, f, and g, respectively. Generally, all of the dendrograms have two main clusters each. In Fig. 3e, cluster 1 is represented by five samples (19, 14, 20, and 1), whereas the cluster 2 is represented by fifteen water samples (9, 8, 17, 16, 7, 15, 12, 6, 11, 3, 18, 5, 4, 2, 13, and 10). For the PLI, four water samples (19, 14, 20, and 1) were identified as the members of the first cluster while the remaining sixteen groundwater samples (3, 5, 4, 7, 6, 9, 8, 11, 10, 13, 12, 16, 15, 2, 18, and 17) formed the second cluster (Fig. 3f). Meanwhile, Fig. 3g showed that four water samples (19, 14, 20, and 1) represent cluster 1 on the basis of SPI, whereas cluster 2 is represented by sixteen groundwater samples (8, 2, 18, 5, 9, 13, 10, 4, 17, 16, 7, 15, 12, 6, 11, and 3).

With respect to the hierarchical classifications of the groundwater quality based on the MHMI, PLI, and SPI (Fig. 3e–g), a high-level agreement was noticed between the produced dendrograms. Generally, the dendrograms suggest that the first clusters of MHMI, PLI, and SPI classifications are typical representatives of the water samples which were highly influenced by PTEs and anthropogenic inputs. As can be recalled, samples 19, 14, 20, 1, and 5 were initially identified to have higher MHMI, PLI, and SPI values than other samples. On the other hand, their second clusters represent those samples with lesser contamination, as their index results were observed to be much lower than those of the first clusters (Fig. 3e–g; Tables 5 and 6). The results of the dendrograms appear to be consistent with the findings of the linear regression models.

K-means partitional clustering

The summarized result of the KMC analysis is shown in Table 7, and the bar and area charts of the extracted clusters are provided in Fig. 4. The bar and area charts were used to show the predominance of the various cluster families. Generally, it was observed that the number of clusters gotten from the KMC analysis for some of the groundwater quality parameters varied from those obtained from the Q-mode hierarchical dendrograms. According to the water quality classification on the basis of pH, five clusters were obtained (Table 7), unlike the hierarchical clustering that presented three clusters. With respect to Table 7, it was noticed that clusters 1 and 5 were the cluster 1 of the hierarchical dendrogram shown in Fig. 3a; cluster 2 of the KMC is the same as cluster 2 of the hierarchical dendrogram (Fig. 3a); and clusters 3 and 4 of the KMC were the components of cluster 3 in the pH hierarchical dendrogram (Fig. 3a).

The KMC classifications of the water quality based on EC and TDS seem to follow the same trend as those reported from the hierarchical clustering. For the EC shown in Table 7, two main clusters were identified. While cluster 1 represents cluster 2 of the hierarchical dendrogram, cluster 2 represents cluster 1 of the dendrogram (compare Table 7 and Fig. 3b). On the other hand, the KMC for TDS has two clusters (Table 7) contrary to the three generated in hierarchical dendrogram (Fig. 3c). Nevertheless, cluster 1 from the KMC represents cluster 1 of the dendrogram, whereas clusters 2 and 3 of the dendrogram merged to form cluster 2 of the KMC. Table 7 presents the KMC for TH. It was realized that cluster 1 formed in the KMC represents cluster 2 of the counterpart hierarchical dendrogram shown in Fig. 3d. However, clusters 2 and 3 of the KMC typically represent cluster 1 of the hierarchical dendrogram (Table 7; Fig. 3d).

Table 7 Result summary of K-means clustering

a. pH			b. EC			c. TDS			d. TH			e. MHMI			f. PLI			g. SPI		
Case no.	C	CD	Case no.	C	CD	Case no.	C	CD	Case no.	C	CD	Case no.	C	CD	Case no.	C	CD	Case no.	C	CD
1	5	0.100	1	1	7.222	1	2	16.222	1	1	1.533	1	1	3.138	1	1	0.017	1	1	10.481
2	3	0.150	2	1	1.222	2	2	5.778	2	1	1.467	2	2	0.269	2	2	0.000	2	2	1.009
3	3	0.050	3	1	9.222	3	2	3.778	3	1	4.467	3	2	0.229	3	2	0.000	3	2	0.511
4	3	0.050	4	2	14.500	4	1	7.000	4	2	3.500	4	2	0.089	4	2	0.000	4	2	0.459
5	3	0.050	5	1	9.778	5	2	9.222	5	1	4.467	5	2	0.072	5	2	0.000	5	2	0.342
6	3	0.050	6	1	7.222	6	2	7.778	6	1	6.467	6	2	0.295	6	2	0.000	6	2	0.519
7	5	8.882E−16	7	2	14.500	7	1	7.000	7	2	3.500	7	2	0.420	7	2	0.000	7	2	0.535
8	2	0.100	8	1	9.222	8	2	6.778	8	1	3.533	8	2	1.031	8	2	0.000	8	2	3.879
9	4	0.100	9	1	3.222	9	2	5.778	9	1	3.533	9	2	1.038	9	2	0.000	9	2	0.350
10	5	0.100	10	1	18.778	10	2	0.222	10	3	5.667	10	2	0.200	10	2	0.000	10	2	0.261
11	2	0.200	11	1	29.778	11	2	13.222	11	3	3.333	11	2	0.232	11	2	0.000	11	2	0.511
12	4	0.200	12	1	13.222	12	2	10.778	12	1	5.533	12	2	0.295	12	2	0.000	12	2	0.519
13	1	0.000	13	1	19.778	13	2	14.222	13	3	2.333	13	2	0.200	13	2	0.000	13	2	0.261
14	3	0.150	14	1	4.778	14	2	6.778	14	1	3.533	14	1	3.579	14	3	0.121	14	1	9.491
15	4	0.100	15	1	8.778	15	2	2.222	15	1	0.467	15	2	0.293	15	2	0.000	15	2	0.519
16	3	0.150	16	1	4.222	16	2	12.222	16	1	2.467	16	2	0.359	16	2	0.000	16	2	0.527
17	2	0.100	17	1	8.222	17	2	7.778	17	1	4.467	17	2	0.483	17	2	0.000	17	2	0.543
18	2	0.200	18	1	15.222	18	2	8.778	18	1	3.533	18	2	0.030	18	2	0.000	18	2	0.285
19	3	0.050	19	1	11.222	19	2	7.778	19	1	1.533	19	1	9.791	19	3	0.121	19	1	30.444
20	4	0.200	20	1	2.222	20	2	4.222	20	1	1.533	20	1	3.074	20	1	0.017	20	1	10.473

C, cluster; *CD*, cluster distance

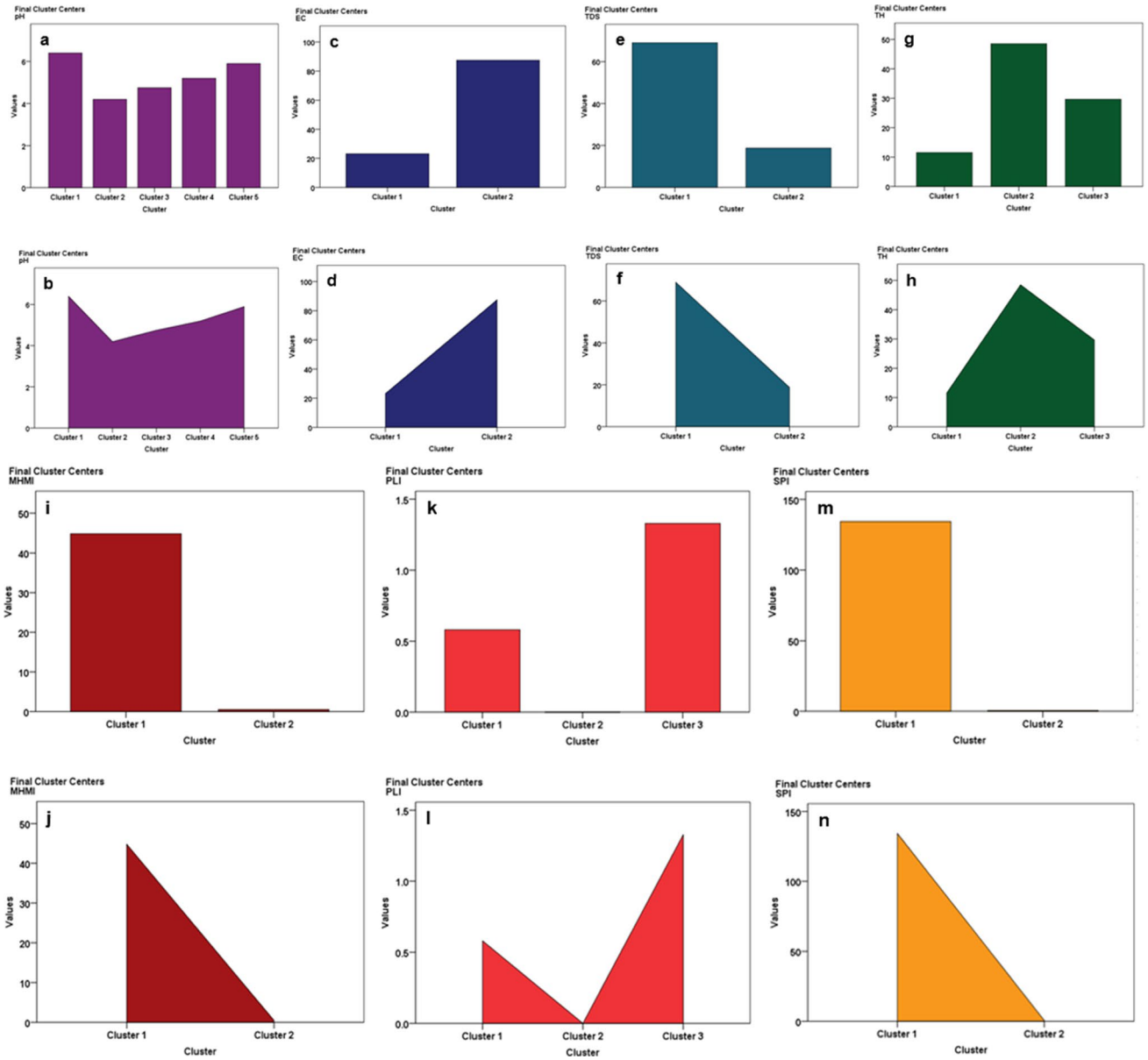

Fig. 4 Bar and area charts showing the shape of distribution and final cluster centers of K-means clustering of water quality: pH (**a**, **b**), EC (**c**, **d**), TDS (**e**, **f**), TH (**g**, **h**), MHMI (**i**, **j**), PLI (**k**, **l**), and SPI (**m**, **n**)

Furthermore, the KMC groupings for the MHMI, PLI, and SPI seem to be consistent with the respective Q-mode hierarchical dendrograms. However, the KMC for PLI (Table 7) showed three main clusters. Clusters 1 and 3 formed cluster 1 of the counterpart hierarchical dendrogram, while cluster 2 typically represents cluster 2 of the same dendrogram (Table 7; Fig. 3f). The KMC for the MHMI and SPI showed two main clusters each in Tables 7, respectively. These clusters are the same as those formed in their hierarchical dendrograms (Fig. 3e and g), respectively. It is pertinent to mention that the implications mentioned earlier for these clusters still hold true accordingly.

Soft computing methods for groundwater contamination source apportionment

Principal component analysis

The groundwater quality datasets were also transformed and reduced for the assessment and prediction of the possible influencers of the water quality using principal components (PCs) and factor loadings. Results of the non-rotated PCA are presented in Table 8. The scree plot for the selection of PCs and factor loadings is shown in Fig. 5. Seven PCs explained about 89.942% of the information regarding the

datasets. Their eigenvalues and percentages of variance were also documented (Table 8). It is realized that Na, K, Zn, Ni, Pb, PLI, MHMI, and SPI have significant loadings in PC 1. This indicates that these parameters could have closer associations amongst themselves. Dissolved Na and K could be attributed to geogenic processes such as the weathering of orthoclase minerals (Egbueri 2019a, b); Zn, Ni, and Pb are attributed to anthropogenic inputs. Ni in water environment has been attributed to mining, fuel combustion, and sewage sludge (Obasi and Akudinobi 2020). Mineral mining, fossil fuel combustion, traffic emissions, agriculture, and industrial manufacturing are some of the human activities that could release Pb in the water environment (Hanfi et al. 2020). Meanwhile, the strong correlation between PLI, MHMI, and SPI, as earlier suggested by the simple linear regression modeling and the cluster techniques, could be why they appeared alongside each other in the same PC. Moreover, in PC 1, Zn, Ni, and Pb had a higher positive loading value compared to Na^+ and K^+. It is understood that they possibly influenced the scores of the MHMI, PLI, and SPI, far more than the other water quality parameters.

In PC 2, significant loadings were observed on the EC, TDS, TH, Ca, and Mg. These parameters seem to influence the occurrences of one another. For instance, TDS measures the concentrations of solids and ions present in water. The divalent Ca and Mg concentrations in water also influence the TDS. The TDS, in turn, influences the EC and TH of water. The Ca and Mg present in the water resources may be linked to geogenic processes, such as silicate rock weathering (Kadam et al. 2021). They could also be attributed to the weathering of carbonate minerals (Egbueri et al. 2021a). Furthermore, PC 3 was noticed to have obvious loadings on pH, NO_3, and Cr (Table 8). While the pH was seen to have positive loading, the NO_3 and Cr were seen to have negative loadings. The variation in the signs could be indicating differences in their possible sources and influencers. The pH could be attributed to acid rain and geochemical processes (Ukah et al. 2020). However, NO_3 and Cr seem to be mostly influenced by anthropogenic exercises. Again, the pH possibly has little or no impact/control on the enrichment of NO_3 and Cr in the groundwater resources. Similarly, NO_3 and Cr seem to have no impact on the acidity of the water resources. Moreover, several reports have shown that NO_3 and Cr in water are influenced by agricultural activities and, therefore, it validates the similarity in the loadings of both elements (Papazotos et al. 2019; Vasileiou et al. 2019; Papazotos et al. 2020).

As the eigenvalues decreased with an increasing number of PCs, PC 4, PC 5, PC 6, and PC 7 were seen to have high loadings on Cl, SO_4, SO_4, and Fe, respectively (Table 8). Only one parameter was represented in each of the four PCs. This suggests that these parameters have little or no relationships with the other variables. This further implies that their sources/origins are different. Although geogenic processes (such as mineral-rock weathering) could be responsible for the occurrences of Cl and SO_4 in the groundwater system, they seem to have some peculiarities in the kind of minerals they originated from. Cl could be attributed to the dissolution of chlorite minerals in mudrocks, whereas SO_4 may be from the weathering of minerals like pyrite and oxidation of sulfide in soil (Egbueri et al. 2019). Fe in water could have both geogenic and human-induced origins. Due to the ease with which Fe reacts with sulphur compounds and oxygen, it is rare to find elemental Fe in free state. Thus, Fe is usually found in combined states in the form of oxides, hydroxides, sulphides, and/or carbonates, with the oxides being the most common in nature (Elinder et al. 1986; Knepper 1981). The weathering of ferruginized mineral-rock deposits, such as ironstones and siltstones, is one of the primary geogenic sources of Fe in water. Also, Fe leachate from corroding water pipes could influence Fe concentration in water. On the other hand, indiscriminate disposal of metallic wastes in dumpsites and wastewater flows from septic tanks could influence the release of Fe into groundwater system (Bader 1973; Barzegar et al. 2019; Meyer 1973; Ballentine 1972).

Varimax-rotated factor analysis

The Varimax-rotated factors extracted in this study for the association and source identification of the water quality parameters are presented in Table 8. Similar to the PCA, 89.942% of the information regarding the water quality data was explained by seven factor classes. The results of the FA seem to provide clearer insights into the associations between the water quality parameters. Factor 1 has high loadings on Zn, Ni, Pb, PLI, MHMI, and SPI (Table 8). This observation seems to validate that Zn, Ni, and Pb could have the same anthropogenic origin and that they impacted the scores of the PLI, MHMI, and SPI, more than other water quality variables. Moreover, the association existing between EC, TDS, TH, Ca, and Mg, which was initially captured by the PCA, was also reported in Factor 2. This appears to be confirmatory to the relationships existing between these parameters. Factor 3, which explained about 12.131% of the total variance, was seen to have loadings on TSS, Cl, Zn, and Ni (Table 8). This association suggests that, although Zn and Ni had been linked to anthropogenic sources, they could also have similar geogenic origin as Cl and TSS.

Factor 4 has high loadings on Na and Cr, while Factor 5 has significant loadings on pH, K, and NO_3 (Table 8). Na in Factor 4 has positive loading, whereas Cr has negative loading. This indicates that both might have different influencers. Na has been attributed to geogenic origin, while Cr concentration could more likely be influenced by human-induced

Table 8 Data dimensionality reduction extractions

a. Unrotated principal components								b. Varimax-rotated factor loadings								Communality
Parameter	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	Parameter	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	
pH	0.098	0.356	**0.530**	0.349	0.327	0.308	−0.279	pH	0.315	0.371	−0.123	0.292	**−0.515**	−0.464	0.035	0.819
EC	−0.155	**0.946**	−0.058	0.119	0.030	0.115	0.063	EC	−0.098	**0.961**	0.076	0.098	0.033	−0.068	0.001	0.955
TDS	−0.135	**0.881**	−0.026	0.314	−0.042	−0.019	−0.142	TDS	0.041	**0.938**	−0.048	−0.072	−0.152	0.012	−0.062	0.916
TSS	0.450	0.316	−0.229	−0.485	0.377	−0.248	−0.139	TSS	0.116	0.112	**0.875**	0.016	−0.030	0.076	−0.119	0.813
TH	−0.276	**0.902**	−0.155	0.134	−0.051	0.022	0.041	TH	−0.197	**0.945**	0.018	−0.039	0.048	0.026	−0.020	0.937
Na	**0.540**	0.357	0.423	−0.442	−0.130	0.029	0.266	Na	0.264	0.137	0.341	**0.771**	0.099	0.265	−0.048	0.882
K	**0.533**	0.245	−0.381	−0.196	−0.452	0.190	0.020	K	0.351	0.171	0.189	0.078	**0.570**	0.194	−0.460	0.768
Ca	−0.192	**0.924**	−0.147	0.078	−0.031	−0.097	0.135	Ca	−0.165	**0.936**	0.124	−0.001	0.071	0.137	0.064	0.946
Mg	−0.073	**0.806**	0.226	−0.076	−0.131	0.225	−0.050	Mg	−0.072	**0.754**	0.001	0.398	−0.068	−0.027	−0.212	0.782
Cl	0.471	0.177	−0.056	**−0.667**	0.293	0.107	−0.163	Cl	0.076	−0.066	**0.774**	0.322	0.040	−0.117	−0.314	0.826
SO_4	0.241	0.181	0.317	−0.148	**−0.579**	**−0.598**	0.079	SO_4	0.178	0.079	−0.034	0.258	−0.113	**0.888**	−0.073	0.912
HCO_3	0.337	0.300	0.739	−0.171	0.069	0.034	0.290	HCO_3	0.238	0.147	0.106	0.820	−0.208	0.100	0.233	0.870
NO_3	0.280	−0.087	**−0.614**	0.050	−0.229	0.459	0.454	NO_3	0.182	−0.023	−0.030	−0.101	**0.926**	−0.179	−0.027	0.935
Fe	−0.252	−0.196	0.323	0.330	0.345	−0.164	**0.612**	Fe	−0.110	−0.106	−0.225	0.108	−0.083	−0.056	**0.860**	0.835
Zn	**0.851**	0.262	−0.197	0.115	0.108	−0.341	−0.022	Zn	**0.757**	0.170	**0.534**	-−0.114	0.073	0.253	0.049	0.973
Ni	**0.803**	0.068	0.006	−0.156	0.316	0.193	0.194	Ni	**0.580**	−0.077	**0.538**	0.330	0.238	−0.206	0.093	0.849
Cr	0.417	0.283	**−0.572**	0.318	0.346	−0.378	0.107	Cr	0.421	0.317	0.487	**−0.551**	0.179	0.068	0.319	0.957
Pb	**0.895**	−0.122	0.159	0.333	−0.136	0.035	−0.119	Pb	**0.968**	−0.151	0.012	0.112	0.014	0.056	−0.092	0.985
PLI	**0.950**	−0.073	−0.121	0.117	0.074	0.101	0.074	PLI	**0.853**	−0.147	0.339	0.083	0.285	−0.061	−0.006	0.957
MHMI	**0.898**	−0.123	0.162	0.334	−0.123	0.035	−0.104	MHMI	**0.969**	−0.153	0.018	0.118	0.017	0.051	−0.075	0.986
SPI	**0.895**	−0.122	0.159	0.333	−0.135	0.035	−0.118	SPI	**0.969**	−0.151	0.014	0.113	0.015	0.055	−0.091	0.985
Eigenvalue	6.337	4.806	2.347	1.827	1.379	1.185	1.005	Eigenvalue	5.159	4.609	2.548	2.198	1.734	1.325	1.316	–
Variance %	30.176	22.887	11.178	8.699	6.569	5.645	4.788	Variance %	24.565	21.947	12.131	10.465	8.258	6.309	6.267	–
Cumulative %	30.176	53.063	64.242	72.941	79.509	85.154	89.942	Cumulative %	24.565	46.512	58.643	69.108	77.366	83.675	89.942	–

Significant loadings are in bold

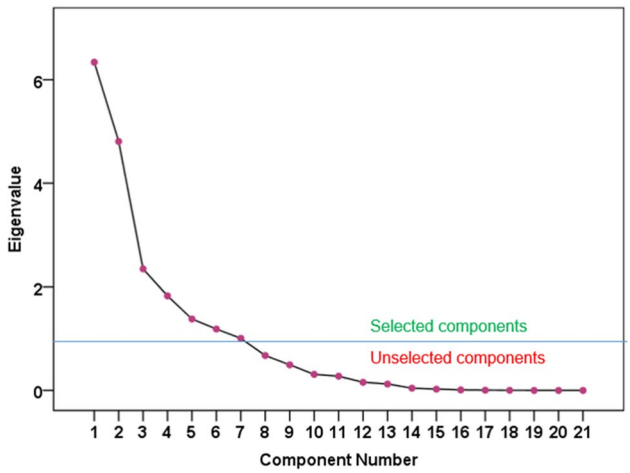

Fig. 5 Scree plot showing the selected components based on eigenvalue ≥ 1

activities. Many industrial activities such as electroplating, tanning, dyeing, painting, wood/paper processing, and batteries are known to release Cr into the environment. However, for both to occur in the same factor class could further be suggesting that Cr might also have some geogenic origins (such as chromite deposits and the weathering of mafic minerals in mudrocks) that contributed to its presence in the groundwater system. Nevertheless, literature has not indicated that there are chromite deposits in the study area. In Factor 5, it is thought that the association between K and NO_3 in water could be attributed to the use of NPK fertilizers and other nutrients from anthropogenic sources (Egbueri 2019a, b; Rahman et al. 2021). While K and NO_3 had positive loading in this factor class, pH has negative loading. This could imply that the pH has no obvious control on the release of K and NO_3 in the groundwater. Finally, Factor 6 has loading on only SO_4, whereas Factor 7 has loading on only Fe. These loadings are seen to be consistent with the reports of the PC 6 and PC 7, respectively.

Soft computing methods for groundwater quality modeling

Multiple linear regression modeling

The statistical metrices of the performance of the MLR modeling of the water quality parameters are given in Table 9. The parity plots of the models are presented in Fig. 6. The results showed that the MLR models generally performed excellently well in predicting all the variables. However, variation in performances was also noticed. Although high R, R^2, and adjusted R^2 values were obtained in all, the SEE varied (Table 9). However, with respect to the R^2 valuation, the performances of the models are in the order of MHMI = PLI =S PI > EC > TH > PH > TDS (MHMI = 1.000, PLI = 1.000, SPI = 1.000, EC = 0.996, TH = 0.992, Ph = 0.958, TDS = 0.946). Moreover, the overall results of the MLR modeling suggest that this modeling technique is suitable for the soft computation and prediction of the considered variables.

Artificial neural network modeling

The soft computation and prediction of the seven parameters were also carried out using the SCG-ANN technique. The performance metrices of the SCG-ANN models produced in this study are presented in Table 9. The parity plots of the models are presented in Fig. 7, whereas the residual error plots are shown in Fig. 8. It was also realized that the SCG-ANN models performed excellently well. Very low modeling errors were observed in the SCG-ANN models. Based on the R^2, the order of performance for the SCG-ANN models seems to be MHMI > PLI > EC > pH > SPI > TH > TDS (Table 9). Nevertheless, just like the MLR models, the SCG-ANN models have proven to be efficient and economical tools for the computation and prediction of the analyzed groundwater quality parameters.

Table 9 Performance summary of the MLR and SCG-ANN models

Predicted parameter	a. MLR modeling				b. SCG-ANN modeling			
	Multiple correlation coefficient (R)	Coefficient of determination (R^2)	Adjusted R^2	Standard error of estimates (SEE)	R^2	Sum of square errors (SOSE)	Relative error (RE)	Residual error plot
pH	0.979	0.958	0.598	0.4088	0.966	0.031	0.077	Fig. 8a
EC	0.998	0.996	0.957	4.8803	0.974	0.012	0.086	Fig. 8b
TDS	0.973	0.946	0.491	12.7611	0.914	1.5978E−05	0.006	Fig. 8c
TH	0.996	0.992	0.928	3.4750	0.950	0.004	0.182	Fig. 8d
MHMI	1.000	1.000	1.000	0.0000021	0.994	0.005	0.013	Fig. 8e
PLI	1.000	1.000	0.999	0.0137088	0.981	0.005	0.073	Fig. 8f
SPI	1.000	1.000	1.000	0.0000025	0.957	0.002	0.004	Fig. 8g

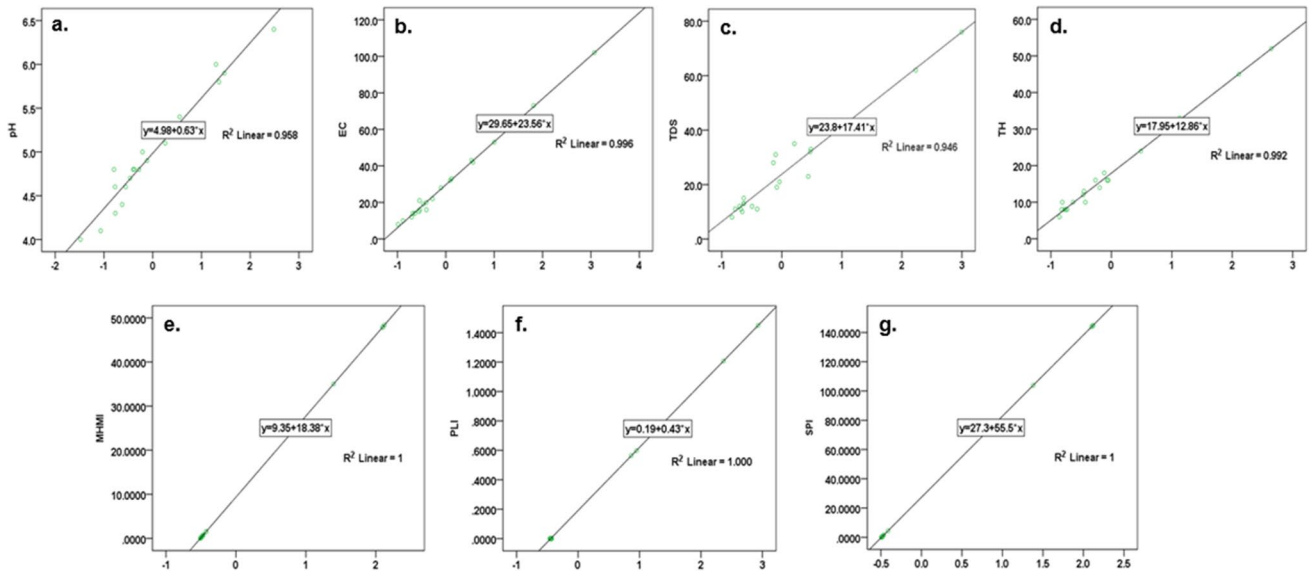

Fig. 6 Parity plots and R^2 values for the MLR prediction of pH (**a**), EC (**b**), TDS (**c**), TH (**d**), MHMI (**e**), PLI (**f**), and SPI (**g**)

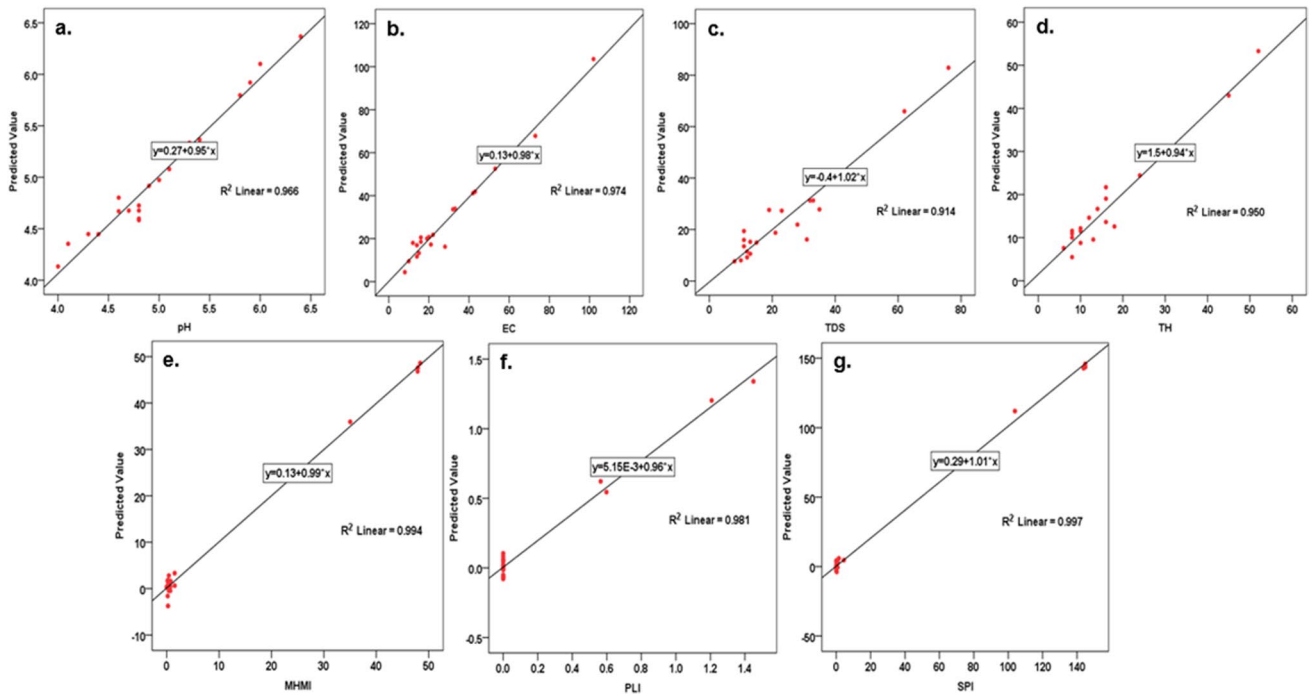

Fig. 7 Parity plots and R^2 values for the SCG-ANN prediction of pH (**a**), EC (**b**), TDS (**c**), TH (**d**), MHMI (**e**), PLI (**f**), and SPI (**g**)

Sensitivity analysis was also carried out to determine the impacts of the input variables in the prediction of the groundwater quality parameters. The results of the analysis are summarized in Fig. 9. In this study, only input variables with sensitivity score (normalized importance) above 50% were considered very important in the SCG-ANN modeling of the various water quality parameters. The sensitivity analysis produced for pH, EC, TDS, and TH showed that ten, two, three, and three input variables, respectively, significantly impacted their SCG-ANN models, respectively (Fig. 9a–d). For the PLI model, Pb, Ni, Cr, and Mg contributed more to its SCG-ANN performance (Fig. 9f). However, for the MHMI and SPI SCG-ANN predictions, only Pb was noticed to be the most important predictor (Fig. 9e and g).

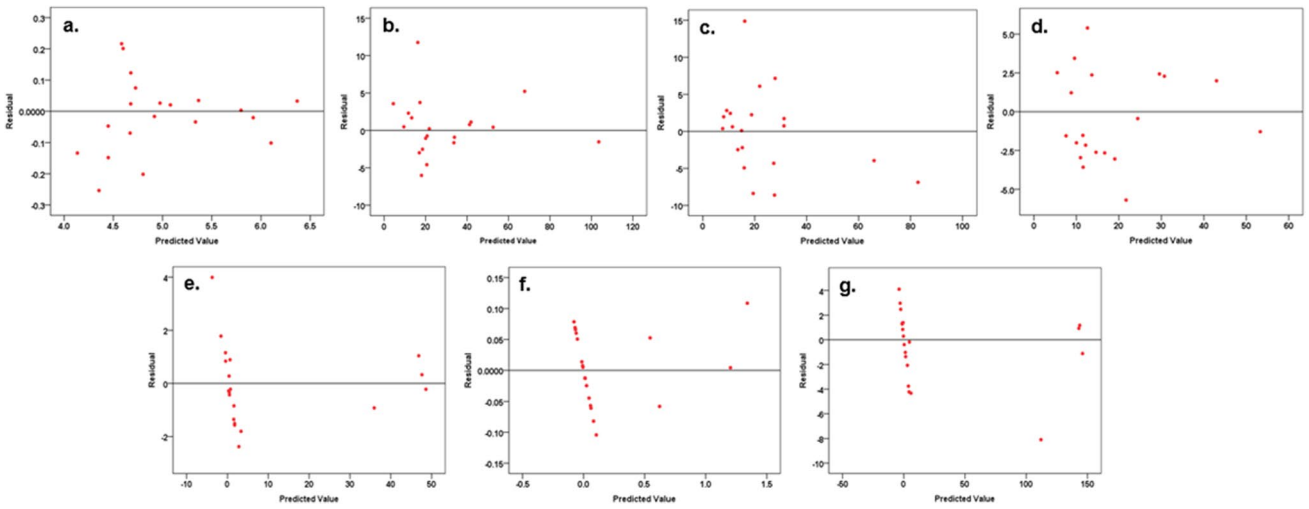

Fig. 8 Parity and residual error plots for the SCG-ANN prediction of pH (**a**), EC (**b**), TDS (**c**), TH (**d**), MHMI (**e**), PLI (**f**), and SPI (**g**)

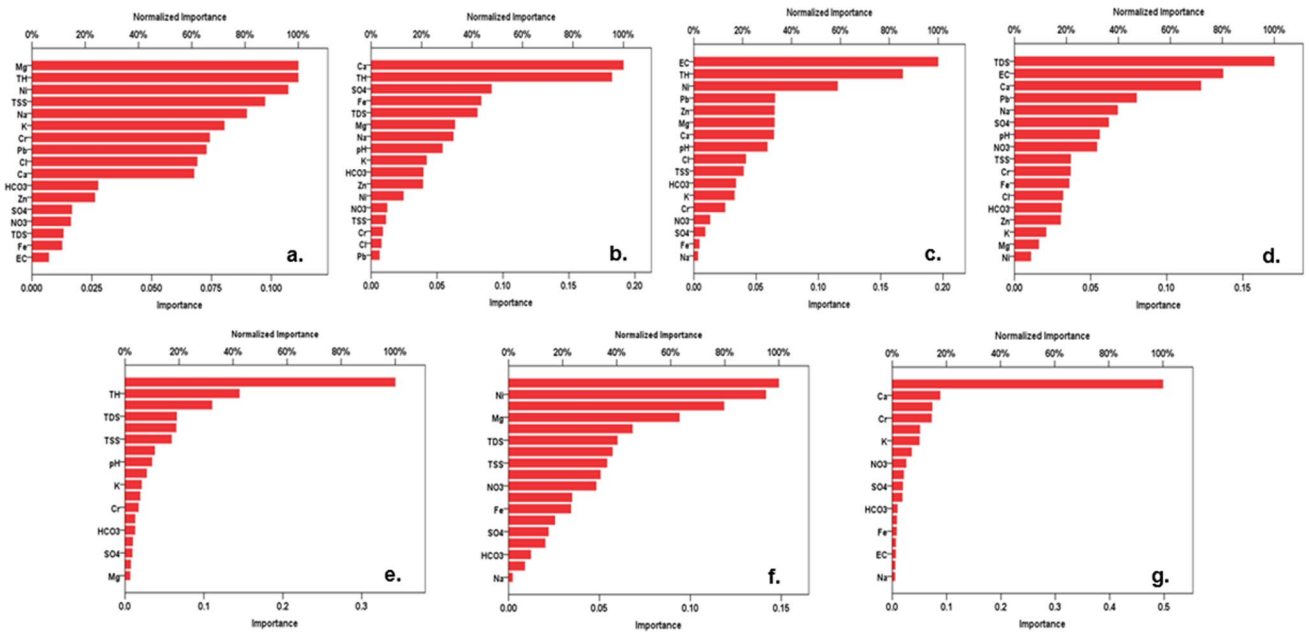

Fig. 9 Bar charts showing the sensitivities of input variables on the SCG-ANN prediction of pH (**a**), EC (**b**), TDS (**c**), TH (**d**), MHMI (**e**), PLI (**f**), and SPI (**g**)

Comparison of accuracies of MLR and SCG-ANN models

It is important to have accurate models that could predict the parameters of interest, in order to save cost of monitoring and assessment of the groundwater quality. In this study, both MLR and ANN techniques have proven to be reliable tools for the computation and prediction of pH, EC, TDS, TH, MHMI, PLI, and SPI. Although both techniques have performed very well, it is also considered necessary to determine the one that outperformed the other. With respect to the R^2 ratings obtained in both MLR and ANN modeling of the groundwater parameters, the SCG-ANN and the MLR models performed equally in the prediction of the water quality indices (MHMI, SPI, and PLI). However, the MLR performed better than the SCG-ANN model in predicting EC, TH, and TDS. Nevertheless, the ANN model predicted the pH better than the MLR model.

Conclusions

This paper has assessed the potentials of several soft computing techniques in the assessment, computation, and prediction of several groundwater quality parameters, including pH, EC, TDS, TH, MHMI, PLI, and SPI, in southeastern Nigeria. The results of the physicochemical analysis, PLI, MHMI, and SPI revealed that about 20–25% of the groundwater is polluted and, thus, unsuitable for drinking. Meanwhile, about 80% of the water samples were found in excellent condition; hence, they are considered suitable for drinking. Principal component and Varimax-rotated factor analyses were utilized for source apportionment, and their findings suggest that both geogenic processes and human-induced exercises are responsible for the water quality deterioration. The current study integrated different soft computing algorithms, and the results have confirmed that using multiple models usually provides robust and better judgments than using standalone model, which could easily lead to bias, and perhaps, costly assumptions. The findings of this paper would enhance groundwater monitoring and assessment programs in the area and could also provide baseline information for future water quality simulation and prediction, locally and internationally, based on the studied parameters. It is recommended that the inhabitants of areas with water samples identified to be unsuitable for human consumption should apply reliable water treatment strategies before use or search for alternative source of water resources. Government should set up policies in these areas to check indiscriminate waste disposal. It is also recommended that future water resources research should apply more advanced methods and also develop new models for computation, assessment, and prediction of the water quality parameters considered in the present paper.

Acknowledgements The author expresses gratitude to the anonymous-reviewers and the editor, for their useful review comments.

Author contribution Johnbosco C. Egbueri: Conceptualization, manuscript design, machine learning modeling, indexical computation, data analysis, manuscript writing, editing, review and revision; Johnson C. Agbasi: machine learning modeling, indexical computation, editing, manuscript review and revision

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

Akpoborie IA, Nfor BN, Etobro AAI, Odagwe S (2011) Aspects of the geology and groundwater conditions of Asaba, Nigeria. Arch Appl Sci Res 3(2):537–550

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 (2017a) 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):e11792

Alizamir M, Sobhanardakani S (2017b) Predicting arsenic and heavy metals contamination in groundwater resources of Ghahavand plain based on an artificial neural network optimized by imperialist competitive algorithm. Environ Health Eng Manag J 4(4):225–231

Alizamir M, Sobhanardakani S (2018) An artificial neural network - particle swarm optimization (ANN- PSO) approach to predict heavy metals contamination in groundwater resources. Jundishapur J Health Sci. 10(2):e67544. https://doi.org/10.5812/jjhs.67544

Alizamir M, Sobhanardakani S, Hasanalipour Shahrabadi A (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 Mil Health Sci Res 17(4):e98554. 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):207–216

Alsubih M, El Morabet R, Khan RA, Khan NA, Khan MU, Ahmed S, Qadir A, Changani F (2021) Occurrence and health risk assessment of arsenic and heavy metals in groundwater of three industrial areas in Delhi, India. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-021-15062-3

Amiri V, Rezaei M, Sohrabi N (2014) Groundwater quality assessment using entropy weighted water quality index (EWQI) in Lenjanat, Iran. Environ Earth Sci 72(9):3479–3490

APHA (2005) Standard methods for the examination of water and wastewater, 21st edn. American Public Health Association, Washington DC

Arnedo-Pena A, Bellido-Blasco J, Puig-Barbera J, Artero-Civera A, Campos-Cruañes JB, Pac-Sa MR, Villamarín-Vázquez JL, Felis-Dauder C (2007) Domestic water hardness and prevalence of atopic eczema in Castellon (Spain) School Children. Salud Pública De México 492(4):295–301

Arua I (1986) Paleoenvironment of Eocene deposits in the Afikpo syncline, southern Nigeria. J Afr Earth Sci 5:279–284

Asadollahfardi G, Taklify A, Ghanbari A (2012) Application of artificial neural network to predict TDS in Talkheh Rud River. J Irrig Drain Eng 138(4):363–370

Avci H, Dokuz UE, Avci AS (2018) Hydrochemistry and groundwater quality in a semiarid calcareous area: an evaluation of major ion chemistry using a stoichiometric approach. Environ Monit Assess 190:641

Bader J (1973) Ground-water contamination. U. S. Geological Survey, Washington DC, The United State of America and Puerto Rica, p 103

Bai J, Cui B, Chen B, Zhang K, Deng W, Gao H, Xiao R (2011) Spatial distribution and ecological risk assessment of heavy metals in surface sediments from a typical plateau lake wetland, China. Ecol Model 222:301–306

Ballentine R (1972) Subsurface pollution problems in the United States. Tech. Studies Rept. TS-00-72-02. U. S. Environmental Protection Agency, Washington DC, p 29

Barzegar R, Asghari Moghaddam A, Adamowski J, Ozga-Zielinski B (2017) Multi-step water quality forecasting using a boosting ensemble. Stoch Env Res Risk Assess. https://doi.org/10.1007/s00477-017-1394-z

Barzegar R, Moghaddam AA, Adamowski J, Nazemi AM (2019) Assessing the potential origins and human health risks of trace elements in groundwater: a case study in the Khoy plain, Iran. Environ Geochem Health 41(2):981–1002. https://doi.org/10.1007/s10653-018-0194-9

Başyiğit B, Tekin-Özan S (2013) Concentrations of some heavy metals in water, sediment, and tissues of pikeperch (Sander lucioperca) from Karataş Lake related to physicochemical parameters, fish size, and seasons. Pol J Environ Stud 22(3):633–644

Belkhiri L, Mouni L, Tiri A, Narany TS, Nouibet R (2018) Spatial analysis of groundwater quality using self-organizing maps. Groundw Sustain Dev 7:121–132

Chatterjee S, Sarkar S, Dey N, Ashour AS, Sen S, Hassanien AE (2017) Application of cuckoo search in water quality prediction using artificial neural network. Int J Comput Intell Stud 6(2–3):229–244

Chen H, Xu L, Ai W, Lin B, Feng Q, Cai K (2020) Kernel functions embedded in support vector machine learning models for rapid water pollution assessment via near infrared spectroscopy. Sci Total Environ 714:136765. https://doi.org/10.1016/j.scitotenv.2020.136765

Chen T, Zhang H, Sun C, Li H, Gao Y (2018) Multivariate statistical approaches to identify the major factors governing groundwater quality. Appl Water Sci. https://doi.org/10.1007/s13201-018-0837-0

Chen W, Liu W (2015) Water quality modeling in reservoirs using multivariate linear regression and two neural network models. Adv Artif Neural Syst. https://doi.org/10.1155/2015/521721

Egbueri JC (2018) Assessment of the quality of groundwaters proximal to dumpsites in Awka and Nnewi metropolises: a comparative approach. Int J Energy Water Res. https://doi.org/10.1007/s42108-018-0004-1

Egbueri JC (2019a) Evaluation and characterization of the groundwater quality and hydrogeochemistry of Ogbaru farming district in southeastern Nigeria. SN Appl Sci. https://doi.org/10.1007/s42452-019-0853-1

Egbueri JC (2019b) Water quality appraisal of selected farm provinces using integrated hydrogeochemical, multivariate statistical, and microbiological technique. Model Earth Syst Environ 5(3):997–1013. https://doi.org/10.1007/s40808-019-00585-z

Egbueri JC (2020) Groundwater quality assessment using pollution index of groundwater (PIG), ecological risk index (ERI) and hierarchical cluster analysis (HCA): a case study. Groundw Sustain Dev 10:100292. https://doi.org/10.1016/j.gsd.2019.100292

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. https://doi.org/10.1007/s11356-021-13678-z

Egbueri JC, Ezugwu CK, Ameh PD, Unigwe CO, Ayejoto DA (2020) Appraising drinking water quality in Ikem rural area (Nigeria) based on chemometrics and multiple indexical methods. Environ Monit Assess 192(5):308. https://doi.org/10.1007/s10661-020-08277-3

Egbueri JC, Ezugwu CK, Unigwe CO, Onwuka OS, Onyemesili OC, Mgbenu CN (2021a) Multidimensional analysis of the contamination status, corrosivity and hydrogeochemistry of groundwater from parts of the Anambra Basin, Nigeria. Anal Lett 54(13):2126–2156. https://doi.org/10.1080/00032719.2020.1843049

Egbueri JC, Mgbenu CN, Chukwu CN (2019) Investigating the hydrogeochemical processes and quality of water resources in Ojoto and environs using integrated classical methods. Model Earth Syst Environ 5(4):1443–1461. https://doi.org/10.1007/s40808-019-00613-y

Egbueri JC, Mgbenu CN, Digwo DC, Nnyigide CS (2021c) A multi-criteria water quality evaluation for human consumption, irrigation and industrial purposes in Umunya area, southeastern Nigeria. Int J Environ Anal Chem. https://doi.org/10.1080/03067319.2021.1907360

Egbueri JC, Unigwe CO (2019) An integrated indexical investigation of selected heavy metals in drinking water resources from a coastal plain aquifer in Nigeria. SN Appl Sci. https://doi.org/10.1007/s42452-019-1489-x

Egbueri JC, Unigwe CO, Omeka ME, Ayejoto DA (2021b) Urban groundwater quality assessment using pollution indicators and multivariate statistical tools: a case study in southeast Nigeria. Int J Environ Anal Chem. https://doi.org/10.1080/03067319.2021.1907359

Elinder C, Iron I, Friberg L, Nordberg G, Vouk V (1986) Handbook on the technology of metals. Elsevier, Amsterdam, pp 276–297

El-Sayed SA, Moussa EMM, El-Sabagh MEI (2017) Evaluation of heavy metal content in Qaroun Lake, El-Fayoum, Egypt. Part I: Bottom sediments. J Radiat Res Appl Sci 8:276–285

Enyigwe MT, Onwuka OS, Egbueri JC (2021) Geochemical distribution, statistical and health risk assessment of toxic elements in groundwater from a typical mining district in Nigeria. Environ Forensics. https://doi.org/10.1080/15275922.2021.1907822

Gad M, Abou El-Safa MM, Farouk M, Hussein H, Alnemari AM, Elsayed S, Khalifa MM, Moghanm FS, Eid EM, Saleh AH (2021) Integration of water quality indices and multivariate modeling for assessing surface water quality in Qaroun Lake, Egypt. Water 13:2258. https://doi.org/10.3390/w13162258

Garg RK, Rao RJ, Uchchariya D, Shukla G, Saksena DN (2010) Seasonal variations in water quality and major threats to Ramsagar reservoir, India, Afr J Environ Sci Technol 4(2)

Gaya MS, Abba SI, Abdu AM, Tukur AI, Saleh AM, Esmaili P, Wahab NA (2020) Estimation of water quality index using artificial intelligence approaches and multi-linear regression. IAES Int. J Artif Intell 9(1):126–134. https://doi.org/10.11591/ijai.v9.i1.pp126-134

Ghobadi A, Cheraghi M, Sobhanardakani S (2021) Groundwater quality modeling using a novel hybrid data-intelligence model based on gray wolf optimization algorithm and multi-layer perceptron artificial neural network: a case study in Asadabad Plain, Hamedan. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-021-16300-4

Goyal RR, Patel H, Mane SJ (2015) Artificial neural network: an effective tool for predicting water quality for Kalyan-Dombivali municipal corporation. Int J Sci Res 4(6):2863–2866

Håkanson L (1980) An ecological risk index for aquatic. Pollution control: a sedimentological approach. Water Research 14:975–1001

Hameed M, Sharqi SS, Yaseen ZM, Afan HA, Hussain A, Elshafie A (2016) Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in tropical region. Malaysia. Neural Comput Applic 28(S1):1–13. https://doi.org/10.1007/s00521-016-2404-7

Hanfi MY, Mostafa MYA, Zhukovsky MV (2020) Heavy metal contamination in urban surface sediments: sources, distribution, contamination control, and remediation. Environ Monit Assess 192(1):32

Haritash AK, Kaushik CP, Kaushik A (2008) Suitability assessment of groundwater for drinking, irrigation and industrial use in some North Indian villages. Environ Monit Assess 145:397–406. https://doi.org/10.1007/s10661-007-0048-x

Hong H, Panahi M, Shirzadi A, Ma T, Liu J, Zhu AX, Kazakis N (2018) Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution. Sci Total Environ 621:1124–1141. https://doi.org/10.1016/j.scitotenv.2017.10.114

Idriss IEA, Abdel-Azim M, Karar KI, Osman S, Idris AM (2020) Isotopic and chemical facies for assessing the shallow water table aquifer quality in Goly Region, White Nile State, Sudan: focusing on nitrate source apportionment and human health risk. Toxin Rev:1–13. https://doi.org/10.1080/15569543.2020.1775255

Kadam A, Wagh V, Jacobs J, Patil S, Pawar N, Umrikar B, Sankhua R, Kumar S (2021) Integrated approach for the evaluation of groundwater quality through hydro geochemistry and human health risk from Shivganga river basin. Environ Sci Pollut Res, Pune, Maharashtra, India. https://doi.org/10.1007/s11356-021-15554-2

Kadam AK, Wagh VM, Muley AA, Umrikar BN, Sankhua RN (2019) Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin. Model Earth Syst Environ, India. https://doi.org/10.1007/s40808-019-00581-3

Kargar K, Samadianfard S, Parsa J, Nabipour N, Shamshirband S, Mosavi A, Chau KW (2020) Estimating longitudinal dispersion coefficient in natural streams using empirical models and machine learning algorithms. Eng Applic Comput Fluid Mech 14(1):311–322. https://doi.org/10.1080/19942060.2020.1712260

Knepper W (1981) Iron. In: Kirk-Othmer encyclopedia of chemical technology, vol 13. Wiley Interscience, New York, pp 735–753

Kogbe CA (1976) Paleographic history of Nigeria from Albian Times. In: Kogbe CA (ed) Geology of Nigeria. Elizabethan Publishers, Lagos

Kožíšek F (2003) *Health significance of drinking water calcium and magnesium*. National Institute of Public Health, Šrobárová, Slovakia

Kükrer S, Mutlu E (2019) Assessment of surface water quality using water quality index and multivariate statistical analyses in Saraydüzü Dam Lake, Turkey. Environ Monit Assess 191:71–87

Kumar S, Shirke KD, Pawar NJ (2008) GIS-based colour composites and overlays to delineate heavy metal contamination zones in the shallow alluvial aquifers, Ankaleshwar industrial estate, south Gujarat, India. Environ Geol 54:117–129. https://doi.org/10.1007/s00254-007-0799-2

Li P, Feng W, Xue C, Tian R, Wang S (2017) Spatiotemporal variability of contaminants in lake water and their risks to human health: a case study of the Shahu Lake tourist area, northwest China. Expo Health 9(3):213–225

Li P, Qian H, Wu J (2010) Groundwater quality assessment based on improved water quality index in Pengyang County, Ningxia, North west China. J Chem 7:209–216

Lower S (2007) Hard water and water softening [online] https://www.chem1.com/CQ/hardwater.html. Accessed 30th August 2021

Luo D, Guo Q, Wang X (2003) Simulation and prediction of underground water dynamics based on RBF neural network. Acta Geosci Sin 24:475–478

Mahmoudi N, Orouji, Fallah-Mehdipour E (2016) Integration of shuffled frog leaping algorithm and support vector regression for prediction of water quality parameters. Water Resour Manage. https://doi.org/10.1007/s11269-016-1280-3

Maier HR, Morgan N, Chow CWK (2004) Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters. Environ Model Softw 19(5):485–494. https://doi.org/10.1016/S1364-8152(03)00163-4

Maroufpoor S, Jalali M, Nikmehr S, Shiri N, Shiri J, Maroufpoor E (2020) Modeling groundwater quality by using hybrid intelligent and geostatistical methods. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-020-09188-z

Marque S, Jacqmin-Gadda H, Dartigues JF, Commenges D (2003) Cardiovascular mortality and calcium and magnesium in drinking water: an ecological study in elderly people. Eur J Epidemiol 18(4):305–309

Matthess G (1982) The properties of groundwater, vol No. 551.49 M38. Wiley, New York

McGowan W (2000) Water processing: residential, commercial, light industrial, 3rd edn. Water Quality Association, Lisle

McNally NJ, Williams HC, Phillips DR, Smallman-Raynor M, Lewis S, Venn A, Britton J (1998) Atopic eczema and domestic water hardness. Lancet 352(9127):527–531

McNeil VH, Cox ME (2000) Relationship between conductivity and analysed composition in a large set of natural surface-water samples, Queensland, Australia. Environ Geol 39(12):1325–1333

Meyer C (1973) Polluted ground: some causes, effects, controls, and monitoring. Environmental Protection Agency, Washington, D. C, U.S, p 282

Miranda J, Krishnakumar G (2015) Microalgal diversity in relation to the physicochemical parameters of some industrial sites in Mangalore. South India. Environ Monit Assess 187(11):664. https://doi.org/10.1007/s10661-015-4871-1

Miyake Y, Yokoyama T, Yura A, Iki M, Shimizu T (2004) Ecological association of water hardness with prevalence of childhood atopic dermatitis in a Japanese urban area. Environ Res 94(1):33–37

Mohammadpour R, Shaharuddin S, Zakaria NA, Ghani AA, Vakili M, Chan NW (2016) Prediction of water quality index in free surface constructed wetlands. Environ Earth Sci 75:139. https://doi.org/10.1007/s12665-015-4905-6

Mukate S, Panaskar D, Wagh V, Muley A, Jangam C, Pawar R (2017) Impact of anthropogenic inputs on water quality in Chincholi industrial area of Solapur, Maharashtra, India. Groundw Sustain Dev 7:359–371. https://doi.org/10.1016/j.gsd.2017.11.001

Mukate S, Wagh V, Panaskar D, Jacobs JA, Sawant A (2019) Development of new integrated water quality index (IWQI) model to evaluate the drinking suitability of water. Ecol Indic 101:348–354

Najafzadeh M, Ghaemi A (2019) Prediction of the five-day biochemical oxygen demand and chemical oxygen demand in natural streams using machine learning methods. Environ Monit Assess 191:380. https://doi.org/10.1007/s10661-019-7446-8

Nwachukwu SO (1972) The tectonic evolution of the southern portion of the Benue Trough, Nigeria. Geol Mag 109:411–419

Nwajide CS (2013) Geology of Nigeria's sedimentary basins. CSS Press, Lagos

Nfor BN, Olobaniyi SB, Ogala JE (2007) Extent and distribution of groundwater resources in parts of Anambra State, Southeastern Nigeria. J Appl Sci Environ Manag 11(2):215–221

Obasi PN, Akudinobi BB (2020) Potential health risk and levels of heavy metals in water resources of lead–zinc mining communities of Abakaliki, southeast Nigeria. Appl Water Sci 10(7):1–23

Okoro EI, Egboka BCE, Anike OL, Enekwechi EK (2010b) Evaluation of groundwater potentials in parts of the Escarpment area of southeastern Nigeria. Int J Geomat Geosci 1(3):544–551

Okoro EI, Egboka BCE, Onwuemesi AG (2010a) Evaluation of the aquifer characteristics of the Nanka Sand using hydrogeological method in combination with vertical electric sounding (VES). J Appl Sci Environ Manag 14(2):5–9

Orouji H, Haddad OB, Fallah-Mehdipour E, Mariño MA (2013) Modeling of water quality parameters using data-driven models. J Environ Eng 139(7):947–957. https://doi.org/10.1061/(ASCE)EE.1943-7870.0000706

Ozel HU, 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. https://doi.org/10.1007/s11356-020-10156-w

Pan C, Ng KTW, Fallah B, Richter A (2019) Evaluation of the bias and precision of regression techniques and machine learning

approaches in total dissolved solids modeling of an urban aquifer. Environ Sci Pollut Res 26(2):1821–1833. https://doi.org/10.1007/s11356-018-3751-y

Papazotos P (2021) Potentially toxic elements in groundwater: a hotspot research topic in environmental science and pollution research. Environ Sci Pollut Res 28(35):47825–47837. https://doi.org/10.1007/s11356-021-15533-7

Papazotos P, Vasileiou E, Perraki M (2019) The synergistic role of agricultural activities in groundwater quality in ultramafic environments: the case of the Psachna basin, central Euboea. Greece. Environ Monit Assess 191(5):317. https://doi.org/10.1007/s10661-019-7430-3

Papazotos P, Vasileiou E, Perraki M (2020) Elevated groundwater concentrations of arsenic and chromium in ultramafic environments controlled by seawater intrusion, the nitrogen cycle, and anthropogenic activities: the case of the Gerania Mountains, NE Peloponnese, Greece. Appl Geochem 121:104697. https://doi.org/10.1016/j.apgeochem.2020.104697

Pham QC, Mohammadpour R, Linh NTT, Mohajane M, Pourjasem A, Sammen SS, Anh DT, Nam VT (2020) Application of soft computing to predict water quality in wetland. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-020-10344-8

Pisciotta A, Cusimano G, Favara R (2015) Groundwater nitrate risk assessment using intrinsic vulnerability methods: a comparative study of environmental impact by intensive farming in the Mediterranean region of Sicily, Italy. J Geochem Explor 156:89–100

Pocock SJ, Shaper AG, Packham RF (1981) Studies of water quality and cardiovascular disease in the United Kingdom. Sci Total Environ 18:25–34

Pourret O, Bollinger JC, Hursthouse A (2021) Heavy metal: a misused term? Acta Geochim 40:466–471. https://doi.org/10.1007/s11631-021-00468-0

Pourret O, Hursthouse A (2019) It's time to replace the term "heavy metals" with "potentially toxic elements" when reporting environmental research. Int J Environ Res Public Health. https://doi.org/10.3390/ijerph16224446

Proshad R, Zhang D, Idris AM, Islam MS, Kormoker T, Sarker MNI, Khadka S, Sayeed A, Islam M (2021) Comprehensive evaluation of chemical properties and toxic metals in the surface water of Louhajang River. Environ Sci Pollut Res, Bangladesh. https://doi.org/10.1007/s11356-021-14160-6

Rahman A, Mondal NC, Tiwari KK (2021) Anthropogenic nitrate in groundwater and its health risks in the view of background concentration in a semi-arid area of Rajasthan. India. Sci Rep 11:9279. https://doi.org/10.1038/s41598-021-88600-1

Reyment RA (1965) Aspects of the geology of Nigeria: the stratigraphy of the cretaceous and Cenozoic deposits. Ibadan University Press, Ibadan

Rezaei Raja O, Sobhanardakani S, Cheraghi M (2016) Health risk assessment of citrus contaminated with heavy metals in Hamedan city, potential risk of Al and Cu. Environ Health Eng Manag J 3(3):131–135

Roy R, Majumder M (2018a) A quick prediction of hardness from water quality parameters by artificial neural network. Int J Environ Sustain Dev 17(2/3):247–257

Roy R, Majumder M (2018b) Empirical modelling of total suspended solids from turbidity by polynomial neural network in north eastern India. Desalin Water Treat 132:75–78

Roy R, Majumder M (2019) Assessment of water quality trends in Loktak Lake, Manipur. India. Environ Earth Sci 78:383. https://doi.org/10.1007/s12665-019-8383-0

Roy R, Majumder M, Barman RN (2017) Assessment of water quality by RSM and ANP: a case study in Tripura. India. Asian J Water Environ Pollut 14(1):51–58. https://doi.org/10.3233/AJW-170006

Rubenowitz E, Axelsson G, Rylander R (1999) Magnesium and calcium in drinking water and death from acute myocardial infarction in women. Epidemiology 10(1):31–36

Rupakheti D, Tripathee L, Kang S, Sharma CM, Paudyal R, Sillanpää M (2017) Assessment of water quality and health risks for toxic trace elements in urban Phewa and remote Gosainkunda lakes. Nepal. Hum Ecol Risk Assess 23:959–973

Saravanakumar K, Kumar RR (2011) Analysis of water quality parameters of groundwater near Ambattur industrial area, Tamil Nadu. India. Indian J Sci Technol 4(5):660–662

Sawyer GN, McCarthy DL (1967) Chemistry of sanitary engineers, 2nd edn. McGraw Hill, New York

Selvaraj A, Saravanan S, Jennifer JJ (2020) Mamdani fuzzy based decision support system for prediction of groundwater quality: an application of soft computing in water resources. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-020-08803-3

Senapati T, Samanta P, Roy R, Sasmal T, Ghosh AR (2021) Artificial neural network: an alternative approach for assessment of biochemical oxygen demand of the Damodar River, West Bengal, India. In: Intelligent Environmental Data Monitoring for Pollution Management. Elsevier Inc. https://doi.org/10.1016/B978-0-12-819671-7.00010-5

Setshedi KJ, Mutingwende N, Ngqwala NP (2021) The use of artificial neural networks to predict the physicochemical characteristics of water quality in three district municipalities, Eastern Cape Province, South Africa. Int J Environ Res Public Health 18:5248. https://doi.org/10.3390/10.3390/ijerph18105248

Shah MI, Alaloul WS, Alqahtani A, Aldrees A, Musarat MA, Javed MF (2021) Predictive modeling approach for surface water quality: development and comparison of machine learning models. Sustainability 13:7515. https://doi.org/10.3390/su13147515

Shamshirband S, Jafari Nodoushan E, Adolf JE, Abdul Manaf A, Mosavi A, Chau KW (2019) Ensemble models with uncertainty analysis for multi-day ahead forecasting of chlorophyll a concentration in coastal waters. Eng Applic Comput Fluid Mech 13(1):91–101. https://doi.org/10.1080/19942060.2018.1553742

Sibanda T, Chigor VN, Koba S, Obi CL, Okoh AI (2014) Characterisation of the physicochemical qualities of a typical rural-based river: Ecological and public health implications. Int J Environ Sci Technol 11(6):1771–1780. https://doi.org/10.1007/s13762-013-0376-z

Sobhanardakani S (2019) Ecological and human health risk assessment of heavy metal content of atmospheric dry deposition, a case study: Kermanshah. Iran. Biol Trace Elem Res 187:602–610. https://doi.org/10.1007/s12011-018-1383-1

Sobhanardakani S, Tayebi L, Hosseini SV (2018) Health risk assessment of arsenic and heavy metals (Cd, Cu, Co, Pb, and Sn) through consumption of caviar of Acipenser persicus from Southern Caspian Sea. Environ Sci Pollut Res 25:2664–2671. https://doi.org/10.1007/s11356-017-0705-8

Sobhanardakani S (2017) Potential health risk assessment of heavy metals via consumption of caviar of Persian sturgeon. Marine Pollut Bull 123(1-2):34–38. https://doi.org/10.1016/j.marpolbul.2017.09.033

Solangi GS, Siyal AA, Babar MM, Siyal P (2019) Evaluation of drinking water quality using the water quality index (WQI), the synthetic pollution index (SPI) and geospatial tools in Thatta district, Pakistan. Desalin Water Treat. https://doi.org/10.5004/dwt.2019.24241

SON (Standard Organization of Nigeria) (2015) Nigerian-standard for-drinking-water-quality-NIS-554-2015 (pp. 1–28)

Srinivas Y, Aghil TB, Hudson Oliver D, Nithya Nair C, Chandrasekar N (2017) Hydrochemical characteristics and quality assessment of groundwater along the Manavalakurichi coast, Tamil Nadu, India. Appl Water Sci 7:1429–1438. https://doi.org/10.1007/s13201-015-0325-8

Subramani T, Elango L, Damodarasamy SR (2005) Groundwater quality and its suitability for drinking and agricultural use in Chithar River Basin, Tamil Nadu. India. Environ Geol 47(8):1099–1110

Sun K, Rajabtabar M, Samadi SZ, Rezaie-Balf M, Ghaemi A, Band SS, Mosavi A (2021) An integrated machine learning, noise suppression, and population-based algorithm to improve total dissolved solids prediction. Eng Applic Comput Fluid Mech 15(1):251–271. https://doi.org/10.1080/19942060.2020.1861987

Swathi L, Lokeshappa B (2015) Artificial neural networks application in prediction of water quality. Int J Innov Res Sci Eng Technol 4(8):6911–6916

Tomlinson DL, Wilson JG, Harris CR, Jeffrey DW (1980) Problems in the assessment of heavy-metals levels in estuaries and the formation of a pollution index. Helgoländer Meeresuntesuchungen. 33(1-4):566–575

Tyagi S, Sharma B, Singh P, Dobhal R (2013) Water quality assessment in terms of water quality index. Am J Water Resour 1(3):34–38

Ukah BU, Ameh PD, Egbueri JC, Unigwe CO, Ubido OE (2020) Impact of effluent-derived heavy metals on the groundwater quality in Ajao industrial area, Nigeria: an assessment using entropy water quality index (EWQI). Int J Energ Water Res. https://doi.org/10.1007/s42108-020-00058-5

Vasileiou E, Papazotos P, Dimitrakopoulos D, Perraki M (2019) Expounding the origin of chromium in groundwater of the Sarigkiol basin, Western Macedonia, Greece: a cohesive statistical approach and hydrochemical study. Environ Monit Assess 191(8):509. https://doi.org/10.1007/s10661-019-7655-1

Verma AK, Singh TN (2013) Prediction of water quality from simple field parameters. Environ Earth Sci 69:821–829. https://doi.org/10.1007/s12665-012-1967-6

Vetrimurugan E, Elango L, Rajmohan N (2013) Sources of contaminants and groundwater quality in the coastal part of a river delta. Int J Environ Sci Technol 10:473–486. https://doi.org/10.1007/s13762-012-0138-3

Wagh V, Mukate S, Muley A, Kadam A, Panaskar D, Varade A (2020) Study of groundwater contamination and drinking suitability in basaltic terrain of Maharashtra, India through PIG and multivariate statistical techniques. J Water Supply Res Technol AQUA 69(4):398–414

Ward JH (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58(301):236–244

Weber-Scannell PK, Duffy LK (2007) Effects of total dissolved solids on aquatic organism: a review of literature and recommendation for salmonid species. Am J Environ Sci. https://doi.org/10.3844/ajessp.2007.1.6

WHO (2011) Hardness in Drinking-water: background document for development of WHO guidelines for drinking-water quality. World Health Organization, Geneva

WHO (2017) Guidelines for drinking water quality, 3rd edn. World Health Organization, Geneva

Yaseen ZM, Ramal MM, Diop L, Jaafar O, Demir V, Kisi O (2018) Hybrid adaptive neuro-fuzzy models for water quality index estimation. Water Resour Manage. https://doi.org/10.1007/s11269-018-1915-7

Yidana SM (2010) Groundwater classification using multivariate statistical methods: Birimian Basin, Ghana. J Environ Eng 136:1379–1388. https://doi.org/10.1061/(ASCE)EE.1943-7870.0000291

Zou H, Zou Z, Wang X (2015) An enhanced K-means algorithm for water quality analysis of the Haihe River in China. Int J Environ Res Public Health 12:14400–14413. https://doi.org/10.3390/ijerph121114400

Zubaidah T, Karnaningroem N, Slamet A (2018) K-means method for clustering water quality status on the rivers of Banjarmasin. Indonesia. ARPN J Eng Appl Sci 13(11):3692–3697