**ORIGINAL ARTICLE**

# Modeling of groundwater quality index by using artificial intelligence algorithms in northern Khartoum State, Sudan

Musaab A. A. Mohammed[1,2] · Nasraldeen A. A. Khleel[3] · Norbert P. Szabó[1,4] · Péter Szűcs[1,4]

## Abstract

In the present study, multilayer perceptron (MLP) neural network and support vector regression (SVR) models were developed to assess the suitability of groundwater for drinking purposes in the northern Khartoum area, Sudan. The groundwater quality was evaluated by predicting the groundwater quality index (GWQI). GWQI is a statistical model that uses sub-indices and accumulation functions to reduce the dimensionality of groundwater quality data. In the first stage, GWQI was calculated using 11 physiochemical parameters collected from 20 groundwater wells. These parameters include pH, EC, TDS, TH, $Cl^-$, $SO_4^{-2}$, $NO_3^-$, $Ca^{+2}$, $Mg^{+2}$, $Na^+$, and $HCO_3^-$. The primary investigation confirmed that all parameters except for EC and $NO_3^-$ are beyond the standard limits of the World Health Organization (WHO). The measured GWQI ranged from 21 to 396. As a result, groundwater samples were classified into three classes. The majority of the samples, roughly 75%, projected into the excellent water category; 20% were considered good water and 5% were classified as unsuitable. GWQI models are powerful tools in groundwater quality assessment; however, the computation is lengthy, time-consuming, and often associated with calculation errors. To overcome these limitations, this study applied artificial intelligence (AI) techniques to develop a reliable model for the prediction of GWQI by employing MLP neural network and SVR models. In this stage, the input data were the detected physiochemical parameters, and the output was the computed GWQI. The dataset was divided into two groups with a ratio of 80% to 20% for models training and validation. The predicted (AI) and actual (calculated GWQI) models were compared using four statistical criteria, namely, mean square error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination ($R^2$). Based on the obtained values of the performance measures, the results revealed the robustness and efficiency of MLP and SVR models in modeling GWQI. Consequently, groundwater quality in the north Khartoum area is evaluated as suitable for human consumption except for BH 18, where highly mineralized water is observed. The developed approach is advantageous in groundwater quality evaluation and is recommended to be incorporated in groundwater quality modeling.

**Keywords** Sudan · Khartoum · Groundwater quality · Artificial intelligence · SVR · MLP

## Introduction

The primary source of water supply in dry and sub-dry areas is groundwater (Kayemah et al. 2021). Groundwater quality has deteriorated worldwide due to population growth, heavy use of chemical fertilizers, climate change, and improper management of groundwater resources (Singh et al. 2015). Groundwater in Sudan is a fundamental source of water supply. It is essential to community settlement and the development of sustainable social activities (Hassan et al. 2017). Khartoum State is the capital of Sudan, and therefore it is the most vibrant and populous city, with nearly 15 million population. The percentage of the population is constantly increasing as a result of the permanent migration from rural

✉ Musaab A. A. Mohammed
  musab20501@gmail.com

1 Faculty of Earth Science and Engineering, University of Miskolc, Miskolc 3515, Hungary

2 College of Petroleum Geology and Minerals, University of Bahri, Khartoum, Sudan

3 Faculty of Mechanical Engineering and Informatics, University of Miskolc, Miskolc 3515, Hungary

4 MTA-ME Geoengineering Research Group, University of Miskolc, Miskolc 3515, Hungary

areas to the capital city. As a result, groundwater demand has rapidly increased to fulfill the strategic plans. However, this has resulted in a variety of challenges, including decreased production and groundwater quality degradation (Abdo and Salih 2012). Groundwater quality degradation leads to an increase in groundwater salinity, caused mainly by natural and anthropogenic activities (Mohammed et al. 2022). In Khartoum, groundwater contributes to more than 52% of the total demand, primarily for agriculture. Since north Khartoum is agricultural land, the communities have maintained stable life due to the efficiency of irrigation and chemical fertilizers. However, their extensive use harmed the quality of the groundwater. Water quality evaluation and management are issues profoundly affecting human health. According to the World Health Organization (WHO) (Edition 2011), 80% of all diseases are water-borne. Therefore, it is critical to periodically assess groundwater quality with appropriate and effective methods to ensure its suitability for human consumption (Ram et al. 2021).

Groundwater quality evaluation necessitates collecting massive physical and chemical data, which can be challenging to analyze and synthesize. The traditional approach of laboratory analysis is time-consuming and requires intensive efforts. Water quality index (WQI) models are one of the techniques that have been created to analyze water quality data. WQI models rely upon an aggregating mechanism that allows the analysis of huge datasets to yield a single value, i.e., the water quality index. Horton (1965) introduced the first WQI. Subsequently, many experts have developed several WQI and groundwater quality index (GWQI) to assess the suitability of surface and groundwater for drinking and irrigation purposes (Gitau et al. 2016; Tian et al. 2019; Asadi et al. 2020; Kanga et al. 2020). GWQI is a complex index that integrates physical, chemical, and biological parameters to provide an easy-to-understand index for policy and decision-makers (Brown et al. 1970). However, assessing groundwater quality using GWQI is time-consuming and costly (Tung et al. 2020). To overcome the limitations of GWQI, some researchers have turned to non-physical methods using artificial intelligence (AI) models (Imneisi 2019; Kadam et al. 2019; Gaya et al. 2020; Agrawal et al. 2021; Asadollah et al. 2021; Elbeltagi et al. 2021). This approach is based on the idea that any system can learn from datasets, create models, and then make decisions with the least amount of manual intervention (Azrour et al. 2022). For modeling GWQI, AI-based models have minimized sub-index calculations and generated GWQI value efficiently. The benefits of AI approaches include solving complex nonlinear problems and the capacity

to manage big datasets (Bui et al. 2020). Researchers have been able to utilize a variety of AI models due to the continual advancement of computational capabilities. Approaches such as artificial neural networks (ANN) and support vector regression (SVR) have been effectively applied by many researchers to predict the quality of water worldwide. For example, Sakizadeh (2016) used ANN to predict GWQI in Andimeshk City. The study indicated the excellent generalization ability of ANN in the modeling of GWQI. Kadam et al. (2019) confirmed the robustness of multi-linear regression (MLR) and ANN in the prediction of WQI. For WQI modeling in Nainital Lake, Koranga et al. (2022) used multiple machine learning techniques such as random forest, support vector regression, and stochastic gradient descent. Wang et al. (2020) combined particle swarm optimization (PSO), wavelet analysis (WA), and support vector regression (SVR) for modeling WQI in China. Their study indicated the robustness of these models in modeling parameter fluctuation. Singha et al. (2021) developed and compared a deep learning model to other conventional methods for modeling WQI. Their research indicated that deep learning is more effective than the traditional GWQI models in groundwater quality assessment. Gholami et al. (2021) operated an AI-based model using a co-active neuro-fuzzy inference system (CANFIS) and ANN, to assess the quality of groundwater in Iran. The study revealed that the fuzzy neural network has the highest performance in simulating water quality parameters over the other techniques. Elbeltagi et al. (2021) applied four AI models including random subspace (RSS), support vector machine (SVM), M5 pruning tree, and additive regression to predict WQI. The research carried out by Sillberg et al. (2021) demonstrated the possibility of applying machine learning tools such as attribute realization (AR) and SVM algorithms to classify WQI. Ahmed et al. (2019) explored a series of machine learning algorithms, including gradient boosting and multilayer perceptron (MLP), to estimate the WQI. The study conducted by Nathan et al. (2017) revealed that ANN models could be considered a powerful and dependable tool for simulating GWQI. The inspiration from previous works demonstrates the great applicability of AI approaches for GWQI simulation. In general, it was found that every study in the reviewed literature had improved upon earlier ones regarding the effectiveness and reliability of observations.

From the prementioned reviews, artificial intelligence techniques have successfully and accurately predicted water quality indices. Thus, this study aims to investigate the accuracy and performance of two models, including support vector regression (SVR) and multilayered perceptron neural

network (MLP-ANN), in the modeling of GWQI in northern Khartoum State, Sudan. The modeling results will help evaluate groundwater quality, thereby contributing to water supply sustainability. To the best of the authors' knowledge, this is the first study to evaluate the groundwater quality in the central Sudan hydrogeologic system using AI methods.

## Study area

The area is located in north Khartoum State, Sudan, and it covers about 350 km$^2$ (Fig. 1). The study area is situated in the Savanna belt, with an average annual precipitation range of 100–200 mm/year. The Savanna belt is associated with a hot climate and low humidity. The research area is associated with flat topography, which progressively rises from 300 m above the sea level in the west to more than 600 m in the east. Figure 2 shows the geological map of the study area. The geological succession is composed of three main units as basement rocks, Nubian formation, and recent deposits. The Precambrian basement rocks are the oldest rocks in the study area. They consist of gneisses, schists, and granites, which crop at the surface mainly to the north and eastern sides of the area. The Precambrian rocks underlie the Cretaceous Nubian formation (Kheiralla 1966; Whiteman 1971; Saeed 1974; Awad 1994). This formation consists of conglomerates, sandstone, and mudstone. The recent deposits are found in the vicinity of the Nile River and are composed of unconsolidated sand, silt, and gravel. In the study area, groundwater is stored in the Nubian sandstone formation under semiconfined to confined conditions due to the occurrence of clay, clayey sand, and mudstone layers above the groundwater aquifers (Abdelsalam et al. 2016). Two aquifers have been recognized in the Nubian formation (Farah et al. 1997): an upper aquifer of variable thickness (10–300 m) and a lower one more than 400 m thick with higher transmissivity values. The water levels vary from15 to 20 m near the Nile River. However, it attains 45 m in the eastern part of the study area. The flow in the Nubian aquifer, as illustrated
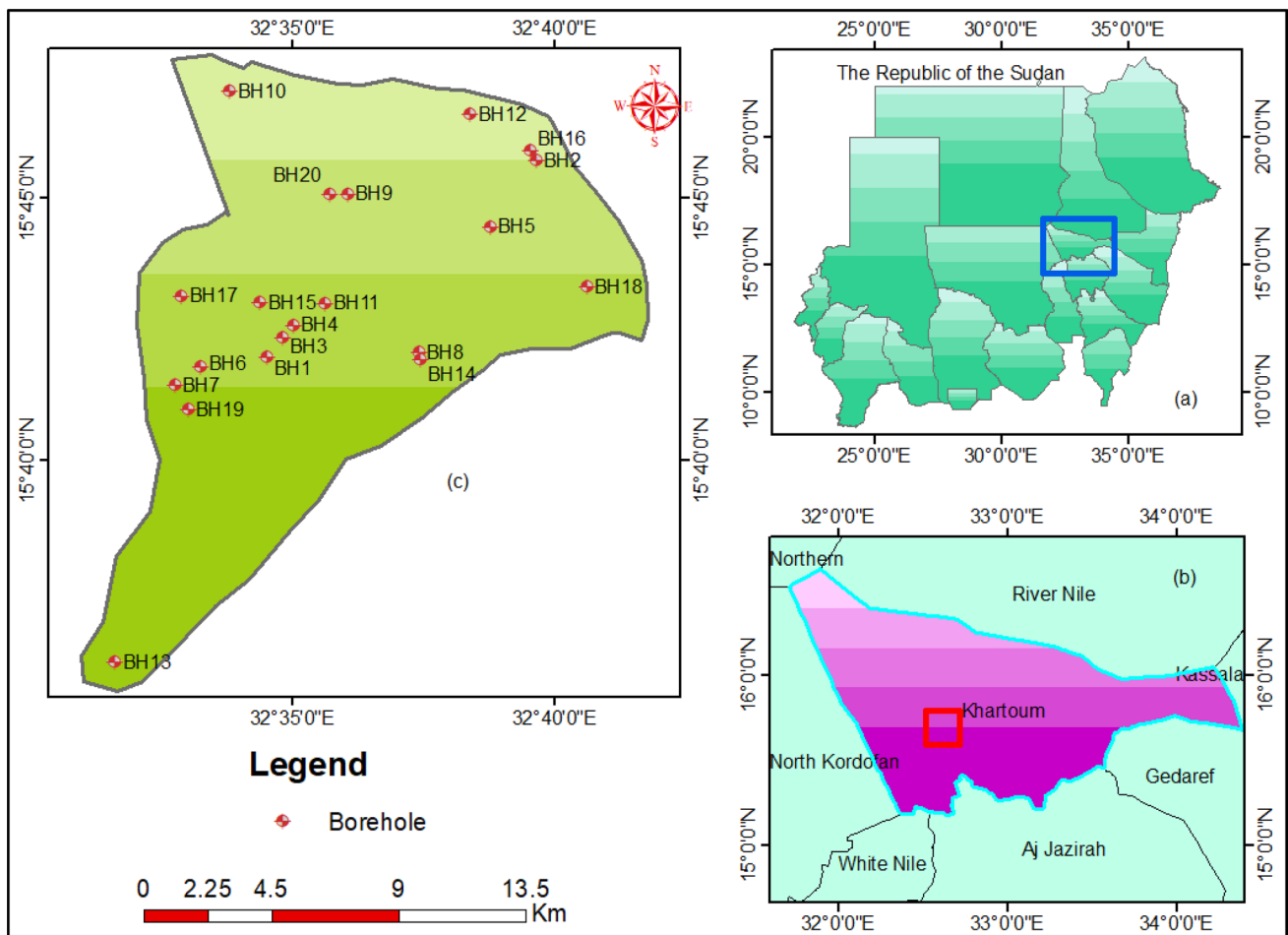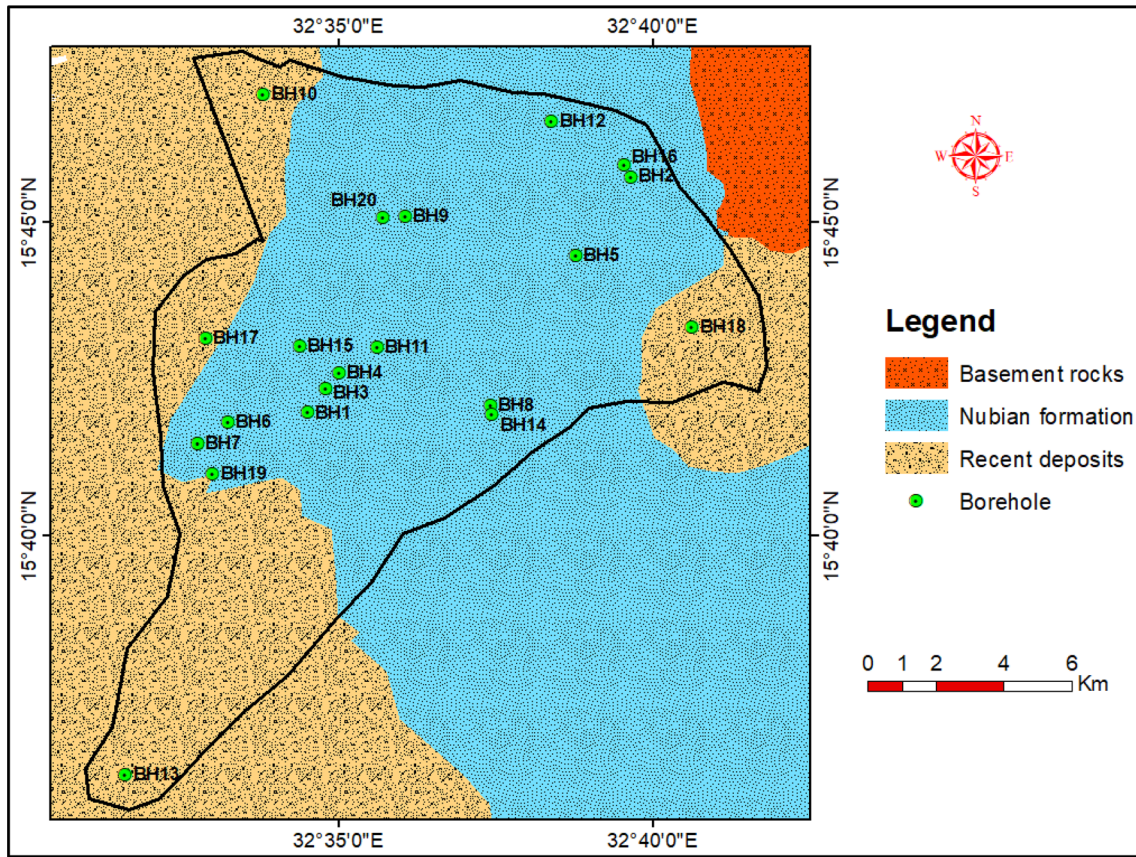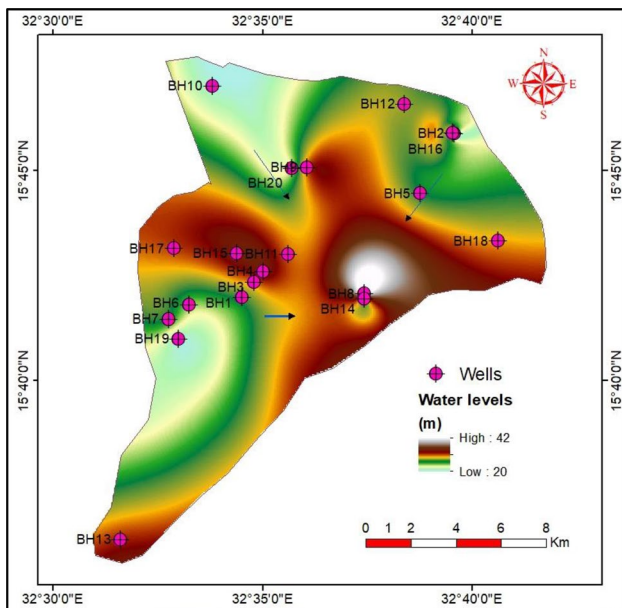


**Fig. 1** Location map of the study area, including Khartoum State showing the location of the collected groundwater samples

**Fig. 2** Geological map of the study area and the surroundings modified after Hussein and Awad (2006)



**Fig. 3** Potentiometric map showing the diverse direction of groundwater flow in the studied area

in Fig. 3, shows diverse directions, but the main direction for groundwater flow is from the west to the southeastern parts. The primary source of groundwater recharge in Khartoum State is the Nile River (Farah et al. 2000). In the areas outside the Nile influence, the groundwater aquifers recharged from the wadies and ephemeral streams.

## Methodology

### Groundwater sampling

Twenty groundwater samples were collected during the postmonsoon season to assess the groundwater quality and its suitability for domestic purposes in the north Khartoum area, Sudan. The groundwater samples were taken from bore wells installed in the study area and ranged in depth from 100 to 150 m. The locations of groundwater samples are selected randomly, aiming at covering vast spaces in the study area. The spatial distribution of groundwater samples is illustrated in Fig. 1. Groundwater samples were also collected based on the accessibility to groundwater boreholes. The collected samples were kept in previously cleaned plastic bottles to

avoid interacting with the atmospheric gases and ions. The location of samples is tracked using the global positioning system (GPS) and subsequently supplied to the geographic information system (GIS) to design base and geographic distribution maps.

Eleven physicochemical parameters for 20 groundwater samples are analyzed in the groundwater and wadies directorate (GWD) laboratory. The analyzed parameters are total hardness (TH), calcium ($Ca^{+2}$), sodium ($Na^{+}$), magnesium ($Mg^{+2}$), chloride ($Cl^{-}$), nitrate ($NO_3^{-}$), sulfate ($SO_4^{-2}$), and bicarbonate ($HCO_3^{-}$). Electrical conductivity (EC), total dissolved solids (TDS), and hydrogen ion activity (pH) were measured using a portable multi-parameter instrument immediately after the sample collection. Appelo and Postma (2005) equation (Eq. 1) is applied to reveal the reliability of the conducted hydrochemical analysis. This formula measures the electrical balance (EB%) between the total of all cations ($\Sigma$ cations) and anions ($\Sigma$ anions) in milliequivalents per liter (meq/L). If the calculated EB is within $+10$ and $-10$, the accuracy of the measurement is indicated as reliable and can be considered for further interpretation. Otherwise, the hydrochemical must be repeated to fulfill the suggested range. Fortunately, in this research, the measured EB for all groundwater samples was within $\pm 5$, indicating high accuracy. The EB formula is as

$$(EB\%) = \frac{\sum \text{cations} - \sum \text{anions}}{\sum \text{cations} + \sum \text{anions}} \times 100. \tag{1}$$

## Groundwater quality index (GWQI)

GWQI is a widely used model in determining the potability of groundwater, considering management strategies. GWQI results from a rating method that uses water quality parameters to create an overall depiction of groundwater quality. This approach is utilized to reduce the dimensionality of the groundwater quality data into a single dependent numerical value. In general, GWQI is created in three steps: assigning weights, computing the rating scale, and aggregating the sub-indices. In this study, 11 physiochemical parameters (i.e., pH, EC, TDS, TH, $Cl^{-}$, $SO_4^{-2}$, $NO_3^{-}$, $Ca^{+2}$, $Mg^{+2}$, $Na^{+}$, $HCO_3^{-}$) for 20 groundwater samples were incorporated in GWQI computation. The lack of microbiological contamination measurements in the study area constrains the definition of the groundwater quality index. However, the routinely analyzed physiochemical parameters can effectively determine the suitability of the groundwater for drinking purposes in the Khartoum area since biological contamination is rare.

Weights are loaded to the selected parameters depending on their influence on the overall groundwater quality. In this study, the weights area was assigned with the aid of correlation analysis to reveal the influence rate of each physiochemical parameter in the overall groundwater quality. A weight of 5 is given to the most significant parameter, while the least significant parameter is given a weight of 2. Consequently, the relative weight ($W_i$) for the parameters is calculated using Eq. 2 (Singh 1992) as

$$W_i = \frac{w_i}{\sum_{i=1}^{n} w_i}, \tag{2}$$

where $W_i$ denotes the relative weight of each parameter, $w_i$ is the weight allocated to each parameter, and $n$ denotes the number of variables used in the GWQI calculation. The assigned weights and the relative weights applied in this study are illustrated in Table 1.

The rating scale is calculated in the second phase. Since the measured hydrochemical parameters have different units and ranges, the goal of scaling is to convert all the selected parameters into a common scale. The rating scale in this study was generated using the standard limits prescribed by WHO (Edition 2011). Equation 3 is applied to create the rating scale as

$$R_i = \frac{X_i}{X_s} \times 100, \tag{3}$$

where $R_i$ is the rating scale value, $X_i$ is the actual parameter value, and $X_s$ is the prescribed standard value.

The final stage in GWQI calculation is aggregating the sub-indices with their weights. In this study, the mean arithmetic with unequal weights approach has been used for sub-index aggregation, and the final index value was calculated by Eqs. 4 and 5 (Tiwari and Mishra 1985). Based on the final

Table 1 The weights and the relative weights of the physiochemical parameters

| Parameters/units | Weight ($w_i$) | Relative weight ($W_i$) |
|---|---|---|
| pH | 2 | 0.05 |
| EC μS/cm | 3 | 0.08 |
| TDS mg/L | 5 | 0.13 |
| TH mg/L | 4 | 0.10 |
| $Ca^{2+}$ mg/L | 4 | 0.10 |
| $Mg^{2+}$ mg/L | 4 | 0.10 |
| $Na^{+}$ mg/L | 4 | 0.10 |
| $Cl^{-}$ mg/L | 4 | 0.10 |
| $SO_4$ mg/L | 4 | 0.10 |
| $HCO_3$ mg/L | 3 | 0.08 |
| $NO_3$ mg/L | 3 | 0.08 |

**Table 2** Classification of groundwater based on GWQI as given by Ramakrishniah et al. (2009)

| GWQI range | Class | Type of water |
|---|---|---|
| < 50 | I | Excellent water |
| 50.1–100 | II | Good water |
| 100.1–200 | III | Poor water |
| 200.1–300 | IV | Very poor water |
| > 300 | V | Unsuitable water |

GWQI, groundwater is categorized into five classes. Table 2 shows the classification of groundwater samples based on GWQI values as given by Ramakrishniah et al. (2009).

$$SI = W_i \times R_i, \tag{4}$$

$$GWQI = \sum SI, \tag{5}$$

where SI is the sub-index values for each parameter.

## Artificial intelligence methods

Multilayer perceptron neural networks (MLP-ANN) and support vector regression (SVR) are employed to predict GWQI in this research. The experiment was performed using Python 3.7 environment with Keras as a high-level application programming interface (API) based on TensorFlow module. A detailed description of MLP and SVR is given in the following sections.

## Multilayer perceptron (MLP)

ANNs are computer programs that use a large number of interlinked neurons to replicate the functioning of the biological nervous system (Akbari and Jalali 2007). They serve as a representation for the nervous system, where neurons act as operating units. Their wide scope of uses comes from the capacity of the networks to simulate the human brain (Tom et al. 2020). The ANN is a decentralized, parallel data processing system with unique operational characteristics similar to the human brain (Momenzadeh et al. 2011). The popular type of ANN applied for environmental problems is multilayer perceptron (MLP) neural networks (Heddam 2016). MLP is the most realistic neural network architecture applied for classification or regression problems (Gholami et al. 2015). MLP neural networks are basic types of feed-forward neural networks (FFNN), which are parallel layered structure networks. A one-layer perception is transformed into an MLP model by adding one or more hidden layers. This suggested the topology is capable of resolving challenging and complex linear and nonlinear problems (Tokar and Markus 2000). In most cases, an MLP network has three layers: the input layer, hidden layers, and output layer (Fig. 4). In this work, the inputs were the physiochemical variables (i.e., pH, EC, TDS, TH, $Cl^-$, $SO_4^{-2}$, $NO_3^-$, $Ca^{+2}$, $Mg^{+2}$, $Na^+$, $HCO_3^-$), and the output is the groundwater quality index (GWQI). The hidden layers consist of neurons for transforming the input data. The neurons in the first layer transmit the signal to the neurons in the following layer until the optimal output is reached. The degree of association between every two neurons in two layers is called weight, and the modification of this weight is called model training (Schaid et al. 1999). In other words, the resulting output is the total of the weighted inputs. In the modeling procedure, the data is divided for model training and validation. The network training aims to evaluate the network capacity to replicate the relationship between inputs and output. The first stage in MLP neural network operation is to link the

**Fig. 4** The architecture of MLP neural network applied in this study
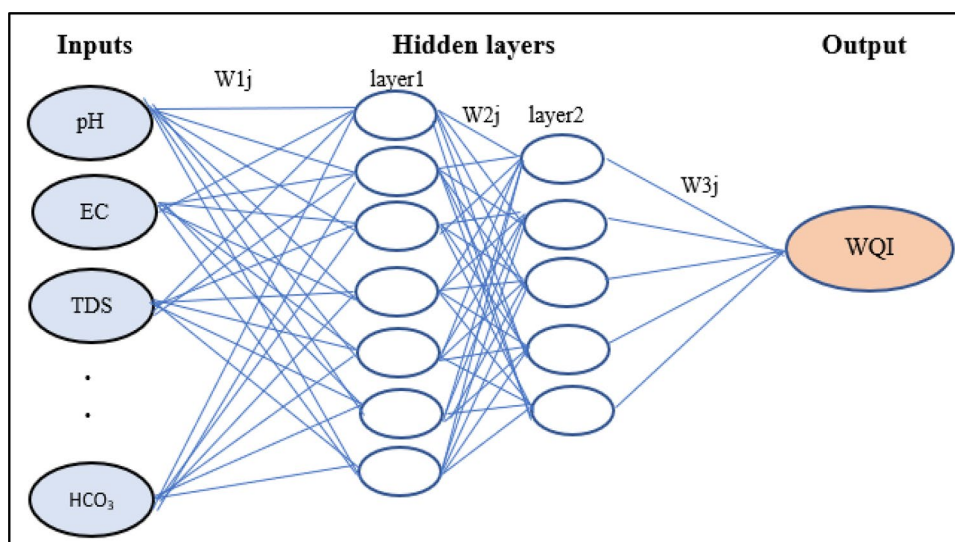
input variables to the hidden layers and weights. For training, MLP employs Bayesian regularization, which adjusts the weight values by optimization (Toprak and Cigizoglu 2008). The weighted parameters are added to the bias of the layer and changed from the $j$th to the $j$th+1 layer. The layer weights and biases are adjusted iteratively throughout the training process to achieve good performance and provide an acceptable correlation coefficient (Nasir et al. 2022). The following equation (Eq. 6) is used to predict the output (GWQI):

$$Y_k = f_k\left(\sum_{i=1}^{m} W_{jk} f_j\left(\sum_{i=1}^{n} X_i W_{ij}\right)\right) + W_0, \tag{6}$$

where $f_k$ and $f_j$ are the transfer functions of the output and hidden layer neurons $k$ and $j$, accordingly, $n$ is the features number, $m$ is the hidden layer neuron number, and the bias is $W_0$. The weight between the $j$th neuron and the $k$th target neuron is $W_{jk}$, whereas the weight between the $i$th and $j$th neurons is $W_{ij}$.

## Support vector regression (SVR)

Support vector machine (SVM) is a machine learning technique that can provide satisfactory solutions to the nonlinear problems of regression, prediction, classification, and function estimation (Haghiabi et al. 2017). An additional feature of support vector machines over the conventional artificial neural network is their capability to enhance the data network functionality (Manzar et al. 2022). The regression model of the SVM is divided into linear support vector regression (L-SVR) and nonlinear support vector regression (N-SVR) (Kaya et al. 2021). Support vector regression (SVR) was first introduced by Boser et al. (1992). It is a machine learning technique that was developed from the SVM. In this study, SVR is employed to predict GWQI. In order to improve the forecasting capability of the model, the primary goal of SVR is to simultaneously minimize the system complication and prediction error (Bagheripour et al. 2015). SVR is a supervised classifier that can quickly and accurately fit and predict samples. The approach effectively finds a hyperplane in the data sets that fits the nearest plane distance. The optimal hyperplane is the line with the maximum margin, which defines the distance separating the hyperplane and the adjacent input variable (Aldhyani et al. 2020). Figure 5 shows how hyperplanes fit the data points. In Fig. 5, the green and blue dots represent two types of data points. Three planes designated as P1, P2, and P3 are projected. The data points are not successfully categorized by P1. Although both P2 and P3 can categorize data points, P2 provides a narrower margin than plane P3. This is the rationale behind choosing P2 for prediction. There are three levels
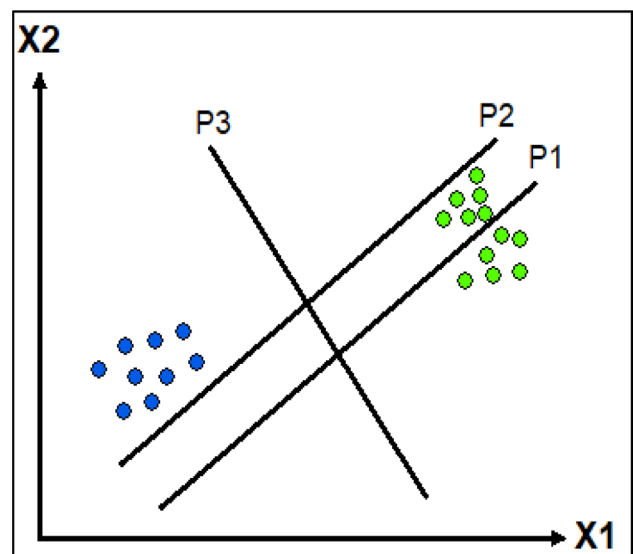


**Fig. 5** Data classification using different hyperplanes

in the SVR framework: inputs, kernel functions, and outputs. In this study, the input is the physiochemical parameters, and the output is the GWQI. The kernel function is employed to map the lower-dimension data into high-dimension data points and, thus, reduce the space between the points. The kernel function enables the separation of the nonlinear data points. There are different types of kernel functions such as sigmoid, polynomial, Gaussian kernel functions. In this analysis, Gaussian kernel function (RBF) was employed for its simplicity and reliability. Gaussian kernel function is an exponential function and expressed in Eq. 7 where $K(x_1, x_2)$ is the kernel function.

$$K(x_1, x_2) = \exp(-x_1 - x_2). \tag{7}$$

## Performance metrics

The functionality of MLP-ANN and SVR models is assessed using four statistical indicators: mean square error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination ($R^2$). These statistical indicators are referred to the variance explored by the predicted model compared to the actual. The performance metrics were calculated by Eqs. 8, 9, 10, and 11 as

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(x(i) - y(i))^2, \tag{8}$$

$$\text{RMSE} = \frac{1}{n} \sqrt{\sum_{i=1}^{n} (x(i) - y(i))^2}, \tag{9}$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y(i) - x(i)|, \tag{10}$$

$$R^2 = \left( \frac{\left[ \sum_{i=1}^{n} \left( y(i) - \overline{y} \right)(x(i) - \mathbb{x}) \right]^2}{\sqrt{\sum_{i=1}^{n} \left( y(i) - \overline{y} \right)^2 \sum_{i=1}^{n} (x(i) - \mathbb{x})^2}} \right)^2, \tag{11}$$

where $n$ is the number of observations and $x(i)$ and $\overline{y}(i)$ are the actual and predicted value for the $i$th observation, respectively. $\overline{y}$ and $\mathbb{x}$ are the mean for the predicted and actual values, respectively.

## Results and discussions

### General hydrochemistry

Physicochemical parameters of the groundwater are considered prime principles in identifying the type and nature of groundwater (Selvakumar et al. 2017). In this study, the detected physiochemical parameters are pH, TDS, EC, TH, $Na^+$, $Ca^{+2}$, $Mg^{+2}$, $HCO_3^-$ $Cl^-$ $SO_4^{2-}$, and $NO_3^-$, and the result of the hydrochemical analysis is illustrated in Table 3. Table 4 shows the descriptive statistics of the analyzed physiochemical parameters (minimum, mean, and maximum) to reveal the deviation of the parameters from the prescribed standards. The pH of the groundwater samples ranged from 7.14 to 8.59, and the greatest pH value was reported in borehole 17. A pH above seven is considered acidic for groundwater, and lower than seven is considered alkaline. The acceptable pH values for groundwater samples range from 6 to 8.5 WHO (Edition 2011). Thus, the groundwater in the study area is neutral to alkaline in nature. TDS is one of the major parameters used to understand the amount of contaminant in the groundwater. Classifying groundwater according to TDS is crucial to assess its suitability for all uses (Freeze and Cherry 1979). It ranges from 190 to 6225 mg/L. WHO (Edition 2011) advises that a TDS level of 600 mg/L is ideal for drinking. In this study, 20% of the groundwater samples exceeded the prescribed limits. Groundwater with TDS concentration below 1000 mg/L is considered fresh, between 1000 and 10,000 mg/L is brackish, and groundwater is considered saline when TDS concentration exceeds 10,000 mg/L WHO (Edition 2011). In this research, 90% of groundwater samples were classified as freshwater, while 10% were defined as brackish water. The EC varies between 317 and 1500 μS/cm. The

permissible limit for the EC of groundwater is 1500 μS/cm WHO (Edition 2011). Thus, all the groundwater samples are suitable for human consumption based on EC. TH concentrations range from 124 to 1172 mg/L. $Na^+$ is the major ion in groundwater chemistry. The maximum concentration (1844 mg/L) is recorded at borehole 18 in the eastern part of the study area, and the minimum (14 mg/L) is at location 19. Temporary hardness is mainly caused by calcium or magnesium carbonates, while calcium and magnesium sulfate or chloride contributes to the TH. Consumption of hard water for drinking purposes may stimulate kidney stones and cardiovascular diseases (Sengupta 2013). According to Sawyer and McCarty (1967), groundwater with TH concentration less than 75 mg/L is regarded as soft water, 75–100 mg/L is considered to be moderately hard water, 150–300 mg/L is hard water, and groundwater with TH higher than 300 mg/L is considered to be very hard. In this analysis, 90% of groundwater samples are hard, whereas 10% of the samples are very hard water. $Na^+$ concentration in groundwater samples ranges from 14 to 1844 mg/L with an average value of 161.5 mg/L. Consumption of groundwater with $Na^+$ concentration higher than 200 mg/L may induce congenital disorders and nervous system problems according to WHO (Edition 2011). Higher $Na^+$ might indicate weathering of silicate minerals or the dissolution of halite (Hem 1985). $Ca^{+2}$ content of the groundwater samples varies from 16 to 132.8 mg/L. Calcium is an essential constituent of many igneous-rock minerals such as pyroxenes, amphiboles, and feldspars. The most common forms of $Ca^{+2}$ in sedimentary rocks are calcite, aragonite, and dolomite gypsum. The maximum concentration of $Ca^{+2}$ is recorded in borehole 18. In the case of $Mg^{+2}$, the concentration varies from 5.8 to 201 mg/L. Water hardness is mainly affected by cations such as $Ca^{+2}$ and $Mg^{+2}$. Generally, the sources of $Mg^{+2}$ are the ferromagnesian minerals, especially pyroxene, amphiboles, and biotite. Common forms in sedimentary rocks include carbonates such as magnesite and dolomite. The concentration of the $HCO_3^-$ varies between 130 and 620 mg/L. $HCO_3^-$ is the dominant anion present in the study area. The maximum concentration is found in borehole 18. As the mineral content increases, the $HCO_3^-$ content also increases. $SO_4^{2-}$ content in groundwater varies from 3 to 1500 mg/L. The permissible limit prescribed by WHO (Edition 2011) for $SO_4^{2-}$ concentration is exceeded in 10% of groundwater samples. The presence of $SO_4^{2-}$ ions in water can affect the taste, and too much sulfate concentration can negatively impact consumers (Rishi et al. 2020). $Cl^-$ concentrations range from 4 to 2120 mg/L. The highest concentration is detected in sample location 18, while the lowest is recorded in borehole 19. 95% of the groundwater samples are below the limit of WHO (Edition 2011). The concentration of $NO_3^-$ ranges from 0.07 to 13.6 mg/L. The maximum concentration is recorded in borehole 2, while the minimum

**Table 3** Physiochemical parameters used in the calculation of GWQI

| Borehole no. | pH | EC (µS/cm) | TDS (mg/L) | TH (mg/L) | Cl⁻ (mg/L) | $SO_4^{-2}$ (mg/L) | $NO_3^-$ (mg/L) | $Ca^{+2}$ (mg/L) | $Mg^{+2}$ (mg/L) | $Na^+$ (mg/L) | $HCO_3^-$ (mg/L) | GWQI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| BH 1 | 7.3 | 601.5 | 330.8 | 234 | 23 | 52 | 3.2 | 35.2 | 35 | 37.6 | 226 | 33.8 |
| BH 2 | 7.5 | 588 | 411.6 | 230.6 | 11.36 | 37 | 13.6 | 48.6 | 26.4 | 88.4 | 396.5 | 42.8 |
| BH 3 | 7.4 | 590 | 354 | 208 | 8 | 28 | 1.2 | 40.3 | 25.4 | 37.2 | 248 | 31.6 |
| BH 4 | 7.2 | 585 | 322 | 200 | 12 | 27.9 | 3 | 38.4 | 25 | 46.8 | 250 | 33.9 |
| BH 5 | 7.9 | 1500 | 1050 | 380 | 193.1 | 321 | 0.07 | 16 | 82.6 | 332.8 | 414.8 | 91.3 |
| BH 6 | 7.4 | 485 | 339.5 | 254 | 12.7 | 98 | 9.24 | 56 | 27.7 | 37 | 414.8 | 42.8 |
| BH 7 | 8.3 | 417 | 292 | 186 | 120.7 | 4 | 0.1 | 32 | 25.8 | 17.6 | 231.8 | 32.8 |
| BH 8 | 8 | 357.3 | 214.4 | 134 | 9 | 14 | 5.3 | 24 | 17.8 | 19.8 | 142 | 23.8 |
| BH 9 | 7.8 | 858 | 600.6 | 190 | 82.5 | 86 | 5.7 | 66.4 | 5.8 | 124.3 | 378.2 | 49.8 |
| BH 10 | 7.4 | 518 | 362.6 | 185 | 23.8 | 23 | 6.2 | 38.4 | 22.4 | 45.8 | 281.6 | 35.3 |
| BH 11 | 7.5 | 650 | 375 | 232 | 18 | 41 | 1.4 | 52.8 | 24 | 46.5 | 276 | 38.1 |
| BH 12 | 7.7 | 929 | 694.4 | 260 | 91.5 | 152 | 7.4 | 57.6 | 28.2 | 123 | 292.8 | 55.5 |
| BH 13 | 7.2 | 590 | 413 | 186 | 25.5 | 47 | 3.9 | 32 | 21.4 | 78.12 | 317 | 39.5 |
| BH 14 | 7.9 | 317 | 190.2 | 124 | 20 | 13 | 3.7 | 27.2 | 13.4 | 24.13 | 130 | 21.0 |
| BH 15 | 7.14 | 626 | 438.2 | 226 | 35.5 | 37 | 0.9 | 20.8 | 42.3 | 36.1 | 353.8 | 37.9 |
| BH 16 | 7.7 | 1318 | 724.9 | 300 | 120 | 100 | 2.8 | 44 | 45.6 | 147.52 | 340 | 59.7 |
| BH 17 | 8.59 | 646 | 452.2 | 256 | 16.3 | 38 | 1.32 | 54.4 | 29.2 | 54.9 | 280.6 | 38.5 |
| BH 18 | 7.4 | 1245 | 6225 | 1172 | 2120 | 1500 | 10.2 | 132.8 | 201.6 | 1844 | 620 | 396 |
| BH 19 | 7.4 | 370 | 222 | 160 | 4 | 3 | 7.3 | 39.2 | 14.88 | 14.4 | 186 | 25.4 |
| BH 20 | 7.4 | 598 | 359 | 148 | 24 | 53 | 3.5 | 30.4 | 17.28 | 74.7 | 212 | 31.9 |

**Table 4** The descriptive statistics of the parameters and their permissible limits prescribed by WHO

| Parameter | Unit | Minimum | Mean | Maximum | WHO standard (Edition 2011) |
|---|---|---|---|---|---|
| pH | – | 7.1 | 7.6 | 8.6 | 8.5 |
| EC | (µS/cm) | 317 | 689.4 | 1500 | 1500 |
| TDS | (mg/L) | 190 | 718.6 | 6225 | 1000 |
| TH | (mg/L) | 124 | 263.3 | 1172 | 500 |
| $Cl^-$ | (mg/L) | 4 | 148.5 | 2120 | 250 |
| $SO_4^{-2}$ | (mg/L) | 3 | 133.7 | 1500 | 250 |
| $NO_3^-$ | (mg/L) | 0.07 | 4.5 | 13.6 | 50 |
| $Ca^{+2}$ | (mg/L) | 16 | 44.3 | 132.8 | 200 |
| $Mg^{+2}$ | (mg/L) | 5.8 | 36.6 | 201 | 150 |
| $Na^+$ | (mg/L) | 14 | 161.5 | 1844 | 200 |
| $HCO_3^-$ | (mg/L) | 130 | 299.6 | 620 | 350 |

concentration is recorded in borehole 5. The essential source of $NO_3^-$ is agricultural activities. High $NO_3^-$ concentrations in drinking water can result in goiter, stomach cancer, and hypertension, in addition to methemoglobinemia in children (Majumdar and Gupta 2000). Figure 6 shows the geographical dispersion of the physical and chemical parameters used in this study. $Na^+$, $Mg^{+2}$, $Cl^-$, $SO_4^{2-}$, TH, and TDS exhibit similar trend with their concentrations increasing from the western to the eastern part of the study area. This suggests the high contribution of these ions on the groundwater quality. The other parameters show diverse trend which suggests different origination source.

## Groundwater quality index (GWQI)

The input variable selection and weight assignment are the most crucial part of developing the GWQI model. the highest weights are assigned to the parameter with the most substantial influence on the overall groundwater quality. In this study, the selection of the relevant weights for each parameter is based on the degree of correlation between the measured physiochemical parameters. Pearson correlation analysis is applied to detect the linearity between groundwater quality parameters. Correlation analysis measures the degree of the association between the selected variables; if the correlation coefficient is nearer to $+1$ or $-1$, the relationship between the two variables, either proportion or inversely proportion, is perfected and vice versa. In this study, the highest weight is assigned to TDS since its concentration determines the suitability of groundwater for domestic purposes (Freeze and Cherry 1979). The high linkage of the dominant ions and TDS reflects the role of mineral dissolution in groundwater chemistry (Singh et al. 2008). The Pearson correlation analysis is illustrated in Fig. 7. It is observed that TDS has a high correlation with TH ($r=0.99$),

$Na^+$ ($r=1$), $Cl^-$ ($r=1$), $SO_4^{-2}$ ($r=0.99$), $Ca^{+2}$ ($r=0.84$), $Mg^{+2}$ ($r=0.96$), and $HCO_3^-$ ($r=0.74$), which indicates the great influence of these parameters on the overall groundwater chemistry, medium correlation with EC ($r=0.53$) and $NO_3^-$ ($r=0.34$), and low association with pH ($r=0.1$), which reflect the least effect of these variables on groundwater quality. Accordingly, the highest weight was assigned to TDS, while the lowest one was given to pH, EC, and $NO_3^-$. Accordingly, the total weights are used to calculate the relative weights.

Weighted arithmetic GWQI is calculated to appraise the groundwater quality in the north Khartoum area. The quantitative results of GWQI are evaluated to determine the suitability of groundwater for domestic purposes based on WHO (Edition 2011) guidelines for drinking water. GWQI aided in comprehending the combined overall effect of the analyzed physiochemical parameters on groundwater quality (Srivastava 2019). The calculated values of GWQI range from 21 to 396 (Table 3); hence, the water samples were classified into three categories. The majority of the samples, around 75%, fall under the excellent water class, 20% are projected in the good water class, and 5% of groundwater samples are considered unsuitable for human consumption. The areal distribution of WQI, represented in Fig. 8, shows that most of the area is occupied by excellent water types and water quality characteristics change gradually from the western to the eastern part of the study area. The lowest value is observed in BH 14, while the highest GWQI is indicated in BH 18. The high WQI at BH 18 is impacted by TDS, TH, $Ca^{+2}$, $Mg^{+2}$, $Na^+$, $Cl^-$, $HCO_3^-$ and $SO_4^{-2}$. As Sharma et al. (2022) suggested, the abundance of these parameters is likely to be influenced by rock–water interaction. The remaining water samples represent an excellent water type. However, some samples are highly influenced by individual physiochemical parameters. For example, borehole two (2) is associated with high $NO_3^-$ concentration, while in borehole 5, high EC is observed. Therefore, caution must be taken when using groundwater samples with a high concentration of individual physiochemical parameters.

According to the measured GWQI, the groundwater in the study area is suitable for drinking purposes Except for the BH 18 sample. The unsuitability of groundwater in BH 18 may significantly influence the present scenario of groundwater quality in the study area since advection and dispersion processes may spread pollution along the groundwater flow paths. Therefore, the concerned authorities should plan proper steps for maintaining and improving the current situation of the groundwater quality in the study area.

## Artificial intelligence models

In this research, 11 routinely analyzed physiochemical parameters were chosen to model GWQI using the
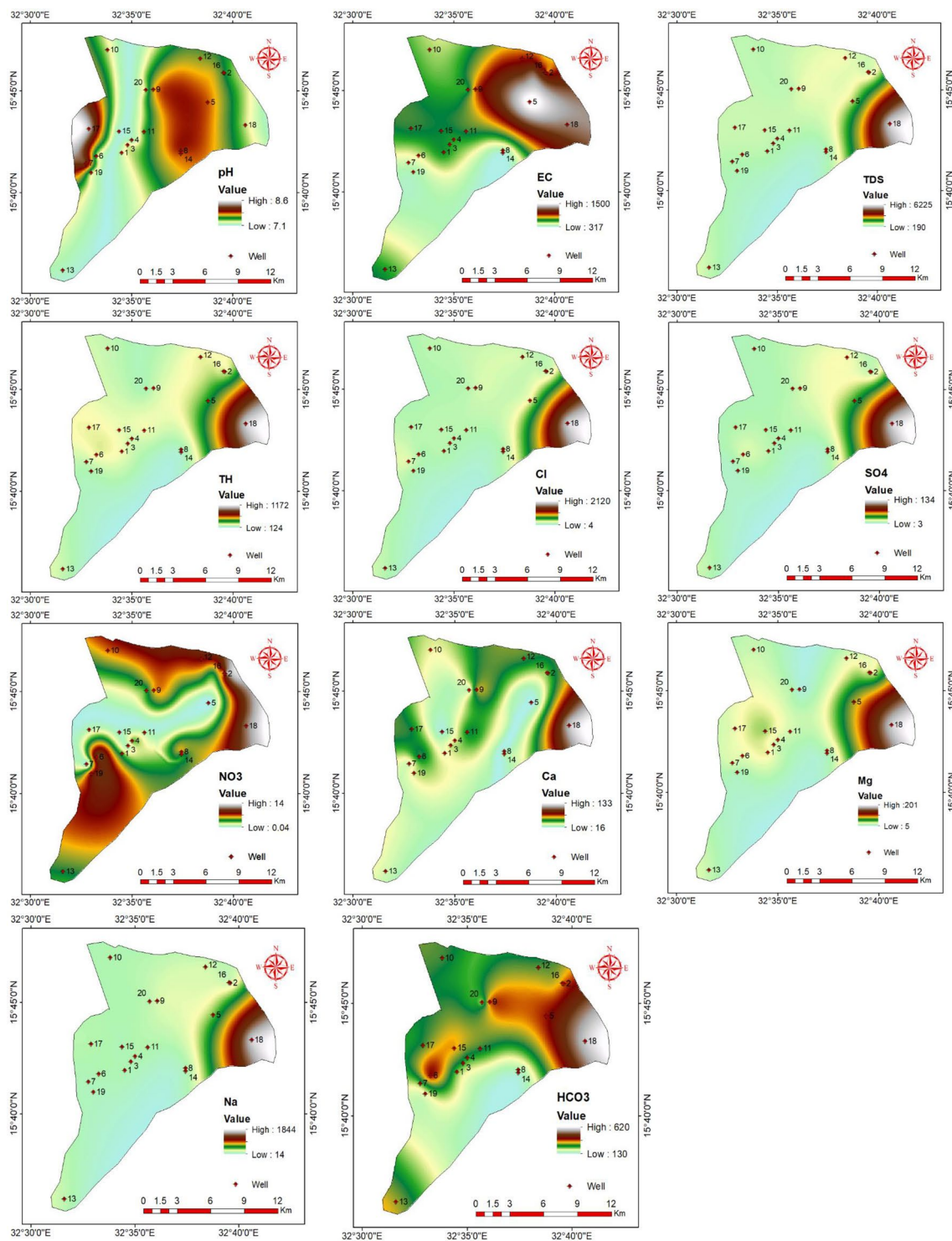
**Fig. 6** The areal distribution of the physiochemical parameters used in the calculation and prediction of the water quality indices

multiLayer perceptron (MLP) of ANN and support vector regression (SVR). These methods are applied to overcome the limitation of the conventional GWQI. The analyzed parameters are considered the input, while the calculated GWQI using a statistical (conventional) approach is considered the output. Experimental data were categorized into training and testing. The training set was employed to generate the ANN and SVR model; validating sets were used to confirm the model's generalization competencies. The
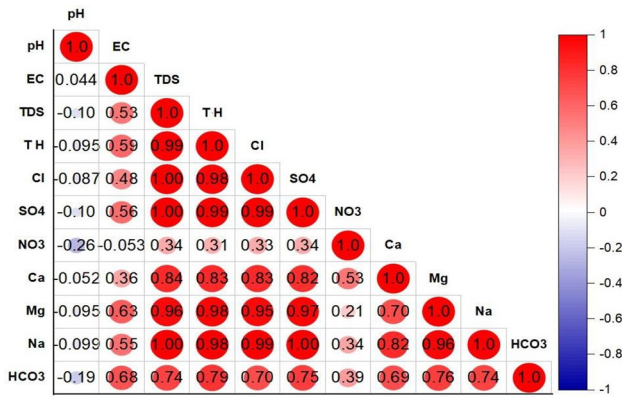
**Fig. 7** Pearson correlation analysis for the physiochemical parameters
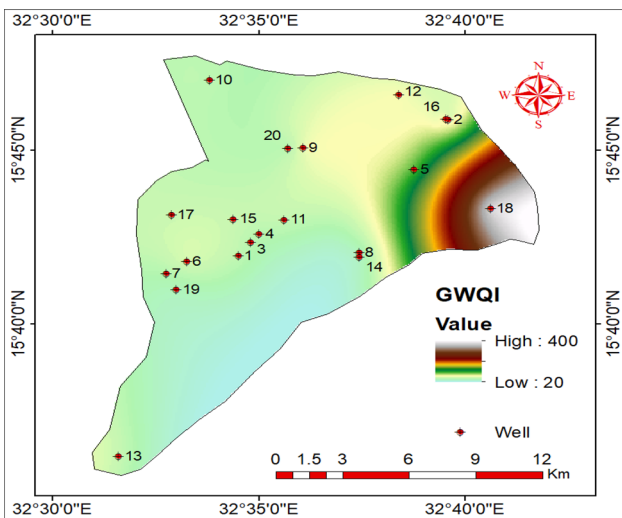


**Fig. 8** The spatial variation of GWQI in the study area

**Table 5** MLP performance during training and validation stages for WQI prediction

| Models | MSE | RMSE | MAE | $R^2$ | Support |
|---|---|---|---|---|---|
| MLP training | 1.4436 | 1.2015 | 0.8999 | 0.9998 | 16 |
| MLP validation | 0.2594 | 0.5093 | 0.4663 | 0.9976 | 4 |

measured water samples are divided into 80% for training and 20% for validation.

MLP exhibits the best performance by applying two hidden layers with 126 and 64 neurons by a trial-and-error procedure in each layer, respectively (Belayneh et al. 2016). So, the most appropriate model structure is 11-128-64-1, and the trial-and-error process led to the selection of learning rates of 0.1. The weights were also updated using RELU function. Providing the right choice in the selection of hidden neurons and the architecture of the network is crucial

to prevent overlearning in the calibration stage. Table 5 shows the effectiveness of the MLP model for predicting WQI during the training and validation stages. The values of MSE, RMSE, MAE, and $R^2$ obtained for MLP training are 1.4436, 1.2015, 0.8999, and 0.9998, respectively, while the performance measures for validation are 0.2594, 0.5093, 0.4663, and 0.9976 respectively. The statistical results of the MLP model for predicting GWQI during the training and validation stages are presented in Fig. 9, which indicates the projected points generally correlated close to the 1:1 line.

SVR modeling was created by using the Gaussian kernel type. Both grid and pattern search and tenfold cross-validation re-sampling methods were employed to find optimal parameter values. The performance measures, including MSE, RMSE, MAE, and $R^2$ for SVR training, as shown in Table 6 are 0.0083, 0.0911, 0.0874, and 0.9999 and 0.0113, 0.1064, 0.0853, 0.9998 for validation, respectively. The representation of the observed and optimal simulated GWQI by SVR model is presented in Fig. 10. It is evident from this figure that the predicted GWQI derived by SVR model is well-matched with the observed GWQI. Based on quantitative performance assessment indicators, the SVR model performed better than the MLP model. The comparison between the predicted and actual GWQI presents a good correlation between the GWQI of SVR model and the conventional GWQI with high values of statistical coefficient. The robustness of SVR could be attributed to the great advantage of handling complex and nonlinear system, unlike that of the MLR models, which is based on the assumptions of linear input–output relationship.

The results of GWQI modeling using artificial intelligence techniques showed a resealable match with the conventional GWQI. Consequently, the quality of groundwater in the north Khartoum area can be evaluated solely with artificial intelligence techniques. It can be concluded that artificial intelligence techniques such as MLP neural network and SVR can effectively simulate GWQI and other hydrochemical parameters in time and cost-effective way in regional assessment when large water quality data is recorded. In order to improve groundwater quality assessments and management, the application of artificial intelligence is recommended for groundwater resource modeling.

## Conclusions

Management of groundwater resources requires a proper assessment of groundwater quality since it provides evidence of the influence of physical and anthropogenic activities on groundwater resources. In this research, the groundwater quality index (GWQI) model as a practical tool is developed to evaluate groundwater quality for domestic uses in the north Khartoum area. GWQI model is constructed by using
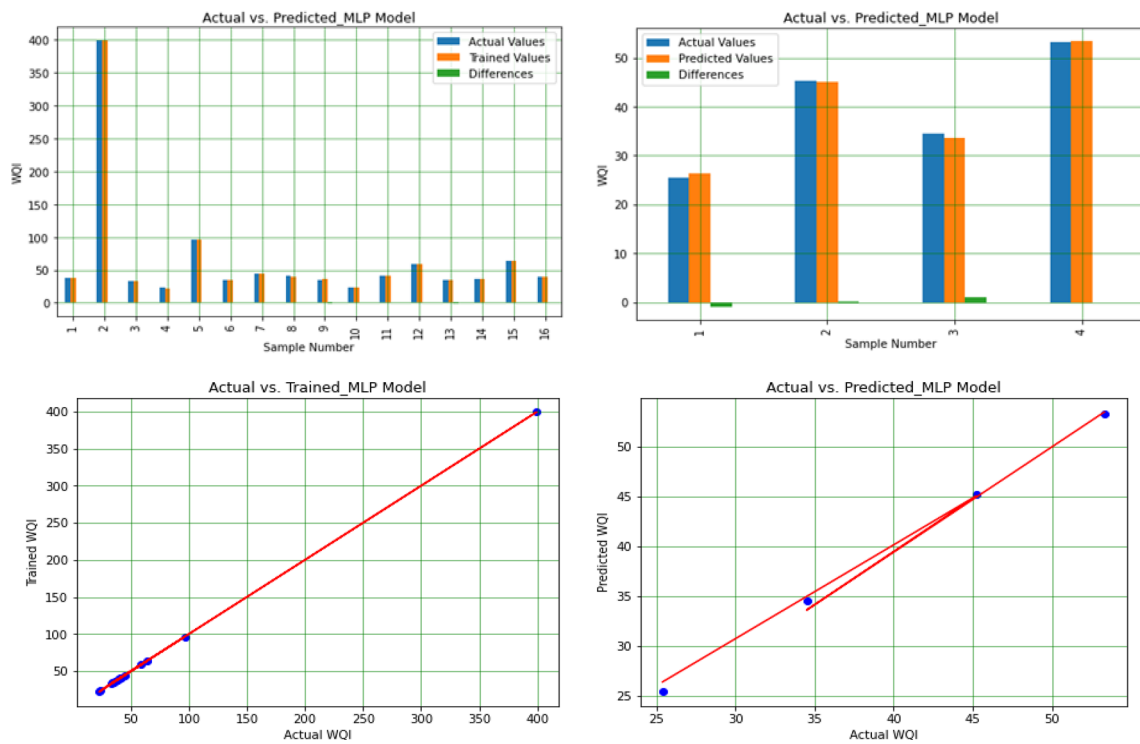
**Fig. 9** Actual versus simulated WQI during training and validation for SVR model

**Table 6** Performance measures for SVR model during training and validation stages for WQI prediction

| Models | MSE | RMSE | MAE | $R^2$ | Support |
|---|---|---|---|---|---|
| SVR training | 0.0113 | 0.1064 | 0.0853 | 0.9998 | 16 |
| SVR validation | 0.0083 | 0.0911 | 0.0874 | 0.9999 | 4 |

11 physiochemical parameters measured in 20 groundwater boreholes scattered over the study area. These parameters are primarily investigated to reveal their deviation from the world health organization (WHO) standard. As a result, all the detected parameters are found to be beyond the applied standard except for EC and $NO_3^-$. The most challenging part of GWQI calculation is the assignment of the weights since there is no consensus on the weight that should be given to each physiochemical parameter. In this study, Pearson correlation analysis is applied to aid in the weight assignment; consequently, GWQI is computed. The measured GQWI indicated that most of the groundwater samples fall in excellent and good categories, and only one sample (BH 18) showed a GWQI of 396, projected in the unsuitable class.

The major limitation of the weighted arithmetic GWQI model is the calculation of the sub-indices since it is time-consuming and prone to calculation errors. Artificial intelligence (AI) techniques are introduced to cope with the limitations associated with conventional GWQI models. Soft computing models such as multilayer perceptron (MLP) neural network and support vector regression (SVR) are proposed to reduce the time for sub-indices calculation. The architecture of MLP network involves inputs, hidden layers, and output. The inputs are the physiochemical parameters, (2) hidden layers are applied, and the output is the GWQI. For SVR, the Gaussian kernel function is applied to find the optimal hyperplane in the data and thus predict the GWQI. In this research, the collected groundwater samples are used for training and validation of the developed model in a ratio of 80% to 20%, respectively. The performance metrics revealed that AI models could be applied successfully for groundwater quality assessment as an alternative to conventional GWQI. Furthermore, they suggested that the prediction capabilities of SVR models are higher than MLP, mainly due to the high ability of SVR to process complex nonlinear data.

The results obtained from GWQI models helped to understand the groundwater's overall quality. It is indicated that groundwater quality in north Khartoum State is generally acceptable for human consumption except
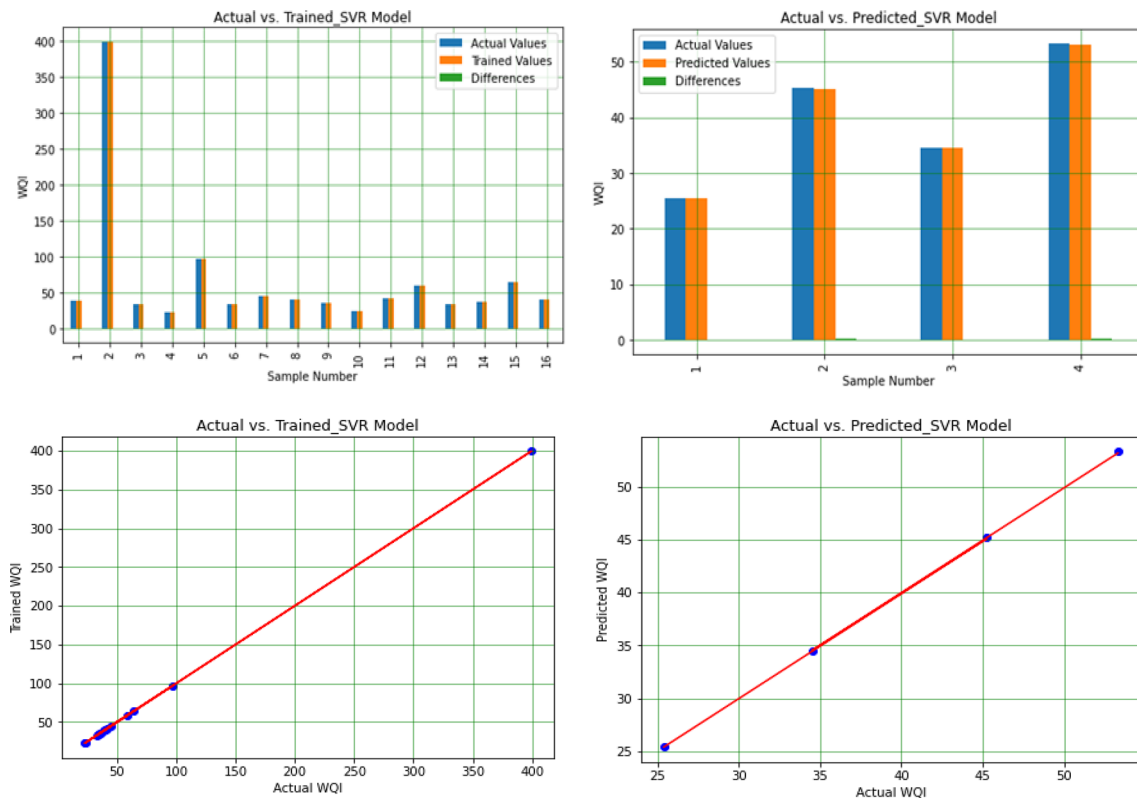
**Fig. 10** Actual versus simulated WQI during training and validation for SVR model

for some samples with high salinity. Consequently, for water supply sustainability, the present study suggested implementing a groundwater quality monitoring program in the study area since pollution spread may affect the suitability of groundwater. The general outcomes of the present research indicate the benefits of using AI techniques in GWQI prediction with enhanced accuracies. The algorithms established in this research can be used for groundwater quality evaluation effectively.

**Author contributions** MAAM: methodology, development of WQI model, statistical analysis, original draft writing. NAAK: development of MLP and SVR models, validation. NPS: writing review and editing, supervision. PS: review and editing, supervision.

**Availability of data and materials** Not applicable.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Ethical approval** The authors confirm that all the research met the ethical guidelines.

**Consent to participate** Not applicable.

**Consent to publish** The authors declare that this work does not contain any material from any individual.

## References

Abdelsalam YE, EA EM, Elhadi H El (2016) Problems and factors which retard the development and the utilization of groundwater for drinking purposes in the Khartoum state-SUDAN. In: 7th international conference on environment and engineering geophysics & summit forum of Chinese Academy of Engineering on Engineering Science and Technology, pp 449–451

Abdo G, Salih A (2012) Challenges facing groundwater management in Sudan

Agrawal P, Sinha A, Kumar S et al (2021) Exploring artificial intelligence techniques for groundwater quality assessment. Water Switz. https://doi.org/10.3390/w13091172

Ahmed U, Mumtaz R, Anwar H et al (2019) Efficient water quality prediction using supervised. Water 11:1–14

Akbari M, Jalali F (2007) Dew point pressure estimation of gas condensate reservoirs, using artificial neural network (ANN). In: Paper SPE 107032 presented at the Society of Petroleum Engineers Europec/EAGE annual conference and exhibition, London, pp 11–14

Aldhyani THH, Al-Yaari M, Alkahtani H, Maashi M (2020) Water quality prediction using artificial intelligence algorithms. Appl Bionics Biomech 2020

Appelo CAJ, Postma D (2005) Geochemistry, groundwater and pollution, 2nd edn. Balkema, Rotterdam

Asadi E, Isazadeh M, Samadianfard S et al (2020) Groundwater quality assessment for sustainable drinking and irrigation. Sustainability 12:177

Asadollah SBHS, Sharafati A, Motta D, Yaseen ZM (2021) River water quality index prediction and uncertainty analysis: a comparative study of machine learning models. J Environ Chem Eng 9:104599. https://doi.org/10.1016/j.jece.2020.104599

Awad AZ (1994) Stratigraphic palyloical and paleoclogical studies in east Central Sudan (Khartoum–Kosti Basin) Late Jurassic to mid tertiary. Berliner Geowiss B161Technical univ Berliner

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

Bagheripour P, Gholami A, Asoodeh M (2015) Support vector regression between PVT data and bubble point pressure. J Pet Explor Prod Technol 5:227–231

Belayneh A, Adamowski J, Khalil B (2016) Short-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet transforms and machine learning methods. Sustain Water Resour Manag 2:87–101

Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. In: Proceedings of the fifth annual workshop on computational learning theory, pp 144–152

Brown RM, McClelland NI, Deininger RA, Tozer RG (1970) A water quality index-do we dare. Water Sew Works 117

Bui DT, Khosravi K, Tiefenbacher J et al (2020) Improving prediction of water quality indices using novel hybrid machine-learning algorithms. Sci Total Environ 721:137612

Edition F (2011) Guidelines for drinking-water quality. WHO Chron 38:104–108

Elbeltagi A, Pande CB, Kouadri S, Islam ARMT (2021) Applications of various data-driven models for the prediction of groundwater quality index in the Akot basin, Maharashtra, India. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-021-17064-7

Farah EA, Abdullatif OM, Kheir OM, Barazi N (1997) Groundwater resources in a semi-arid area: a case study from central Sudan. J Afr Earth Sci 25:453–466

Farah EA, Mustafa EMA, Kumai H (2000) Sources of groundwater recharge at the confluence of the Niles, Sudan. Environ Geol 39:667–672

Freeze RA, Cherry JA (1979) Groundwater. Prentice-Hall, Hoboken

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

Gholami V, Darvari Z, Mohseni Saravi M (2015) Artificial neural network technique for rainfall temporal distribu-tion simulation (case study: Kechik region). Casp J Environ Sci 13:53–60

Gholami V, Khaleghi MR, Pirasteh S, Booij MJ (2021) Comparison of self-organizing map, artificial neural network, and co-active neuro-fuzzy inference system methods in simulating groundwater quality: geospatial artificial intelligence. Water Resour Manag. https://doi.org/10.1007/s11269-021-02969-2

Gitau MW, Chen J, Ma Z (2016) Water quality indices as tools for decision making and management. Water Resour Manag 30:2591–2610

Haghiabi AH, Azamathulla HM, Parsaie A (2017) Prediction of head loss on cascade weir using ANN and SVM. ISH J Hydraul Eng 23:102–110

Hassan I, Elhassan BM, Mustafa MA (2017) Heavy metals and refractory organic compounds in Khartoum State's groundwater resources. Eur J Eng Technol Res 2:13–16

Heddam S (2016) Simultaneous modelling and forecasting of hourly dissolved oxygen concentration (DO) using radial basis function neural network (RBFNN) based approach: a case study from the Klamath River, Oregon, USA. Model Earth Syst Environ 2:1–18

Hem JD (1985) Study and interpretation of the chemical characteristics of natural water. Department of the Interior, US Geological Survey

Horton RK (1965) An index number system for rating water quality. J Water Pollut Control Fed 37:300–306

Hussein MT, Awad HS (2006) Delineation of groundwater zones using lithology and electric tomography in the Khartoum basin, central Sudan. C R Geosci 338:1213–1218

Imneisi IB (2019) Using algorithm (Levenberg marquardt) as activation function to prediction Water Quality Index (WQI) in Kastamonu City-Turkey

Kadam AK, Wagh VM, Muley AA et al (2019) Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India. Model Earth Syst Environ 5:951–962. https://doi.org/10.1007/s40808-019-00581-3

Kanga IS, Naimi M, Chikhaoui M (2020) Groundwater quality assessment using water quality index and geographic information system based in Sebou River Basin in the North-West region of Morocco. Int J Energy Water Resour 4:347–355

Kaya YZ, Zelenakova M, Üneş F et al (2021) Estimation of daily evapotranspiration in Košice City (Slovakia) using several soft computing techniques. Theor Appl Climatol 144:287–298. https://doi.org/10.1007/s00704-021-03525-z

Kayemah N, Al-Ruzouq R, Shanableh A, Yilmaz AG (2021) Evaluation of groundwater quality using Groundwater Quality Index (GWQI) in Sharjah, UAE. In: E3S web of conferences

Kheiralla MK (1966) Study of the Nubian Sand stone Formation of the Nile Vally between 14 N and 17 42 N, with reference to groundwater geology. University of Khartoum

Koranga M, Pant P, Kumar T et al (2022) Efficient water quality prediction models based on machine learning algorithms for Nainital Lake, Uttarakhand. Mater Today Proc. https://doi.org/10.1016/j.matpr.2021.12.334

Majumdar D, Gupta N (2000) Nitrate pollution of groundwater and associated human health disorders. Indian J Environ Health 42:28–39

Manzar MS, Benaafi M, Costache R et al (2022) New generation neurocomputing learning coupled with a hybrid neuro-fuzzy model for quantifying water quality index variable: a case study from Saudi Arabia. Ecol Inform 101696

Mohammed MAA, Szabó NP, Szűcs P (2022) Multivariate statistical and hydrochemical approaches for evaluation of groundwater quality in north Bahri city-Sudan. Heliyon 8:e11308. https://doi.org/10.1016/j.heliyon.2022.e11308

Momenzadeh L, Zomorodian A, Mowla D (2011) Experimental and theoretical investigation of shelled corn drying in a

microwave-assisted fluidized bed dryer using artificial neural network. Food Bioprod Process 89:15–21

Nasir N, Kansal A, Alshaltone O et al (2022) Water quality classification using machine learning algorithms. J Water Process Eng 48:102920

Nathan NS, Saravanane R, Sundararajan T (2017) Application of ANN and MLR models on groundwater quality using CWQI at Lawspet, Puducherry in India. J Geosci Environ Prot 05:99–124. https://doi.org/10.4236/gep.2017.53008

Ram A, Tiwari SK, Pandey HK et al (2021) Groundwater quality assessment using water quality index (WQI) under GIS framework. Appl Water Sci 11:1–20

Ramakrishniah CR, Sadashivaiah C, Ranganna G (2009) Assessment of water quality index for the groundwater in Tumkur Taluk. E-J Chem 6:523–530

Rishi MS, Kaur L, Sharma S (2020) Groundwater quality appraisal for non-carcinogenic human health risks and irrigation purposes in a part of Yamuna sub-basin, India. Hum Ecol Risk Assess 26:2716–2736. https://doi.org/10.1080/10807039.2019.1682514

Saeed EM (1974) Geological and hydrogeological studies of Khartoum Province, Sudan. Ph.D. thesis, Cairo University

Sakizadeh M (2016) Artificial intelligence for the prediction of water quality index in groundwater systems. Model Earth Syst Environ 2:1–9

Sawyer CN, McCarty PL (1967) Chemistry for sanitary engineers

Schaid DJ, Buetow K, Weeks DE et al (1999) Discovery of cancer susceptibility genes: study designs, analytic approaches, and trends in technology. JNCI Monogr 1999:1–16

Selvakumar S, Ramkumar K, Chandrasekar N et al (2017) Groundwater quality and its suitability for drinking and irrigational use in the Southern Tiruchirappalli district, Tamil Nadu, India. Appl Water Sci 7:411–420

Sengupta P (2013) Potential health impacts of hard water. Int J Prev Med 4:866

Sillberg CV, Kullavanijaya P, Chavalparit O (2021) Water quality classification by integration of attribute-realization and support vector machine for the Chao Phraya river. J Ecol Eng 22:70–86. https://doi.org/10.12911/22998993/141364

Singh DF (1992) Studies on the water quality index of some major rivers of Pune, Maharashtra. In: Proceedings of the national academy of environmental biology, pp 61–66

Singh AK, Mondal GC, Kumar S et al (2008) Major ion chemistry, weathering processes and water quality assessment in upper catchment of Damodar River basin, India. Environ Geol 54:745–758

Singh SP, Tripathi SK, Vimal K et al (2015) Hydrochemical investigation and groundwater quality evolution for irrigation purpose in some blocks of Varanasi district, Uttar Pradesh, India. Int J Trop Agric 33:1653–1660

Singha S, Pasupuleti S, Singha SS et al (2021) Prediction of groundwater quality using efficient machine learning technique. Chemosphere 276:130265. https://doi.org/10.1016/j.chemosphere.2021.130265

Srivastava SK (2019) Assessment of groundwater quality for the suitability of irrigation and its impacts on crop yields in the Guna district, India. Agric Water Manag 216:224–241

Tian Y, Jiang Y, Liu Q et al (2019) Using a water quality index to assess the water quality of the upper and middle streams of the Luanhe River, northern China. Sci Total Environ 667:142–151

Tiwari TN, Mishra MA (1985) A preliminary assignment of water quality index of major Indian rivers. Indian J Env Prot 5:276–279

Tokar AS, Markus M (2000) Precipitation-runoff modeling using artificial neural networks and conceptual models. J Hydrol Eng 5:156–161

Tom I, Okon A, Okologume W (2020) Volumetric approach based DOE and ANN models for estimating reservoirs oil in place. Int J Sci Eng Res 11:873–883

Toprak ZF, Cigizoglu HK (2008) Predicting longitudinal dispersion coefficient in natural streams by artificial intelligence methods. Hydrol Process Int J 22:4106–4129

Tung TM, Yaseen ZM et al (2020) A survey on river water quality modelling using artificial intelligence models: 2000–2020. J Hydrol 585:124670

Wang Y, Yuan Y, Pan Y, Fan Z (2020) Modeling daily and monthly water quality indicators in a canal using a hybrid wavelet-based support vector regression structure. Water. https://doi.org/10.3390/w12051476

Whiteman AJ (1971) Geology of the Sudan Republic