

# Prediction of Abidjan Groundwater Quality Using Machine Learning Approaches: An Exploratory Study

# **Dion Gueu Edith Kressy**

School of Resources and Environmental Engineering, Hefei University of Technology, Hefei, China Email: kressy.dion@gmail.com

How to cite this paper: Kressy, D.G.E. (2024) Prediction of Abidjan Groundwater Quality Using Machine Learning Approaches: An Exploratory Study. *Intelligent Control and Automation*, **15**, 215-248. https://doi.org/10.4236/ica.2024.154010

Received: October 5, 2024 Accepted: November 25, 2024 Published: November 28, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

http://creativecommons.org/licenses/by/4.0/

CC Open Access

# Abstract

Continuous groundwater quality monitoring poses significant challenges affecting the environment and public health. Groundwater in Abidjan, specifically from the Continental Terminal (CT), is the primary supply source. Therefore, ensuring safe drinking water and environmental protection requires a thorough evaluation and surveillance of this resource. Our present research evaluates the quality of the CT groundwater in Abidjan using the water quality index (WQI) based on the analytical hierarchy process (AHP). This study also explores the application of machine learning predictions as a timeefficient and cost-effective approach for groundwater resource management. Therefore, three Machine Learning regression algorithms (Ridge, Lasso, and Gradient Boosting (GB)) were executed and compared. The AHP-based WQI results classified 98.98% of samples as "good" water quality, while 0.68% and 0.34% of samples were respectively categorized as "excellent" and "poor" water. Afterward, the prediction performance evaluation highlighted that the GB outperformed the other models with the highest accuracy and consistency (MSE = 0.097, RMSE = 0.300, r = 0.766,  $r_s = 0.757$ , and  $\tau = 0.804$ ). In contrast, the Lasso model recorded the lowest prediction accuracy, with an MSE of 148.921, an RMSE of 6.828, and consistency parameters of r = 0.397,  $r_s = 0.079$ , and  $\tau$  = 0.082. Gradient Boosting regression effectively learns nonlinear events and interactions by iteratively fitting new models to errors of previous models, enabling a more realistic groundwater quality prediction. This study provides a novel perspective for improving groundwater quality management in Abidjan, promoting real-time tracking and risk mitigations.

# **Keywords**

Groundwater, AHP Weight-Based WQI, Machine Learning Prediction,

**Regression Models** 

## **1. Introduction**

Worldwide, Groundwater quality presents urgent and complex challenges impacting the environment, health, and well-being of individuals in communities. Being a major source of fresh water, the quality of groundwater is essential for sustaining life, supporting ecosystems, and driving economic activities [1]. In Abidjan the capital of Côte d'Ivoire, Groundwater from the Continental Terminal (CT) accounts for about 68% of the total drinking water production [2]. Unfortunately, this resource quality is subject to continuous deterioration, as reported by the national water supply company [3]. The key reported causes of deterioration include the dispersion of solid and liquid waste materials, the sanitation equipment and system shortage, the improper disposal of used oils, and the discharge of non-compliant waste from households, industries, or automobile workshops [4]. Moreover, over 79% of the district's groundwater shows unsafe hydrogeological protection, while 12% is vulnerable [5]. Developing nations like Côte d'Ivoire strive to raise and regulate water quality monitoring systems. However, Abidjan aquifers lack unified approaches and resource management policy as there are several decision-making centers with opposing stands [6]. Besides these issues, another challenge in groundwater quality monitoring is the complex and variable nature of the hydrogeological system, which can generate fuzziness in the assessment process [7]-[9]. Furthermore, water quality assessment generates an intricate database of various parameter types, often overlooked by most stakeholders. The primary interest of these assessments lies in the interpretation derived with a significantly heightened focus on the inferences drawn regarding the suitability of the water resources for different purposes (drinking, agriculture, or industry) [10] [11]. Consequently, setting up a detailed approach to offer insights into ground and surface water quality regarding their intended functions is important [12] as applied by the WQI. The Water quality index (WQI) defines the overall state of water bodies by converting concentrations of water quality parameters into a numerical score. This is achieved via calculation techniques [13], minimizing the significant amount of data in a single value and enabling their straightforward interpretation [14] [15]. The WQI is widely used to classify ground and surface water quality [16]-[18]. The applied WQI model comprises five key steps, including the selection of water quality parameters, the conversion of parameter values into dimensionless sub-indices or parameters standardization, the weight attribution to individual parameters, the aggregation of the weighting factors with standardized values to obtain the final WQI value, and the categorization of the resource quality based on the WQI classification ranges [19] [20].

The primary drawback is the variability of the conventional weight assignment to parameters in the WQI calculation [21] [22]. In contrast, multi-criteria decision

analysis methods (MCDM), such as the analytical hierarchy process (AHP), represent a consistent weighting system that avoids misjudgments. AHP is a robust method for determining parameter weights through matrix comparisons that minimizes errors and improper weight distributions [20]. In water quality assessment, the need for large amounts of data requires considerable time and resources (technical and financial) [23] [24]. Additionally, handling dense hydrogeological data using traditional water quality evaluation methods can lead to information losses or model inaccuracies [25]. Consequently, scholars considered machine learning (ML) a more convenient approach due to its quick, cost-effective, and accurate forecasting capabilities [23] [24] [26] [27]. Machine learning is a multifaceted method that enables insights and interpretation of system behaviors from input data, enhancing holistic water resource management and strategic planning [24] [26] [27]. Moreover, ML can be a relevant alternative in situations lacking exhaustive hydrogeological data for detailed modeling with physics-based models [28].

For two decades, Scholars have significantly employed ML prediction models for their studies in the water management system, including regression techniques such as decision tree regression (DT), boosting ensembles, ridge regression (RR), lasso regression (LR), and the artificial neural network (ANN). The DT regression model has been applied in several studies, such as river water quality prediction [29] or applied in combination with the support vector regression (SVR) for wastewater quality indicator prediction [30]. The boosting ensembles have been used for discharge coefficient estimation [31], arsenic adsorption in water treatment [32], among others. Ridge regression was also used for medium and longterm runoff forecasting [33], groundwater quality forecasting with other models, such as lasso [29], and other techniques. Artificial neural network prediction has been extensively utilized in various studies, including flood susceptibility [34], nitrate or fluoride contamination [35] [36], etc.

The integration of ML in water quality assessment using the WQI method has also gained traction recently [37]. For example, the prediction of some river's quality index in India was optimized through eight individual machine-learning regression methods, which include DT, Ridge, Lasso, and ANN [29]. The water quality rating scale and water quality weight score were derived as feature sets. The results showed that linear regression (LR) and Ridge trained using the scale accurately predicted WQI, with MSE = 0 and r = 1, outperforming the existing models. Ahmed *et al.* [38] applied eight algorithms, such as gradient boosting and polynomial regression, to predict WQI. The Gradient boosting algorithm performs better with MAE = 1.9642, while the polynomial regression has MAE = 2.7273. Yilma *et al.* [39] forecasted the WQI of the Akaki River using the artificial neural network. The model involved eight hidden layers and 15 hidden neurons predicting WQI with more than 90% accuracy. Gupta *et al.* [40] proposed an ANN model for predicting WQI using a cascade forward network with the best predictability. Leong *et al.* [41] developed two models for WQI forecasting: the support vector machine (SVM) with three kernel functions (linear, polynomial, and Radial Basis Function) and the least squares support vector machine (LS-SVM). The polynomial kernel demonstrates the best performance for the SVM, with an  $R^2$  of 0.8796. Moreover, the LS-SVM that used pertinent predictors for training had higher accuracy, with an  $R^2$  of 0.9227, while the SVM that used all the predictors yielded an  $R^2$  of 0.9184.

Past literature reveals that different ML models perform optimally under diverse hydrogeological conditions. Nonetheless, most studies on WQI prediction have employed conventional weight assignment methods. To the best of our knowledge, no research has evaluated Abidjan CT groundwater quality using a WQI based on AHP MCDM. In addition, no investigation into the performance of ML models for groundwater quality prediction in Abidjan has been recorded. To address this gap, this exploratory study aims to predict an AHP-based WQI for Abidjan CT groundwater through three regression algorithms (Ridge, Lasso, and Gradient Boosting). The Continental Terminal groundwater is a vital source for the Abidjan population, and if its quality can be forecasted, this can control and mitigate significant risks. Hence, this research provides novel perspectives on managing Abidjan groundwater through machine learning prediction. Integrating AHP in the WQI method also simplifies and improves the determination of the relative importance of groundwater quality indicators. Lastly, this study allows a regular evaluation framework without the necessity of extensive sampling, laboratory analysis time, and expenses.

# 2. Study Area

The Continental terminal aquifer considered in this study is predominantly found in the district of Abidjan, the southeastern part of Côte d'Ivoire (Figure 1). The district covers the area between latitude 5°20'10"N and longitude 4°01'39"W, with an elevation ranging from 0 to 200m [42]. This region is characterized by an equatorial transitional climate known as the Attieen climate, with two wet seasons (April to July and October to November) and two dry seasons (August to September and December to March) [43]. The geology of Abidjan corresponds to the coastal sedimentary basin, representing 2.5% of Côte d'Ivoire's surface area and measuring 350 to 400 km long and 40 to 50 km wide [43]. The sedimentary basin in Abidjan is composed of three aquifers: the Quaternary, the Continental Terminal, and the Maastrichtian. The CT aquifer, further north in the basin, covers its entire surface aside from the Quaternary coastal area [44]. The CT aquifer is unconfined from the Mio-Pliocene age formations in the form of high plateaus [42]. These formations are characterized by a lenticular stratification of coarse sands, varicolored clays, ferruginous sandstones, and iron ores [45]. Known as the Abidjan aquifer, CT exploitation accounts for more than 50% of the national drinking water production. The CT groundwater is mainly recharged by two processes: the direct infiltration through the Quaternary formation (upper layer) and the long rainy season, which recharges the groundwater in approximately



**Figure 1.** CT groundwater sampling location in Abidjan, Côte d'Ivoire: Anonkoua Kouté (AK), Niangon Nord (NN), Djibi (DJI), Nord Riviera (NR), Riviera Centre (RC), Zone Ouest (ZO), Zone Est (ZE), Zone Nord (ZN), Adjamé Nord (AN).

two months and lasts 2 to 4 months [43] [45].

# 3. Methodology

# 3.1. Data Collection

In this study, nine sampling sites in the CT groundwater of Abidjan were studied. Given the impact of the rainy season on CT groundwater recharge and imploring the representative sampling method, the National Public Hygiene Laboratory in Abidjan provided data (98 samples/site) from their sampling campaigns from May 2nd to September 1st, 2022. These samples were considered sufficiently representative to approximate the characteristics of the broader population from which they were extracted [46]. A total of twelve indicators have been assessed in compliance with World Health Organization (WHO) specifications, notably pH, electric conductivity (EC), total dissolved solids (TDS), chloride (Cl<sup>-</sup>), sulfate ( $SO_4^{2-}$ ), bicarbonate ( $HCO_3^{-}$ ), sodium (Na<sup>+</sup>), potassium (K<sup>+</sup>), magnesium (Mg<sup>2+</sup>), calcium (Ca<sup>2+</sup>), nitrate ( $NO_3^{-}$ ), and aluminum (Al<sup>3+</sup>). The methodology comprises two key phases designed to achieve the research objectives. Initially, the computation of the WQI was conducted using the Analytic Hierarchy Process (AHP) to

determine the relative weight of parameters. Subsequently, the prediction of the AHP-based WQI was executed, applying the twelve physicochemical parameters as features and the WQI as label.

#### 3.2. Water Quality Index Calculation

The Water Quality Index is a popular method with straightforward mathematical features and a user-friendly interface. This approach uses data or indicators from a water body to determine a unique number. The obtained number is then classified into a category that reflects the water quality status [19]. The analytical hierarchy process (AHP) is a multi-criteria decision analysis method (MCDM) developed by Saaty [47]. In recent years, numerous researchers have used AHP to calculate the relative weight of parameters for different groundwater studies [48]-[50]. This method has the advantage of limiting the possibility of errors and incorrect distribution of weight to a specific parameter [50] [51]. Additionally, the consistency of the approach can be verified after computation, and the pair-wise comparison matrix generated in the calculation is a promising approach for providing relative weights [20].

#### 3.2.1. Parameters Relative Weights Calculation

The AHP-weighting was implemented through a four-step process.

**Step 1:** The parameters are hierarchically organized considering their nature or function.

**Step 2:** Scales of 1 to 9 (**Table 1**) are attributed to the parameters based on their significance in the overall water quality.

Table 1. Saaty's lineal scale of preferences in the pair-wise comparison process.

Numerical rating	Judgments of preferences between factor <i>i</i> and <i>j</i>
1	factor $i$ is equally important to factor $j$
3	factor $i$ is slightly more important than factor $j$
5	factor <i>i</i> is clearly more important than <i>j</i>
7	factor $i$ is strongly more important than factor $j$
9	factor $i$ is extremely more important than factor $j$
2, 4, 6, 8	Intermediate values

**Step 3:** The pair-wise comparison matrix is constructed. Considering *n* evaluation factors and the importance intensity of factor *i* over *j*. The pair-wise comparison matrix A gives:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$
(1)

 $a_{ij} = W_i / W_j$ ; where  $W_i$  and  $W_j$  represent the assigned scores to factors *i* and *j*,

respectively. This value is supposed to approximate the relative importance of *i* to *j*.

The factors eigenvectors are then calculated using the average of the normalized column, presented as

$$w_i = \frac{1}{n} \sum_{j=1}^n \frac{a_{ij}}{\sum_{i=1}^n a_{ij}}; (i, j = 1, 2, \dots, n)$$
(2)

**Step 4:** The consistency index (*CI*) and the consistency ratio (*CR*) are calculated to validate the judgment of the different values assigned to parameters.

$$CR = \frac{CI}{RCI}$$
(3)

where

$$CI = \frac{\lambda_{\max} - n}{n - 1} \tag{4}$$

With *n*, the pair-wise comparison matrix size, and  $\lambda_{\max}$ , the principal eigenvalue of the matrix *A*. The consistency is observed only when matrix *A* is a positive reciprocal one and  $\lambda_{\max}$  is closer to *n*.

$$\lambda_{\max} = \sum_{i=1}^{n} \frac{(A\vec{w})_i}{nw_i}$$
(5)

where  $A\vec{w} = A_iW_n$  is the product of the geometrical mean of the pair-wise comparison matrix rows and the normalized weight;  $nw_i$  is the normalized weight of each variable  $W_n$ .

*RCI* is the average random consistency index, computed and tabulated as shown in the following Table 2. If CR < 0.1, the numerical judgments will be considered acceptable.

Table 2. Average random consistency index values according to matrix size.

п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
RCI	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49	1.52	1.48	1.56	1.57	1.59

#### 3.2.2. Calculation of Quality Rating

The calculation is performed using the normalized rating scale  $(q_i)$ , which is obtained by subtracting the measured value of a parameter from the acceptable level and then dividing it all by the permissible level. The equation is formulated as follows:

$$q_i = \frac{A_l - V_m}{A_l} \tag{6}$$

where  $A_l$  is the acceptable level set by the WHO [52] for each parameter in drinking water and  $V_m$  is the measured value of this parameter.

### 3.2.3. Determination of the Index

For computing WQI, the sub-index SI for each parameter gives:

$$SI_i = W_i \times q_i \tag{7}$$

where  $W_i$  is the normalized weight of the  $t^{\text{th}}$  parameter.

Then, WQI = 
$$\sum_{i=1}^{n} SI_i$$
 (8)

The computed WQI values are finally classified into five categories (**Table 3**): excellent, good, poor, very poor, and unfit water for drinking purposes [53].

Table 3. Water quality classification ranges.

Range	Water quality
<50	Excellent water
50 - 100.1	Good water
100 - 200.1	Poor water
200 - 300.1	Very Poor water
>300	Unfit for drinking

#### 3.3. Machine Learning Prediction

#### 3.3.1. Data Pre-Processing

Our study experiment was conducted in the Windows 11 system, using Python (version 3.11.0) as the programming language for code programming. The data reading and preprocessing via Pandas (version 2.1.4). Machine learning models were built through scikit-learn (version 1.2.2), and matplotlib (version 3.8.0) was used to draw graphics. First, the data preprocessing consisted of filling the missing values with 0 using the Pandas tool. The second step focused on feature scaling, with the min-max normalization technique commonly applied for standardization tasks [54]. These features range between 0 and 1, helping to improve the convergence speed and performance of specific machine-learning algorithms.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{9}$$

Here, x is the original data, x' represents the normalized data,  $\max(x)$  represents the maximum value of the original data while  $\min(x)$  represents the minimum value of the original data.

The third step encompasses data splitting into training and test sets. The training set allows ML algorithm data to learn how to generate predictions and uncover insights within the dataset, while the test set is used to execute the model. The current experiment implies time series data, where data before August 1, 2022, were used as the training set and those after August 1, 2022, as the test set. Based on the slide window projection, data features for seven consecutive days have been input in each model's windows (window size = 10) to enable water quality predictions after seven days.

#### 3.3.2. Learning Algorithms

• Ridge Regression

The basic principle of ridge regression analysis introduced by Hoerl and Kennard [55], established research to estimate biases with a smaller Mean squared error (MSE). By adding a small constant 'k' to the diagonal of the correlation matrix of independent variables, the error variance can be considerably reduced to manage the overall instability of ordinary least square (OLS) estimates. The calculation formula gives:

$$\hat{\beta}_{\text{ridge}} = \arg_{\beta} \left\{ \sum_{i=1}^{N} \left( d_i - \beta_0 - \sum_{j=1}^{p} x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}$$
(10)

With  $\hat{\beta}$  as the ridge estimator, *N* samples, the constant '*k*,' the *p* covariates, and  $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$  the covariate vector, as well as  $\lambda$  which refers to the shrinkage parameter indicating the larger the value of  $\lambda$ , the regression coefficients decline further as they approach zero. Before implementing Ridge regression, the inputs need to be standardized because the obtained solutions are not homogeneous when the input variables are scaled. Also, the intercept  $\beta_0$  has not been penalized in the Ridge regression.

Lasso Regression

Lasso is a linear model with regularization, initially introduced in the context of least squares. Tibshirani [56], presented the model as an innovative variable selection mechanism for regression, optimizing the residual sum of squares while constraining the total absolute coefficient magnitudes. Lasso is a well-known regression method that regularizes the parameter  $\beta$  under the sparse assumption [57]. Like ridge regression, the model considers *N* sample, *p* covariates, and a single outcome. Supposing  $y_i$  is the response variable, and  $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ is the covariate vector for the *t*<sup>th</sup> case  $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ , the following relation gives the mathematical expression:

$$\sum_{i=1}^{n} \left(\beta_{0} + \sum_{k=j}^{K} \beta_{k} x_{k,i} - y_{i}\right)^{2}$$
(11)

In Lasso regression,  $\lambda$  subject is computed to minimize the residual sum of the square:

$$\sum_{i=1}^{n} \left(\beta_{0} + \sum_{k=j}^{K} \beta_{k} x_{k,i} - y_{i}\right)^{2} + \lambda \sum_{k=1}^{k} \beta_{k}$$
(12)

#### Gradient Boosting

The Gradient Boosting (GB) approach is an ensemble learning technique [27] that integrates multiple weak models to enhance overall performance. The GB algorithm consists of three key elements: the loss function, the weak learner, and the additive model. The loss function component must constantly be optimized to reduce the prediction error. Gradient Boosting is based on sequential learning using weak learners like the decision trees model to mitigate the losses without altering the existing tree. Instances mistakenly identified in one phase are assigned additional weight in the subsequent step. This way, each subsequent weak learner is designed to address and rectify the errors made by the previous models [58] [59]. Friedman provided the entire calculation process [60], explained as follows:

Considering the predictors x, the targe y, and the number of observations  $n = 1, \dots, n$ .

• Model initialization with a constant value:

$$F_0(x) = \arg\min_{\gamma} \sum_{i=1}^n L(y_i, \gamma)$$
(13)

where  $F_0$  is the initial constant value prediction and *L*, the loss function.

• For *m* = 1 to *M* with *M*, the number of created trees and *m*, the index of each tree.

Residuals computation: 
$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$
 (14)

for  $i = 1, \dots, n$  and  $F_{m-1}$ , previous prediction.

- Regression tree training and terminal node reasons creation  $R_{im}$
- for  $j = 1, \dots, J_m$ ; with the features x against  $\gamma$  j, the terminal or leave node and J, the total number of leaves.

Computation of 
$$\gamma_{jm} = \arg\min_{\gamma} \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma)$$
 (15)

for  $j = 1, \dots, J_m$ 

Model update: 
$$F_m(x) = F_{m-1}(x) + v \sum_{j=1}^{J_m} \gamma_{jm} l(x \in R_{jm})$$
 (16)

## 3.4. Prediction Performance Assessment

Verifying the accuracy of the constructed models is an essential phase in predictions; otherwise, the predictive model would lack scientific significance [61]. The performance assessment indicators for our study's prediction include the mean square error (MSE), the root mean square error (RMSE), the Pearson, the Spearman, and the Kendall correlations. The Pearson, Spearman, and Kendall correlation coefficients represent the consistency evaluation metrics. These indicators are calculated as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2$$
(17)

RMSE = 
$$\sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2}$$
 (18)

where,  $x_i$  is the  $t^{h}$  observed value,  $y_i$  is the corresponding predicted value, and n the number of observations.

MSE and RMSE values approaching 0 imply higher prediction accuracy, whereas values approaching 1 reflect lower prediction performance [62].

The correlation coefficients indicate the degree of linear association between two variables, with 1 indicating a positive relationship, -1 indicating a negative relationship, and 0 indicating no linear association [63] [64]. For variables *a* and *b* with *n* data sets (each data set can be labeled  $a_i$  and  $b_i$ , where  $i = 1, 2, \dots, n$ ), the coefficient is used to estimate the correlation *r* between *a* and *b*.

Pearson: 
$$r_p = \frac{\sum_{i=1}^n (a_i - \overline{a}) (b_i - \overline{b})}{\sqrt{\sum_{i=1}^n (a_i - \overline{a})^2} \sqrt{\sum_{i=1}^n (b_i - \overline{b})^2}}$$
 (19)

Spearman: 
$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$
 (20)

Kendall: 
$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}$$
 (21)

where  $\overline{a}$  and  $\overline{b}$  are the means of *a* and *b*, respectively,  $d_i = R(a_i) - R(b_i)$  is the difference between the two ranks of each variable and  $n_c$  is the number of concordance and  $n_d$  the number of discordances.

## 4. Results and Discussion

#### 4.1. Physico-Chemical Parameters Evolution

**Table 4** presents the descriptive statistics of all the analytical data of the groundwater samples collected from the study area, majorly during the long rainy season. Most parameters were found within permissible limits of the WHO [52], except for K<sup>+</sup> and Al<sup>3+</sup>, which exceeded the permissible thresholds in 4.76% and 3.4% of samples, respectively. In the CT aquifer, water is particularly acidic, with pH values between 4 and 5.9. The primary factor predicting this pH is the acidity of the leached ferrallitic soil of the southern coastal plateau landscapes, which is 3.9 according to the soil record of the Côte d'Ivoire sedimentary basin [65]. Acidic rainwater recharge also influences the pH of the resource [66]. The observed EC varies from 4.1 to 469.0  $\mu$ S/cm, indicating a low conductivity due to groundwater mixing, as the CT is an unconfined aquifer. After surface recharge enters the aquifer, a significant dilution occurs due to less mineralization and more acidic water [43]. The relatively high values correspond to areas that lack a functional sanitation infrastructure [67]. The observed TDS varies from 0.0629 mg/L to 214.315 mg/L, which ascertains the low salinity of the resource and the groundwater mixture.

The anions order in the overall samples is  $HCO_3^- > Cl^- > NO_3^- > SO_4^{2-}$ . The recorded  $HCO_3^-$  level varies between 0 and 240.5 mg/L. In some samples, higher concentrations of  $HCO_3$  can be justified by the weathering of silicates [68], occurring when they react with  $CO_2$  gas from root respiration and aqueous  $CO_2$  from the water, leading to a potential release of Ca, Mg, and  $HCO_3$  into the groundwater [69]. The measured concentration of Cl<sup>-</sup> varies from 1.68 mg/L to 16.908 mg/L, while the nitrate ranges from 0 to 48.73 mg/L. The values of  $NO_3^-$  close to permissible limits are symptomatic of sewage effluent drainage and decomposition of organic material with precipitation in the unconfined CT groundwater. The sulfate content ranges from 0 to 18.4 mg/L as the origins of  $SO_4^{2-}$  in groundwater are mainly from the dissolution or oxidation of sulfate minerals and acidic rainwater recharge [70].

The cationic order in the CT groundwater samples is  $Ca^{2+} > Na^+ > Mg^{2+} > K^+ >$ 

		EC	pН	Cl-	$NO_3^-$	$SO_4^{2-}$	$HCO_3^-$	Na <sup>+</sup>	<b>K</b> <sup>+</sup>	Mg <sup>2+</sup>	Ca <sup>2+</sup>	Al <sup>3+</sup>	TDS
	Min	19.56	4.5	1.82	2.02	0.025	0.01	1.066	0.01	0.002	0.5	0.013	7.773
	Max	46.54	4.96	3.83	5.72	2.86	2.46	2.173	0.99	1.42	0.993	4.209	15.186
AK	Mean	32.297	4.651	2.629	3.934	1.291	1.070	1.854	0.272	0.588	0.713	0.164	10.911
	SD	7.913	0.112	0.532	0.978	0.766	0.664	0.168	0.188	0.376	0.145	0.482	1.539
	Min	4.1	4.3	3.61	4.5	1	0.1	1.6	0	0	0	0.003	13.988
	Max	51.55	5.19	4.89	7.3	2.7	0.402	2.73	2.12	1.999	0.99	0.086	20.534
NN	Mean	20.904	4.841	4.260	5.985	1.827	0.249	2.167	0.864	0.918	0.776	0.041	16.748
	SD	15.132	0.175	0.357	0.795	0.498	0.092	0.301	0.613	0.589	0.215	0.028	1.313
	Min	5.77	4.3	2.8	4.8	0.7	0	2.5	0.2	0.4	0.4	0.001	11.353
DII	Max	55	5.5	5.001	13.2	2.401	9.4	4.9	1.8	1.9	1.2	0.13	25.615
DJI	Mean	20.722	4.901	3.858	7.807	1.556	3.574	3.620	0.720	0.944	0.779	0.060	17.557
	SD	12.918	0.373	0.647	2.432	0.463	2.679	0.726	0.319	0.338	0.229	0.039	3.225
	Min	21.07	4	3.7	4.9	0.071	0	1.056	0.1	0.5	0.33	0.001	9.069
ND	Max	109.7	4.99	8.6	27.4	4.9	11.7	9.9	3.8	1.602	4.2	3.088	25.2
INK	Mean	74.483	4.456	6.059	12.109	2.223	4.257	5.100	2.102	0.935	2.303	0.153	16.593
	SD	18.437	0.283	1.287	5.430	1.256	3.373	2.297	0.974	0.301	1.120	0.362	3.517
	Min	83.6	4.07	2.5	0	0.4	29.3	5.13	0.32	1.011	2.401	0.001	0.063
RC	Max	151.8	5.73	6.97	8.2	1.9	67.32	11.76	3.99	3.593	11.7	0.017	71.227
ĸc	Mean	107.99	5.022	4.608	3.464	1.216	50.792	8.343	2.185	2.328	7.252	0.009	33.223
	SD	22.289	0.349	1.109	2.610	0.456	11.239	1.954	1.070	0.597	2.653	0.005	25.117
	Min	41.7	4.3	4.44	0	0	0	3.11	0.506	0.12	0.4	0.001	4.115
70	Max	159.4	5	16.908	48.73	14.57	78.1	12.55	33.61	4.9	19.3	0.48	103.70
20	Mean	83.2	4.623	10.066	10.331	5.362	18.139	8.254	11.013	3.045	9.515	0.048	48.564
	SD	30.215	0.197	3.643	13.752	4.324	21.621	2.770	11.008	1.207	5.534	0.109	25.261
	Min	174.96	4.17	2.21	0	0.053	56.92	9.2	0.1	2.3	5.5	0.002	0.420
ZE	Max	236.3	5.84	15.3	27.3	4.55	135.2	18.7	1.29	5.81	39.9	0.103	129.91
22	Mean	201.30	4.846	7.413	9.069	2.028	96.021	13.943	0.605	4.085	14.565	0.012	62.489
	SD	18.107	0.349	3.274	9.336	1.236	25.202	2.846	0.344	0.942	8.826	0.022	43.873
	Min	57.2	4.8	1.9	0	0.6	122.54	12	0.4	5.2	18.1	0.001	118.63
ZN	Max	289.33	5.5	4.93	20.48	8.8	175.6	21.48	9.81	17.07	35.06	0.052	182.68
ZN	Mean	205.28	5.191	3.603	4.432	3.576	148.131	17.105	4.946	9.112	26.271	0.014	143.12
	SD	82.370	0.211	0.775	6.963	2.804	16.114	2.416	2.824	2.923	4.756	0.015	13.119
	Min	298	4.5	1.68	0	0	169.98	6.8	0.2	10.4	38.1	0.001	154.89
AN	Max	469	5.9	7.32	2.2	18.4	240.5	22.41	4.3	15.4	51.3	0.008	214.32
	Mean	385.89	5.161	4.220	0.893	4.981	205.446	15.646	2.252	12.88	44.071	0.004	187.67
	SD	52.575	0.256	1.604	0.618	5.970	18.453	4.002	1.240	1.587	3.975	0.002	12.603

 Table 4. Statistics of groundwater physicochemical parameters for each site.

DOI: 10.4236/ica.2024.154010

Al<sup>3+</sup>. The Ca<sup>2+</sup> value indicates variations from 0 to 51.3 mg/L. The primary factors influencing the presence of Ca<sup>2+</sup> in groundwater are the calcium-rich silicate weathering and the process of ion exchange [71]. The Na<sup>+</sup> concentration fluctuates between 1.056 mg/L and 22.4 mg/L, while the Mg<sup>2+</sup> fluctuates between 0 and 17.07 mg/L. The recorded K<sup>+</sup> varies between 0 and 33.61 mg/L. High levels of K<sup>+</sup> observed in a few samples are due to alteration products such as kaolin in clay formations [72], of the aquifer. For the Al<sup>3+</sup> content, the measured concentration ranges between 0.001 mg/L and 4.209 mg/L and the highest Al<sup>3+</sup> values potentially come from the geological formation of the CT aquifer [73].

## 4.2. Water Quality Index

Applying the AHP methodology, the structure was organized into three hierarchies given major ions percentage in samples (**Figure 2**): (1) physical parameters such as pH, EC, and TDS, (2) chemical parameters like  $Cl^-$ ,  $SO_4^{2-}$ ,  $HCO_3^-$ ,  $Na^+$ ,  $K^+$ ,  $Mg^{2+}$ , and  $Ca^{2+}$ , and (3) health-threatening chemical parameters such as  $NO_3^-$  and  $Al^{3+}$ . Scores of 5 and 9 were respectively given to the first and third structures. The physical variables are undoubtedly essential; however, the health-threatening parameters are paramount in groundwater. Additionally, structure (2) was divided into two subgroups, (a) and (b), with respective rates of 6 and 4 based on anions and cations orders in the samples. The consistency ratio calculation gives CR = 0.024 < 0.1 (**Table 5**), validating the generated pair-wise matrix. This result of CR enables the application of normalized weights as relative weights for WQI determination.



Figure 2. Hierarchical structure of the selected parameters.

CR	0.024												
CI	0.035												
б = (A2/Wn)	0.872	0.872	0.872	1.047	1.570	0.698	1.047	1.047	0.698	1.047	1.047	1.570	
Consistency Check (A2 = A1*Wn)	0.061	0.061	0.061	0.088	0.199	0.039	0.088	0.088	0.039	0.088	0.088	0.199	$\lambda_{ m max}=12.385$
Normalized Weights (Wn)	0.070	0.070	0.070	0.085	0.127	0.056	0.085	0.085	0.056	0.085	0.085	0.127	$\sum\nolimits_{i=1}^{n} W_{n} = 1.000$
Geo. Mean (A1)*	0.872	0.872	0.872	1.047	1.570	0.698	1.047	1.047	0.698	1.047	1.047	1.570	
EC TDS CI- NO $_3^\circ$ SO $_4^{2-}$ HCO $_3^\circ$ Na+ K+ Mg^{2+} Ca^{2+} Al^3+	1.000  1.000  0.833  0.556  1.250  0.833  0.833  1.250  0.833  0.833  0.556	1.000  1.000  0.833  0.556  1.250  0.833  0.833  1.250  0.833  0.833  0.556	1.000  1.000  0.833  0.556  1.250  0.833  0.833  1.250  0.833  0.833  0.556	$1.200 \ 1.200 \ 1.000 \ 0.667 \ 1.500 \ 1.000 \ 1.000 \ 1.500 \ 1.000 \ 1.000 \ 0.667$	1.800  1.800  1.500  1.000  2.250  1.500  1.500  2.250  1.500  1.500  1.000	0.800  0.800  0.667  0.444  1.000  0.667  0.667  1.000  0.667  0.667  0.444	1.200  1.200  1.000  0.667  1.500  1.000  1.000  1.500  1.000  1.000  0.667	$1.200 \ 1.200 \ 1.000 \ 0.667 \ 1.500 \ 1.000 \ 1.000 \ 1.500 \ 1.000 \ 1.000 \ 0.667$	0.800  0.800  0.667  0.444  1.000  0.667  0.667  1.000  0.667  0.667  0.444	$1.200 \ 1.200 \ 1.000 \ 0.667 \ 1.500 \ 1.000 \ 1.000 \ 1.500 \ 1.000 \ 1.000 \ 0.667$	$1.200 \ 1.200 \ 1.000 \ 0.667 \ 1.500 \ 1.000 \ 1.000 \ 1.500 \ 1.000 \ 0.667$	1.800  1.800  1.500  1.000  2.250  1.500  1.500  2.250  1.500  1.500  1.000	
Hq	1.000	1.000	1.000	1.200	1.800	0.800	1.200	1.200	0.800	1.200	1.200	1.800	
Assigned weights	5	5	ß	9	6	4	9	9	4	9	9	6	
WHO limits	8.5	2500	1500	250	50	250	500	200	12	50	200	0.2	
Parameters	Hq	EC	TDS	CI-	$NO_{3}^{-}$	$\mathrm{SO}_4^{2-}$	$HCO_{3}^{-}$	$\mathrm{Na}^{+}$	$\mathbf{K}^+$	${\rm Mg}^{2+}$	$Ca^{2+}$	$\mathrm{Al}^{3+}$	

 Table 5. Physico-chemical parameters pair-wise comparison matrix.

The obtained index values fluctuate on average between 83.40 and 90.79, with a standard deviation between 0.867 and 14.639 (Table 6). The classification of these indices based on the categories for human consumption indicates good groundwater quality for most samples, corresponding to 98.98%. Moreover, 0.68% of the samples are under the excellent water quality category (<50), while 0.34% of the samples are under the poor water quality category (100 - 200.1) (Ta**ble** 7). Considering that  $Al^{3+}$  and  $K^+$  exceed the permissible thresholds for 4.76% and 3.4% of the samples, respectively. The WQI results and the CT groundwater quality are deemed protected with a minor degree of threat or impairment. The resource conditions represent areas with suitable groundwater for various uses. The Abidjan quaternary groundwater quality assessment presented similar results supporting our study's outcome [74]. The GWQI also shows that the resource is generally suitable for drinking purposes. In dry season, 70.49% of the samples fall into the category of excellent water quality, 16.39% fall into good water quality class. 55.73% of the samples are classified as excellent in rainy seasons, while 31.13% are good. Since the Quaternary layer is located above the CT layer, the modified classification of some samples can be attributed to the effects of precipitation on the chemical process of water, as observed in CT groundwater.

Table 6. Statistics of the computed WQI for each site.

	AK	NN	DJI	NR	RC	ZO	ZE	ZN	AN
min	14.430	87.248	83.404	3.781	88.359	58.864	79.705	76.128	82.218
max	172.587	93.536	93.856	103.707	93.553	91.957	92.007	88.668	85.905
mean	89.123	90.649	89.123	83.606	90.788	84.991	87.935	84.062	83.821
SD	14.639	1.852	2.621	13.341	1.049	7.754	2.832	2.481	0.867

Table 7. CT groundwater quality evaluation summary.

Model	Total monitoring sites		Wate	er quality s	tatus	
AHP	9	Excellent	Good	Poor	Very poor	Unfit
weight-based		0.68% of	98.98% of	0.34% of	0% of	0% of
WQI		samples	samples	samples	sample	sample

## 4.3. Prediction Results and Performances

This section provides the outcomes of real-time prediction of CT groundwater quality using Ridge, Lasso, and GB regression models. The twelve physicochemical parameters (pH, EC, TDS, Cl<sup>-</sup>,  $SO_4^{2-}$ ,  $HCO_3^-$ ,  $Na^+$ ,  $K^+$ ,  $Mg^{2+}$ ,  $Ca^{2+}$ ,  $NO_3^-$  and  $Al^{3+}$ ) represent the features, and the AHP-based WQI are the label. The limited data enables a prediction of resource quality seven days ahead. However, all the models were successfully executed. The algorithm accuracy indicators present performances varying from one model to another. For the interpretation, we only considered the testing set.

#### 4.3.1. Ridge Regression

The real and predicted WQI plots (**Figure 3**) indicate that the predicted values oscillate within a similar range to the measured values for locations AK, RC, ZO, ZE, and AN. Conversely, the differences between the actual and predicted values are apparent in the remaining sites. In contrast to previous observations, the performance metrics delineated in **Table 8** indicate that at these sites (AK, RC, ZO, ZE, and AN), the model has a suboptimal performance (MSE and RMSE close to 1), while the other sites yield good performance (MSE and RMSE close to 0). Additionally, the correlation metrics imply a lack of consistency in the model's execution across most monitored sites with *r*, *r*<sub>s</sub>, and  $\tau < 0.5$ .





Intelligent Control and Automation



**Figure 3.** Actual vs predicted WQI plots for Ridge model at each site (a) Site 1 AK, (b) Site 2 NN, and (c) Site 3 DJI, (d) Site 4 NR, (e) Site 5 RC, (f) Site 6 ZO, (g) Site 7 ZE, (h) Site 8 ZN, (i) Site 9 AN.

Sites	MSE	RMSE	Pearson r	Kendall <i>r</i> s	Spearman $ au$	
AK	0.704	0.839	0.538	0.314	0.406	
NN	0.192	0.438	0.703	0.193	0.298	
DJI	0.255	0.505	0.546	0.116	0.093	
NR	0.303	0.551	0.420	0.010	0.012	
RC	1.792	1.339	0.426	0.363	0.332	
ZO	1.001	1.000	0.457	0.159	0.234	
ZE	1.318	1.148	0.456	0.152	0.093	
ZN	0.160	0.401	0.264	0.166	0.298	
AN	2.659	1.631	0.340	0.606	0.783	
mean	0.924	0.834	0.427	0.204	0.252	
std	0.774	0.418	0.136	0.167	0.216	
						1

 Table 8. Groundwater quality prediction performance indicators for Ridge regression for each site.

#### 4.3.2. Lasso Regression

The obtained plots (**Figure 4**) suggested that the predicted WQI values vary within the same range as the measured values for locations AK, NN, NR, RC, ZO,



Intelligent Control and Automation



DOI: 10.4236/ica.2024.154010



**Figure 4.** Actual vs predicted WQI plots for Lasso model at each site (a) Site 1 AK, (b) Site 2 NN, and (c) Site 3 DJI, (d) Site 4 NR, (e) Site 5 RC, (f) Site 6 ZO, (g) Site 7 ZE, (h) Site 8 ZN, (i) Site 9 AN.

and ZN. Meanwhile other sites, DJI, ZE, and ZE, discrepancies are identified between the actual and predicted values at the same time. However, the performance metrics outlined in **Table 9** indicate that the model has low performance (MSE and RMSE close to 1) at these sites, while the remaining sites showcase good performance (MSE and RMSE close to 0). Moreover, the indicators exhibit inconsistent model execution for most monitored sites like Ridge regression, with r, r<sub>s</sub>, and  $\tau$  close to 0.

 Table 9. Groundwater quality prediction performance indicators for Lasso regression for each site.

Sites	MSE	RMSE	Pearson r	Kendall <i>r</i> <sub>s</sub>	Spearman $\tau$	
AK	7.954	2.820	0.357	0.190	0.161	
NN	1.032	1.016	0.629	0.109	0.107	
DJI	0.217	0.466	0.004	0.001	0.002	
NR	141.417	11.892	0.630	0.002	0.002	
RC	763.142	27.625	0.419	0.030	0.058	
ZO	1.377	1.173	0.513	0.094	0.093	
ZE	0.245	0.495	0.359	0.022	0.021	
ZN	16.626	4.077	0.641	0.042	0.059	
AN	0.235	0.485	0.078	0.099	0.092	
mean	148.921	6.828	0.397	0.079	0.082	
std	246.561	8.653	0.213	0.063	0.059	

#### 4.3.3. Gradient Boosting Regression

 Table 10 presents the results of the gradient-boosting regression performance

 evaluation for the CT groundwater quality prediction. The computed metrics at

each site display satisfactory performance and consistency in the prediction. This is translated by the low error levels (close to 0) and high correlations (close to 1), with values ranging from 0.097 to 0.112 and 0.311 to 0.334 for MSE and RMSE, respectively. Also, r ranges from 0.439 to 0.966,  $r_s$  from 0.632 to 0.904, and  $\tau$  from 0.836 to 0.945. The real and predicted WQI plots (**Figure 5**) variations corroborate

 Table 10. Groundwater quality prediction performance indicators for GB regression for each site.

Sites	MSE	RMSE	Pearson r	Kendall <i>rs</i>	Spearman $ au$
AK	0.097	0.311	0.880	0.904	0.945
NN	0.106	0.325	0.966	0.853	0.836
DJI	0.097	0.311	0.880	0.904	0.945
NR	0.104	0.323	0.879	0.858	0.890
RC	0.112	0.334	0.439	0.632	0.836
ZO	0.106	0.325	0.966	0.853	0.836
ZE	0.106	0.325	0.966	0.853	0.836
ZN	0.112	0.334	0.439	0.632	0.836
AN	0.106	0.325	0.966	0.853	0.836
mean	0.097	0.300	0.766	0.757	0.804
std	0.023	0.069	0.257	0.198	0.189

![](_page_21_Figure_4.jpeg)

![](_page_22_Figure_1.jpeg)

Intelligent Control and Automation

![](_page_23_Figure_1.jpeg)

**Figure 5.** Actual vs predicted WQI plots for GB model at each site (a) Site 1 AK, (b) Site 2 NN, and (c) Site 3 DJI, (d) Site 4 NR, (e) Site 5 RC, (f) Site 6 ZO, (g) Site 7 ZE, (h) Site 8 ZN, (i) Site 9 AN.

the performance metrics results, showing fluctuation of predicted values in line with the trend of actual values in all sites. This confirms the accuracy of the Gradient-boosting model.

## 4.4. Discussion

In the last decade, WQI models have been widely used to assess water quality for various contexts. Nonetheless, a major challenge faced while using WQIs is the

uncertainties from incorrect weight assignment in the methodology [20]. Our study agrees that the AHP weight-based WQI model represents the ideal approach, highlighting its advantages over alternative models like weighted arithmetic WQI or entropy weight-based WQI. The AHP model allows for a systematic and structured approach to decision-making, ensuring that the most critical parameters are emphasized in the evaluation process [75]. This model focuses on the uncertainty in the data as well as considers the contextual significance of certain parameters that are better evaluated through expert input [20] [76]. Additionally, the AHP model's flexibility allows for incorporating local knowledge and specific health impacts associated with groundwater quality parameters, making it more suitable for targeted assessments in diverse environments [77].

Regarding the prediction phase, the three regression models (Ridge, Lasso, and Gradient Boosting) were executed. Choosing these models over others like random forest, artificial neural network, or deep learning can be attributed to different factors. First, the interpretability of ridge and lasso regression is significantly higher. For instance, Deep learning might give accurate predictions, however, hardly show how predictions are made, leading to its identification as "black boxes" [78], posing limitations as researchers and stakeholders need to trust or understand the results. The ability to interpret prediction outcomes is essential for effective management and policymaking [29]. Next, these complex models typically require large datasets to train accurately and avoid overfitting. In the current framework, the available datasets do not effectively support such models. In contrast, traditional regression models can perform well even with smaller datasets, making them more practical for our scenario. Then, simpler models, such as ridge, lasso, and gradient boosting, not only require less computational power, but, they also allow for quicker deployment in real-time monitoring systems [38].

Models	MSE	RMSE	Pearson r	Kendall <i>r</i> s	Spearman $\tau$
Ridge	0.924	0.834	0.427	0.204	0.252
Lasso	148.921	6.828	0.397	0.079	0.082
GB	0.097	0.300	0.766	0.757	0.804

Table 11. Average values of the model performance metrics.

To interpret the forecast results, the mean values of the performance metrics for each model were calculated (Table 11). This enabled their comparison and the identification of the best model. The performance metrics lie in their ability to quantitatively assess prediction model accuracy and inform stakeholders about prediction reliability. In groundwater quality scenarios, a model with low error rates (MSE) and high correlations is reliable, providing a clear view of resource quality variation over time [79]. This is essential for understanding the strengths and limitations of different models, allowing their recalibration, the integration of multiple approaches, or additional monitoring to improve overall prediction robustness. These error metrics extend beyond academic interest as RMSE values can be used to establish confidence intervals around predictions [80], while correlation coefficients serve as the measure of the strength and direction of the model. The correlations also capture real-world relationships in the groundwater quality parameters [79]. In practical decision-making, performance metrics are crucial indicators for resource allocation, pollution prevention, and the establishment of informed remediation strategies to inform plans for future water demands [81]. Ultimately, they develop evidence-based policies to justify investments and create sustainable decision-support systems.

Our findings indicate that the GB regression has the lowest values for MSE and RMSE (MSE = 0.097; RMSE = 0.300). Conversely, both Lasso and Ridge regression display significant errors, with Lasso regression recording the highest MSE and RMSE figures (148.921; 6.828). These metric outcomes suggest that GB potentially outperforms Lasso and Ridge regarding predictive accuracy. Nonetheless, obtaining smaller RMSE and MSE does not necessarily indicate excellent models due to their exclusive representation of the mean of squared variances between the forecast and observation failing to reveal the inherent logical bias [28]. The computation of Pearson, Spearman, and Kendall correlation coefficients was performed to identify the variance in model performance and ascertain their consistency [64]. In agreement with the performance metrics, the correlation coefficients reveal that GB demonstrates the most significant level of uniformity while Lasso indicates the lowest, highlighting the differences in predicted and actual WQI values between Lasso and Ridge compared to GB. This outcome was supported by Chen and Guestrin [82], who stated that GB encapsulates complex patterns in a way that other models are unable to. Discrepancies between the performance rates and the calculated correlations were also observed. For instance, in Lasso regression prediction, the metrics suggested an excellent performance in the ZE site (MSE = 0.245 and RMSE = 0.485), while the consistency evaluation presented weak consistency for the same site (r = 0.359;  $r_s = 0.022$ ;  $\tau = 0.021$ ). Thus, the Lasso and Ridge prediction potentially involves systematic biases.

Other researchers have argued that components such as model fitting can lead to systematic instability and bias [28]. An effective data-splitting approach should support the models in grasping the statistical patterns of the input and prevent the introduction of excessive bias. Our study divided sample data into training and testing subsets rather than training, testing, and validation subsets due to the limited sample size (4 months) and the significance of avoiding underestimation of the observed groundwater quality. As a result, the conventional training and testing window selection technique was employed for data splitting [83]. Therefore, three-quarters of the sample size was selected as the training set (data before August 1, 2022) and the rest as the testing set (data after August 1, 2022). The test set included daily sampling data for over one month, representing a short period that can hinder the ability of the model to capture the overall dynamic trend. However, identifying changes in groundwater quality over four months of observation under the same seasonal conditions was a complex challenge. Hence, both Lasso and

Ridge regression experienced limitations due to the small sample size and the absence of diversity in the data categories, which used solely physicochemical parameters. Furthermore, Lasso and Ridge models exhibit limited capabilities in forecasting nonlinear occurrences compared to GB regression [28]. This outcome implores the hydrogeochemical process, which embodies nonlinear relationships and a diverse array of interactions within the subsurface environment [84] [85]. By iteratively fitting new models to the errors of the previous models, GB can learn complex patterns and interactions between variables [58] [59], implying the model's suitability for our study.

## **5.** Conclusions

This research was designed to evaluate the quality of the CT groundwater in Abidjan using the water quality index (WQI) based on the analytical hierarchy process (AHP) and multi-criteria decision analysis methodology (MCDM). Additionally, this study aimed to improve its proposed frameworks for continuous resource quality monitoring by applying machine learning predictions. The efficacy of the models was assessed through a comprehensive analysis using a range of statistical approaches comprising of prediction performance metrics (RMSE, MSE) and model consistency (Pearson, Spearman, and Kendall correlation). First, the results of the research reveal that most of the groundwater quality parameters across the sampling sites are within the World Health Organization (WHO) guideline values except for the aluminum (Al<sup>3+</sup>) and potassium (K<sup>+</sup>) that exceeded the permissible thresholds of 4.76% and 3.4% of samples, respectively. Next, the AHP-based WQI model classified the CT groundwater quality averagely into "good" category, with 0.68% in "excellent" and 0.34% in "poor" categories. The groundwater quality is deemed protected with a minor degree of threat or impairment with conditions comparable to areas with suitable groundwater for various uses. Then, the performance evaluation indicated that the gradient boosting (GB) regression model optimally predicted the CT groundwater quality. The GB outperformed the other models by providing the highest accuracy with MSE = 0.097 and RMSE = 0.300, in addition to a remarkable consistency with Pearson, Spearman, and Kendall correlations respectively stated as, r = 0.766,  $r_s = 0.757$ , and  $\tau = 0.804$ . Conversely, the lowest prediction accuracy was recorded for the Lasso model with an MSE and RMSE of 148.921 and 6.828 respectively, and consistency parameters of r = 0.397,  $r_s = 0.079$ , and  $\tau = 0.082$  respectively. Both Lasso and Ridge regression faced limitations due to the small sample size and the lack of data diversity. Our study considered a single rainy season dataset and major ions, however, multiple-season data and parameters may have provided greater insight into the groundwater quality of this study. Also, these models struggle with forecasting nonlinear events compared to GB regression, considering that the hydrogeochemical process embodies nonlinear relationships and diverse interactions within the subsurface environment. Through weak learners, GB can learn intricate patterns and interactions between variables by iteratively fitting new models to errors of previous models. With continuous data updates and the application of ensemble models like GB, a groundwater quality prediction monitoring system may be established for efficient water resource management in Abidjan.

In summary, this exploratory study not only provided the assessment and prediction of groundwater quality but also provided novel actionable insights for the effective management of water resources. The implications for groundwater management policy in Abidjan and the whole country are as follows; the integration of real-time monitoring systems, enhanced pollution risk assessment frameworks, regulatory frameworks for groundwater protection, data-driven decision-making, investment in infrastructure, long-term planning and sustainability, collaboration across sectors, and continuous improvement.

## **Conflicts of Interest**

The author declares no conflicts of interest regarding the publication of this paper.

## References

- Kazakis, N., Mattas, C., Pavlou, A., Patrikaki, O. and Voudouris, K. (2017) Multivariate Statistical Analysis for the Assessment of Groundwater Quality under Different Hydrogeological Regimes. *Environmental Earth Sciences*, **76**, Article No. 349. <u>https://doi.org/10.1007/s12665-017-6665-y</u>
- [2] Direction de l'hydraulique humaine (DHH) (2001) Hydraulique Humaine en Côte d'Ivoire. Ministère des Infrastructures Economiques, Direction de l'Hydraulique Humaine, 66.
- [3] SODECI (2005) Campagne de relevé piézométrique réalisée au niveau du District d'Abidjan. Rapport d'activité, 13 p.
- [4] Soro, N., Ouattara, L., Dongo, K., Kouadio, E., Ahoussi, E., Soro, G., et al. (2011) Déchets municipaux dans le District d'Abidjan en Côte d'Ivoire: Sources potentielles de pollution des eaux souterraines. *International Journal of Biological and Chemical Sciences*, 4, 2203-2219. <u>https://doi.org/10.4314/ijbcs.v4i6.64952</u>
- [5] Deh, S., Kouame, K., Saley, M., Tanoh, K., Anani, E., Signo, K., et al. (2012) Evaluation de la vulnérabilité spécifique aux nitrates (NO<sub>3</sub>) des eaux souterraines du District d'Abidjan (Sud de la Côte d'Ivoire). International Journal of Biological and Chemical Sciences, 6, 1390-1408. <u>https://doi.org/10.4314/ijbcs.v6i3.40</u>
- [6] Jourda, J.P., Kouamé, K.J., Saley, M.B., Kouadio, B.H., Oga, Y.S. and Deh, S. (2006) Contamination of the Abidjan Aquifer by Sewage: An Assessment of Extent and Strategies for Protection. In: Xu, Y.X. and Usher, B., Eds., *Groundwater Pollution in Africa*, Taylor & Francis/Balkema, 291-300.
- [7] Qin, Q.S., Huang, G.H. and Chakma, A. (2008) Modeling Groundwater Contamination under Uncertainty: A Factorial-Design-Based Stochastic Approach. *Journal of Environmental Informatics*, **11**, 11-20. <u>https://doi.org/10.3808/jei.200800106</u>
- [8] Ahn, H. and Chon, H. (1999) Assessment of Groundwater Contamination Using Geographic Information Systems. *Environmental Geochemistry and Health*, 21, 273-289. <u>https://doi.org/10.1023/a:1006697512090</u>
- [9] Massmann, J., Freeze, R.A., Smith, L., Sperling, T. and James, B. (1991) Hydrogeological Decision Analysis: 2. Applications to Ground-Water Contamination. *Ground-water*, 29, 536-548. <u>https://doi.org/10.1111/j.1745-6584.1991.tb00545.x</u>
- [10] Khan, H., Khan, A. A. and Hall, S. (2005) The Canadian Water Quality Index: A Tool

for Water Resources Management. MTERM International Conference, Vol. 8.

- [11] Ma, Z., Li, H., Ye, Z., Wen, J., Hu, Y. and Liu, Y. (2020) Application of Modified Water Quality Index (WQI) in the Assessment of Coastal Water Quality in Main Aquaculture Areas of Dalian, China. *Marine Pollution Bulletin*, **157**, Article ID: 111285. <u>https://doi.org/10.1016/j.marpolbul.2020.111285</u>
- [12] Uddin, M.G., Rahman, A., Nash, S., Diganta, M.T.M., Sajib, A.M., Moniruzzaman, M., et al. (2023) Marine Waters Assessment Using Improved Water Quality Model Incorporating Machine Learning Approaches. Journal of Environmental Management, 344, Article ID: 118368. https://doi.org/10.1016/j.jenvman.2023.118368
- [13] Ramesh, S., Sukumaran, N., Murugesan, A.G. and Rajan, M.P. (2010) An Innovative Approach of Drinking Water Quality Index—A Case Study from Southern Tamil Nadu, India. *Ecological Indicators*, **10**, 857-868. <u>https://doi.org/10.1016/j.ecolind.2010.01.007</u>
- [14] Poonam, T., Tanushree, B. and Sukalyan, C. (2013) Water Quality Indices-Important Tools for Water Quality Assessment: A Review. *International Journal of Advances in chemistry*, 1, 15-28.
- [15] Rupal, M., Tanushree, B. and Sukalyan, C. (2012) Quality Characterization of Groundwater Using Water Quality Index in Surat City, Gujarat, India. *International Research Journal of Environment Sciences*, 1, 14-23.
- [16] Hui, T., Du, J., Sun, Q., Liu, Q., Kang, Z. and Jin, H. (2020) Using the Water Quality Index (WQI), and the Synthetic Pollution Index (SPI) to Evaluate the Groundwater Quality for Drinking Purpose in Hailun, China. *Sains Malaysiana*, **49**, 2383-2401. https://doi.org/10.17576/jsm-2020-4910-05
- [17] Mamun, M. and An, K. (2021) Application of Multivariate Statistical Techniques and Water Quality Index for the Assessment of Water Quality and Apportionment of Pollution Sources in the Yeongsan River, South Korea. *International Journal of Environmental Research and Public Health*, **18**, Article No. 8268. <u>https://doi.org/10.3390/ijerph18168268</u>
- [18] Chidiac, S., El Najjar, P., Ouaini, N., El Rayess, Y. and El Azzi, D. (2023) A Comprehensive Review of Water Quality Indices (WQIS): History, Models, Attempts and Perspectives. *Reviews in Environmental Science and Bio/Technology*, 22, 349-395. https://doi.org/10.1007/s11157-023-09650-7
- [19] Uddin, M.G., Nash, S. and Olbert, A.I. (2021) A Review of Water Quality Index Models and Their Use for Assessing Surface Water Quality. *Ecological Indicators*, **122**, Article ID: 107218. <u>https://doi.org/10.1016/j.ecolind.2020.107218</u>
- [20] Rajkumar, H., Naik, P.K. and Rishi, M.S. (2022) A Comprehensive Water Quality Index Based on Analytical Hierarchy Process. *Ecological Indicators*, 145, Article ID: 109582. <u>https://doi.org/10.1016/j.ecolind.2022.109582</u>
- [21] Amiri, V., Rezaei, M. and Sohrabi, N. (2014) Groundwater Quality Assessment Using Entropy Weighted Water Quality Index (EWQI) in Lenjanat, Iran. *Environmental Earth Sciences*, 72, 3479-3490. <u>https://doi.org/10.1007/s12665-014-3255-0</u>
- [22] Gorgij, A.D., Kisi, O., Moghaddam, A.A. and Taghipour, A. (2017) Groundwater Quality Ranking for Drinking Purposes, Using the Entropy Method and the Spatial Autocorrelation Index. *Environmental Earth Sciences*, 76, Article No. 269. <u>https://doi.org/10.1007/s12665-017-6589-6</u>
- [23] Islam Khan, M.S., Islam, N., Uddin, J., Islam, S. and Nasir, M.K. (2022) Water Quality Prediction and Classification Based on Principal Component Regression and Gradient Boosting Classifier Approach. *Journal of King Saud University—Computer and*

Information Sciences, 34, 4773-4781. https://doi.org/10.1016/j.jksuci.2021.06.003

- [24] Singha, S., Pasupuleti, S., Singha, S.S., Singh, R. and Kumar, S. (2021) Prediction of Groundwater Quality Using Efficient Machine Learning Technique. *Chemosphere*, 276, Article ID: 130265. <u>https://doi.org/10.1016/j.chemosphere.2021.130265</u>
- [25] Singha, S., Pasupuleti, S., Durbha, K.S., Singha, S.S., Singh, R. and Venkatesh, A.S. (2019) An Analytical Hierarchy Process-Based Geospatial Modeling for Delineation of Potential Anthropogenic Contamination Zones of Groundwater from Arang Block of Raipur District, Chhattisgarh, Central India. *Environmental Earth Sciences*, 78, Article No. 694. <u>https://doi.org/10.1007/s12665-019-8724-z</u>
- [26] Elbeltagi, A., Pande, C.B., Kouadri, S. and Islam, A.R.M.T. (2021) Applications of Various Data-Driven Models for the Prediction of Groundwater Quality Index in the Akot Basin, Maharashtra, India. *Environmental Science and Pollution Research*, 29, 17591-17605. <u>https://doi.org/10.1007/s11356-021-17064-7</u>
- [27] Haggerty, R., Sun, J., Yu, H. and Li, Y. (2023) Application of Machine Learning in Groundwater Quality Modeling—A Comprehensive Review. *Water Research*, 233, Article ID: 119745. <u>https://doi.org/10.1016/j.watres.2023.119745</u>
- [28] Xiao, S. (2024) Machine Learning Methods for Providing Groundwater Management Insights in Semi-Arid Alluvial Aquifers. Ph.D. Thesis, UNSW.
- [29] Mohd Zebaral Hoque, J., Ab. Aziz, N.A., Alelyani, S., Mohana, M. and Hosain, M. (2022) Improving Water Quality Index Prediction Using Regression Learning Models. *International Journal of Environmental Research and Public Health*, **19**, Article No. 13702. <u>https://doi.org/10.3390/ijerph192013702</u>
- [30] Granata, F., Papirio, S., Esposito, G., Gargano, R. and De Marinis, G. (2017) Machine Learning Algorithms for the Forecasting of Wastewater Quality Indicators. *Water*, 9, Article No. 105. <u>https://doi.org/10.3390/w9020105</u>
- [31] Azma, A., Tavakol Sadrabadi, M., Liu, Y., Azma, M., Zhang, D., Cao, Z., *et al.* (2022) Boosting Ensembles for Estimation of Discharge Coefficient and through Flow Discharge in Broad-Crested Gabion Weirs. *Applied Water Science*, **13**, Article No. 45. <u>https://doi.org/10.1007/s13201-022-01841-x</u>
- [32] Abdi, J. and Mazloom, G. (2022) Machine Learning Approaches for Predicting Arsenic Adsorption from Water Using Porous Metal-Organic Frameworks. *Scientific Reports*, **12**, Article No. 16458. <u>https://doi.org/10.1038/s41598-022-20762-y</u>
- [33] Chen, B., Chen, Z., Song, C. and Song, Y. (2024) Integrated Forecasting Method of Medium- and Long-Term Runoff by Ridge Regression Based on Optimal Sub-Model Selection. *Water Supply*, 24, 799-811. <u>https://doi.org/10.2166/ws.2024.033</u>
- [34] Falah, F., Rahmati, O., Rostami, M., Ahmadisharaf, E., Daliakopoulos, I.N. and Pourghasemi, H.R. (2019) Artificial Neural Networks for Flood Susceptibility Mapping in Data-Scarce Urban Areas. In: Pourghasemi, H.R. and Gokceoglu, C., Eds., *Spatial Modeling in GIS and R for Earth and Environmental Sciences*, Elsevier, 323-336. <u>https://doi.org/10.1016/b978-0-12-815226-3.00014-4</u>
- [35] Cao, H., Xie, X., Wang, Y. and Liu, H. (2022) Predicting Geogenic Groundwater Fluoride Contamination throughout China. *Journal of Environmental Sciences*, 115, 140-148. <u>https://doi.org/10.1016/j.jes.2021.07.005</u>
- [36] Foddis, M.L., Montisci, A., Uras, G., Matzeu, A., Seddaiu, G. and Carletti, A. (2012) Prediction of Nitrate Concentration in Groundwater Using an Artificial Neural Network (ANN) Approach.
- [37] Abba, S.I., Hadi, S.J., Sammen, S.S., Salih, S.Q., Abdulkadir, R.A., Pham, Q.B., et al. (2020) Evolutionary Computational Intelligence Algorithm Coupled with Self-Tuning Predictive Model for Water Quality Index Determination. *Journal of Hydrology*,

587, Article ID: 124974. https://doi.org/10.1016/j.jhydrol.2020.124974

- [38] Ahmed, U., Mumtaz, R., Anwar, H., Shah, A.A., Irfan, R. and García-Nieto, J. (2019) Efficient Water Quality Prediction Using Supervised Machine Learning. *Water*, 11, Article No. 2210. <u>https://doi.org/10.3390/w11112210</u>
- [39] Yilma, M., Kiflie, Z., Windsperger, A. and Gessese, N. (2018) Application of Artificial Neural Network in Water Quality Index Prediction: A Case Study in Little Akaki River, Addis Ababa, Ethiopia. *Modeling Earth Systems and Environment*, 4, 175-187. https://doi.org/10.1007/s40808-018-0437-x
- [40] Gupta, R., Singh, A.N. and Singhal, A. (2019) Application of ANN for Water Quality Index. *International Journal of Machine Learning and Computing*, 9, 688-693. <u>https://doi.org/10.18178/ijmlc.2019.9.5.859</u>
- [41] Leong, W.C., Bahadori, A., Zhang, J. and Ahmad, Z. (2019) Prediction of Water Quality Index (WQI) Using Support Vector Machine (SVM) and Least Square-Support Vector Machine (LS-SVM). *International Journal of River Basin Management*, 19, 149-156. <u>https://doi.org/10.1080/15715124.2019.1628030</u>
- [42] Jourda, J.P. (1987) Contribution à l'étude géologique et hydrogéologique de la région du Grand Abidjan (Côte d'Ivoire). Doctoral Dissertation, Université Scientifique et Médicale de Grenoble.
- [43] Adiaffi, B. (2008) Apport de la Géochimie Isotopique, de l'Hydrochimie et de la Télédétection a la Connaissance des Aquifères de la Zone de Contact "Socle-Bassin Sédimentaire" du Sud-Est de la Côte d'Ivoire. Doctoral Dissertation, Université Paris Sud-Paris XI.
- [44] Loroux, B.F.E. (1978) Contribution à l'étude hydrogéologique du bassin sédimentaire côtier de Côte d'Ivoire. These de Doctorat, Université de Bordeaux I.
- [45] Kouamé, A.A., Jaboyedoff, M., Goula Bi Tie, A., Derron, M., Kouamé, K.J. and Meier, C. (2019) Assessment of the Potential Pollution of the Abidjan Unconfined Aquifer by Hydrocarbons. *Geosciences*, 9, Article No. 60. <u>https://doi.org/10.3390/geosciences9020060</u>
- [46] Othmer Jr, E.F. and Berger, B.J. (2002) Future Monitoring Strategies with Lessons Learned on Collecting Representative Samples. Storm Water Program. Office of Water Programs Sacramento State.
- [47] Saaty, R.W. (1987) The Analytic Hierarchy Process—What It Is and How It Is Used. Mathematical Modelling, 9, 161-176. <u>https://doi.org/10.1016/0270-0255(87)90473-8</u>
- [48] Tallar, R.Y. and Suen, J. (2015) Aquaculture Water Quality Index: A Low-Cost Index to Accelerate Aquaculture Development in Indonesia. *Aquaculture International*, 24, 295-312. <u>https://doi.org/10.1007/s10499-015-9926-3</u>
- [49] Singh, L.K., Jha, M.K. and Chowdary, V.M. (2018) Assessing the Accuracy of GIS-Based Multi-Criteria Decision Analysis Approaches for Mapping Groundwater Potential. *Ecological Indicators*, 91, 24-37. <u>https://doi.org/10.1016/j.ecolind.2018.03.070</u>
- [50] Kumar, P., Thakur, P.K., Bansod, B.K. and Debnath, S.K. (2017) Multi-Criteria Evaluation of Hydro-Geological and Anthropogenic Parameters for the Groundwater Vulnerability Assessment. *Environmental Monitoring and Assessment*, 189, Article No. 564. <u>https://doi.org/10.1007/s10661-017-6267-x</u>
- [51] Ismaïl, H., Zeaiter, Z., Farkh, S. and Abou-Hamdan, H. (2015) Comparative Analysis of Results Obtained from 3 Indexes (SEQ-Eau IBD IPS) Used to Assess Water Quality of the Berdawni, A Mediterranean Stream at the Beqaa Region-Lebanon. *International Journal of Scientific & Technology Research*, 4, 34-40.
- [52] World Health Organization (2017) Guidelines for Drinking Water Quality. 4th

Edition.

- [53] Sahu, P. and Sikdar, P.K. (2007) Hydrochemical Framework of the Aquifer in and around East Kolkata Wetlands, West Bengal, India. *Environmental Geology*, 55, 823-835. <u>https://doi.org/10.1007/s00254-007-1034-x</u>
- [54] Alkindi, K.M., Mukherjee, K., Pandey, M., Arora, A., Janizadeh, S., Pham, Q.B., *et al.* (2021) Prediction of Groundwater Nitrate Concentration in a Semiarid Region Using Hybrid Bayesian Artificial Intelligence Approaches. *Environmental Science and Pollution Research*, 29, 20421-20436. <u>https://doi.org/10.1007/s11356-021-17224-9</u>
- [55] Hoerl, A.E. and Kennard, R.W. (1970) Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, **12**, 55-67. <u>https://doi.org/10.1080/00401706.1970.10488634</u>
- [56] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58, 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
- [57] Wang, S., Ji, B., Zhao, J., Liu, W. and Xu, T. (2018) Predicting Ship Fuel Consumption Based on LASSO Regression. *Transportation Research Part D: Transport and Environment*, 65, 817-824. <u>https://doi.org/10.1016/j.trd.2017.09.014</u>
- [58] Shams, M.Y., Elshewey, A.M., El-kenawy, E.M., Ibrahim, A., Talaat, F.M. and Tarek, Z. (2023) Water Quality Prediction Using Machine Learning Models Based on Grid Search Method. *Multimedia Tools and Applications*, 83, 35307-35334. <u>https://doi.org/10.1007/s11042-023-16737-4</u>
- [59] Li, X., Li, W. and Xu, Y. (2018) Human Age Prediction Based on DNA Methylation Using a Gradient Boosting Regressor. *Genes*, 9, Article No. 424. <u>https://doi.org/10.3390/genes9090424</u>
- [60] Friedman, J.H. (2001) Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29, 1189-1232. https://doi.org/10.1214/aos/1013203451
- [61] Chung, C.F. and Fabbri, A.G. (2003) Validation of Spatial Prediction Models for Landslide Hazard Mapping. *Natural Hazards*, **30**, 451-472. <u>https://doi.org/10.1023/b:nhaz.0000007172.62651.2b</u>
- [62] Jha, M.K. and Sahoo, S. (2014) Efficacy of Neural Network and Genetic Algorithm Techniques in Simulating Spatio-Temporal Fluctuations of Groundwater. *Hydrological Processes*, 29, 671-691. <u>https://doi.org/10.1002/hyp.10166</u>
- [63] Ratner, B. (2009) The Correlation Coefficient: Its Values Range between +1/-1, or Do They? *Journal of Targeting, Measurement and Analysis for Marketing*, 17, 139-142. <u>https://doi.org/10.1057/jt.2009.5</u>
- [64] Complete Dissertation by Statistics Solutions. Correlation (Pearson, Kendall, Spearman). <u>https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/correlation-pearson-kendall-spearman/</u>
- [65] Roose, E., Chéroux, M., Humbel, F.X. and Perraud, A. (1966) Les sols du bassin sédimentaire de Côte d'Ivoire. *Cahiers ORSTOM*, 4, 51-92.
- [66] Zhang, C., Li, X., Ma, J., Wang, Z. and Hou, X. (2022) Stable Isotope and Hydrochemical Evolution of Shallow Groundwater in Mining Area of the Changzhi Basin, Northern China. *Environmental Earth Sciences*, 81, Article No. 294. <u>https://doi.org/10.1007/s12665-022-10416-7</u>
- [67] Riess, M. (2015) Assainissement et Gestion Intégrée du Bassin Versant du Gourou (Côte d'Ivoire). These de Doctorat, École Nationale du Génie et de l'Eau et de

l'Environnement de Strasbourg.

- [68] Oga, Y., Lasm, T., Massault, M., Baka, D., Ake, G., Marlin, C., et al. (2011) Caractérisation et suivi isotopique des eaux de la nappe semi-captive du Maestrichtien de la Côte d'Ivoire. International Journal of Biological and Chemical Sciences, 5, 991-1004. https://doi.org/10.4314/ijbcs.v5i3.72195
- [69] Kamtchueng, B.T., Fantong, W.Y., Ueda, A., Tiodjio, E.R., Anazawa, K., Wirmvem, M.J., et al. (2014) Assessment of Shallow Groundwater in Lake Nyos Catchment (Cameroon, Central-Africa): Implications for Hydrogeochemical Controls and Uses. Environmental Earth Sciences, 72, 3663-3678. https://doi.org/10.1007/s12665-014-3278-6
- [70] Kadam, A.K., Wagh, V.M., Muley, A.A., Umrikar, B.N. and Sankhua, R.N. (2019) Prediction of Water Quality Index Using Artificial Neural Network and Multiple Linear Regression Modelling Approach in Shivganga River Basin, India. *Modeling Earth Systems and Environment*, 5, 951-962. <u>https://doi.org/10.1007/s40808-019-00581-3</u>
- [71] Zhang, Y., Dai, Y., Wang, Y., Huang, X., Xiao, Y. and Pei, Q. (2021) Hydrochemistry, Quality and Potential Health Risk Appraisal of Nitrate Enriched Groundwater in the Nanchong Area, Southwestern China. *Science of The Total Environment*, **784**, Article ID: 147186. <u>https://doi.org/10.1016/j.scitotenv.2021.147186</u>
- [72] Naseem, S., Rafique, T., Bashir, E., Bhanger, M.I., Laghari, A. and Usmani, T.H. (2010) Lithological Influences on Occurrence of High-Fluoride Groundwater in Nagar Parkar Area, Thar Desert, Pakistan. *Chemosphere*, **78**, 1313-1321. <u>https://doi.org/10.1016/j.chemosphere.2010.01.010</u>
- [73] Kouamé, A. (2018) Apport de la modélisation hydrogéologique dans l'étude des risques de contamination de la nappe d'Abidjan par les hydrocarbures: Cas du benzène dans le District d'Abidjan (Côte d'Ivoire). Doctoral Dissertation, Université de Lausanne, Faculté des géosciences et de l'environnement.
- [74] Douagui, A.G., Kouamé, I.K., Mangoua, J.M.O., Kouassi, A.K. and Savané, I. (2019) Using Water Quality Index for Assessing of Physicochemical Quality of Quaternary Groundwater in the Southern Part of Abidjan District (Côte D'ivoire). *Journal of Water Resource and Protection*, 11, 1278-1291. https://doi.org/10.4236/jwarp.2019.1110074
- [75] Tiwari, A., Singh, P. and Mahato, M. (2014) GIS-Based Evaluation of Water Quality Index of Groundwater Resources in West Bokaro Coalfield, India. *Current World Environment*, 9, 843-850. <u>https://doi.org/10.12944/cwe.9.3.35</u>
- [76] Feng, Y., Fanghui, Y. and Li, C. (2019) Improved Entropy Weighting Model in Water Quality Evaluation. *Water Resources Management*, **33**, 2049-2056. <u>https://doi.org/10.1007/s11269-019-02227-6</u>
- [77] Batabyal, A.K. and Chakraborty, S. (2015) Hydrogeochemistry and Water Quality Index in the Assessment of Groundwater Quality for Drinking Uses. *Water Environment Research*, 87, 607-617. <u>https://doi.org/10.2175/106143015x14212658613956</u>
- [78] Alshehri, F. and Rahman, A. (2023) Coupling Machine and Deep Learning with Explainable Artificial Intelligence for Improving Prediction of Groundwater Quality and Decision-Making in Arid Region, Saudi Arabia. *Water*, 15, Article No. 2298. https://doi.org/10.3390/w15122298
- [79] Feng, F., Ghorbani, H. and Radwan, A.E. (2024) Predicting Groundwater Level Using Traditional and Deep Machine Learning Algorithms. *Frontiers in Environmental Science*, **12**, Article ID: 1291327. <u>https://doi.org/10.3389/fenvs.2024.1291327</u>
- [80] Morovvati Zarajabad, A., Hadi, M., Nabizadeh Nodehi, R., Moradi, M., Rezvani Ghalhari, M., Zeraatkar, A., et al. (2024) Providing Predictive Models for Quality

Parameters of Groundwater Resources in Arid Areas of Central Iran: A Case Study of Kashan Plain. *Heliyon*, **10**, e31493. <u>https://doi.org/10.1016/j.heliyon.2024.e31493</u>

- [81] Jibrin, A.M., Al-Suwaiyan, M., Aldrees, A., Dan'azumi, S., Usman, J., Abba, S.I., et al. (2024) Machine Learning Predictive Insight of Water Pollution and Groundwater Quality in the Eastern Province of Saudi Arabia. *Scientific Reports*, 14, Article No. 20031. https://doi.org/10.1038/s41598-024-70610-4
- [82] Chen, T. and Guestrin, C. (2016) XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, 13-17 August 2016, 785-794. <u>https://doi.org/10.1145/2939672.2939785</u>
- [83] Sharafati, A., Asadollah, S.B.H.S. and Neshat, A. (2020) A New Artificial Intelligence Strategy for Predicting the Groundwater Level over the Rafsanjan Aquifer in Iran. *Journal of Hydrology*, 591, Article ID: 125468. https://doi.org/10.1016/j.jhydrol.2020.125468
- [84] Gibbs, R.J. (1970) Mechanisms Controlling World Water Chemistry. *Science*, 170, 1088-1090. <u>https://doi.org/10.1126/science.170.3962.1088</u>
- [85] Li, C., Gao, Z., Chen, H., Wang, J., Liu, J., Li, C., et al. (2021) Hydrochemical Analysis and Quality Assessment of Groundwater in Southeast North China Plain Using Hydrochemical, Entropy-Weight Water Quality Index, and GIS Techniques. Environmental Earth Sciences, 80, Article No. 523. https://doi.org/10.1007/s12665-021-09823-z