

Article

Groundwater Quality Assessment and Irrigation Water Quality Index Prediction Using Machine Learning Algorithms

Enas E. Hussein ¹, Abdessamed Derdour ^{2,3}, Bilel Zerouali ⁴, Abdulrazak Almaliki ⁵, Yong Jie Wong ^{6,*}, Manuel Ballesta-de los Santos ⁷, Pham Minh Ngoc ⁸, Mofreh A. Hashim ¹ and Ahmed Elbeltagi ⁹

- ¹ National Water Research Center, Shubra El-Kheima 13411, Egypt; enas_el-sayed@nwrc.gov.eg (E.E.H.); mofrehhashim00@gmail.com (M.A.H.)
 - ² Artificial Intelligence Laboratory for Mechanical and Civil Structures, and Soil, University Center of Naama, P.O. Box 66, Naama 45000, Algeria; derdour@cuniv-naama.dz
 - ³ Laboratory for the Sustainable Management of Natural Resources in Arid and Semi-Arid Zones, University Center Salhi Ahmed Naama (Ctr Univ Naama), P.O. Box 66, Naama 45000, Algeria
 - ⁴ Vegetal Chemistry-Water-Energy Research Laboratory, Faculty of Civil Engineering and Architecture, Department of Hydraulic, Hassiba Benbouali, University of Chlef, B.P. 78C, Ouled Fares, Chlef 02180, Algeria; b.zerouali@univ-chlef.dz
 - ⁵ Department of Civil Engineering, College of Engineering, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia; a.almaliki@tu.edu.sa
 - ⁶ Department of Bioenvironmental Design, Faculty of Bioenvironmental Sciences, Kyoto University of Advanced Science, Kyoto 606-8501, Japan
 - ⁷ Field in Agricultural Chemistry and Soil Science, Scientific R&D Department, Fertilizantes y Nutrientes Ecológicos S.L. (FYNECO), Industrial Estate Ceutí, C/Río Taibilla S/N, 30562 Ceutí, Spain; m.ballesta@fyneco.es
 - ⁸ Research Center for Environmental Quality Management, Graduate School of Engineering, Kyoto University, Kyoto 520-0811, Japan; m.ngoc.pham.ktu@gmail.com
 - ⁹ Agricultural Engineering Department, Faculty of Agriculture, Mansoura University, Mansoura 35516, Egypt; ahmedelbeltagi81@mans.edu.eg
- * Correspondence: wong.yongjie@kuas.ac.jp or wong.yongjie.p80@kyoto-u.jp



Citation: Hussein, E.E.; Derdour, A.; Zerouali, B.; Almaliki, A.; Wong, Y.J.; Ballesta-de los Santos, M.; Minh Ngoc, P.; Hashim, M.A.; Elbeltagi, A. Groundwater Quality Assessment and Irrigation Water Quality Index Prediction Using Machine Learning Algorithms. *Water* **2024**, *16*, 264. <https://doi.org/10.3390/w16020264>

Academic Editors: Chi-Feng Chen and Chia-Ling Chang

Received: 24 November 2023

Revised: 1 January 2024

Accepted: 8 January 2024

Published: 11 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The evaluation of groundwater quality is crucial for irrigation purposes; however, due to financial constraints in developing countries, such evaluations suffer from insufficient sampling frequency, hindering comprehensive assessments. Therefore, associated with machine learning approaches and the irrigation water quality index (IWQI), this research aims to evaluate the groundwater quality in Naama, a region in southwest Algeria. Hydrochemical parameters (cations, anions, pH, and EC), qualitative indices (SAR, RSC, Na%, MH, and PI), as well as geospatial representations were used to determine the groundwater's suitability for irrigation in the study area. In addition, efficient machine learning approaches for forecasting IWQI utilizing Extreme Gradient Boosting (XGBoost), Support vector regression (SVR), and K-Nearest Neighbours (KNN) models were implemented. In this research, 166 groundwater samples were used to calculate the irrigation index. The results showed that 42.18% of them were of excellent quality, 34.34% were of very good quality, 6.63% were good quality, 9.64% were satisfactory, and 4.21% were considered unsuitable for irrigation. On the other hand, results indicate that XGBoost excels in accuracy and stability, with a low RMSE (of 2.8272 and a high R of 0.9834. SVR with only four inputs (Ca²⁺, Mg²⁺, Na⁺, and K) demonstrates a notable predictive capability with a low RMSE of 2.6925 and a high R of 0.98738, while KNN showcases robust performance. The distinctions between these models have important implications for making informed decisions in agricultural water management and resource allocation within the region.

Keywords: groundwater; hydrochemical parameters; irrigation water quality index (IWQI); machine learning; agricultural water management

1. Introduction

For the long-term development of many sectors, groundwater is a valuable resource, especially in arid regions [1]. However, the fast population increase, industrial and agricultural expansion, climate change, and other factors have led to the severe degradation of and threat to groundwater quality in recent decades [2,3]. Due to the complexity of protecting groundwater resources for future generations while also meeting the needs of many economic activities, most notably agricultural activities, groundwater sustainability has become an important issue [4,5]. Agriculture is the main user of water resources [6–8]. More than 70% of freshwater is used for agriculture in most of the world's areas [9]. To feed a planet of 9 billion people by 2050, agricultural production is anticipated to increase by 50%, and water withdrawals are expected to increase by 15% [10,11]. In Algeria, the water resources sector mobilizes nearly 11.2 billion m³/year, of which 7.3 billion m³ are devoted to agriculture, i.e., more than 70%, and 3.6 billion m³ per year is allocated to drinking water [12]. Large agricultural perimeters are irrigated in the north using boreholes and dams. At the same time, large aquifers in the south are used to irrigate perimeters through deep boreholes. The country's irrigated areas have evolved from 905,293 ha in 2007 to 1,640,000 ha in 2020 [13]. Small-scale irrigation systems have also grown significantly due to official subsidies and aid given to farmers and the liberalization of drilling and well digging [14]. The excessive use of groundwater in such regions, where the influence of climate variability is very pronounced, has been the cause of the degradation of this resource both from a quantitative and qualitative point of view. Water quality is a limiting element for life quality worldwide [15]. Therefore, it is crucial to consider the quality of these resources while using them for irrigation due to their impact on human health, where salinity levels and soluble salt compositions are the main issues with water quality in most irrigation situations [16]. Frequently, the unwise use of salty water leads to groundwater pollution, sodicity, soil salinity, and ion toxicity [17]. In addition, excessive salinity levels can negatively impact crop productivity, fertility requirements, physical soil conditions, and irrigation systems [17]. Consequently, increasing the quality of water is necessary to guarantee the development of excellent crops and the maintenance of soil integrity [16]. Around the world, research on the sustainability of water quality used for irrigation is expanding. Numerous studies have been conducted to carefully construct hydrochemical indicators for assessing irrigation water quality. For example, Tlili-Zrelli et al. [18] evaluated the quality of groundwater in the region of Grombalia in Tunisia using graphical and multivariate statistical methods. It has been shown that sodium adsorption ratios (SAR) are an effective method for evaluating irrigation water quality in many studies [19–22]. Furthermore, other research evaluated water quality using a statistical method for irrigation [23–27]. Research has also been conducted in Algeria on the quality of groundwater in many aquifers for irrigation. Many researchers worldwide have developed several other indicators to represent water quality for irrigation [28,29]. Among these indicators is the water quality index (WQI), developed by Brown and McClelland [30]. It was first defined by Horton [31]. A multivariate statistical analysis of water was employed by Meireles et al. [32] to develop a new water quality indicator (WQI) for irrigation, called the irrigation water quality index (IWQI). The variables included in the index were Electrical Conductivity, the sodium adsorption ratio, Bicarbonate, sodium, and chloride. The authors also reclassified the WQI for irrigation, taking into account soil salinity and infiltration rates. Decision-makers can easily use this method to evaluate a water type's quality and potential risks based on a wide range of parameters [33]. Additionally, IWQI enables the assessment and comparison of different water samples to prevent adverse effects on soil and plants [34]. Drilling wells for agricultural use in areas with significant groundwater salinization is made more affordable with the IWQI forecast [35]. This intelligent method is becoming more frequently employed for monitoring the quality of water in many research projects due to its usefulness in identifying a solution to a complicated issue and highlighting the input and output data relationship. Artificial neural networks (ANN) models, among other machine learning models, have been used to forecast outputs as computer technology has

advanced. Data-driven models called ANNs are products of the evolution of artificial intelligence [33,34]. This study aims to estimate the IWQI to evaluate the groundwater of Naama's arid region in southwest Algeria for irrigation purposes. Additionally, we propose a methodology for forecasting IWQI using Extreme Gradient Boosting (XGBoost), Support vector regression (SVR), and K-Nearest Neighbours (KNN) models. The results were classified into different classes from excellent to unsuitable in order to facilitate its consideration. The accuracy of computed and forecasted IWQI values is then evaluated. The results will be helpful for predicting changes in water quality, enabling better water resource management, planning, and decision-making concerning available resources, especially in arid locations.

2. Materials and Methods

2.1. The Description of the Study Area

We are focusing on the Wilaya of Naama (29,514.14 km²) located in southwestern Algeria, between the latitudes of 33°22'7.84" N and 33°22'7.84" N and longitudes 0°21'25.05" E and 0°21'25.05" E (Figure 1). The study area's northern boundary is Wilaya of Tlemcen and Sidi Belabbes, west of the Algerian–Moroccan border, east of Wilaya of El-Bayadh, and south of Wilaya of Bechar. In terms of agricultural activities, the northern section of this research region is marked by pastoral activities and livestock, occupying about 74% of the total area. At the same time, the southern part is characterized by small-scale irrigation systems, with the cultivation of vegetables, cereal, and olive trees [36]. The primary water resources refer to groundwater with four principal aquifer systems: the quaternary alluvial aquifer, the tertiary limestones aquifer, the Jurassic sandstone reservoir, and the Albian aquifer. Among these, the flow rates vary between 5 and 80 L/s. These aquifers provide water for 1,893,122 animals grazing in the study area and 208,136 residents living there [37]. This region's surface water resources are severely stressed due to climatic conditions [38]. The research area is an arid region with a mean annual rainfall of 287 mm and a maximum yearly rate of evapotranspiration of 2000 mm [39]. Mineral soils, saline soils, and limestone magnetic soils make up the majority of the soil types in this region [40]. According to a land use analysis, the research area's surface comprises 29.14% steppe ranges, 24.06% severely degraded ranges, 17.94% wind accumulations, only 6.74% dunes, 15.06% rocky outcrops, and 7.05% of forests [41]. Geologically, the tertiary sediments cover the study's northern area, while the south is composed of cretaceous and Jurassic sediments [42]. The first important economic sector in the wilaya of Naama is agriculture, specifically pastoralism, with more than 2,203,460 Ha of the agricultural area, of which 28,283 Ha is irrigated [43]. The main crops cultivated are cereals and market gardening. All the efforts made by the state contribute to the consolidation of the various actions included in the framework of the national FNRDA program and the upgrading of all the farms on the one hand, and on the other hand, the increase in the area useful as agricultural land through the development of new lands. The main objective is the intensification of agricultural pockets by tree planting as a means of combating desertification and the promotion of fodder crops to meet the needs of livestock.

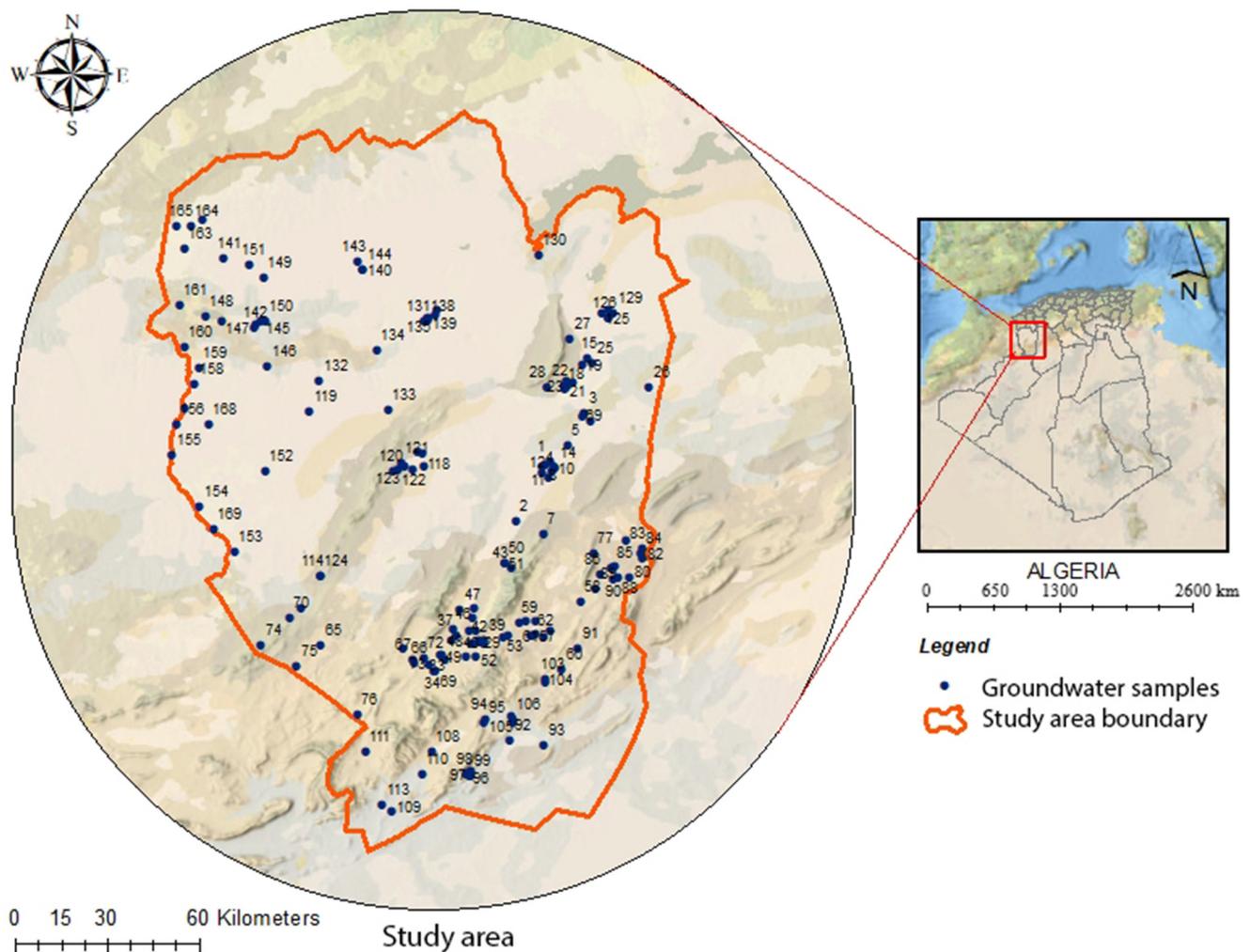


Figure 1. Study area.

2.2. Collection of Data, Analysis, and Calculation

From the study area, 166 samples were collected, and ten elements were evaluated (Ca^{2+} , K^+ , Na^+ , Mg^{2+} , Cl^- , NO_3^- , SO_4^{2-} , HCO_3^- , Electrical Conductivity, and Hydrogen power (pH)). A multiparameter portable quality type HANNA (HI98194) was utilized to assess Hydrogen power (pH), Electric Conductivity (EC), and temperature (T) in situ. To evaluate the cation and anion elements, samples were collected and transported to the Sustainable Management of Natural Resources laboratory in Arid and Semi-Arid Zones at Naama and reserved at four degrees Celcius. The chemical analyses were completed using the procedures for the chemical examination of wastes and water (EPA-600/4-79-020) [44]. Magnesium (Mg), Calcium (Ca), and Bicarbonate (HCO_3^-) were evaluated using the method of titration. Atomic absorption spectrometry was used to determine sodium (Na) and potassium (K). Chloride (Cl) concentrations were dosed by the Mohr method. In addition to sulphate (SO_4), Nitrate (NO_3) was also differentiated by the UV-Vis spectrophotometer. The assessment of the suitability of groundwater in the Naama region for irrigation was established with the international standard (FAO).

2.3. Suitability Indices for Irrigation

Various groundwater indices are frequently used to evaluate groundwater suitability for agricultural use. Sodium percentage, sodium adsorption ratio, Magnesium hazard, Permeability Index, Potential salinity, and Kelly's ratio. The conventional formulas from

(1) to (6), shown in Table 1, were used to determine the (SAR), (MH), (Na%), (PI), (KR), and (PS) correspondingly, where all ions are given in meq/L.

Table 1. The irrigation water's qualitative formulas.

Parameter	Formula Adopted	References
Sodium adsorption ratio	$SAR = \frac{Na^+}{\sqrt{\frac{Ca^{2+} + Mg^{2+}}{2}}}$ (1)	[45]
Sodium percentage	$Na\% = \frac{(Na^+ + K^+)}{(Ca^{2+} + Mg^{2+} + Na^+)}$ (2)	[28]
Permeability Index	$PI = \frac{(Na^+ + \sqrt{HCO_3^-})}{(Ca^{2+} + Mg^{2+} + Na^+)} \times 100$ (3)	[29]
Magnesium hazard	$MH = \frac{Ca^{2+}}{Ca^{2+} + Mg^{2+}} \times 100$ (4)	[45]
Kelly's ratio	$KR = \frac{Na^+}{Ca^{2+} + Mg^{2+}}$ (5)	[46]
Potential salinity	$PS = Cl^- + \frac{SO_4^{2-}}{2}$ (6)	[47]

2.4. Irrigation Water Quality Index (IWQI)

It is recognized that the quality of irrigation water and many other factors, such as the nature of the soil, the type of crops, the climatic conditions, and the methods of irrigation, play a significant role in profitability and agricultural yield. The increase in the salinity of irrigation water negatively affects the soil and the plants. The mineral salts present in the irrigation water can cause changes in the structure of the soil, thus modifying its permeability and aeration, which leads to a disturbance in the development of plants [48]. To obtain a clear view of the overall quality of irrigation water, the IWQI was employed to reflect the composite influence of numerous water quality parameters on that water's overall quality [33,34]. Equation (7) is used in this model to calculate the irrigation water quality parameter (qi), which is determined by the tolerance limits of the parameters listed in Table 2:

$$qi = qi_{max} - \frac{[(x_{ij} - x_{inf}) \times qi_{amp}]}{X_{amp}} \quad (7)$$

where qi is the quality of each parameter, qi_{max} stands for the maximum value of qi for every class, x_{ij} stands for every parameter's observed value, and x_{inf} stands for the value corresponding to the parameter's lower limit class and where qi_{amp} is the amplitude of quality measurement class, and X_{amp} is the amplitude class.

Table 2. Limiting values for parameters used in quality assessments (qi).

qi	EC(uS cm ⁻¹)	SAR	Na ⁺	Cl ⁻	HCO ₃ ⁻
0–35	EC < 750 or EC ≥ 3000	SAR < 2 or SAR ≥ 12	Na < 2 or Na ≥ 12	Cl < 1 or Cl ≥ 10	HCO ₃ ⁻ < 1 or HCO ₃ ⁻ ≥ 8.5
35–60	1500 ≤ EC < 3000	6 ≤ SAR < 12	6 ≤ Na < 12	7 ≤ Cl < 10	4.5 ≤ HCO ₃ ⁻ < 8.5
60–85	750 ≤ EC < 1500	3 ≤ SAR < 6	3 ≤ Na < 6	4 ≤ Cl < 7	1.5 ≤ HCO ₃ ⁻ < 4.5
85–100	200 ≤ EC < 750	2 ≤ SAR < 3	2 ≤ Na < 3	1 ≤ Cl < 4	1 ≤ HCO ₃ ⁻ < 1.5

Finally, using Equation (8), the IWQI was calculated. Table 3 presents the relative weight of every parameter according to Meireles et al. [32].

$$IWQI = \sum_{i=1}^n qi w_i \quad (8)$$

The IWQI ranges between 0 and 100. Five (05) categories were used to classify the irrigation water quality index IWQI from excellent to inappropriate, as shown in Table 4.

Table 3. Relative weights used to calculate IWQI.

Parameters	Wi
SAR	0.189
EC	0.211
Cl	0.194
Na	0.204
HCO ₃	0.202
Total	1

Table 4. IWQI classification.

IWQI Type	IWQI
Unsuitable	0–40
Satisfying	40–55
Good	55–70
Very Good	70–85
Excellent	85–100

2.5. Extreme Gradient Boosting (XGBoost) Algorithm

XGBoost, which stands for Extreme Gradient Boosting, is a powerful and popular machine learning algorithm that is primarily used for supervised learning tasks, including classification and regression [49–51]. It was introduced by Tianqi Chen, a computer scientist and machine learning researcher, in a research paper [52]. XGBoost is an ensemble learning method, meaning it combines the predictions from multiple models (usually decision trees) to create a more robust model. An overview of XGBoost’s work was summarised as follows:

1. Gradient Boosting: XGBoost is a gradient boosting algorithm, which means it combines multiple decision trees to create a stronger predictive model.
2. Decision Trees as Base Learners: Decision trees are used as the base or “weak” learners in XGBoost. These trees are trained to minimize a specified loss function (e.g., mean squared error for regression or log-loss for classification).
3. Iterative Training: XGBoost iteratively adds trees to the model. It starts with an initial prediction (e.g., the mean of the target variable) and then fits a tree to the residuals (the differences between predictions and actual values).
4. Regularization: XGBoost includes regularization techniques (L1 and L2 regularization) to control overfitting and enhance model generalization.
5. Ensemble Learning: The predictions from multiple trees are combined to create the final model. Each new tree is weighted and added to the previous predictions.
6. Parallel and Distributed Computing: XGBoost is designed for efficiency and can leverage parallel and distributed computing to handle large datasets and complex models.
7. Feature Importance: XGBoost provides feature importance scores, helping identify the most influential features in the model’s decisions.
8. Hyperparameter Tuning: To optimize model performance, users can fine-tune hyperparameters like learning rate, tree depth, and subsampling.

2.6. Support Vector Regression (SVR) Algorithm

SVR is a machine learning technique used for regression tasks. It is a variation of the Support Vector Machine (SVM) algorithm primarily used for classification. The second objective of SVR is to predict a continuous target variable (real numbers) based on input features [53–55]. First developed by Vladimir N. Vapnik and Alexey Ya Chervonenkis [56], the SVR is used when you have data that do not necessarily follow a linear pattern and may involve complex relationships.

SVR introduces the concept of a margin of tolerance (ϵ) around the predicted hyperplane. The margin represents the acceptable prediction error, where data points outside this margin contribute to the loss function. SVR seeks to minimize this loss while respecting

the margin constraints to find a hyperplane that best fits the data. Support vectors are the data points closest to the margin (inside or on the margin boundary), where these points are the most influential data points, as they have the largest impact on defining the hyperplane [57]. SVR uses a loss function that captures the trade-off between minimizing the error (the difference between simulated and observed values) and maximizing the margin around the hyperplane. The loss function typically has two terms: minimizing the error and controlling the margin. The radial basis function (RBF) is the most common kernel function used to transform the input data into a higher-dimensional space [58].

It controls the trade-off between maintaining a large margin and fitting the data based on regularization parameter, often denoted as “C”. A smaller C value results in a larger margin but may allow more errors, and conversely, training SVR based on Sequential Minimal Optimization (SMO) involves finding the optimal hyperplane and the support vectors. After training, SVR can make predictions for new data points by calculating their position relative to the hyperplane, where the predicted value is influenced by the position and distance from the hyperplane and the margin [59].

2.7. K-Nearest Neighbours (KNN) Algorithm

KNN is a simple yet effective supervised machine learning algorithm for classification and regression tasks. KNN is a non-parametric, instance-based algorithm, meaning it does not make any underlying assumptions about the data distribution and instead relies on the data itself to make predictions. The prediction process in KNN is briefly explained. It mentions that the prediction for a given sample is based on information from its K-Nearest Neighbours in the feature space. K represents the number of neighbours considered when making predictions [60]. The KNN highlights the importance of choosing an appropriate value for k. It is emphasized that the choice of k should depend on the dataset’s specific characteristics and the desired accuracy level. The value of k significantly impacts the algorithm’s performance [61,62].

2.8. Performance Criteria

The dataset consisting of 166 samples was divided into separate training and testing subsets to evaluate the proposed algorithms’ effectiveness. The training subset, comprising approximately 70% of the data, equivalent to around 116 samples, was employed for optimizing model parameters and achieving peak performance. The remaining 30% of the data, roughly 50 samples, formed the testing set for conclusive model assessment. This 70/30 data split ratio conforms to a widely recognized practice supported by the existing literature [63,64]. We used multiple performance metrics to assess the model’s effectiveness, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Nash–Sutcliffe Efficiency (NSE), and the Pearson Correlation Coefficient (R). Concise explanations of each performance measure are presented below [65,66].

$$MAE = \sqrt{\frac{1}{n} \sum_{i=1}^n [(IWQI_{O,i}) - (IWQI_{S,i})]} \quad (9)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [(IWQI_{O,i}) - (IWQI_{S,i})]^2} \quad (10)$$

$$NSE = 1 - \frac{\sum_{i=1}^n [(IWQI_{O,i}) - (IWQI_{S,i})]^2}{\sum_{i=1}^n [(IWQI_{O,i}) - (IWQI_{O})]^2} \quad (11)$$

$$R = \frac{\sum_{i=1}^n (IWQI_{o,i} - \overline{IWQI_o}) (IWQI_{s,i} - \overline{IWQI_s})}{\sum_{i=1}^n (IWQI_{obs,i} - \overline{IWQI_{obs}})^2 \sum_{i=1}^n (IWQI_{s,i} - \overline{IWQI_s})^2} \quad (12)$$

where $IWQI_{o,i}$ and $IWQI_{s,i}$ represent the actual and simulated observations, respectively. N is the sample size of the database. $\overline{IWQI_o}$ and $\overline{IWQI_s}$ represent the mean values of the actual and simulated samples, respectively.

3. Results

3.1. Descriptive Statistics of Physico-Chemical Parameters of Irrigation Water

Electrical Conductivity (EC), Hydrogen power (pH), main ions (K^+ , Na^+ , Ca^{2+} , Mg^{2+} , NO_3^- , HCO_3^- , SO_4^{2-} , Cl^-) are all listed in Table 5, along with their respective minimums, maximums, means, and standard deviations for all 166 samples. The pH of the research area averages at 7.71. Consequently, the groundwater is alkaline in the studied area. Water salinity and total dissolved solids in water can be measured using the practical and trustworthy index of Electrical Conductivity (EC). The values of EC of the study area varied from 290.00 $\mu\delta/cm$ to 6200.00 $\mu\delta/cm$. According to FAO guidelines, 92.16% of Electrical Conductivity (EC) values of the research area are within an acceptable range (<3000 $\mu\delta/cm$) [67]. On the other hand, Calcium concentrations in the groundwater of the research area range from 0.60 to 56.10 meq/L (Figure 2a). About 94.6% of the Calcium concentrations are within the permissible range of the FAO recommendations, which set a maximum value of 20 meq/L [67]. While 53.61% of Magnesium concentrations are within the FAO guidelines, as shown in Figure 2b (<5 meq/L) [67]. All sodium results (except only three samples) are within FAO standard limits (<40 meq/L), as these levels varied from 0.22 mg/L to 48.48 mg/L [67]. Among the water sampling points in our study area, potassium values range from 0.03 to 6.69 meq/L (Figure 2d), and the maximum concentration allowed in irrigation water stipulated by FAO is 2.00 meq/L [67]. As a result, all potassium results are within FAO standard limits (except one sample). The mean sulphate value in this study is 7.64 meq/L, ranging from 0.79 to 49.38 meq/L (Figure 3a). One hundred fifty samples of sulphate concentrations are within the guideline range established by the FAO (20 meq/L) [67]. The mean Nitrate concentration ranges between 0.02 and 6.29 meq/L in the groundwater samples (Figure 3b), with a mean value of 0.44 meq/L. Bicarbonate concentrations range from 0.33 to 8.67 meq/L (Figure 3c). We remark that all samples fall within the FAO's permitted range, with a maximum value of 10 meq/L [67]. Our study region's average chloride concentration ranges from 0.28 to 79.41 meq/L (Figure 3d). According to FAO guidelines, 94% of chloride values of the research area are within an acceptable range (<30 meq/L) [67].

Table 5. Descriptive statistics of physico-chemical parameters of irrigation water.

	Min Value	Max Value	Mean Value	Standard Deviation
EC ($\mu\delta/cm$)	290.00	6200.00	1464.42	1100.87
pH	6.58	10.60	7.71	0.51
Ca^{2+} (meq/L)	0.60	56.10	7.01	7.05
Mg^{2+} (meq/L)	0.25	46.67	6.29	5.76
Na^+ (meq/L)	0.22	48.48	6.81	8.65
K^+ (meq/L)	0.03	6.69	0.25	0.54
Cl^- (meq/L)	0.28	79.41	7.87	12.39
SO_4^{2-} (meq/L)	0.79	49.38	7.64	8.99
HCO_3^- (meq/L)	0.33	8.67	3.90	1.04
NO_3^- (meq/L)	0.02	6.29	0.44	0.60
SAR	0.12	14.56	2.46	2.48
Na%	6.78	83.35	29.49	14.93
MH	2.86	91.74	48.60	13.00
KR	0.03	4.94	0.50	0.55
PS	1.35	83.47	10.34	13.09
PI	13.41	99.07	43.47	13.31

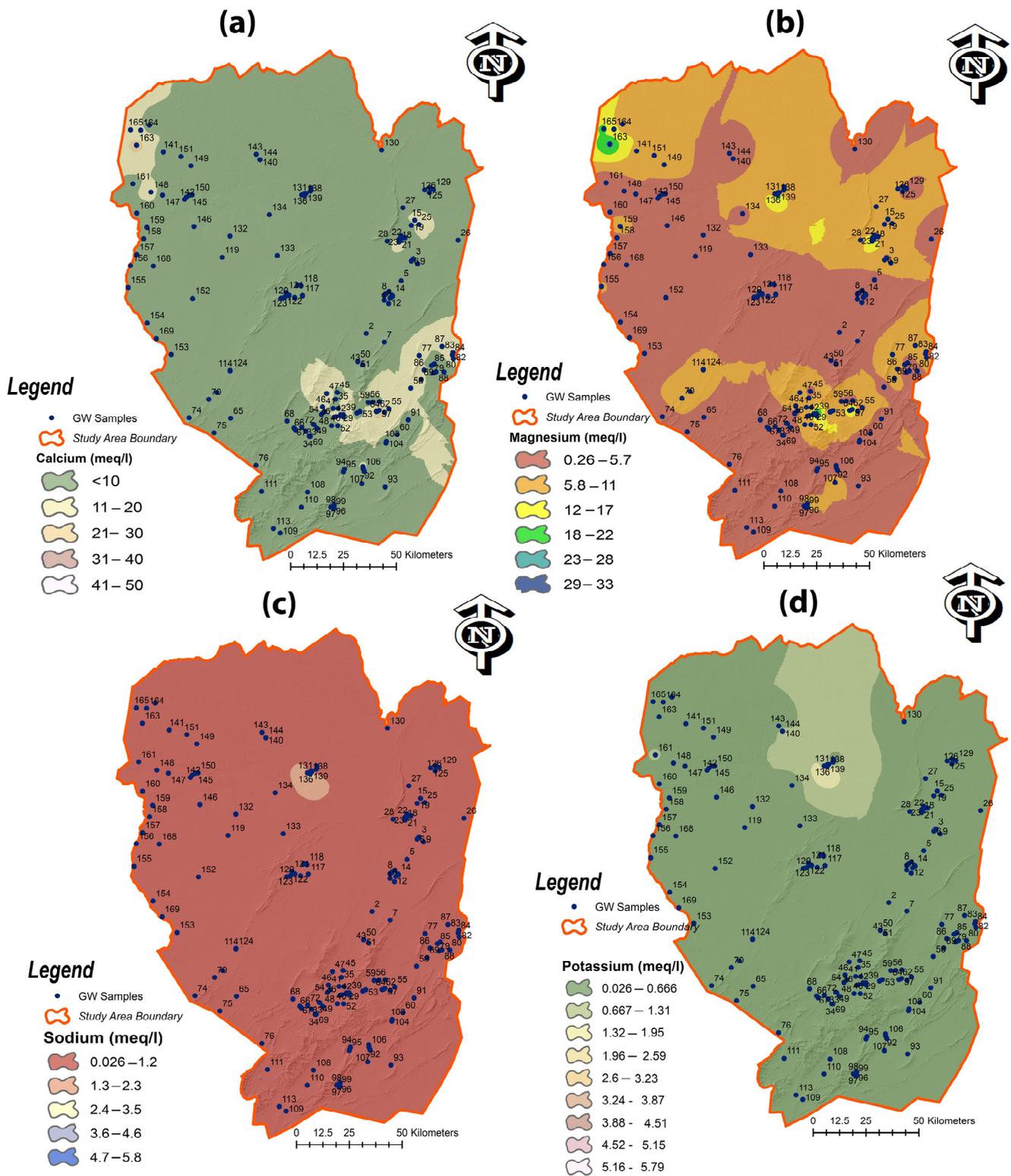


Figure 2. Geospatial distribution of (a) calcium; (b) magnesium; (c) sodium; (d) potassium.

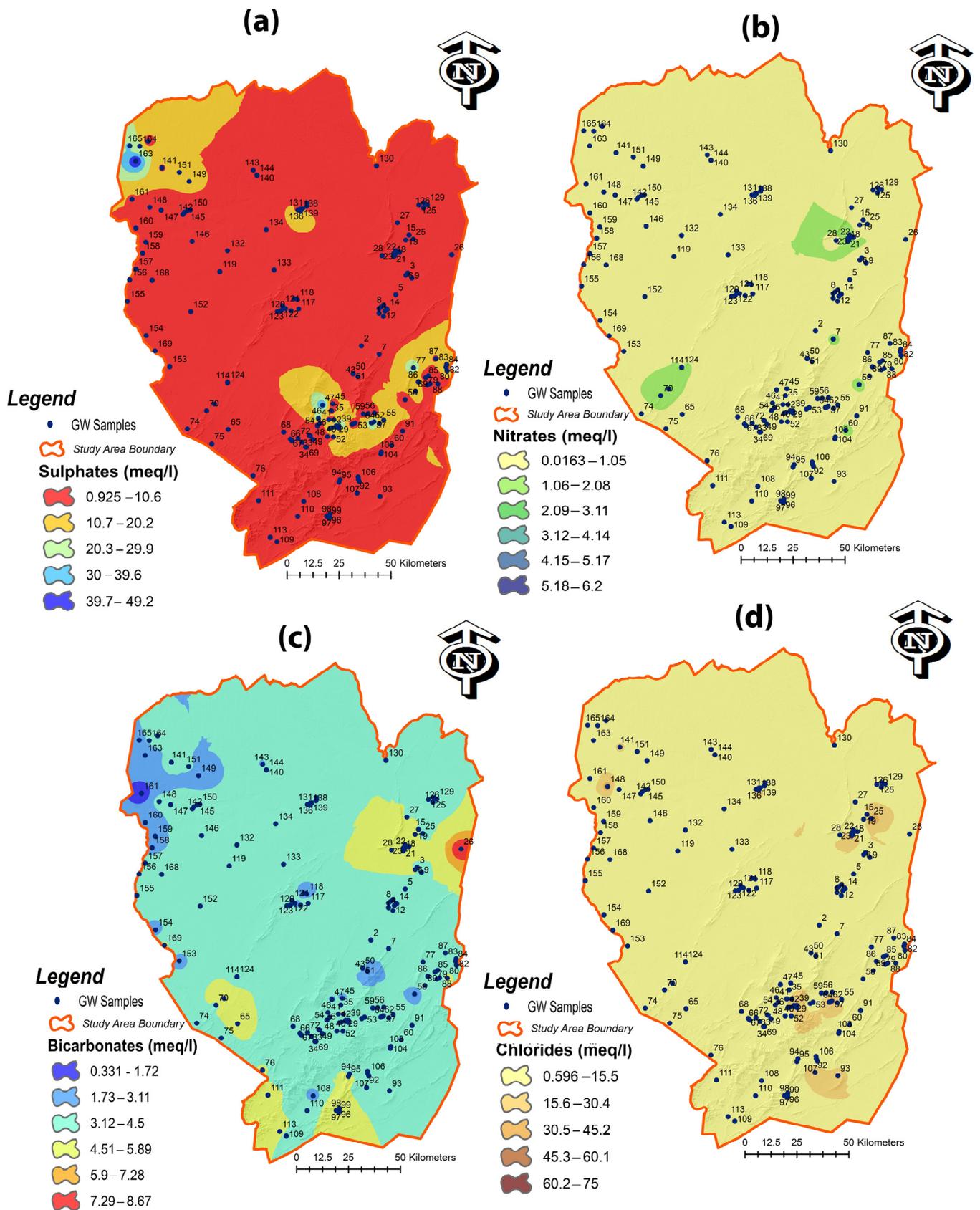


Figure 3. Geospatial distribution of (a) sulphates, (b) nitrates, (c) bicarbonates, and (d) chlorides.

The irrigation water's qualitative parameters like Magnesium hazard (MH), sodium percentage (Na%), sodium adsorption ratio (SAR), Permeability Index (PI), Potential salinity (PS), and Kelly's ratio (KR) are also reported in Table 3. The values of the sodium adsorption ratio (SAR) ranged from 0.12 to 14.56. SAR values between 0 and 18 placed all samples in excellent and good irrigation categories. SAR findings indicate that 98.8% of these samples are excellent and suitable for irrigation (Table 6). As a result of SAR analysis, the irrigation water quality can be classified into four primary categories: "excellent" for water with SARs less than ten meq/L, "good" for water with SARs between 10 and 18 meq/L, "doubtful" for water with SARs between 18 and 26 meq/L, and "unsuitable" for water with SARs more than 26 meq/L. The calculated values for sodium percentage (Na%) spanned from 6.78 to 83.35%. The percent sodium index increases the "permissible" category of the samples to 15.66% while 4.82% are doubtful for irrigation, and decreases the excellent category to 32.53% compared to the SAR classification. According to Aravinthasamy et al. [68], a greater sodium concentration (>60%) may cause the deterioration of the physical properties of soil. It is possible to determine the suitability of groundwater for irrigation based on its Permeability Index (PI) values. PI values, on the other hand, range from 13.41 to 99.07 meq/L, with a mean value of 43.77 meq/L. According to Doneen [47], there are three different categories for the Permeability Index (PI): Class 1, acceptable when $PI > 75$; Class 2, which is good when PI is between 25 and 75%; and Class 3, which is unsuitable when $PI < 25$ %. It is advised to use water under classes I and II for irrigation [47]. Only 3.61% of the samples in our study area have PI values greater than 25, making them unsuitable for irrigation. As shown in Table 6, all other samples are appropriate for irrigation according to Permeability Index Classification. The Magnesium hazard computed values range from 2.86 to 91.74%. It is not advised to use water for irrigation when the Magnesium hazard value is superior to 50 [45]. Moreover, 46.99% of the samples in the area under study have Magnesium hazard values higher than 50%, making them unsuitable for irrigation (Table 6). It is considered appropriate to irrigate water with a Kelly's index of less than one due to the assumption that water with a Kelly's index greater than one contains excessive sodium [46]. The values of KI vary between 0.03 and 4.94, with an average of 0.50. As shown in Table 6, 89.16% of groundwater samples in the study area were suitable for irrigation based on Kelly's ratio. The values of the Potential salinity of the study area ranged from 1.35 to 83.47 meq/L, while the average value was 10.34 meq/L. Potential salinity (PS) results revealed that 43.98% of samples are "excellent" to "good" for irrigation, 24.10% of samples are "Good" to "Injurious", and 31.93% are "Injurious" to "Unsatisfactory" for irrigation.

Table 6. Classification of irrigation water's qualitative parameters.

Irrigation Indices	Classification	Type	N° of Samples	Percentage (%)
SAR	SAR > 26	Unsuitable	0	0
	18 < SAR < 26	Doubtful	0	0
	10 < SAR < 18	Good	2	1.2
	SAR < 10	Excellent	164	98.8
Na%	80–100	Unsuitable	1	0.60
	60–80	Doubtful	8	4.82
	40–60	Permissible	26	15.66
	20–40	Good	77	46.39
	<20	Excellent	54	32.53
PI	<25%	Unsuitable	6	3.61
	>75%	Good	4	2.41
	25–75%	Suitable	156	93.98
MH	>50%	Unsuitable	78	46.99
	<50%	Suitable	88	53.01
KR	<1	Unsuitable	18	10.84
	>1	Suitable	148	89.16
PS	>10	Injurious to Unsatisfactory	53	31.93
	5–10	Good to Injurious	40	24.10
	<5	Excellent to good	73	43.98

3.2. Irrigation Water Quality Index Assessments

Table 7 shows the IWQI results for the groundwater samples of the Wilaya of Naama. Among the IWQI values, there was a wide range of values, ranging from 1.81 to 97.64, with an average value of 78.15. It has been concluded that the 75 samples falling into the excellent category represent 45.18% of the total samples (Figure 4). About 34.34% of samples fell into the very good category, and 6.63% fell into the good category. A total of 16 samples fall into the satisfactory category, representing 9.64% of the total samples. Meanwhile, the IWQI in seven samples in the study area was characterized as unsuitable for irrigation.

Table 7. IWQI results.

N°	IWQI	Type	N°	IWQI	Type	N°	IWQI	Type	N°	IWQI	Type
GW150	97.64	Excellent	GW117	88.79	Excellent	GW80	82.38	V. Good	GW82	72.50	V. Good
GW44	96.89	Excellent	GW1	88.71	Excellent	GW48	82.32	V. Good	GW64	72.44	V. Good
GW43	96.52	Excellent	GW111	88.51	Excellent	GW163	82.20	V. Good	GW158	72.38	V. Good
GW34	95.71	Excellent	GW66	88.41	Excellent	GW141	82.19	V. Good	GW160	72.25	V. Good
GW112	95.17	Excellent	GW115	88.37	Excellent	GW79	81.71	V. Good	GW30	71.82	V. Good
GW68	95.17	Excellent	GW69	88.27	Excellent	GW94	81.64	V. Good	GW15	70.34	V. Good
GW45	94.91	Excellent	GW56	88.26	Excellent	GW20	81.36	V. Good	GW92	69.88	Good
GW107	94.47	Excellent	GW166	88.17	Excellent	GW124	81.17	V. Good	GW125	68.79	Good
GW106	93.91	Excellent	GW8	88.14	Excellent	GW104	80.74	V. Good	GW138	68.65	Good
GW120	93.65	Excellent	GW129	87.92	Excellent	GW72	80.67	V. Good	GW122	68.44	Good
GW137	93.25	Excellent	GW140	87.91	Excellent	GW131	80.24	V. Good	GW123	67.92	Good
GW75	93.20	Excellent	GW7	87.79	Excellent	GW78	80.07	V. Good	GW110	67.43	Good
GW76	92.84	Excellent	GW4	87.55	Excellent	GW97	79.86	V. Good	GW148	67.17	Good
GW113	92.83	Excellent	GW32	87.19	Excellent	GW86	79.72	V. Good	GW109	66.62	Good
GW2	92.78	Excellent	GW144	87.09	Excellent	GW128	79.67	V. Good	GW52	66.34	Good
GW105	92.70	Excellent	GW132	87.06	Excellent	GW70	79.58	V. Good	GW83	65.87	Good
GW74	92.52	Excellent	GW19	86.83	Excellent	GW17	79.42	V. Good	GW126	65.54	Good
GW90	91.78	Excellent	GW10	86.65	Excellent	GW5	79.16	V. Good	GW159	64.22	Satisfactory
GW73	91.78	Excellent	GW33	86.53	Excellent	GW84	78.48	V. Good	GW46	62.59	Satisfactory
GW11	91.65	Excellent	GW157	86.48	Excellent	GW53	78.03	V. Good	GW55	61.02	Satisfactory
GW114	91.41	Excellent	GW127	86.47	Excellent	GW162	78.03	V. Good	GW61	60.10	Satisfactory
GW67	91.18	Excellent	GW65	86.35	Excellent	GW28	77.57	V. Good	GW77	59.94	Satisfactory
GW154	90.47	Excellent	GW142	86.21	Excellent	GW24	77.42	V. Good	GW47	58.52	Satisfactory
GW152	90.47	Excellent	GW121	86.06	Excellent	GW27	77.31	V. Good	GW146	58.42	Satisfactory
GW89	90.26	Excellent	GW63	85.70	Excellent	GW59	76.84	V. Good	GW41	57.40	Satisfactory
GW151	90.19	Excellent	GW139	85.69	Excellent	GW81	76.40	V. Good	GW57	55.52	Satisfactory
GW134	90.07	Excellent	GW155	85.55	Excellent	GW161	76.11	V. Good	GW98	55.31	Satisfactory
GW119	90.07	Excellent	GW143	85.30	Excellent	GW23	75.82	V. Good	GW40	52.76	Satisfactory
GW87	89.84	Excellent	GW156	85.18	Excellent	GW12	75.62	V. Good	GW26	51.54	Satisfactory
GW133	89.76	Excellent	GW38	85.16	Excellent	GW85	75.52	V. Good	GW29	45.46	Satisfactory
GW165	89.69	Excellent	GW96	85.15	Excellent	GW14	75.34	V. Good	GW145	43.94	Satisfactory
GW130	89.45	Excellent	GW102	85.14	Excellent	GW21	75.23	V. Good	GW62	43.70	Satisfactory
GW93	89.30	Excellent	GW136	85.03	Excellent	GW35	75.05	V. Good	GW99	42.36	Satisfactory
GW149	89.20	Excellent	GW60	84.80	V. Good	GW36	74.66	V. Good	GW91	36.76	Unsuitable
GW164	89.20	Excellent	GW16	84.43	V. Good	GW118	74.29	V. Good	GW3	33.83	Unsuitable
GW54	89.14	Excellent	GW71	84.25	V. Good	GW6	73.79	V. Good	GW42	29.98	Unsuitable
GW116	88.98	Excellent	GW147	84.23	V. Good	GW22	73.76	V. Good	GW25	28.23	Unsuitable
GW37	88.96	Excellent	GW103	84.03	V. Good	GW101	73.56	V. Good	GW39	27.78	Unsuitable
GW108	88.96	Excellent	GW58	83.67	V. Good	GW100	73.37	V. Good	GW18	26.76	Unsuitable
GW13	88.95	Excellent	GW50	83.61	V. Good	GW31	73.30	V. Good	GW135	1.81	Unsuitable
GW51	88.87	Excellent	GW49	82.97	V. Good	GW9	73.29	V. Good			
GW153	88.80	Excellent	GW95	82.89	V. Good	GW88	72.74	V. Good			

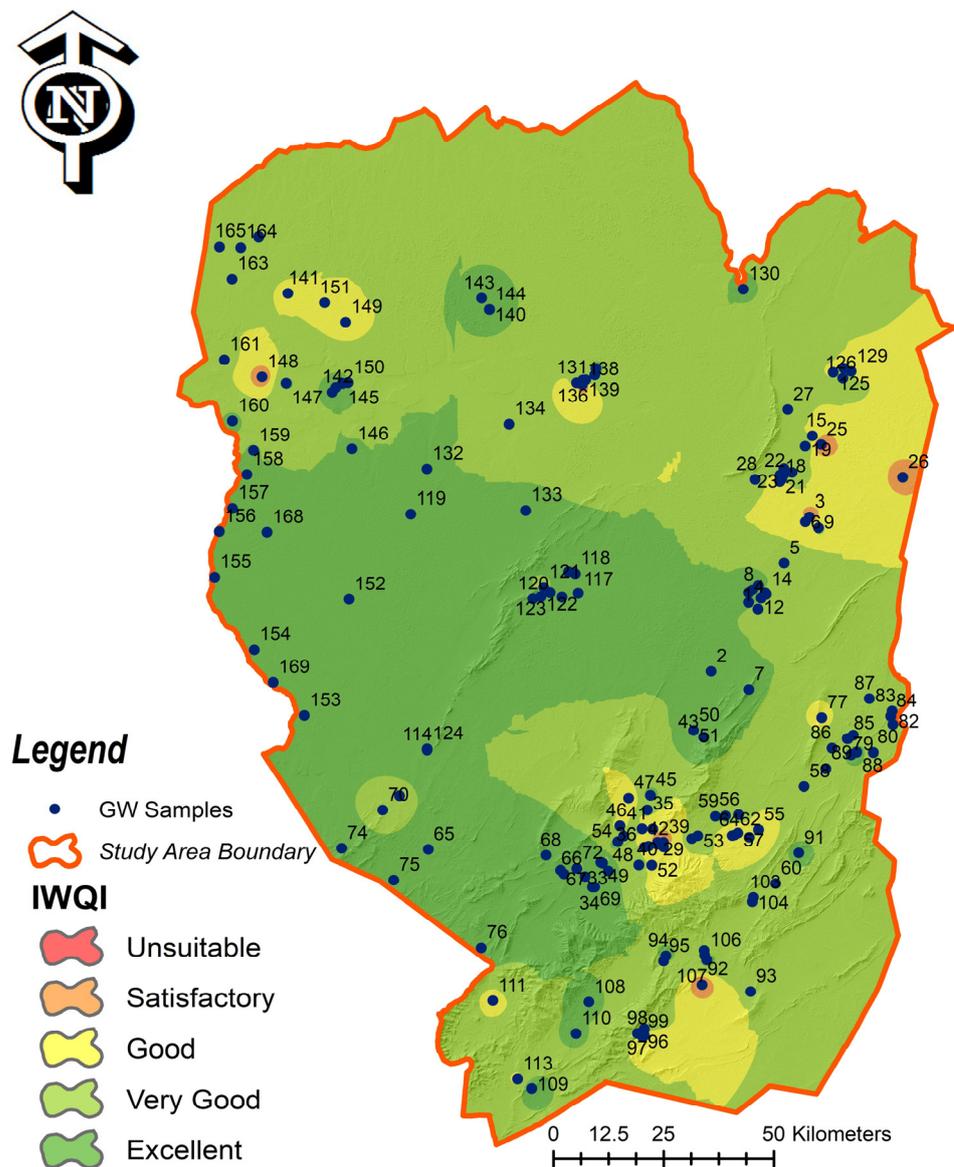


Figure 4. Geospatial distribution of IWQI of the study area.

3.3. Machine Learning Analysis and Modelling

A total of 166 data samples of five quality parameters and their actual outputs were used to train three machine learning models: XGBoost, SVR, and KNN algorithms. The input parameters are Ca^{2+} , Mg^{2+} , Na^+ , K^+ , Cl^- , SO_4^{2-} , HCO_3^- , NO_3^- , EC, Mineralisation, and PH, and IWQI is the output response. The 166 data samples were categorized into five categories expressing irrigation water quality: excellent, very good, good, satisfactory, and unsuitable. Table 8 shows the IWQI status distribution.

Table 8. Distribution of the IWQI data samples.

	Excellent	Very Good	Good	Satisfactory	Unsuitable	Total
No. of data samples	75	57	11	16	7	166

Before performing the analysis, eleven input combinations (models) were proposed and assessed based on five performance criteria.

- Model 1: Ca^{2+} ;

- Model 2: Ca^{2+} , and Mg^{2+} ;
- Model 3: Ca^{2+} , Mg^{2+} , and Na^+ ;
- Model 4: Ca^{2+} , Mg^{2+} , Na^+ , and K^+ ;
- Model 5: Ca^{2+} , Mg^{2+} , Na^+ , K^+ , and Cl^- ;
- Model 6: Ca^{2+} , Mg^{2+} , Na^+ , K^+ , Cl^- , and SO_4^{2-} ;
- Model 7: Ca^{2+} , Mg^{2+} , Na^+ , K^+ , Cl^- , SO_4^{2-} , and HCO_3^- ;
- Model 8: Ca^{2+} , Mg^{2+} , Na^+ , K^+ , Cl^- , SO_4^{2-} , HCO_3^- , and NO_3^- ;
- Model 9: Ca^{2+} , Mg^{2+} , Na^+ , K^+ , Cl^- , SO_4^{2-} , HCO_3^- , NO_3^- , and EC;
- Model 10: Ca^{2+} , Mg^{2+} , Na^+ , K^+ , Cl^- , SO_4^{2-} , HCO_3^- , NO_3^- , EC, and Mineralisation;
- Model 11: Ca^{2+} , Mg^{2+} , Na^+ , K^+ , Cl^- , SO_4^{2-} , HCO_3^- , NO_3^- , EC, Mineralisation, and PH.

The models XGBoost, SVR, and KNN have been implanted using Python programming, which iteratively performs training and testing with the specified applied data. Figure 5 shows the performance of the XGBoost, SVR, and KNN models. The RMSE decreased after the second epochs for SVR and KNN, reaching 4.852 and 3.745 for training and 3.595 and 2.692 for testing, respectively, where the RMSE of XGBoost continued to decrease for both training and validation to epochs or an iteration number of 45, giving its minimum 0.00089029 in training and 2.8272 in testing. Therefore, the best network performance can be considered when the validation error is the lowest.

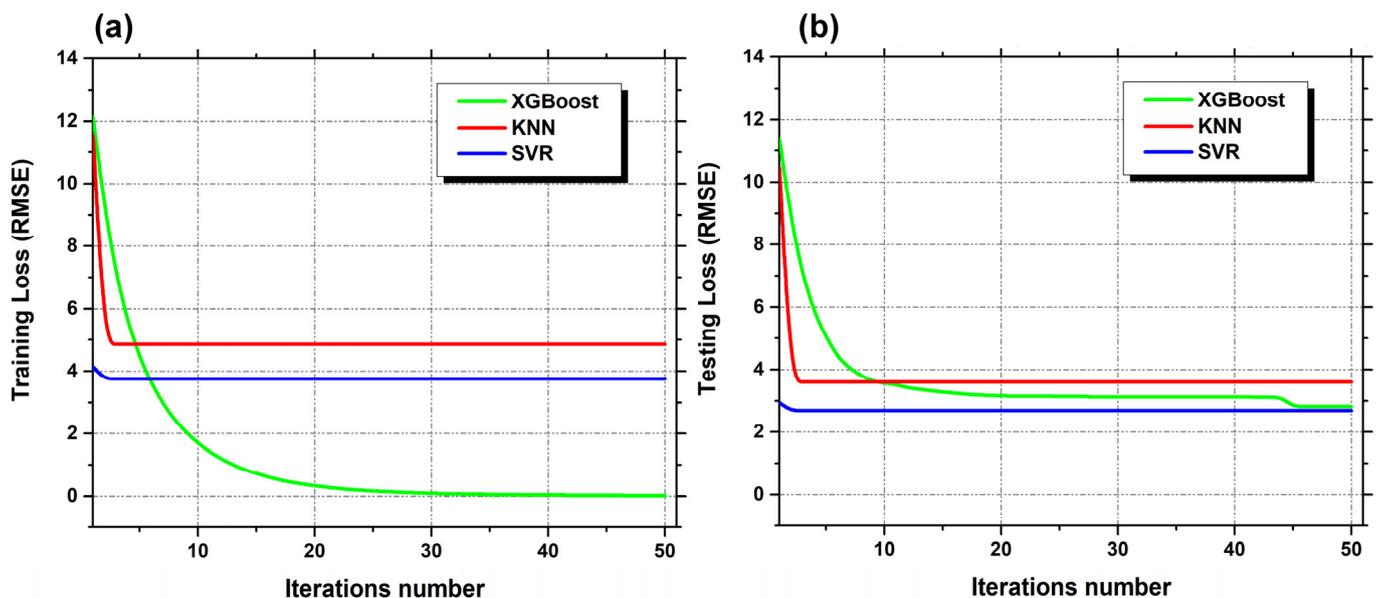


Figure 5. (a) Training and (b) validation loss based on RMSE performance of the XGBoost, SVR, and KNN.

The training, testing, and validation curves were reduced, as in Figure 5, and the errors decreased, indicating that the XGBoost, SVR, and KNN models are reliable. It is important to know how the three models identify the quality parameters and the output response relationship (IWQI), as well as what the accuracy of the prediction model to obtain the correct prediction of IWQI with the variation of the quality inputs is.

This section presents the results of the analytical investigation. In initial assessments, we measured the effectiveness of basic models using RMSE and NSE as the evaluation metrics. Figure 6 illustrates the comparative RMSE and NSE values for the XGBoost, SVR, and KNN models during the testing phase. Among the 11 models analysed, several models stand out as top performers. Model 7 of XGBoost exhibits the best predictive performance with an impressively low RMSE of 2.99 and the highest NSE of 0.954, demonstrating exceptional accuracy and goodness of fit. Additionally, XGBoost models 9 (Ca^{2+} , Mg^{2+} ,

Na⁺, K⁺, Cl⁻, SO₄²⁻, HCO₃⁻, NO₃⁻, and EC) and 10 (Ca²⁺, Mg²⁺, Na⁺, K⁺, Cl⁻, SO₄²⁻, HCO₃⁻, NO₃⁻, and EC) also showcase strong performance, boasting low RMSE values of 2.96 and 2.83, along with high NSE values of 0.953 and 0.957, respectively. Model 4 (Ca²⁺, Mg²⁺, Na⁺, and K⁺) of SVR is another standout, featuring an RMSE of 2.69 and an exceptional NSE of 0.961, making it one of the top-performing models in the SVR category. In contrast, the SVR model 1 ranks as the least accurate among the models, with an RMSE of 10.52 and an NSE of 0.483, indicating moderate predictive performance. Similarly, SVR models 9 and 10 also exhibit lower predictive performance with RMSE values of 4.42 and 4.39, accompanied by NSE values of 0.911 and 0.912, respectively. When selecting a model, it is crucial to prioritize predictive accuracy, and in this context, the XGBoost models, particularly model 7 and SVR model 4, emerge as the top choices, while SVR model 1 may require further improvement to achieve better results. When selecting a model, it is crucial to prioritize predictive accuracy, and in this context, XGBoost model 10, KNN model 5 (Ca²⁺, Mg²⁺, Na⁺, K⁺, and Cl⁻), and other high-performing models are excellent choices, while other models may require further improvement to achieve better results.

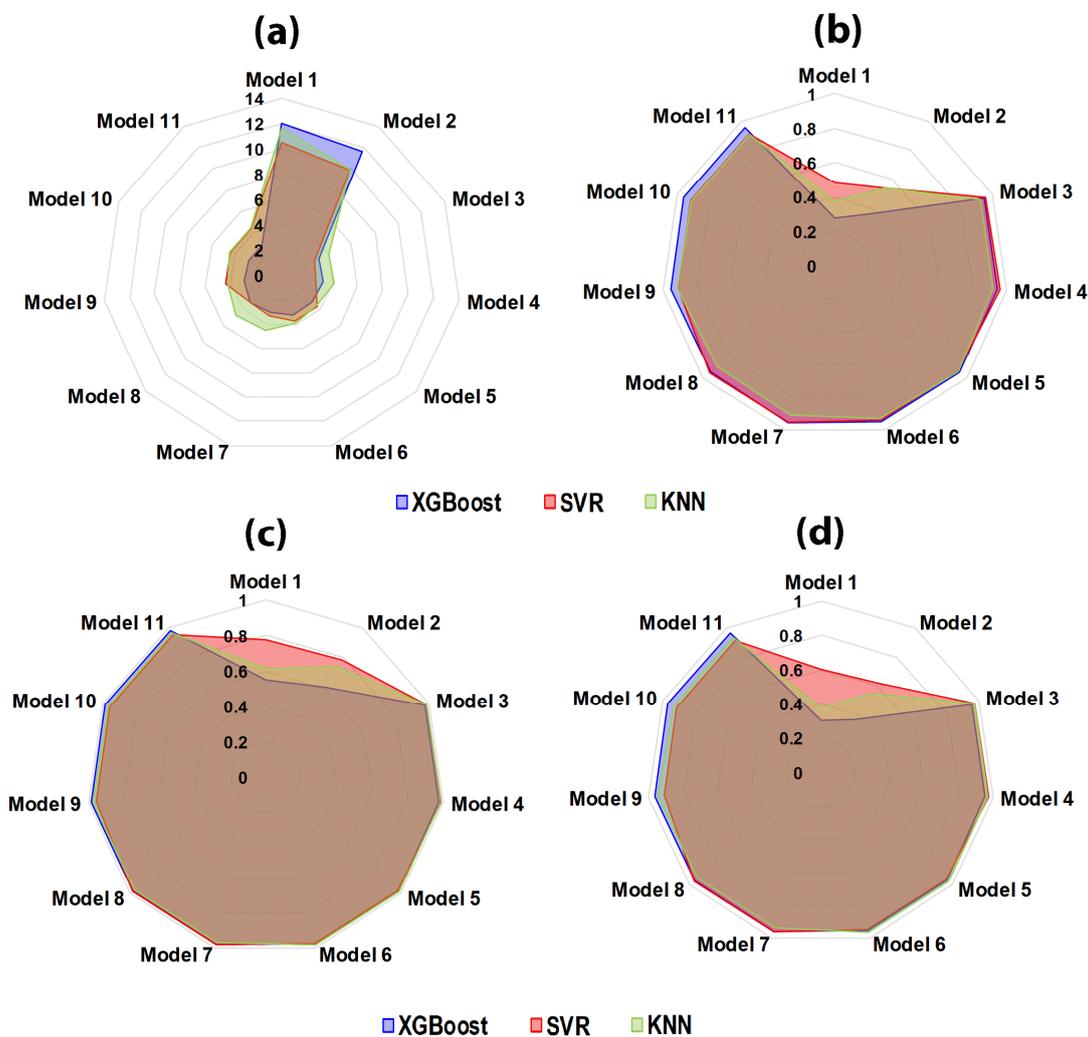


Figure 6. (a) RMSE,(b) NSE, (c) R, and (d) R² performance for the 11 models using XGBoost, SVR, and KNN.

In this comprehensive analysis, we have assessed the performance of 11 models across three machine learning algorithms, XGBoost, SVR, and KNN, based on R and R², to gauge their predictive accuracy, goodness of fit, and overall performance (Figure 6).

XGBoost model 10: Among the XGBoost models, model 10 emerges as the top choice. It not only demonstrated the lowest RMSE but also excelled in other metrics, including NSE, R, and R2. With an R2 value of 0.971, this model showcases exceptional predictive capabilities. Its consistently high rankings across multiple metrics underscore its robustness, making it an excellent candidate for accurate predictions. KNN model 3 and KNN model 4: The KNN models, represented by models 3 and 4, also stand out as top-performing models. Both models exhibited R2 values of 0.971 and 0.970, indicating strong predictive accuracy and goodness of fit. These models consistently outperformed other KNN models and demonstrated competitive performance across various metrics. Within the SVR models, model 4 displays noteworthy performance. With an R2 value of 0.975, it exhibits a high level of predictive accuracy and goodness of fit. This model is a strong contender among the SVR models, showcasing its ability to provide reliable predictions. In conclusion, the choice of the best model depends on the specific project requirements. XGBoost model 10 and SVR model 4 are top contenders for tasks demanding the utmost accuracy and goodness of fit. KNN models 3 and 4 offer competitive performance and may be preferred for their simplicity and interpretability.

Figure 7 provides a comprehensive comparative analysis of the best-performing models, specifically XGBoost model 10, SVR model 4, and KNN model 5, based on essential evaluation metrics. This figure serves as a visual representation of the model's predictive capabilities and their suitability for different application scenarios.

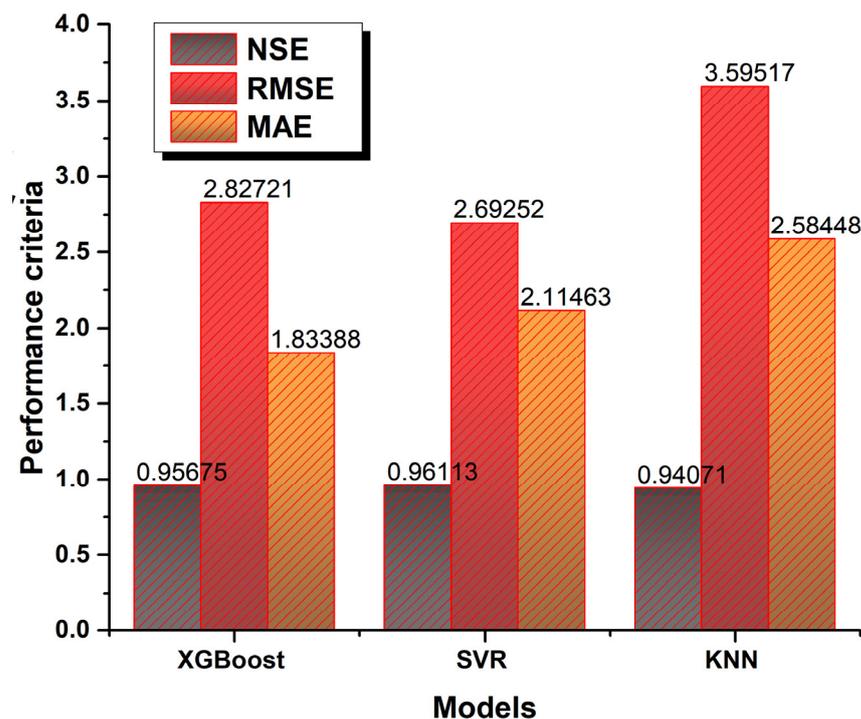


Figure 7. RMSE, NSE, and MAE performance for the best models XGBoost (model 10), SVR (model 4), and KNN (model 5).

In the graph, we observe that XGBoost model 10 stands out with the highest Nash–Sutcliffe Efficiency (NSE) of 0.957, emphasizing its superior predictive accuracy. This model also demonstrates the lowest Root Mean Square Error (RMSE) of 2.827 and an MAE of 1.834, indicating its ability to minimize prediction errors effectively. SVR model 4 maintains strong performance, with an NSE of 0.961 and competitive error measures, including an RMSE of 2.693 and an MAE of 2.115. This model strikes a balance between accuracy and goodness of fit. KNN model 5 highlights robust predictive capabilities with an NSE of 0.941, although it incurs a slightly higher RMSE of 3.595 and an MAE of 2.584. It offers a reliable option with a focus on simplicity.

Figure 8 indicates the correlation between the actual and predicted IWQI output using (a) SVR (model 4), (b) XGBoost (model 10), and (c) KNN (model 5). When comparing the input data of each model within the context of making predictions with limited data, several considerations come into play. The ability to generate accurate predictions with limited data is crucial, and it often depends on the complexity and adaptability of the model. Here is a comparison of the input data for each of the best models, considering their suitability for such scenarios:

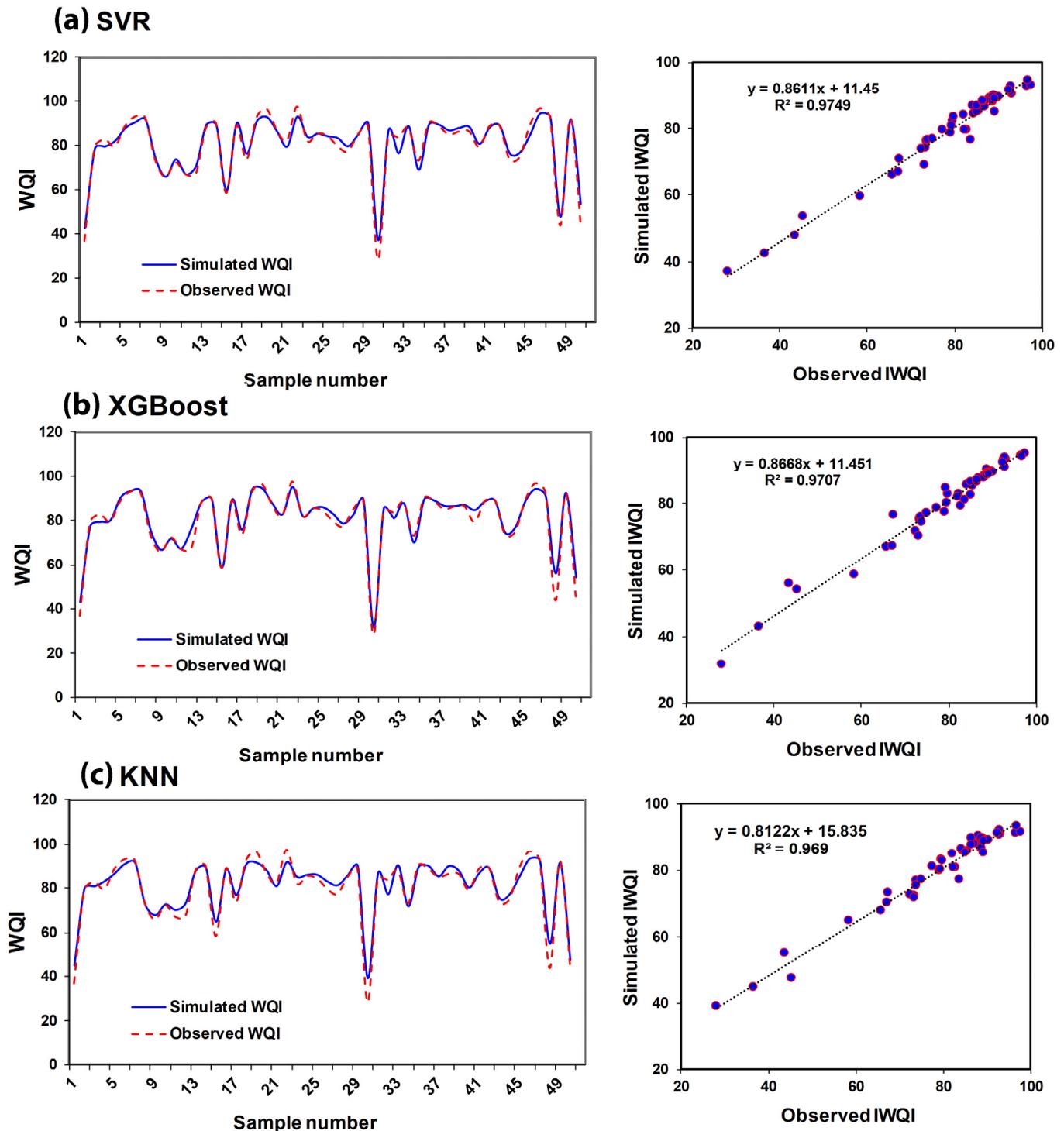


Figure 8. Correlation between the actual and predicted IWQI output using (a) SVR, (b) XGBoost, and (c) KNN.

When addressing the challenge of making predictions with limited data, the selection of the right modelling approach and input variables becomes pivotal. In this regard, the three best models, XGBoost model 10, SVR model 4, and KNN model 5, each offer unique advantages. XGBoost, represented by model 10, is renowned for its adaptability and robust predictive capabilities. With a comprehensive set of input variables, it can effectively handle sparse datasets. However, success hinges on careful feature selection and parameter tuning to avoid overfitting. In contrast, SVR model 4, with its focused input variables, provides a simpler yet robust solution. Its reduced risk of over-parameterization makes it well-suited for limited data scenarios. KNN model 5, known for its simplicity and adaptability, relies on nearest neighbours and can perform effectively even when data are scarce.

Figure 9 compares the statistical parameters of observed and prediction values based on mean, minimum, maximum, variance, and standard deviation (STD). Mean: All three models have means that are very close to or slightly higher than the reference mean, indicating that they capture the central tendency of the data well. For instance, the mean of XGBoost’s predictions (80.80293) is just slightly higher than the reference (80.01060). Min and Max: The minimum and maximum values of the predicted data by each model are generally within or near the range of the observed data. For example, KNN’s minimum value (39.16076) and maximum value (83.10271) are well within the observed data range. The variance of the predicted data by all three models is lower than the variance of the observed data, indicating that the models provide predictions with reduced variability. For instance, XGBoost and SVR have lower variances compared to the reference data. The STD of the predicted data by each model is also lower than the standard deviation of the observed data, indicating that the models offer predictions with less spread and are less variable than the observed data. XGBoost and SVR have lower standard deviations compared to the reference data.

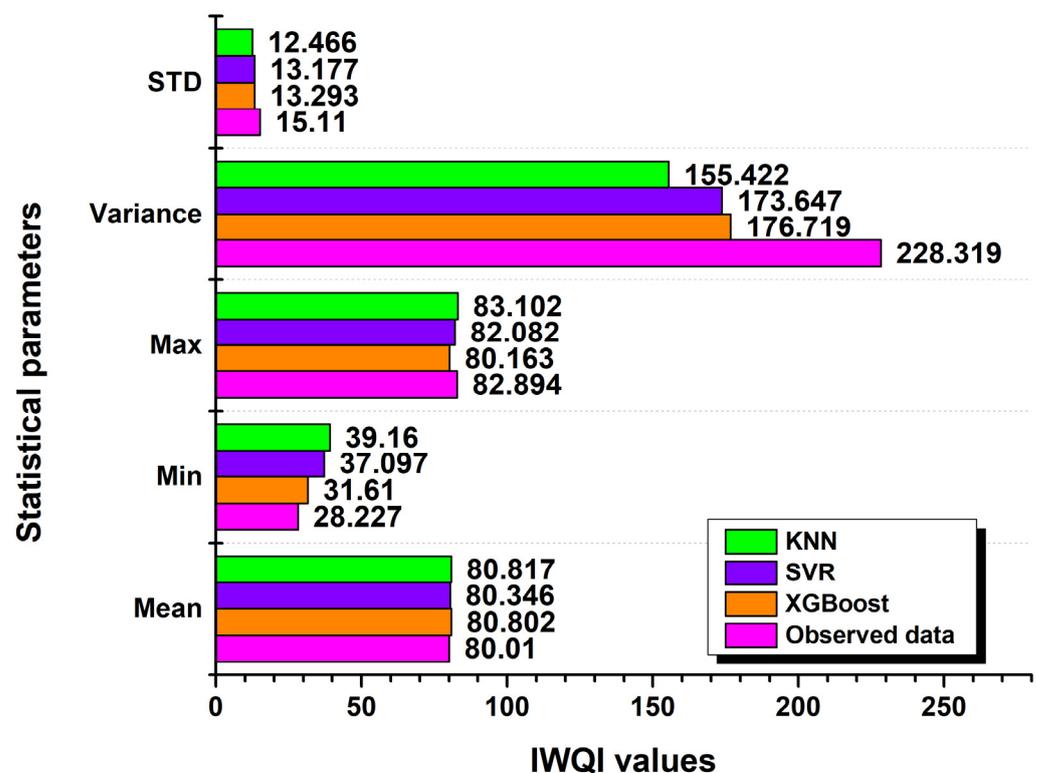


Figure 9. Statistical comparison of observed and predicted values.

4. Discussion

In the discussion section, we have comprehensively evaluated the performance of three predictive models, XGBoost (model 10), SVR (model 4), and KNN (model 5), through

a range of performance metrics and statistical parameters. XGBoost exhibited strong predictive accuracy with high NSE, low RMSE, and MAE values, closely aligning with the reference data's central tendency and demonstrating stability with lower variance and standard deviation. SVR showcased notable predictive capability, maintaining a high NSE and low RMSE (2.692) and MAE (2.1146) values, in addition to closely matching the reference data's central tendency and offering stable and consistent predictions. KNN, while having slightly lower R and R² values, presented strong predictive performance with a closely aligned central tendency and stable, less variable predictions supported by a lower variance and standard deviation. These findings collectively highlight the models' potential for accurate, stable, and consistent predictions across various applications. The choice of the best model should be influenced by specific application requirements and priorities, taking into account predictive accuracy, model complexity, and stability. The results obtained from this study agree with research studies that have applied similar approaches for IWQI, which indicated the high performance and stability of machine learning models for IWQI prediction [35,69–71]. Lap et al. [72] indicated that the random forest (RF) model excels in accurately forecasting WQI values for the An Hai irrigation system in Vietnam, achieving a good Similarity score of 0.94. This analysis identifies four crucial parameters—Coliform, Dissolved Oxygen (DO), Turbidity, and Total Suspended Solids (TSS)—that exert the most significant influence on water quality. In El Kharga Oasis in the Western Desert of Egypt, Ibrahim et al. [73] found that both the ANFIS and SVM models demonstrated the capability to accurately simulate IWQIs, as evidenced by high determination coefficients (R²) in both the training phase (R² = 0.99 and 0.97) and the testing phase (R² = 0.97 and 0.76). In a study by Nguyen et al. [74] for WQI calculations in the Red River Delta, Vietnam, two types of machine learning models were employed. The results revealed that the machine learning model outperformed the deep learning model in terms of prediction accuracy, where the gradient boosting model demonstrated the highest level of predictive accuracy, followed by the XGBoost, RNN, and LSTM models. The accuracy of each of these models was notably high, with predictions ranging from 84% to 96%. Trabelsi and Bel Hadj Ali [75] applied RF, SVR, ANN, and AdaBoost for predicting IWQI in the downstream Medjerda river basin in Tunisia. The findings indicate that the AdaBoost model stands out as the most suitable choice for predicting all parameters, with correlation coefficients (*r*) ranging between 0.88 and 0.89. On the other hand, the random forest model is well-suited for predicting four specific parameters, namely TDS, SAR, PS, and ESP, with R in the range of 0.65 to 0.87.

Finally, despite the uncertainty and limitations of machine learning algorithms, these results highlight the potential of XGBoost, SVR, and KNN as valuable tools for ground-water quality prediction. They can provide essential insights and serve as a basis for further research and monitoring. However, their utility should be complemented by expert knowledge and traditional hydrogeological methods for more robust decision-making and practical applications. Finally, this section emphasizes the need for further research to validate the models under different data conditions, including more dispersed groundwater quality data. We also highlight potential avenues for future research aimed at refining the models in response to varying data characteristics.

5. Conclusions

Arid and semi-arid regions often rely solely on groundwater for irrigation. The management of water resources for drinking and irrigation can be enhanced by understanding and evaluating the irrigation water quality index. Based on data collected from 166 boreholes in Naama, located in southwestern Algeria, the study aims to determine the irrigation water quality index (IWQI), which consists of many physico-chemical parameters. Furthermore, this research rigorously evaluated three predictive models, namely XGBoost, SVR, and KNN, for estimating the IWQI variable. The models were thoroughly assessed using multiple performance metrics, including NSE, RMSE, MAE, R,

The results of the irrigation water qualitative parameter analysis of the groundwater samples of the study area revealed that most of them were “suitable.” Based on the findings of IWQI, we found that 45.18% of samples were categorized as “excellent”, 34.34% of samples were considered as “very good”, 6.63% of samples fell into the good category, 9.64% of the total samples were categorized as “satisfactory”, and 4.21% of samples in the study area were characterized as “unsuitable” of irrigation.

In IWQI modelling, XGBoost (model 10) emerged as a strong performer, with high NSE and low RMSE and MAE values, signifying its predictive accuracy. It is closely aligned with the reference data in terms of mean, minimum, and maximum predictions while offering reduced variability.

SVR (model 4) demonstrated a notable predictive capability, boasting high NSE (0.96112) and low RMSE (2.6925) and MAE (2.11462) values. It closely matched the reference data’s mean and exhibited consistent predictions within the observed data range. Lower variance and standard deviation values emphasized its stability. We found four important parameters that have the greatest impact on water quality, including Ca^{2+} , Mg^{2+} , Na^+ , and K.

KNN (model 5) showcased strong predictive performance with competitive NSE, RMSE, and MAE values. Although it had slightly lower R and R2 values, its predictions closely followed the reference data, with reduced variance and standard deviation, indicating stability.

This study may offer useful and valuable information for decision-makers to comprehend the present state of the water quality for irrigation in the Wilaya of Naama. This will allow for a better and more sustainable management of water resources in the study area and similar regions. Finally, the application of metaheuristic algorithms in conjunction with machine learning is a promising avenue for future research and practical implementations.

Author Contributions: Conceptualization, E.E.H., A.D., B.Z., A.A., M.B.-d.I.S., Y.J.W. and M.A.H.; data curation, E.E.H. and A.D.; formal analysis, E.E.H., A.D., B.Z., M.B.-d.I.S. and A.A.; funding acquisition, A.A.; investigation, E.E.H., A.D., A.A. and M.A.H.; methodology, E.E.H., A.D. and M.A.H.; project administration, E.E.H. and A.D.; resources, E.E.H., A.D., A.E., and A.A.; software, E.E.H., A.D. and B.Z.; supervision, E.E.H., P.M.N., and A.D.; validation, A.E., E.E.H., Y.J.W., A.D., A.A., M.B.-d.I.S. and M.A.H.; visualization, A.A. and P.M.N., A.E., M.A.H.; writing—original draft, E.E.H., A.D., Y.J.W., A.E., and B.Z.; writing—review and editing, E.E.H., A.A., M.B.-d.I.S. and M.A.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work is funded and supported by the Deanship of Scientific Research, Taif University.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author, [Y.J.W.].

Acknowledgments: The researchers would like to acknowledge the Deanship of Scientific Research, Taif University, for funding this work.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Gleeson, T.; Alley, W.M.; Allen, D.M.; Sophocleous, M.A.; Zhou, Y.; Taniguchi, M.; VanderSteen, J. Towards sustainable groundwater use: Setting long-term goals, backcasting, and managing adaptively. *Groundwater* **2012**, *50*, 19–26. [[CrossRef](#)]
- Laube, W.; Schraven, B.; Awo, M. Smallholder adaptation to climate change: Dynamics and limits in Northern Ghana. *Clim. Chang.* **2012**, *111*, 753–774. [[CrossRef](#)]
- Maja, M.M.; Ayano, S.F. The impact of population growth on natural resources and farmers’ capacity to adapt to climate change in low-income countries. *Earth Syst. Environ.* **2021**, *5*, 271–283. [[CrossRef](#)]
- Xanke, J.; Liesch, T. Quantification and possible causes of declining groundwater resources in the Euro-Mediterranean region from 2003 to 2020. *Hydrogeol. J.* **2022**, *30*, 379–400. [[CrossRef](#)]
- Li, Q.; Lu, L.; Zhao, Q.; Hu, S. Impact of Inorganic Solutes’ Release in Groundwater during Oil Shale In Situ Exploitation. *Water* **2023**, *15*, 172. [[CrossRef](#)]
- Molajou, A.; Afshar, A.; Khosravi, M.; Soleimani, E.; Vahabzadeh, M.; Variani, H.A. A new paradigm of water, food, and energy nexus. *Environ. Sci. Pollut. Res.* **2023**, *30*, 107487–107497. [[CrossRef](#)]

7. Mekonnen, M.M.; Gerbens-Leenes, W. The water footprint of global food production. *Water* **2020**, *12*, 2696. [CrossRef]
8. Tian, H.; Huang, N.; Niu, Z.; Qin, Y.; Pei, J.; Wang, J. Mapping Winter Crops in China with Multi-Source Satellite Imagery and Phenology-Based Algorithm. *Remote Sens.* **2019**, *11*, 820. [CrossRef]
9. Abobatta, W. Impact of hydrogel polymer in agricultural sector. *Adv. Agric. Environ. Sci. Open Access* **2018**, *1*, 59–64. [CrossRef]
10. Gerten, D.; Heck, V.; Jägermeyr, J.; Bodirsky, B.L.; Fetzer, I.; Jalava, M.; Kummu, M.; Lucht, W.; Rockström, J.; Schaphoff, S. Feeding ten billion people is possible within four terrestrial planetary boundaries. *Nat. Sustain.* **2020**, *3*, 200–208. [CrossRef]
11. Wang, X. Managing Land Carrying Capacity: Key to Achieving Sustainable Production Systems for Food Security. *Land* **2022**, *11*, 484. [CrossRef]
12. Abdessamed, D.; Jodar-Abellan, A.; Ghoneim, S.S.M.; Almaliki, A.; Hussein, E.E.; Pardo, M.Á. Groundwater quality assessment for sustainable human consumption in arid areas based on GIS and water quality index in the watershed of Ain Sefra (SW of Algeria). *Environ. Earth Sci.* **2023**, *82*, 510. [CrossRef]
13. FAO (Food and Agriculture Organization). AQUASTAT—FAO’s Global Information System on Water and Agriculture. Available online: <https://www.fao.org/aquastat/en/geospatial-information/global-maps-irrigated-areas/irrigation-by-country/country/DZA> (accessed on 16 October 2023).
14. Amichi, H.; Bouarfa, S.; Kuper, M.; Ducourtieux, O.; Imache, A.; Fusillier, J.L.; Bazin, G.; Hartani, T.; Chehat, F. How does unequal access to groundwater contribute to marginalization of small farmers? The case of public lands in Algeria. *Irrig. Drain.* **2012**, *61*, 34–44. [CrossRef]
15. Hounslow, A.W. *Water Quality Data: Analysis and Interpretation*; CRC Press: Boca Raton, FL, USA, 2018.
16. Zaman, M.; Shahid, S.A.; Heng, L. Irrigation water quality. In *Guideline for Salinity Assessment, Mitigation and Adaptation Using Nuclear and Related Techniques*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 113–131.
17. Turdaliev, A.; Darmonov, D.Y.; Teshaboyev, N.; Saminov, A.; Abdurakhmonova, M. Influence of irrigation with salty water on the composition of absorbed bases of hydromorphic structure of soil. *IOP Conf. Ser. Earth Environ. Sci.* **2022**, *1068*, 012047. [CrossRef]
18. Tlili-Zrelli, B.; Hamzaoui-Azaza, F.; Gueddari, M.; Bouhlila, R. Geochemistry and quality assessment of groundwater using graphical and multivariate statistical methods. A case study: Grombalia phreatic aquifer (Northeastern Tunisia). *Arab. J. Geosci.* **2013**, *6*, 3545–3561. [CrossRef]
19. Nishanthiny, S.C.; Thushyanthy, M.; Barathithasan, T.; Saravanan, S. Irrigation water quality based on hydro chemical analysis, Jaffna, Sri Lanka. *Am.-Eurasian J. Agric. Environ. Sci.* **2010**, *7*, 100–102.
20. Cymes, I.; Glińska-Lewczuk, K. The use of water quality indices (WQI and SAR) for multipurpose assessment of water in dam reservoirs. *J. Elem.* **2016**, *21*, 1211–1224.
21. Chaganti, V.N.; Crohn, D.M.; Šimůnek, J. Leaching and reclamation of a biochar and compost amended saline–sodic soil with moderate SAR reclaimed water. *Agric. Water Manag.* **2015**, *158*, 255–265. [CrossRef]
22. Rengasamy, P. Irrigation water quality and soil structural stability: A perspective with some new insights. *Agronomy* **2018**, *8*, 72. [CrossRef]
23. Misaghi, F.; Delgosha, F.; Razzaghmanesh, M.; Myers, B. Introducing a water quality index for assessing water for irrigation purposes: A case study of the Ghezel Ozan River. *Sci. Total Environ.* **2017**, *589*, 107–116. [CrossRef] [PubMed]
24. Koklu, R.; Sengorur, B.; Topal, B. Water quality assessment using multivariate statistical methods—A case study: Melen River System (Turkey). *Water Resour. Manag.* **2010**, *24*, 959–978. [CrossRef]
25. Zhang, B.; Song, X.; Zhang, Y.; Han, D.; Tang, C.; Yu, Y.; Ma, Y. Hydrochemical characteristics and water quality assessment of surface water and groundwater in Songnen plain, Northeast China. *Water Res.* **2012**, *46*, 2737–2748. [CrossRef] [PubMed]
26. Prasanna, M.; Praveena, S.; Chidambaram, S.; Nagarajan, R.; Elayaraja, A. Evaluation of water quality pollution indices for heavy metal contamination monitoring: A case study from Curtin Lake, Miri City, East Malaysia. *Environ. Earth Sci.* **2012**, *67*, 1987–2001. [CrossRef]
27. Barakat, A.; El Baghdadi, M.; Rais, J.; Aghezzaf, B.; Slassi, M. Assessment of spatial and seasonal water quality variation of Oum Er Rbia River (Morocco) using multivariate statistical techniques. *Int. Soil Water Conserv. Res.* **2016**, *4*, 284–292. [CrossRef]
28. Eaton, F.M. Significance of carbonates in irrigation waters. *Soil Sci.* **1950**, *69*, 123–134. [CrossRef]
29. Doneen, L. *Notes on Water Quality in Agriculture*; Published as a Water Sciences and Engineering; Department of Water Sciences and Engineering, University of California: Davis, CA, USA, 1964; Volume 4001.
30. Brown, R.M.; McClelland, N.I.; Deininger, R.A.; O’Connor, M.F. A water quality index—Crashing the psychological barrier. In *Indicators of Environmental Quality*; Springer: Berlin/Heidelberg, Germany, 1972; pp. 173–182.
31. Horton, R.K. An index number system for rating water quality. *J. Water Pollut. Control. Fed.* **1965**, *37*, 300–306.
32. Meireles, A.C.M.; Andrade, E.M.d.; Chaves, L.C.G.; Frischkorn, H.; Crisostomo, L.A. A new proposal of the classification of irrigation water. *Rev. Ciência Agronômica* **2010**, *41*, 349–357. [CrossRef]
33. Şener, Ş.; Varol, S.; Şener, E. Evaluation of sustainable groundwater utilization using index methods (WQI and IWQI), multivariate analysis, and GIS: The case of Akşehir District (Konya/Turkey). *Environ. Sci. Pollut. Res.* **2021**, *28*, 47991–48010. [CrossRef]
34. Batarseh, M.; Imreizeeq, E.; Tilev, S.; Al Alaween, M.; Suleiman, W.; Al Remeithi, A.M.; Al Tamimi, M.K.; Al Alawneh, M. Assessment of groundwater quality for irrigation in the arid regions using irrigation water quality index (IWQI) and GIS-Zoning maps: Case study from Abu Dhabi Emirate, UAE. *Groundw. Sustain. Dev.* **2021**, *14*, 100611. [CrossRef]
35. El Bilali, A.; Taleb, A. Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment. *J. Saudi Soc. Agric. Sci.* **2020**, *19*, 439–451. [CrossRef]

36. Moussaoui, T.; Derdour, A.; Hosni, A.; Ballesta-de los Santos, M.; Legua, P.; Pardo-Picazo, M.Á. Assessing the Quality of Treated Wastewater for Irrigation: A Case Study of Ain Sefra Wastewater Treatment Plant. *Sustainability* **2023**, *15*, 11133. [[CrossRef](#)]
37. Abdessamed, D.; Abderrazak, B. Coupling HEC-RAS and HEC-HMS in rainfall–runoff modeling and evaluating floodplain inundation maps in arid environments: Case study of Ain Sefra city, Ksour Mountain. SW of Algeria. *Environ. Earth Sci.* **2019**, *78*, 586. [[CrossRef](#)]
38. Derdour, A.; Abdo, H.G.; Almohamad, H.; Alodah, A.; Al Dughairi, A.A.; Ghoneim, S.S.; Ali, E. Prediction of Groundwater Water Quality Index Using Classification Techniques in Arid Environments. *Sustainability* **2023**, *15*, 9687. [[CrossRef](#)]
39. Derdour, A.; Bouarfa, S.; Kaid, N.; Baili, J.; Al-Bahrani, M.; Menni, Y.; Ahmad, H. Assessment of the impacts of climate change on drought in an arid area using drought indices and Landsat remote sensing data. *Int. J. Low-Carbon Technol.* **2022**, *17*, 1459–1469. [[CrossRef](#)]
40. Bouarfa, S.; Derdour, A.; Okkacha, Y.; Almaliki, A.H.; Jodar-Abellan, A.; Hussein, E.E. Sedimentological investigation of the potential origin and provenance of sand deposits in an arid area: A case study of the Ksour Mountains Region in Algeria. *Arab. J. Geosci.* **2022**, *15*, 1460. [[CrossRef](#)]
41. Derdour, A.; Bouanani, A.; Babahamed, K. Modelling rainfall runoff relations using HEC-HMS in a semi-arid region: Case study in Ain Sefra watershed, Ksour Mountains (SW Algeria). *J. Water Land Dev.* **2018**, *36*, 45–55. [[CrossRef](#)]
42. Derdour, A.; Benkaddour, Y.; Bendahou, B. Application of remote sensing and GIS to assess groundwater potential in the transboundary watershed of the Chott-El-Gharbi (Algerian–Moroccan border). *Appl. Water Sci.* **2022**, *12*, 136. [[CrossRef](#)]
43. Lachache, S.; Derdour, A.; Maazouzi, I.; Amroune, A.; Guastaldi, E.; Merzougui, T. Statistical Approach Of Groundwater Quality Assessment At Naama Region, South-West Algeria. *LARHYSS J.* **2023**, *55*, 125–145.
44. Khodapanah, L.; Sulaiman, W.; Khodapanah, N. Groundwater quality assessment for different purposes in Eshtehard District, Tehran, Iran. *Eur. J. Sci. Res.* **2009**, *36*, 543–553. [[CrossRef](#)]
45. Richards, L.A. *Diagnosis and Improvement of Saline and Alkali Soils*; LWW: Washington, DC, USA, 1954; Volume 78.
46. Kelly, W. Permissible composition and concentration of irrigated waters. *Proc. ASCF* **1940**, *66*, 607–613.
47. Wong, Y.J.; Shimizu, Y.; He, K.; Nik Sulaiman, N.M. Comparison among different ASEAN water quality indices for the assessment of the spatial variation of surface water quality in the Selangor river basin, Malaysia. *Environ. Monit. Assess* **2020**, *192*, 644. [[CrossRef](#)] [[PubMed](#)]
48. Doneen, L.D. Salination of soil by salts in the irrigation water. *Eos Trans. Am. Geophys. Union* **1954**, *35*, 943–950.
49. Lu, H.; Ma, X. Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere* **2020**, *249*, 126169. [[CrossRef](#)] [[PubMed](#)]
50. Lee, J.; Lee, J.; Lee, M.; Lee, M.; Kim, Y.; Hyung, J.; Kim, K.; Cha, Y.; Koo, J. Development of a short-term water quality prediction model for urban rivers using real-time water quality data. *Water Supply* **2022**, *22*, 4082–4097. [[CrossRef](#)]
51. Wong, Y.J.; Nakayama, R.; Shimizu, Y.; Kamiya, A.; Shen, S.; Muhammad Rashid, I.Z.; Nik Sulaiman, N.M. Toward industrial revolution 4.0: Development, validation, and application of 3D-printed IoT-based water quality monitoring system. *J. Clean. Prod.* **2021**, *324*, 129230. [[CrossRef](#)]
52. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
53. Wong, Y.J.; Shimizu, Y.; Kamiya, A.; Maneechot, L.; Bharambe, K.P.; Fong, C.S.; Nik Sulaiman, N.M. Application of artificial intelligence methods for monsoonal river classification in Selangor river basin, Malaysia. *Environ. Monit. Assess* **2021**, *193*, 438. [[CrossRef](#)] [[PubMed](#)]
54. Najafzadeh, M.; Niazmardi, S. A novel multiple-kernel support vector regression algorithm for estimation of water quality parameters. *Nat. Resour. Res.* **2021**, *30*, 3761–3775. [[CrossRef](#)]
55. Su, X.; He, X.; Zhang, G.; Chen, Y.; Li, K. Research on SVR water quality prediction model based on improved sparrow search algorithm. *Comput. Intell. Neurosci.* **2022**, *2022*, 7327072. [[CrossRef](#)]
56. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1999.
57. Sakaa, B.; Elbeltagi, A.; Boudibi, S.; Chaffai, H.; Islam, A.R.M.T.; Kulimushi, L.C.; Choudhari, P.; Hani, A.; Brouziyne, Y.; Wong, Y.J. Water quality index modeling using random forest and improved SMO algorithm for support vector machine in Saf-Saf river basin. *Environ. Sci. Pollut. Res.* **2022**, *29*, 48491–48508. [[CrossRef](#)]
58. Li, J.; Abdulmohsin, H.A.; Hasan, S.S.; Kaiming, L.; Al-Khateeb, B.; Ghareb, M.I.; Mohammed, M.N. Hybrid soft computing approach for determining water quality indicator: Euphrates River. *Neural Comput. Appl.* **2019**, *31*, 827–837. [[CrossRef](#)]
59. Wang, X.; Zhang, F.; Ding, J. Evaluation of water quality based on a machine learning algorithm and water quality index for the Ebinur Lake Watershed, China. *Sci. Rep.* **2017**, *7*, 12858. [[CrossRef](#)] [[PubMed](#)]
60. Tahraoui, H.; Toumi, S.; Hassein-Bey, A.H.; Bousselma, A.; Sid, A.N.E.H.; Belhadj, A.-E.; Triki, Z.; Kebir, M.; Amrane, A.; Zhang, J. Advancing Water Quality Research: K-Nearest Neighbor Coupled with the Improved Grey Wolf Optimizer Algorithm Model Unveils New Possibilities for Dry Residue Prediction. *Water* **2023**, *15*, 2631. [[CrossRef](#)]
61. Juna, A.; Umer, M.; Sadiq, S.; Karamti, H.; Eshmawi, A.A.; Mohamed, A.; Ashraf, I. Water quality prediction using KNN imputer and multilayer perceptron. *Water* **2022**, *14*, 2592. [[CrossRef](#)]
62. Kim, M.; Kim, Y.; Kim, H.; Piao, W.; Kim, C. Evaluation of the k-nearest neighbor method for forecasting the influent characteristics of wastewater treatment plant. *Front. Environ. Sci. Eng.* **2016**, *10*, 299–310. [[CrossRef](#)]

63. Budiarti, R.P.N.; Sukaridhoto, S.; Hariadi, M.; Purnomo, M.H. Big data technologies using SVM (case study: Surface water classification on regional water utility company in Surabaya). In Proceedings of the 2019 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), Jember, Indonesia, 16–17 October 2019; pp. 94–101.
64. Wong, Y.J.; Arumugasamy, S.K.; Chung, C.H.; Selvarajoo, A.; Sethu, V. Comparative study of artificial neural network (ANN), adaptive neuro-fuzzy inference system (ANFIS) and multiple linear regression (MLR) for modeling of Cu (II) adsorption from aqueous solution using biochar derived from rambutan (*Nephelium lappaceum*) peel. *Environ. Monit. Assess* **2020**, *192*, 439. [[CrossRef](#)] [[PubMed](#)]
65. Abda, Z.; Zerouali, B.; Alqurashi, M.; Chettih, M.; Santos, C.A.G.; Hussein, E.E. Suspended sediment load simulation during flood events using intelligent systems: A case study on semiarid regions of Mediterranean Basin. *Water* **2021**, *13*, 3539. [[CrossRef](#)]
66. Zerouali, B.; Santos, C.A.G.; de Farias, C.A.S.; Muniz, R.S.; Difi, S.; Abda, Z.; Chettih, M.; Heddami, S.; Anwar, S.A.; Elbeltagi, A. Artificial intelligent systems optimized by metaheuristic algorithms and teleconnection indices for rainfall modeling: The case of a humid region in the mediterranean basin. *Heliyon* **2023**, *9*, e15355. [[CrossRef](#)]
67. Ayers, R.S.; Westcot, D.W. *Water Quality for Agriculture*; Food and Agriculture Organization of the United Nations Rome: Rome, Italy, 1985; Volume 29.
68. Aravinthasamy, P.; Karunanidhi, D.; Subramani, T.; Roy, P.D. Demarcation of groundwater quality domains using GIS for best agricultural practices in the drought-prone Shanmuganadhi River basin of South India. *Environ. Sci. Pollut. Res.* **2021**, *28*, 18423–18435. [[CrossRef](#)]
69. M'nassri, S.; El Amri, A.; Nasri, N.; Majdoub, R. Estimation of irrigation water quality index in a semi-arid environment using data-driven approach. *Water Supply* **2022**, *22*, 5161–5175. [[CrossRef](#)]
70. Mokhtar, A.; Elbeltagi, A.; Gyasi-Agyei, Y.; Al-Ansari, N.; Abdel-Fattah, M.K. Prediction of irrigation water quality indices based on machine learning and regression models. *Appl. Water Sci.* **2022**, *12*, 76. [[CrossRef](#)]
71. Omeka, M.E. Evaluation and prediction of irrigation water quality of an agricultural district, SE Nigeria: An integrated heuristic GIS-based and machine learning approach. *Environ. Sci. Pollut. Res.* **2023**, 1–26. [[CrossRef](#)] [[PubMed](#)]
72. Lap, B.Q.; Du Nguyen, H.; Hang, P.T.; Phi, N.Q.; Hoang, V.T.; Linh, P.G.; Hang, B.T.T. Predicting water quality index (WQI) by feature selection and machine learning: A case study of An Kim Hai irrigation system. *Ecol. Inform.* **2023**, *74*, 101991. [[CrossRef](#)]
73. Ibrahim, H.; Yaseen, Z.; Scholz, M.; Ali, M.; Gad, M.; Elsayed, S.; Khadr, M.; Hussein, H.; Ibrahim, H.; Eid, M. Evaluation and prediction of groundwater quality for irrigation using an integrated water quality indices, machine learning models and GIS approaches: A representative case study. *Water* **2023**, *15*, 694. [[CrossRef](#)]
74. Nguyen, D.P.; Ha, H.D.; Trinh, N.T.; Nguyen, M.T. Application of artificial intelligence for forecasting surface quality index of irrigation systems in the Red River Delta, Vietnam. *Environ. Syst. Res.* **2023**, *12*, 24. [[CrossRef](#)]
75. Trabelsi, F.; Bel Hadj Ali, S. Exploring machine learning models in predicting irrigation groundwater quality indices for effective decision making in Medjerda River Basin, Tunisia. *Sustainability* **2022**, *14*, 2341. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.