Article

Exploring Machine Learning Models in Predicting Irrigation Groundwater Quality Indices for Effective Decision Making in Medjerda River Basin, Tunisia

Fatma Trabelsi * and Salsebil Bel Hadj Ali

Research Unit Sustainable Management of Water and Soil Resources, Higher School of Engineers of Medjez El Bab (ESIM), University of Jendouba, Jendouba 8189, Tunisia; belhadjali.salsebil@esim.u-jendouba.tn

* Correspondence: fatma.trabelsi@esim.u-jendouba.tn

Abstract: Over the last years, the global application of machine learning (ML) models in groundwater quality studies has proved to be a robust alternative tool to produce highly accurate results at a low cost. This research aims to evaluate the ability of machine learning (ML) models to predict the quality of groundwater for irrigation purposes in the downstream Medjerda river basin (DMB) in Tunisia. The random forest (RF), support vector regression (SVR), artificial neural networks (ANN), and adaptive boosting (AdaBoost) models were tested to predict the irrigation quality water parameters (IWQ): total dissolved solids (TDS), potential salinity (PS), sodium adsorption ratio (SAR), exchangeable sodium percentage (ESP), and magnesium adsorption ratio (MAR) through low-cost, in situ physicochemical parameters (T, pH, EC) as input variables. In view of this, seventy-two (72) representative groundwater samples have been collected and analysed for major cations and anions during pre-and post-monsoon seasons of 3 years (2019–2021) to compute IWQ parameters. The performance of the ML models was evaluated according to Pearson's correlation coefficient (r), the root means square error (RMSE), and the relative bias (RBIAS). The model sensitivity analysis was evaluated to identify input parameters that considerably impact the model predictions using the one-factor-at-time (OFAT) method of the Monte Carlo (MC) approach. The results show that the AdaBoost model is the most appropriate model for predicting all parameters (r was ranged between 0.88 and 0.89), while the random forest model is suitable for predicting only four parameters: TDS, PS, SAR, and ESP (r was with 0.65 to 0.87). Added to that, this study found out that the ANN and SVR models perform well in predicting three parameters (TDS, PS, SAR) and two parameters (PS, SAR), respectively, with the most optimal value of generalization ability (GA) close to unity (between 1 and 0.98). Moreover, the results of the uncertainty analysis confirmed the prominent superiority and robustness of the ML models to produce excellent predictions with only a few physicochemical parameters as inputs. The developed ML models are relevant for predicting cost-effective irrigation water quality indices and can be applied as a DSS tool to improve water management in the Medjerda basin.

Keywords: groundwater; irrigation water quality indices; machine learning; RF; SVR; ANN; AdaBoost; Medjerda river basin; Tunisia

Citation: 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. <https://doi.org/10.3390/su14042341>

Academic Editor: Fernando António Leal Pacheco

Received: 24 January 2022

Accepted: 12 February 2022

Published: 18 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 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/>).

1. Introduction

Water is a critical input for agricultural production and plays an important role in food security [1]. Due to population growth, urbanization, and climate change (CC), competition for water resources has excessively increased, with adverse effects on agriculture. In particular, groundwater resources rapidly depleted in many parts of the world, especially in the Mediterranean region, notably Tunisia, referenced as one of the most responsive regions to CC and a primary “Hot-spot” [2,3]. This is an emerging threat to agriculture-led rural development. To achieve sustainable development goals (SDGs) related to the efficient

use of water as well as eliminating hunger, it is crucial to improve water management, rationalize the water irrigation [4,5] uses and improve the tools of groundwater quality assessment. Indeed, the suitability of groundwater for irrigation purposes depends on the nature of the mineral elements present in water and their impacts on soil and crops [6,7]. It is based on the concentration of cations and anions present in the groundwater. Quality indices such as the sodium adsorption ratio (SAR), residual sodium carbonate (RSC), magnesium adsorption ratio (MAR), Kelly ratio (KR), and percentage of sodium (%Na) are frequently used in assessing the suitability of waters for irrigation [8–10]. Furthermore, one of the main challenges of qualitative assessment methods is their subjectivity, as they require expert knowledge in assigning weights of variables for calculating the index score, which means that the actual result is not clear [11,12]. However, some parameters require a sampling protocol, laboratory analysis, and at a larger scale, testing and data management [13] which increase the cost and study time of water quality assessment and affects the decision-making on water quality management planning. To cope with these issues, it is crucial to develop a powerful and cost-effective approach for quick and accurate assessment of irrigation water quality. Thus, several contemporary studies have opted for a non-physical tool, successfully predicting groundwater quality using ‘Machine Learning’ models [14,15]. The ML technique is a promising and capable multi-functioning approach in all scientific fields [16,17]. Globally, several researchers have applied ML techniques in various water research studies. They were applied [18,19] for nitrate groundwater contamination [20,21], Manganese removal prediction [13], a flood susceptibility study [22], pollution source identification in water supply network [23], wastewater heavy metal removal [24], heavy metal pollution prediction [25], water level forecasting [26], and, in the last decades, artificial intelligence (AI) techniques have been investigated and showed great ability to predict and monitor water quality [15,27]. These techniques include machine learning (ML), deep learning (DL) and artificial neural networks (ANN).

For example, ML models (supervised machine learning, gradient boosting, and multi-layer perceptron) have been studied by [28,29], who demonstrated the relevance of this technique in predicting water quality [30,31] for drinking use. The support vector machine (SVR) model was applied by [12] to predict the water quality index that showed its accurate prediction. The authors of [32] have compared deep learning (DL) models with three other ML models: random forest (RF), eXtreme Gradient Boosting (XGBoost), and ANN to predict groundwater quality.

However, few research studies have applied AI models to predict irrigation water quality. Recently, the ANN model was used by [33] to predict the suitability of groundwater for irrigation purposes in India using physicochemical parameters as input variables. Similarly, [15] predicted groundwater quality in Morocco using ANN, AdaBoost, Random Forest (RF), ANN, and support vector regression (SVR) models based on irrigation water quality indices as inputs. It is important to note that all published studies have proved the good performance of ML models in the prediction of the suitability of groundwater quality for irrigation purposes using few datasets of physicochemical parameters measured in situ or by smart sensor technologies.

This study is performed for the lower and middle sub-basins of the Medjerda catchment known as the basin downstream from the Sidi Salem dam (DMB). This basin is part of the largest watershed of Tunisia, where it supplies about half of the country’s drinking water. The DMB basin, subject of this study, is essentially agricultural, where irrigation water supply depends on surface water in conjunction with groundwater resources. In recent decades, the study area has experienced water scarcity problems due to the increased frequency of droughts that have led to the increased exploitation of groundwater resources, mainly by the agricultural and agro-industrial sectors [34,35]. Nevertheless, despite the importance of groundwater in the Medjerda basin, there is currently a huge lack of data regarding its quality that undermines the ability of decision makers and users to manage it properly. The few studies that have been conducted are limited geographically and, in a time, where few groundwater sampling campaigns and analyses were conducted, and

they are therefore insufficient to fill the existing data gap and to give a real time information about suitability of groundwater use. Thus, improving the water quality evaluation process based on non-cost data using an objective tool with reliability and flexibility in its decision-making capacity for water management and planning is essential in the DMB basin.

Against this backdrop, the main objectives of this research are: (i) to evaluate the effectiveness of machine learning (ML) models to predict the suitability of groundwater for irrigation purposes in the DMB basin using four ML models (random forest, support vector regression (SVR), ANN, and adaptive boosting (AdaBoost)), (ii) to evaluate the accuracy of the implemented models, and (iii) to analyse the uncertainty and sensitivity of the tested models. Concerning the scientific interest, this study is original, as no previous similar studies were carried out in the pilot area using machine learning methods. Then, the focus of this study was to test the performance of the novel approach and to provide spatial information and guidance to support decision-making processes concerning groundwater management in the Medjerda basin.

2. Materials and Methods

2.1. Study Area

The DMB basin is located in the northern part of Tunisia, it expands from the “Sidi Salem” dam to the outlet of the river into the Mediterranean Sea. It is situated between 4,117,516–4,040,248 m in the north and 527,822 m–613,659 m in the east (zone 32 North of the east of the Universal Transverse Mercator (UTM) coordinate system) (Figure 1). It covers a total geographical area of about 1773 km². The average annual precipitation calculated between the period of 1991 and 2020 is about 448.6 mm/year.

From the geological framework, the study area is a subsidence zone belonging to the Tellian domain. It consists of a Quaternary depression limited by the nappes zone in the north [36,37] and the diapirs zone or Triassic province in the south [38,39]. The sedimentary distribution of the basin is essentially controlled by two NE-trending master faults, which are associated with outcrops of Triassic evaporites. From west to east, there is the El Alia-Teboursouk fault (ETF) and the Tunis-Elles fault (TEF) [40]. The Lithostratigraphy of the study area shows geological formations ranging from Triassic to late Quaternary. The Triassic outcrops have often abnormal contact with Jurassic and Cretaceous outcrops in several localities. The thick lithostratigraphic sequences formed by the Cretaceous, Eocene, Miocene, Pliocene, and Quaternary deposits host the shallow and deep aquifers of the study area such as the aquifer of Bled Guenima, the aquifer of the Anti-Pliocene Medjerda, the plio-quaternary aquifer of Medjerda, the Campanian limestone aquifer of Medjerda, Medjerda aquifer of marls, and Barremian limestones. The alluvial aquifers known as the aquifer of the middle valley of the Medjerda, the aquifer of the lower valley of the Medjerda and the aquifer of Ousja Ghar El Meleh (OGM) are hosted in the colluvial series of the mountains and the alluvial fillings of the deltaic plain. The groundwater of DMB aquifers is primarily used for irrigation and agroindustry and it knew, in last years, severe exploitation, especially in the drought seasons. Moreover, they suffer from salinization, largely caused by natural processes such as evaporation, water-rock interaction, saltwater intrusion, and up-coning of saline waters from deep layers in addition to anthropogenic causes related to irrigation return flow [35,41,42]. The hydromorphic nature of soils at the level of DMB is a rather important problem, observed at the level of irrigated areas of Kalâat El Andalous accompanied by drainage that worsens it, noting, moreover, the clogging and stagnation at the level of Garaâ. This phenomenon enhances the problem of salinity of groundwater due to the excessive use of chemical fertilizers at the level of irrigated areas. Moreover, the coastal aquifer of OGM is affected by saltwater intrusion due to the communication between the lagoon of Ghar El Melh and the sea [30]. Saline groundwater used in irrigation adversely affects soil as well as crop yields. The most harmful associated effects on the irrigated areas are sodification, salinization, and alkalization, which may

alter soil structure [43,44]. Consequently, the quality of groundwater is deteriorated, and it is crucial to evaluate its suitability, especially for irrigation purposes [45,46].

Figure 1. Location map of the downstream Medjerda River Basin (DMB).

2.2. Methodology and Datasets

The methodology adopted in this work is based on five steps (Figure 2): (i) data development (data checking reliability and data exploration); (ii) development of machine learning models (ANN, AdaBoost, SVR, and RF) based on the training datasets; (iii) validation of the models performance based on the validation datasets; (iv) generalization ability; (v) uncertainty and sensitivity analysis of the performed models. This allowed us to evaluate whether the developed models are useful to predict irrigation groundwater quality parameters to help farmers and decision makers to manage irrigation strategies.

Figure 2. Flowchart of adopted methodology.

2.2.1. Input Data

- Physico-chemical parameters

The input data for the used models are the results of physico-chemical analyses of groundwater taken from the DMB basin. It is important to respect the standards of sampling and analysis to have reliable data to be used as input variables of the ML models. In this study, groundwater samples were collected in September 2020, during the dry season, to have water samples less affected by the dilution processes and that present the highest concentrations of solutes during a year. A total of 72 groundwater samples were collected from surface wells and piezometers. The samples were analysed (Figure 1) at the “LandcareMed” laboratory of water and soil analysis at the Higher School of Engineers of Medjez El Bab (ESIM) by adopting the standard procedures [46,47]. The measurement of filtrate dry residue or TDS (total dissolved salts) was performed by evaporating 100 mL of groundwater sample at 105 °C for 24 h. Alkalinity was analysed by titration with 0.1 HCl acid. Measurement of major elements, cations (Na^+ , NH_4^+ , K^+ , Mg^{2+} , and Ca^{2+}) and anions, (Cl^- , NO_3^- , SO_4^{2-} , F^- , Br^-) was performed by means of ion chromatography system. Table 1 summarizes the statistical analysis of the groundwater samples analysis.

Table 1. Statistical summary of physico-chemical parameters of groundwater samples.

Parameter	Unit	Min	Max	Mean	Standard Deviation	Skew	Kurtosis
TDS	mg/L	282.20	15,818	3167.72	2525.07	3.39	13.61
T °C	°C	5.80	26	18.65	0.67	3.40	13.06
pH		3.70	10.1	7.66	3996.26	3.08	−0.41
EC	µs/cm	348	24,300	4974.97	3996.26	3.40	13.69
% O ₂		0.70	44.10	6.10	6.83	3.08	13.06
HCO ₃ [−]	mg/L	6.32	820.01	329.67	174.31	0.00	−0.41
F [−]	mg/L	0.12	9.44	1.62	1.55	2.84	11.98
Cl [−]	mg/L	30.70	8492.67	1211.08	1306.23	4.18	19.80
NO ₂ [−]	mg/L	0.03	22.94	7.81	7.88	0.64	−1.18
Br [−]	mg/L	0.08	123.33	45.37	31.35	−0.03	−0.53
NO ₃ [−]	mg/L	0.38	805.43	124.90	125.99	3.01	12.56
PO ₄ ^{2−}	mg/L	0.38	80.04	39.78	20.22	0.31	−0.42
SO ₄ ^{2−}	mg/L	1.85	2173.01	530.36	505.27	1.76	2.99
Na ⁺	mg/L	19.52	4649	708.52	730.89	4.06	18.65
NH ₄ ⁺	mg/L	3.79	25.44	12.39	8.02	1.30	2.21
K ⁺	mg/L	0.03	119.28	13.95	21.46	3.52	13.39
Mg ²⁺	mg/L	0.54	521.53	150.72	92.03	1.50	3.98
Ca ²⁺	mg/L	2.76	659.46	149.59	144.15	1.68	2.54

- Irrigation water quality Indices (IWQ)

Irrigation water chemistry varies depending on its source, reservoir aquifer lithology, and climatic trends. Poor irrigation water quality adversely affects plant growth, agricultural production, soil deterioration, and human health. Generally, the assessment of groundwater suitability for irrigation purposes is evaluated through various agricultural water quality indicators such as percent sodium (%Na), sodium adsorption ratio (SAR), Kelley ratio (KR), magnesium hazard (MH), residual sodium carbonate (RSC), residual sodium bicarbonate (RSBC), permeability index (PI), and potential salinity (PS). In this study, we focus on SAR, PS, TDS, ESP, RSC, and MAR parameters which are calculated according to Table 2.

Table 2. Irrigation water quality indices (IWQ).

Index Formula	Description
$TDS = \sum(cations + anions)$ [48]	The TDS is the sum of the ion concentrations in the water.
$SAR = \frac{Na^+}{\sqrt{\frac{Mg^{2+} + Ca^{2+}}{2}}}$ [49]	SAR (sodium adsorption ratio) is a measure that determines the degree of hazard to crops by measuring the alkali/sodium risk.
$PS = Cl^- + \frac{SO_4^{2-}}{2}$ [50]	The potential salinity or Doneen is used for risk assessment of cations (calcium, sodium, and magnesium) and bicarbonates present in water that can affect soil permeability if used for long-term irrigation.
$ESP = \frac{Na^+}{Ca^{2+} + Mg^{2+} + Na^+ + K^+} \times 100$ [9]	The percent exchangeable sodium parameter (ESP in %) is used to evaluate the effect of sodium on soil texture.
$RSC = (CO_3^{2-} + HCO_3^-) - (Ca^{2+} + Mg^{2+})$ [51]	Residual sodium carbonates RSC indicate excess bicarbonate and carbonate in the irrigation water
$MAR = \frac{Mg^{2+}}{Mg^{2+} + Ca^{2+}} \times 100$ [52]	The excess of the concentration of magnesium, compared with the sum of the concentration of calcium and magnesium in water, affects the quality of soils that can translate into low crop yield.

2.2.2. Data Pre-Processing and Explanatory Data Analysis (EDA)

Data pre-processing and EDA are the most important part of the machine learning project. It is the operation that transforms raw data into clean data (Figure 3).

The verification of the reliability of physicochemical and IWQ datasets was performed using the ionic balance, the ionic scatter plot, and the boxplot.

Firstly, the data cleaning processing was performed to correct mistakes and errors in the quality dataset by checking the accuracy of physico-chemical datasets.

As a first step, the reliability of the analytic procedures used was checked using the ionic balance (IB). Water samples whose IB exceeds 5% were eliminated.

$$IB(\%) = \left| \frac{\sum Cations - \sum Anions}{\sum Cations + \sum Anions} \times 100 \right|, \quad (1)$$

Then, the elaborated scatter plot between the sum of anions and cations (Figure 4) was built and shows a very good correlation ($R^2 = 0.98$), which confirms the reliability of the used data. Secondly, the IWQ were calculated, (Table 3), and their accuracy was checked using correlation matrix. The box plot of the distribution of IWQ and physicochemical variables (Figure 5) was used to screen the outliers' values for a group of variables. Only few outliers were detected for the majority of variables. Thus, 69 samples were retained and normalized to an interval of 0 to 1 to improve the prediction performance by reducing the influence of extreme and lower values.

$$x_{normalized} = \frac{(x - x_{min})}{(x_{max} - x_{min})} \quad (2)$$

Finally, the dataset of computed Irrigation water quality parameters (IWQ) was divided into two sub-sets for model training and model validation (80:20).

Figure 3. Scatter plots showing the correlation of major cations/anions.

Table 3. Descriptive statistics of the Irrigation Water Quality Indices (IWQ).

	Te	EC	TDS	pH	SAR	PS	ESP	MAR
Mean	18.65	4.97	31.68	7.66	9.60	39.68	57.32	63.17
Standard error	0.38	0.47	3.00	0.08	0.81	4.73	1.51	2.84
Median	18.45	3.91	26.00	7.75	7.76	28.52	56.16	71.26
Mode	14.70	3.71	50.38	7.61	10.65	13.78	56.05	85.25
Standard deviation	d	4.02	25.43	0.67	6.91	40.12	12.77	24.08
Variance	10.66	16.20	646.58	0.45	47.75	1609.57	163.18	579.90
Kurtosis (kurtosis coefficient)	2.28	13.69	13.61	18.36	13.01	16.94	0.62	−0.64
Skewness coefficient	−0.53	3.40	3.39	−2.37	3.23	3.82	0.08	−0.62
Range	20.20	23.95	155.36	6.40	42.66	246.14	64.90	93.03
Minimum	5.80	0.35	2.82	3.70	0.72	1.31	22.05	5.48
Maximum	26.00	24.30	158.18	10.10	43.38	247.44	86.95	98.51

Figure 4. Boxplots of IWQ parameters and physico-chemical variables.

Figure 5. Distribution of the raw values of parameters by sample.

2.2.3. Machine Learning Modelling

The ML models were developed in the Jupyter Lab using the open-source tool of the anaconda platform (www.anaconda.com/products/individual, accessed on 8 November 2021) to perform the python package of data science and machine learning.

- Artificial Neural Network (ANN)

ANN is commonly used as an ML model in groundwater modelling [53]. It is a well-established and long-standing machine learning technique that is designed to evaluate the processes (represented by the data) that have high complexities and reduced availability of information for the purpose of regression [54]. In this study, a feed forward multilayer perceptron (MLP) architecture was used for training the ANN committee model. A MLP, which is a specific case of ANN, consists of an input layer, one or more hidden layers, and an output layer [55,56]. The authors of [57] have stated as follows: It consists of a weighted input layer, hidden layers, and an output layer. These layers are interconnected by neurons. Hence, designing ANN requires the transformation from the j th to the $(j + 1)$ th layer through an activation function (f) and so on until the target layer [57]. The iterative training process is repeated for the layers until good preliminary performance.

In this study, only three layers were developed to obtain an output y_i following the Equation (3):

$$Y_i = f \left(\sum_{i=1}^N W_{ij} x_i + b_j \right) \quad (3)$$

with N , x_i , y_j , b_j and w_{ij} showing the number of nodes in the previous layer, the i th nodal in the previous layer, the j th nodal in the present layer, the bias of j th nodal in the present layer, and a weight connecting x_i and y_j [58].

- Adaptive boosting model (AdaBoost)

AdaBoost is an ensemble learning algorithm developed by [46]. It can be used in combination with many other types of learning algorithms to improve ability.

It integrates multiple weak learners into an individual strong learner and initializes an equal weight for all datasets. Then, the weights of the samples misclassified by the previous weak learner are improved. Finally, the samples with the updated weights are used to train the next weak learner. With this approach, new learners are trained to decrease the weighted error produced by previous learners (Figure 6).

Figure 6. Flow chart of the AdaBoost algorithm.

- Support vector machine

The SVM is a machine learning algorithm [59] based on statistical learning theory. It is extensively used in resolving issues related to classification (SVC) and regression (SVR) which also diminishes the algorithm over-fitting [60].

For an observational data set (Ds) $D_s = (x_i, y_i)_{i=1}^n$, the optimal function is the minimization of the function (4) (subject to (5)). Hence, the loss functions such as ϵ -insensitive, quadratic, and Hubber methods can be used [44].

$$\min(\omega, b, \epsilon^-, \epsilon^+) = \frac{1}{2} \times ||\omega^2|| + C \times \sum_{i=1}^n (\epsilon i + \epsilon i^*) \quad (4)$$

with ϵ_i and ϵ_i^* as the lower and upper constraints on the output

$$S.t \begin{cases} y_i - \omega^T \times \phi(x) - b \leq \epsilon - \epsilon_i \\ -y_i + \omega^T \times \phi(x) + b \leq \epsilon - \epsilon_i^* \\ \epsilon_i, \epsilon_i^* \geq 0 \\ i = 1, \dots, n \end{cases} \quad (5)$$

with ω , b , and C representing weight, bias vectors, and the prespecified value to penalize the training error, while $\phi(x)$ is a Kernel function (k) (polynomial, radial basis, and linear functions).

In this study, a radial basis function (RBF) was adopted as Kernel function.

$$k(x_i, x_j) \exp(-\gamma |x_i - x_j|^2) \quad (6)$$

- Random forest

The random forest algorithm proposed by [45] is a general-purpose classification and regression method. It builds an ensemble of weighted average of decision trees in training by swapping and changing the covariates to improve the prediction performance.

In this study, the k -fold ($k = 5$) cross-validation method was used during the learning process to further prevent model overfitting [61]. The optimal architectures, functions, and hyperparameters of each model were determined by trial-and-error analysis based on their evolution during the training process. All models' parameters used for prediction of IWQ parameters are summarized in the Table 4.

Table 4. Optimal parameters and functions used for IWQ indices prediction.

Model	Description of Parameters and Functions
ANN	3 layers 12 neurons in hidden layer algorithm: Levenberg–Marquardt Function activation: sigmoid identity in output layer Epoch number: 1000 Learning rate: 0.01 Momentum coefficient: 0.85
SVR	$C = 200$ Kernel function: RBF ($\gamma = 1.2$) ϵ -function loss, $\epsilon = 0.002$ Gamma = 0.1
Random Forest	Number of trees: 20 Loss function: exponential
AdaBoost	Estimator number: 50 Learning rate: 0.5

2.2.4. Validation of Models Performance

- Metric validation

This step consists of evaluating the developed models. During it, their robustness is tested in order to assess if the results obtained can be trusted.

In this study, three statistical criteria were used to validate the above models (Table 5): (i) Pearson's correlation coefficient[®], (ii) the root mean square error (RMSE), and (iii) the relative bias (RBIAS).

Table 5. Statistical criteria to validate the models.

Designation	Formula	Description
Pearson's correlation coefficient (r)	$r = \frac{\sum_{i=1}^n (X_{0i} - \bar{X}_0)(X_{pi} - \bar{X}_p)}{\left[\sum_{i=1}^n (X_{0i} - \bar{X}_0)^2 \sum_{i=1}^n (X_{pi} - \bar{X}_p)^2 \right]^{0.5}}$	<p>- $r = 1$: best correlation between the observed and predicted values, but it does not indicate the best model.</p> <p>- $r < 1$ indicates a less fit model.</p>
The root mean square error (RMSE)	$RMSE = \sqrt{\frac{\sum (X_{pi} - X_{0i})^2}{n}}$	A lower value of RMSE compared with the values of the results indicates a better fit of the model
The relative bias (RBIAS).	$RBIAS = \frac{\sum_{i=1}^n (X_{pi} - X_{0i})}{\sum_{i=1}^n X_{0i}}$	<p>- $RBIAS > 0$: the model tended to underestimate</p> <p>- $RBIAS < 0$: overestimate the target magnitude</p> <p>- $RBIAS = 0$: the model is perfect,</p> <p>higher absolute value of RBIAS indicates that the model is biased</p>

- Generalization ability

Good performance in the testing phase is believed to be evidence for an algorithm's practical plausibility, where this performance provides an evaluation of the model's generalization capability. Achievement of this objective is typically measured by the generalization ability (GA) of the models [52]. The author of [62] defined GA in groundwater level prediction by:

$$GA = \frac{RMSE \text{ pendant la phase de validation}}{RMSE \text{ pendant la phase d'apprentissage.}} \quad (7)$$

GA values equal to unity indicate that the ML model is perfect. If the GA is less than unity, the models are under-trained, while if it is greater than unity, the models are over-trained.

- Uncertainty and Sensitivity Analysis

In this study, uncertainties of the fitted models were assessed by comparing the observed and simulated values and calculating the standard error and confidence Bound as explained in Equations (8) and (9)

$$SD = \sqrt{\frac{\sum_{i=1}^n (e_i - \bar{e})^2}{(n-1)}} \quad (8)$$

$$CB = z \times \frac{SD}{\sqrt{n}} \quad (9)$$

with $e_i = (X_{0i} - X_{pi})$, z is the z-score of the confidence level (for 95%, it is about 1.96), and $\bar{e}$ is the mean prediction error.

Finally, the model sensitivity analysis was [63,64] performed to identify input parameters that considerably impact the model predictions of IWQ. This analysis was performed using the one-factor-at-time (OFAT) method based on the Monte Carlo approach, which is used to estimate the possible outcomes of an uncertain event [65,66]; an input variable was generated randomly while keeping other variables constant. Then, the absolute value

of the difference in RMSE ($|\Delta\text{RMSE}|$) was calculated to assess the impact of each input variable. Therefore, the sensitivity of the model to an input increases the absolute value of the difference in RMSE.

3. Results

3.1. Statistical Analysis

For further exploration of the variables, a correlation matrix analysis and an assessment of the importance of the input variables [66] were performed.

The correlation matrix is performed since it illustrates the importance of each parameter independently and their effect on the hydrochemistry [67,68]. If the values of (r) are +1 or −1 in the Pearson's correlation matrix, they are treated as strong correlation coefficients values and signify total correlation. If the values are closer to zero, it means there is no significant interaction between two variables at the $p < 0.05$ level [19,55]. If r is bigger than 0.7, the parameters are highly correlated, and if r is between 0.4 and 0.7, the parameters are moderately correlated. In this study, a correlation matrix is used to consider the correlation between chemical parameters and IWQ values. The results reported in Figure 7 show that electrical conductivity (EC) has a high correlation with TDS ($r = 0.99$), PS ($r = 0.99$), and SAR ($r = 0.86$)), while it has a low correlation with ESP ($r = 0.30$) and MAR ($r = 0.05$) indices. The pH has low correlations with all parameters. The temperature has the lowest correlations with all parameters. These results show that electrical conductivity (EC) is a more correlated input variable with the predicted parameters than pH and temperature. Nevertheless, high correlations do not imply causality since complex combinations of the features can have influences on the target variable. According to [15], the lowest correlations between T, pH, and EC prove that these parameters are separable and non-redundant and, therefore, are useful for improving the predictive accuracy of machine learning.

Figure 7. Matrix correlation.

3.2. Implementation and Evaluation of Models

This study included the results of performing four different methods of predicting the irrigation water quality parameters (IWQ). The models used were as follows: artificial neural network (ANN), adaptive boosting (AdaBoost), support vector machine for regression (SVR) and random forest (RF). Three metric criteria were used to validate the above models: Pearson's correlation coefficient (r), RMSE, and RBIAS.

The results of the training and validation processes of the developed models are illustrated in Figures 8 and 9, respectively.

The results of the training process reveal that the SVR model has significant values of RBIAS and RMSE compared with the other models for predicting the TDS parameter. The ANN, RF, and AdaBoost models revealed high accuracy in predicting the TDS parameter during the learning process with values of r equal to 0.94, RMSE equal to 500.07 mg L⁻¹, and RBIAS of 1% on average. It showed that all developed models performed very well with average correlation coefficients of 0.90, RBIAS less than 3% in absolute value, and average RMSE around 5 meq L⁻¹. Based on the training results (Figure 8), the four models perform satisfactory for the prediction of the sodium absorption ratio (SAR) and the percent exchangeable sodium (ESP). In fact, the correlation coefficients are 0.61 and 0.62, respectively. Similarly, the coefficients RMSE and RBIAS proved acceptable results for the two IWQs. As for the magnesium adsorption ratio (MAR), two of the statistical parameters (RBIAS and RMSE) showed that all models performed it moderately well, and only AdaBoost has a good person's coefficient (r). Hence, it was inferred that the AdaBoost model had a good performance in predicting all the IWQs parameters. However, random forest and artificial neural network models were unable to predict the MAR parameter. Overall, we can notice that there is no significant superiority between the ensemble models in the training process.

Figure 8. Results of training model performance.

Figure 9. Results of validation model performance.

Yet, the validation process, evaluation of generalizability, sensitivity, and uncertainty analysis are essential issues to evaluate the above models. Therefore, model validation was performed using same algorithm with twenty percent of the data that were simulated to assess the validation (Figure 9) and generalization ability. The Pearson's coefficient values range from 0.65 to 0.94 for the four parameters TDS, PS, SAR, and ESP over ANN and SVR models. However, RMSE showed an unacceptable performance for all models for the simulation of the TDS and MAR parameters, and RBIAS showed a lowest performance for the SVR model for the simulation of the TDS and MAR parameters. When comparing the performance results, two of the simulated models (AdaBoost and RF) had lower performance in the training process while the ANN and SVR models presented very close results during the two processes for the prediction of all IWQs parameters. All models, except ANN for the SAR parameter, have RBIAS values less of than 6% in absolute value, indicating that the fitted models are unbiased.

The scatter plot (Figure 10) shows the relationship between observed and simulated variables over all IWQs parameters for all developed models. It identifies a better distribution on the $X = Y$ line for the random forest for all models. Moreover, it shows that the predicted values are very close to the observed values for the AdaBoost model except for the MAR parameter. In fact, the accuracy of the models is satisfactory when the values are distributed on or uniformly across the two diagonals of the $X = Y$ line, showing that the errors obey the Gaussian distribution [15]. Even though the SVR and ANN models showed a satisfactory performance during the training phase, they failed to reproduce the ESP parameter due to an RMSE which was very high (greater than 10%).

Figure 10. Scatterplots of observed and simulated values for the prediction of IWQs parameters during the validation process.

Therefore, it can be deduced that the SVR model has the weakest performance in predicting PS and SAR parameter, whereas the AdaBoost model has the best performance in predicting all parameters. After follows the ANN and the RF in predicting TDS, PS, and SAR parameters and TDS, PS, SAR, and ESP parameters, respectively. These results are in accordance with previous findings [15,69]. The researchers found that the AdaBoost model is superior to the support vector machine and artificial neural network models. To have useful models to predict new data sets, while avoiding errors, it is necessary to test its generalization capability. This way, once the model is developed, the end-users could test it with any new dataset coming, for example, from real-time measurement sensors. Therefore,

the stability of machine learning models in forecasting real-time water quality parameters is essential, especially when policy makers and researchers have strategies to develop this approach in irrigation water management [15]. In this study, the generalization ability to different input variables was evaluated. Figure 11 indicates that the ANN model for TDS model is overfitted while all other models are underfitted. However, the generalization ability of the random forest and AdaBoost model are weaker than the ANN and SVR models.

Figure 11. Generalization ability (GA) indices of the models.

3.3. Uncertainty and Sensitivity Analysis

The issue of uncertainties in conceptual models in water quality modelling is inevitable and has been discussed in many studies [42,45,70,71]. In this study, the uncertainty was analysed and showed that the SVR model has the highest (95%) confidence bound values, followed by the ANN, RF, and AdaBoost models (Table 6).

Table 6. Model uncertainty analysis.

Parameter	Error	ANN	SVR	RF	AdaBoost
TDS (mg L ⁻¹)	E	−27.01	412.48	4.79	11.57
	CB (95%)	55.07	142.56	50.65	27.55
PS (meq L ⁻¹)	E	−0.27	0.45	0.21	−0.09
	CB (95%)	1.00	1.96	0.97	0.91
SAR (meq ^{0.5} L ^{−0.5})	E	−0.36	0.04	−0.01	−0.02
	CB (95%)	0.47	0.57	0.09	0.04
ESP (%)	E	−1.31	−1.45	0.13	0.56
	CB (95%)	1.69	1.89	1.14	0.74
MAR (%)	E	−0.05	0.27	−0.02	0.19
	CB (96%)	2.01	2.47	1.47	0.69

The sensitivity of the model provides an overview of the impact of input variables on the output. This analysis is necessary to assess how the model acts according to shifts in input values (data quality, noise tolerance, etc.). Therefore, in this study, sensitivity analysis of built models (Figure 12) was performed by simulating the models after adding a random Gaussian noise to the input variables (EC, pH and T).

Figure 12. Sensitivity analysis results.

Sensitivities of the models to the inputs differ based on type of inputs, IWQs parameters and models. In fact, the results of sensitive analysis show that the models are more sensitive to: (i) electrical conductivity followed by temperature and pH, respectively for predicting TDS and MAR; (ii) pH for predicting ESP parameter; (iii) electrical conductivity followed by the pH and the temperature, respectively for predicting PS and SAR.

Moreover, the AdaBoost was found to be the most sensitive model since it has the highest values of the absolute value of the difference in RMSE. However, the overall results of the sensitivity analysis show that the models are quite stable in predicting IWQ.

4. Discussion

In this research, four models: random forest (RF), support vector regression (SVR), artificial neural networks (ANN), and adaptive boosting (Adaboost) were used to predict the irrigation water quality parameters (IWQ): total dissolved solids (TDS), potential salinity (PS), sodium adsorption ratio (SAR), exchangeable sodium percentage (ESP), and magnesium adsorption ratio (MAR) through low-cost in situ physicochemical [72,73] parameters (T, pH, EC) as input variables. The performance of the tested models was evaluated according to Pearson's correlation coefficient (r), the root means square error

(RMSE), and the relative bias (RBIAS). The model sensitivity was evaluated to identify [74] input parameters that considerably impact the model prediction using the one-factor-at-time (OFAT) method of the Monte Carlo (MC) approach. In accordance with the reviewed literature, [30,69,75] the results show that the AdaBoost model is the most appropriate for predicting all parameters, with R ranged between 0.88 and 0.89, and that the random forest model is suitable for predicting only four parameters: TDS, PS, SAR, and ESP, with R ranged between 0.65 and 0.87. Added to that, as found by [22,76], this study identifies that The ANN and SVR models perform well in predicting three parameters (TDS, PS, SAR) and two parameters (PS, SAR), respectively, with most optimal value of generalization ability (GA) close to the unity.

Furthermore, MAR is the worst predictive parameter. This unproductive prediction accuracy is probably due to the low relationship between the EC and the pH used as input variables. Additionally, as explained by [7,9,22,27,29,61,74], the more significant the correlation between the input and output variables, the higher the performances of the models. Hence, the accurate prediction highly depends on the number of input variables and their impact.

In general, the methodology of the proposed models for prediction of the irrigation water quality parameters (IWQ) has proved its effectiveness. The effectiveness of ML models does not only depend on the accuracy of the prediction but also on the nature and number of predictors used. It is noteworthy that the use of physicochemical parameters such as EC, pH, and T could significantly enhance the performance of machine learning models [15,77]. Consequently, it is important to explore ML models for water quality index prediction using only physicochemical parameters as input variables without decreasing the efficiency of the models. Accordingly, this provides an incentive for decision makers to apply artificial intelligence for water quality planning and management.

However, the stability of the ML models in the forecasting of the IWQ parameters in real time is crucial, mainly when it is closely linked with the decision maker. Therefore, while ML models are fairly stable in forecasting the IWQ parameters, it should be highlighted that the selection of the models must be based on deeper sensitivity analysis by using smart technologies based on the Internet of Things (IoT) as a more secure and regular data alternative as explained by [60]. Moreover, the generalization of these models must be deeply studied because there are other variables that may interfere and influence water quality.

5. Conclusions and Future Trends

The key goal of this research is to evaluate the ability of machine learning (ML) models to predict the quality of groundwater for irrigation purposes in the downstream Medjerda river basin (DMB), in Tunisia. Therefore, Adaboost, random forest, ANN, and SVR models were developed and evaluated to predict TDS, PS, SAR, ESP, and MAR parameters using physico-chemical parameters as input variables. This study confirmed that the AdaBoost model is appropriate for predicting all parameters while the random forest model is suitable for predicting only four parameters: TDS, PS, SAR, and ESP.

Added to that, this study found out that the ANN and SVR models perform well in predicting 3 parameters (TDS, PS, SAR) and 2 parameters (PS, SAR) of 5 parameters, respectively. However, the SVR and ANN models showed better generalization ability than the AdaBoost and random forest models. Then, the sensitivity analysis showed that the developed models are less sensitive to the input variables used compared with the range of each predicted parameter. The ML models characterized by physical parameters are effective tools and should be recommended for predicting water quality parameters.

This research presents an effective use of machine learning models in forecasting the irrigation groundwater quality indices through low-cost data and can be used as a decision support systems (DSS) tool for sustainable water management in DMB. In fact, the traditional simulation modelling approaches are dependent on datasets that involve a large amount of unknown or unspecified input data and generally consist of high-cost

time-consuming processes. Therefore, setting up a DSS based on machine learning models will boost the efficient use of water and rationalize its use by all water stakeholders at watershed level.

Author Contributions: Conceptualization, F.T. and S.B.H.A.; methodology, F.T.; software, S.B.H.A.; validation, F.T.; formal analysis, F.T.; investigation, S.B.H.A.; resources, F.T.; data curation, F.T.; writing—original draft preparation, F.T. and S.B.H.A.; writing—review and editing, F.T.; visualization, F.T. and S.B.H.A.; supervision, F.T.; project administration, F.T.; funding acquisition, F.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the United States Agency for International Development (USAID) through Partnerships for Enhanced Engagement in Research program of the National Academies of Sciences, Engineering, and Medicine (grant number: PEER 7_ Tunisia project 7-289).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The study did not report any data.

Acknowledgments: The authors are greatly thankful to the four Regional Commissariats for Agricultural Development (CRDA) of the Bèjà, Mannouba, Ariana, and Bizerte regions for providing some data and facilitating the groundwater sampling campaigns. We thank all reviewers and the editors for their kind reviews and comments that improved the clarity of the final manuscript.

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

References

1. FAO. *Water for Sustainable Food and Agriculture*; Food and Agriculture Organization of the United Nations: Caracalla, Rome, 2017; ISBN 978-92-5-109977-3.
2. Knaepen, H. *Climate Risks in Tunisia Challenges to Adaptation in the Agri-Food System*; European Centre for Development Policy Management (ECDPM): Maastricht, The Netherlands, 2021.
3. Hssaisoune, M.; Bouchaou, L.; Sifeddine, A.; Bouimetarhan, I.; Chehbouni, A. Moroccan Groundwater Resources and Evolution with Global Climate Changes. *Geosciences* **2020**, *10*, 81. [CrossRef]
4. Aureli, A.; Ganoulis, J.; Margat, J. Groundwater Resources in the Mediterranean Region: Importance, Uses and Sharing. *Water Mediterr.* **2008**, 96–105. Available online: <https://www.iemed.org/publication/groundwater-resources-in-the-mediterranean-region-importance-uses-and-sharing> (accessed on 8 November 2021).
5. Berhail, S. The impact of climate change on groundwater resources in northwestern Algeria. *Arab. J. Geosci.* **2019**, *12*, 770. [CrossRef]
6. Rahmati, O.; Pourghasemi, H.R.; Melesse, A.M. Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran Region, Iran. *CATENA* **2016**, *137*, 360–372. [CrossRef]
7. Yang, L.; Hua, G.; Caoab, L.; Wang, X.; Chen, M.-H. A comparison of Monte Carlo methods for computing marginal likelihoods of item response theory models. *J. Korean Stat. Soc.* **2019**, *48*, 503–512. [CrossRef]
8. Kopittke, P.M.; So, H.B.; Menzies, N.W. Effect of ionic strength and clay mineralogy on Na–Ca exchange and the SAR–ESP relationship. *Eur. J. Soil Sci.* **2006**, *57*, 626–633. [CrossRef]
9. Wang, L.; Long, F.; Liao, W.; Liu, H. Prediction of anaerobic digestion performance and identification of critical operational parameters using machine learning algorithms. *Bioresour. Technol.* **2020**, *298*, 122495. [CrossRef]
10. Paliwal, K.V. *Irrigation with Saline Water*; Water Technology Centre, Indian Agriculture Research Institute: New Delhi, India, 1972; p. 198.
11. Amiri, V.; Rezaei, M.; Sohrabi, N. Groundwater quality assessment using entropy weighted water quality index (EWQI) in Lenjanat, Iran. *Environ. Earth Sci.* **2014**, *72*, 3479–3490. [CrossRef]
12. Gorgij, A.D.; Kisi, O.; Moghaddam, A.A.; Taghipour, A. Groundwater quality ranking for drinking purposes, using the entropy method and the spatial autocorrelation index. *Environ Earth Sci.* **2017**, *76*, 269. [CrossRef]
13. Bhagat, S.K.; Tiyyasha, T.; Tung, T.M.; Mostafa, R.R.; Yaseen, Z.M. Manganese (Mn) removal prediction using extreme gradient model. *Ecotoxicol. Environ. Saf.* **2020**, *204*, 111059. [CrossRef]
14. Leong, Y.C.; Hughes, B.L.; Wang, Y.; Zaki, J. Neurocomputational mechanisms underlying motivated seeing. *Nat. Hum. Behav.* **2019**, *3*, 1. [CrossRef]
15. El Bilali, A.; Taleb, A.; Brouziyne, Y. Groundwater quality forecasting using machine learning algorithms for irrigation purposes. *Agric. Water Manag.* **2021**, *245*, 106625. [CrossRef]
16. Evangelos, R. Machine learning, urban water resources management and operating policy. *Resources* **2019**, *8*, 173.

17. Kim, H.; Kim, S.; Hwang, J.Y.; Seo, C. Efficient Privacy-Preserving Machine Learning for Blockchain Network. *IEEE Access* **2019**, *7*, 136481–136495. [\[CrossRef\]](#)
18. Nolan, B.T.; Fienen, M.N.; Lorenz, D.L. A statistical learning framework for groundwater nitrate models of the Central Valley, California, USA. *J. Hydrol.* **2015**, *531*, 902–911. [\[CrossRef\]](#)
19. Ransom, K.M.; Nolan, B.T.; Traum, J.A.; Faunt, C.C.; Bell, A.M.; Gronberg, J.A.M.; Wheeler, D.C.; Rosecrans, C.Z.; Jurgens, B.; Schwarz, G.E.; et al. A hybrid machine learning model to predict and visualize nitrate concentration throughout the Central Valley aquifer, California, USA. *Sci. Total Environ.* **2017**, *601–602*, 1160–1172. [\[CrossRef\]](#)
20. Rodriguez-Galiano, J.A.V.F.; Luque-Espinar, M.; Chica-Olmo, M.P. Mendes, Feature selection approaches for predictive modelling of groundwater nitrate pollution: An evaluation of filters, embedded and wrapper methods. *Sci. Total Environ.* **2018**, *624*, 661–672. [\[CrossRef\]](#)
21. Ouedraogo, I.; Defourny, P.; Vanclooster, M. Application of random forest regression and comparison of its performance to multiple linear regression in modeling groundwater nitrate concentration at the African continent scale. *Hydrogeol. J.* **2019**, *27*, 1081–1098. [\[CrossRef\]](#)
22. Chen, H.K.; Chen, C.; Zhou, Y.; Huang, X.; Qi, R.; Shen, F.; Liu, M.; Zuo, X.; Zou, J.; Wang, Y.; et al. Ren Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. *Water Res.* **2020**, *171*, 115454. [\[CrossRef\]](#)
23. Grbčić, L.; Lučin, I.; Kranjčević, L.; Družeta, S. Water supply network pollution source identification by random forest algorithm. *J. Hydroinform.* **2020**, *22*, 1521–1535. [\[CrossRef\]](#)
24. Bhagat, S.K.; Tung, T.M.; Yaseen, Z.M. Development of artificial intelligence for modeling wastewater heavy metal removal: State of the art, application assessment and possible future research. *J. Clean. Prod.* **2020**, *250*, 119473. [\[CrossRef\]](#)
25. Lal, R.; Stewart, B.A. *Soil Processes and Water Quality*, 1st ed.; CRC Press: Boca Raton, FL, USA, 1994. [\[CrossRef\]](#)
26. Zhu, S.; Hrnjica, B.; Ptak, M.; Choiński, A.; Sivakumar, B. Forecasting of water level in multiple temperate lakes using machine learning models. *J. Hydrol.* **2020**, *585*, 124819. [\[CrossRef\]](#)
27. Ahmed, U.; Mumtaz, R.; Anwar, H.; Shah, A.A.; Irfan, R. Efficient water quality prediction using supervised machine learning. *Water* **2019**, *11*, 2210. [\[CrossRef\]](#)
28. Fijani, E.; Barzegar, R.; Deo, R.; Tziritis, E.; Skordas, K. Design and implementation of a hybrid model based on two-layer decomposition method coupled with extreme learning machines to support real-time environmental monitoring of water quality parameters. *Sci. Total Environ.* **2019**, *648*, 839–853. [\[CrossRef\]](#)
29. Lu, H.; Ma, X. Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere* **2020**, *249*, 126169. [\[CrossRef\]](#)
30. Bel Hadj Ali, S.; Trabelsi, F. CAJG-2020-P527: Saltwater Intrusion Vulnerability Mapping Using Multi-Model Ensemble of Machine Learning Algorithms: A Case Study of the Aousja Ghar El Melh Coastal Aquifer, Northeast of Tunisia; Advances in Science, Technology & Innovation (ASTI); Springer: Berlin/Heidelberg, Germany, 2022.
31. Bel Hadj Ali, S.; Trabelsi, F. Impact of Anthropogenic Activities on the Groundwater Quality Using Machine Learning Algorithms: A Case Study of the Aousja Ghar El Melh Coastal Aquifer, Northeast of Tunisia. In Proceedings of the Mediterranean Geosciences Union Annual Meeting (MedGU-21), Istanbul, Turkey, 25–28 November 2021.
32. Singh, R.; Kumar, S.; Nangare, D.D.; Meena, M.S. Drip irrigation and black polyethylene mulch influence on growth. *Yield-Water-Use Effic. Tomato* **2009**, *4*, 1427–1430. [\[CrossRef\]](#)
33. Wagh, V.M.; Panaskar, D.B.; Muley, A.A.; Mukate, S.V.; Lolage, Y.P.; Aamalawar, M.L. Prediction of groundwater suitability for irrigation using artificial neural network model: A case study of Nanded tehsil, Maharashtra, India. *Model. Earth Syst. Environ.* **2016**, *2*, 1–10. [\[CrossRef\]](#)
34. Trabelsi, F.; LEE, S. GIS-based groundwater potential mapping using Machine learning models: Case of Medjerda aquifer, North of Tunisia. In Proceedings of the IAH2019, the 46th Annual Congress of the International Association of Hydrogeologists, Málaga, Spain, 22–27 September 2019.
35. Trabelsi, F.; Ali, S.B.; Mukherjee, S.; Sipolya, R. Integrated Use of Satellite Remote Sensing and Hydraulic Modeling for the flood Risk Assessment at the middle valley of Medjerda. In Proceedings of the International Conference & Exhibition. Advanced Geospatial Science & Technology (TeaGeo 2016), Tunis, Tunisia, 26–28 September 2016.
36. Ayed, B.N. *Evolution Tectonique de l'Avant-Pays de la Chaîne Alpine de Tunisie du Début du Mésozoïque à l'Actuel Thèse d'Etat*; Université de Paris Sud—Centre d'Orsay: Gif-sur-Yvette, France, 1986.
37. Rouvier, H. *Géologie de l'Extrême Nord-Tunisien: Tectonique et Paléogéographie Superposées à l'Extrémité Orientale de la Chaîne Nord-Maghrébine*. Thèse d'Etat, Paris, France, 1977; p. 307.
38. Perthuisot, V. *Dynamique et Pétrogenèse des Extrusions Triasiques en Tunisie Septentrionale*. Thèse Doct, ès Science, Travelling Laboratory Geology Ecole North Superior, Paris, France, 1978; p. 312.
39. Ghanmi, M. *Etude géologique du J. Kebbouch (Tunisie septentrionale)*. Ph.D. Thesis, Thèse 3 ème Cycle, Toulouse, France, 1980; p. 141.
40. Melki, F.; Zouaghi, T.; Chelbi, M.B.; Bédir, M.; Zargouni, F. Role of the NE-SW Hercynian Master Fault Systems and Associated Lineaments on the Structuring and Evolution of the Mesozoic and Cenozoic Basins of the Alpine Margin, Northern Tunisia. In *Tectonics—Recent Advances*; IntechOpen: London, UK, 2012. Available online: <https://www.intechopen.com/chapters/37864> (accessed on 8 November 2021).

41. Trabelsi, F.; Mukherjee, S. Remote Sensing and GIS Techniques for Evaluation of Groundwater Quality in middle valley of Medjerda, Tunisia. In Proceedings of the 1st Euro-Mediterranean Conference for Environmental Integration (EMCEI), Sousse, Tunisia, 22–25 November 2017; p. 526.
42. Trabelsi, F.; Mammou, A.B.; Tarhouni, J.; Piga, C.; Ranieri, G. Delineation of saltwater intrusion zones using the time domain electromagnetic method: The Nabeul–Hammamet coastal aquifer case study (NE Tunisia). *Hydrol. Process.* **2013**, *27*, 2004–2020. [\[CrossRef\]](#)
43. Hachicha, M.; Cheverry, C.; Mhiri, A. The impact of long-term irrigation on change of groundwater level and soil salinity in northern Tunisia. *Arid. Soil Res. Rehabil.* **2010**, *14*, 175–182. [\[CrossRef\]](#)
44. Chatti, A.; Trabelsi, F.; Arfaoui, A. *Qualité et Vulnérabilité des Ressources en eau Souterraine de la Basse Vallée de la Medjerda*; University of Jendouba: Jendouba, Tunisia, 2018.
45. Breiman, L. Random Forests. *Mach. Learn. USA* **2001**, *45*, 5–32. [\[CrossRef\]](#)
46. APHA. *Standard Methods for the Examination of Water and Wastewater*, 21st ed.; American Public Health Association/American Water Works Association/Water Environment Federation: Washington, DC, USA, 2005.
47. Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [\[CrossRef\]](#)
48. Sorensen, D.L. *Suspended and Dissolved Solids Effects on Freshwater Biota: A Review*; US Environmental Protection Agency, Office of Research and Development: Washington, DC, USA, 1977.
49. Richards, L.A. *Diagnosis and Improvement of Saline Alkali Soils, Agriculture, 160, Handbook 60*; US Department of Agriculture: Washington, DC, USA, 1954.
50. Freeze, R.A.; Cherry, J.A. *Groundwater*; Prentice-Hall: Hoboken, NJ, USA, 1979.
51. Raghunath, H.M. *Groundwater*; Wiley Eastern Ltd.: Delhi, India, 1987; p. 563.
52. Barzegar, R.; Moghaddam, A.A.; Baghban, H. A supervised committee machine artificial intelligent for improving DRASTIC method to assess groundwater contamination risk: A case study from Tabriz plain aquifer, Iran. *Stoch. Env. Res. Risk A.* **2016**, *30*, 883–899. [\[CrossRef\]](#)
53. Barzegar, R.; Adamowski, J.; Moghaddam, A.A. Application of wavelet-artificial intelligence hybrid models for water quality prediction: A case study in Aji-Chay River, Iran. *Stoch. Env. Res. Risk Assess.* **2016**, *30*, 1797–1819. [\[CrossRef\]](#)
54. Barzegar, R.; Moghaddam, A.A. Combining the advantages of neural networks using the concept of committee machine in the groundwater salinity prediction. *Model. Earth Syst. Environ.* **2016**, *2*, 26. [\[CrossRef\]](#)
55. Belayneh, A.; Adamowski, J.; Khalil, B.; Quilty, J. Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. *Atmos. Res.* **2016**, *172*, 37–47. [\[CrossRef\]](#)
56. Dawson, C.W.; Wilby, R. An Artificial Neural Network Approach to Rainfall-Runoff Modelling. *Hydrol. Sci. J.* **1998**, *43*, 47–66. [\[CrossRef\]](#)
57. Robert, J.S. *Artificial Neural Networks by (1997-06-01) Hardcover–January 1*; McGraw-hill Companies: New York, NY, USA, 1997.
58. Castrillo, M.; García, A.L. Estimation of high frequency nutrient concentrations from water quality surrogates using machine learning methods. *Water Res.* **2020**, *172*, 115490. [\[CrossRef\]](#)
59. Chen, K.; Chen, H.; Zhou, C.; Huang, Y.; Qi, X.; Shen, R.; Liu, F.; Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [\[CrossRef\]](#)
60. Gayen, A.; Pourghasemi, H.R.; Saha, S.; Keesstra, S.; Bai, S. Gully erosion susceptibility assessment and management of hazard-prone areas in India using different machine learning algorithms. *Sci. Total Environ.* **2019**, *668*, 124–138. [\[CrossRef\]](#)
61. Arlot, S.; Celisse, A. A survey of cross-validation procedures for model selection. *Stat. Surv.* **2010**, *4*, 40–79. [\[CrossRef\]](#)
62. Rajae, T.; Ebrahimi, H.; Nourani, V. A review of the artificial intelligence methods in groundwater level modeling. *J. Hydrol.* **2019**, *572*, 336–351. [\[CrossRef\]](#)
63. Khalil, A.; Almasri, M.N.; McKee, M.; Kaluarachchi, J.J. Applicability of statistical learning algorithms in groundwater quality modelling. *Water Resour. Res.* **2005**, *41*, W05010. [\[CrossRef\]](#)
64. Yoon, H.; Jun, S.C.; Hyun, Y.; Bae, G.O.; Lee, K.K. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J. Hydrol.* **2011**, *396*, 128–138. [\[CrossRef\]](#)
65. Qiu, Y.; Aufiero, M.; Wang, K.; Fratoni, M. Development of sensitivity analysis capabilities of generalized responses to nuclear data in Monte Carlo code RMC. *Ann. Nucl. Energy* **2016**, *97*, 142–152. [\[CrossRef\]](#)
66. Patil, R.; Bellary, S. Machine learning approach in melanoma cancer stage detection. *J. King Saud Univ.-Comput. Inf. Sci.* **2020**. [\[CrossRef\]](#)
67. Islam, M.M.S.; Ferdous, Z.; Potenza, M.N. Panic and generalized anxiety during the COVID-19 pandemic among Bangladeshi people: An online pilot survey early in the outbreak. *J. Affect. Disord.* **2020**, *276*, 30–37. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Zhao, X.; Ning, B.; Liu, L.; Song, G. A prediction model of short-term ionospheric foF2 based on AdaBoost. *Adv. Space Res.* **2014**, *53*, 387–394. [\[CrossRef\]](#)
69. Kardos, J.S.; Obropta, C.C. Water quality model uncertainty analysis of a pointpoint source phosphorus trading program. *J. Am. Water Resour. Assoc.* **2011**, *47*, 1317–1337. [\[CrossRef\]](#)
70. Moreno-Rodenas, A.M.; Tschekner-Gratl, F.; Langeveld, J.G.; Clemens, F.H.L.R. Uncertainty analysis in a large-scale water quality integrated catchment modelling study. *Water Res.* **2019**, *158*, 46–60. [\[CrossRef\]](#)

71. Radwan, M.; Willems, P.; Berlamont, J. Sensitivity and uncertainty analysis for river quality modelling. *J. Hydroinform.* **2004**, *6*, 83–99. [[CrossRef](#)]
72. Saghafi, H.; Arabloo, M. Modeling of CO₂ solubility in MEA, DEA, TEA, and MDEA aqueous solutions using adaboost-decision tree and artificial neural network. *Int. J. Greenh. Gas Control* **2017**, *58*, 256–265. [[CrossRef](#)]
73. Zhou, Z.; Feng, J. Deep Forest. *Natl. Sci. Rev.* **2019**, *6*, 74–86. [[CrossRef](#)] [[PubMed](#)]
74. Di, M.Z.; Chang, P. Guo Water quality evaluation of the Yangtze River in China using machine learning techniques and data monitoring on different time scales. *Water* **2019**, *11*, 339. [[CrossRef](#)]
75. Shojaei, M.; Nazif, S.; Kerachian, R. Joint uncertainty analysis in river water quality simulation: A case study of the Karoon River in Iran. *Environ. Earth Sci.* **2015**, *73*, 3819–3831. [[CrossRef](#)]
76. Ayadi, A.; Ghorbel, O.; BenSalah, M.S.; Abid, M. A framework of monitoring water pipeline techniques based on sensors technologies. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**. [[CrossRef](#)]
77. Chowdury, M.S.U.; Emran, T.; Ghosh, S.B.; Pathak, A.; Alam, M.M.; Absar, N.; Andersson, K.; Hossain, M.S. IoT based real-time river water quality monitoring system. *Procedia Comput. Sci.* **2019**, *155*, 161–168. [[CrossRef](#)]