Journal of the Saudi Society of Agricultural Sciences 19 (2020) 439–451

Contents lists available at ScienceDirect

Journal of the Saudi Society of Agricultural Sciences

journal homepage: www.sciencedirect.com

Full length article

Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment

Ali El Bilali *, Abdeslam Taleb

University of Hassan II, Casablanca, Faculty of Sciences and Techniques of Mohammedia, Morocco

ARTICLE INFO

Article history:
Received 22 March 2020
Revised 1 August 2020
Accepted 2 August 2020
Available online 7 August 2020

Keywords:
Sodium absorption ratio
Irrigation water quality
Farmers
Prediction accuracy

ABSTRACT

Evaluation of the water suitability for irrigation purposes using conventional approaches is generally expensive because it requires several parameters, particularly in developing countries. Therefore, developing accurate and reliable models may be valuable to overcome this issue in the management of the water used in agriculture. To achieve this purpose, 8 Machine Learning (ML) models namely: Artificial Neural Network (ANN), Multiple Linear Regression (MLR), Decision Tree, Random Forest (RF), Support Vector Regression (SVR), k-Nearest Neighbour (kNN), Stochastic Gradient Descent (SGD) and Adaptive Boosting (AdaBoost) have been developed and validated for predicting of 10 Irrigation Water Quality (IWQ) parameters such as Sodium absorption ratio (SAR), adjusted SAR_a, Exchangeable Sodium Percentage (ESP), percentage of Sodium (%Na), Residual Sodium Carbonate (RSC), Permeability Index (PI), Kelly Ratio (KR), Chloride Cl^-, Magnesium Absorption Ratio (MAR), and TDS dissolved in water surface of Bouregreg watershed in Morocco using electrical conductivity (EC) and pH as input variables. 300 samples are analysed at 9 monitoring stations across four main rivers, processed and selected to train and validate the models. The results have revealed that, except for SVR and k-NN models and MAR and PI parameters, all other models are highly accurate in predicting the other parameters with coefficients of correlations (r) with ranges of [0.56, 0.99], and [0.64, 0.99] for training and validation processes sequentially. Furthermore, this study attempts to generalize the 6 ML models developed and validated to the Cherrate and Nfifikh watersheds that are different from Bouregreg watershed. The results of the generalization attempt have shown that the ML models are fairly generalized for TDS, SAR, and SAR_a parameters to Cherrate watershed and for TDS, chloride, ESP and %Na parameter for Nfifikh watershed. The results of this study have also demonstrated that the machine learning models are efficient tools for accurately predicting the quality of irrigation water by only using the parameters that can be directly measured in a short time. Consequently, the implementation of the automated sensor technologies coupled with ML models improve the control of water quality and will help the farmers to manage the irrigation water quality.

1. Introduction

Water resources play an important role in supplying drinking, industrial and agriculture domains. Indeed, the good quality of water resources significantly reduces the cost of water treatment for drinking and industrial purposes and improves the agricultural production. Water demand is increasing due to population growth, intense agriculture, urbanization and industrial activities. Since the anthropogenic activities and natural pollution sources threaten water resources by going beyond making it suitable for drinking, irrigation, industrial and other purposes (Busico et al., 2020; Mbizvo et al., 2019), water quality assessment and prediction are necessary to assess its suitability to serve a specific purpose and to determine appropriate treatments or precautions if it is not healthy. This is mainly determined by the analysis of an important number of water quality parameters to quantify the dissolved substance. However, monitoring all parameters involved in a river or a groundwater is often insufficient in developing countries as it is laborious and expensive. Hence, reducing the subjectivity and the effective cost for assessing water quality is a great challenge. How-

* Corresponding author.
E-mail address: Ali1gpee@gmail.com (A. El Bilali).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jssas.2020.08.001
1658-077X/

ever, in recent years, many water quality indexes (WQI) have been proposed and developed by several national and international organizations. Thus, Tyagi and Sharma (2014) describes the most fruitful index-based approaches for assessing water quality for drinking purposes including US National Sanitation Foundation WQI, Florida Stream WQI, British Columbia WQI, Canadian WQI and the Oregon WQI.

In developing countries such as Morocco, agriculture is considered as a main source of employment and the pillar of the economy for non-petroleum countries (FAO, 2002). Agriculture is also the most popular consumer of water with a percentage that reaches 80 as well as it is one of sources of water pollution. Therefore, water planning and management using an effective cost for irrigation purpose is required to ensure a sustainable agriculture. In general, irrigation water quality index (IWQI) is developed based on the parameters adopted by FAO guidelines 29 (Ayers and Westcot, 1994). However, several investigations have been carried out for assessing water quality using an index-based approach. Singh et al. (2018) have already used this approach to minimize the subjectivity in assessing the suitability of water quality for irrigation in India through Saaty's Analytic Hierarchy Process (SAHP) and have demonstrated that the IWQI is useful for irrigation water managers. Ewaid et al. (2019) develops the irrigation water quality guide (IWQG) software based on the FAO guidelines and Meireles et al. (2010) works. The case study has shown that the software is an efficient tool for assessing the quality of water for irrigation purposes. In addition to that, a statistical-based approach is broadly applied to develop the WQI. Jahin et al. (2020) applied the multivariate analysis to develop IWQI for surface water in Egypt. The results have shown that using Principal Component Analysis (PCA) and Factor Analysis (FA) rapidly and cheaply control water quality. Ewaid and Abed (2017) used the PCA and Hierarchical Cluster Analysis (CA) to investigate the potential pollution sources of Al Gharaf River in Iraq. Decision-makers can use these results to decrease the number of samples analyzed and to prioritize the actions to improve the quality of the river. Although these models have been developed and considered efficient tools to assess IWQ, their applicability needs considerable parameters and studies that take into consideration the cost and the time of the study. Therefore, the model-based prediction can be used as an alternative for farmers to optimize assessing water quality.

Moreover, nowadays, there are a few developed models to predict the suitability of water for irrigation purposes in an arid environment. However, Ewaid et al. (2018) has developed a model to rapidly predict the water quality of Tgirs River in Iraq for drinking purposes using the water quality index and Multiple Linear Regression (MLR). The model developed consists of five parameters turbidity (Tur), electrical conductivity (EC), Total Hardness (TH), Chemical Oxygen Demand (COD) and pH with a significant value of R^2, 0.974. Gidey (2018) conducts a study to determine and map the groundwater suitability for irrigation using the WQI and GIS tool and interpolation methods. The model provides spatial prediction of water suitability for irrigation purposes. In recent years, artificial intelligence (AI) techniques have been investigated and showed a high capability for predicting and monitoring water quality (Ahmed et al., 2019a; Fijani et al., 2019; Liu et al., 2019; Lu and Ma, 2020). These techniques include Machine Learning (ML), Deep Learning (DL) and Artificial Neural Network (ANN). For example, the ML models have been investigated by Ahmed et al. (2019) who shows how accurate this technique is in water quality prediction for drinking use. Leong et al. (2019) used Support Vector Machine (SVM) model to predict the water quality index and showed its precise prediction. Nowadays, there is few investigation of AI for evaluating and predicting the quality of irrigation water. However, Wagh et al. (2016) used ANN to predict the groundwater suitability for irrigation in India using 13 physicochemical parameters as input variables of the model. They have showed that the data-based model performs well in water suitability for irrigation use.

Importantly all published studies have demonstrated good performances of ML models for predicting the water quality. However, some parameters (inputs of the models) require sampling protocol and analysis in the laboratory which increases, on the one hand, the cost and the time of studying water quality assessment. Surprisingly, there is no published study focused on the prediction of IWQ parameters using ML models and the physical parameters as input variables so far. On the other hand, particularly in the developing countries, a large amount of low-water quality (wastewater) could be used or reused for food crop irrigation to overcome water scarcity in these areas. It requires monitoring water quality to reduce the impact of this practice on soil and food crop production (Bortolini et al., 2018; Chaoua et al., 2018). Consequently, developing accurate and reliable ML models to predict the suitability of water for irrigation purposes using physical parameters that can be directly measured in situ by the sensor technologies rather than a large number of parameters measured in the laboratory is one of the main aims of this study. These models can support the farmers in arid/semi-arid countries to manage the low quality of water and to improve the crop yield.

Morocco, as an arid region, has adopted dam construction policies to supply water for different purposes since 1960. Due to the climate change effects and the socio-economic development (Droogers et al., 2012), the country has adopted a new integrated strategy of water which includes new dam constructions, desalinization, implementation of wastewater treatment plants, reuse of treated wastewater for agriculture irrigation as well as legislation approaches. Thus, assessing water quality is a pivotal tool for water management and allocation for specific purposes. Since the water surface quality is deteriorated by the evaporation in addition to the anthropogenic activities in the country, predicting and monitoring water parameters can be an efficient tool to manage its quality. The present study attempts to develop and evaluate 8 machine learning (ML) models to numerically predict the quality of water irrigation parameters used for evaluating its suitability in agricultural purposes using the electrical conductivity and pH as input variables in the semi-arid region of Bouregreg in Morocco. The study focuses on 10 IWQ parameters used to assess the salinity hazard, Toxicity, Carbonate and Bicarbonate Hazard, Sodium Hazard and Magnesium Hazard. Therefore, the physicochemical parameters analysis is conducted to determine the Sodium Adsorption Ratio (SAR), adjusted SAR_a, and Magnesium Absorption Ratio (MAR), Kellys Ration (KR), Sodium percentage $\%Na^+$, exchangeable sodium percentage (ESP), permeability index (PI), Total Dissolved Solids (TDS) and Chloride Cl^- as main indexes and parameters commonly used for the evaluation of the water of suitable quality in irrigation. The ML models are developed to predict previous parameters using the electrical conductivity (EC) and pH as input variables. This work is going to support the farmers in arid and semi-arid countries to predict water suitability in irrigation within a short time and at a low cost.

2. Irrigation water quality parameters

Assessing the quality of water used in irrigation is based on a large number of parameters and indexes adopted by several organization and agencies. This study focuses on the sodium, carbonate, bicarbonate, magnesium and the salinity hazards as well as the toxicity (chloride). Indeed, the salinity hazard of water is an important indicator in irrigation, as high salt in water renders the soil saline that affects the salt intake capacity of the crops through the roots. It is generally evaluated using electrical conductivity

(EC) or the Total Dissolved Solids (TDS in mg/L), according to the FAO guideline (Ayers and Westcot, 1994): for TDS < 400 mg/L, 400 mg/L < TDS < 2000 mg/L and TDS > 2000 mg/l the degrees of restriction on use is none, slight to moderate and severe successively. However, the TDS is the sum of anions and cations in the solution as given by Eq. (1).

$$TDS = \sum(\text{cations} + \text{anions}) \tag{1}$$

For the sodium hazard assessment, there are five parameters that are commonly used in evaluating the suitability of water for irrigation purposes such as SAR, SAR_a, KR, %Na, and ESP. The Sodium Absorption Ratio (SAR) in $(meq/L)^{0.5}$ is given by Eq. (2) (Richards, 1954). The high SAR value degrades soil texture by reducing the hydraulic conductivity and, therefore it decreases the irrigation performance. However, water is considered unsuitable when the SAR is more than 10 according to FAO guideline

$$SAR = \frac{Na}{\sqrt{\frac{Mg+Ca}{2}}} \tag{2}$$

In addition to that, Kelly ratio (Kelley, 1963) defines the hazardous impact of sodium on IWQ. The value of KR is more than 1 which is considered unsuitable for irrigation. This index is calculated as:

$$KR = \frac{Na}{Ca + Mg} \tag{3}$$

The exchangeable sodium percentage (ESP) in % evaluates the effect of Na on soil structure and was calculated using Eq. (6) (Kopittke et al., 2006).

$$ESP = \frac{Na}{Ca + Mg + Na + K} * 100 \tag{4}$$

The percentage of sodium is calculated as follows:

$$Na\% = \frac{Na + K}{Ca + Mg + Na + K} * 100 \tag{5}$$

The adjusted SAR (SAR_a) in $(meq/L)^{0.5}$ is generally used when water has an excessive concentration of Ca^{2+}, and/or HCO_3^- is employed for irrigation. So, a variable fraction of Ca^{2+} will precipitate in soil as $CaCO_3$. Then, the Ca in Eq. (2) is replaced by the expected Ca_e after equilibrium with $CaCO_3$ in soil. However, this index is calculated according to Lesch and Suarez (2009) technical note.

The permeability index (PI) in % is defined as water suitability for irrigation index by Doneen (1964) and given by Eq. (4):

$$PI = \frac{Na + \sqrt{HCO3}}{Ca + Mg + Na} * 100 \tag{6}$$

The residual sodium carbonate (RSC meq/l) is calculated as follow:

$$RSC = (CO3 + HCO) - (Ca + Mg) \tag{7}$$

The magnesium absorption ratio is calculated by Eq. (8):

$$MAR = \frac{Mg}{Mg + Ca} \tag{8}$$

All ionic concentrations are given in (meq/l) except for TDS parameter that is given in mg/l.

3. Materials and methods

3.1. Study area description

The present study area is Bouregreg watershed (9000 km^2) located in the middle of Morocco. The soil types in this area are dominated by Chromic Luvisols and Eutric Planosols which make up 45% 18% of the basin respectively and are characterized by loamy or light clay nature (Fischer et al., 2008). Regarding agricultural activities, this watershed is characterized by pastoral activities in the East part especially in mountainous areas, the cultivation of olive trees, cereal, vegetables as well as livestock. The main water resources refer to four rivers particularly Kourifla, Machraa, Grou and Bouregreg where water availability is replenished by rainfall with an annual average of 450 mm. The main reservoir dam commissioned in 1975 with current capacity of up to 900 million m^3 is located 10 km south of Rabat and supplies up to 200 million m^3 per year for both drinking and industrial purposes in coastal cities between Rabat and Casablanca. Thus, decision-makers have adopted the construction of 6 new dams (one of them is under construction) in this area particularly for agricultural purposes. On the other hand, water pollution that is due to discharge of waste water without appropriate treatment and the poor waste management threatens the ecosystem, health and the quality of water resources. Therefore, the pollution is a serious challenge in this area; water quality must be assessed and pollution sources must be controlled. Furthermore, the salinization of the water surface in these areas is increased because of semi-arid conditions with an average annual evaporation of up to 1800 mm, in addition to the geological and geomorphological sources. However, Fig. 1 shows the monitoring of water quality stations and the main pollution sources. These sampling locations were selected in this study as they are optimal monitoring stations in term of pollution control and watershed divisions adopted by the River Basin Agency of Bouregrge and Chaouia. Also, the sampling was conducted during the critical periods where the water quality deteriorates under semi-arid conditions.

The Nfifikh and Cherrate watersheds have surface areas of 510 and 650 km^2. The quarries, for schist extraction, are the main activities that can deteriorate water quality in the Cherrate watershed. In the Nfifikh watershed, the cereal crop is the most dominant activity. Climatically speaking, summer is commonly hot with a temperature that ranges between 35 °C and 45 °C, and for winter, it is cold between 5 °C and 15 °C. As for the rainfall, annually speaking, its average is 350 mm. These watersheds are different from Bouregreg in terms of pollution sources and the socio-economic activity types in addition to geological and morphological conditions. However, these watersheds are monitored by Nfifikh and Cherrate stations as presented in Fig. 1.

3.2. Methods and data pre-processing

Three hundred samples were collected at Bouregreg watershed using 9 monitoring stations illustrated in Fig. 1. The parameters measured and analyzed are: Electrical conductivity (EC), pH, Chloride Cl^-, SO_4^{2-}, CO_3^{2-}, HCO_3^-, NO_3^-, NH_4^+, Na^+, K^+, Ca^{2+}, and Mg^{2+}. The EC and pH were measured during the sampling protocol using EC and pH meter. while the other parameters were analysed in the laboratory that uses the flame photometer for Na^+ and K^+, UV spectrophotometer for NO_3^- and SO_4^{2-} and titration method for Ca^{2+}, Mg^{2+}, Cl^-, CO_3^{2-} and HCO_3^-. Using the same methods, the 40 samples were collected and analyzed for the Cherrate and Nfifikh watersheds (Fig. 1). These samples were used to evaluate the ML model generalizations.

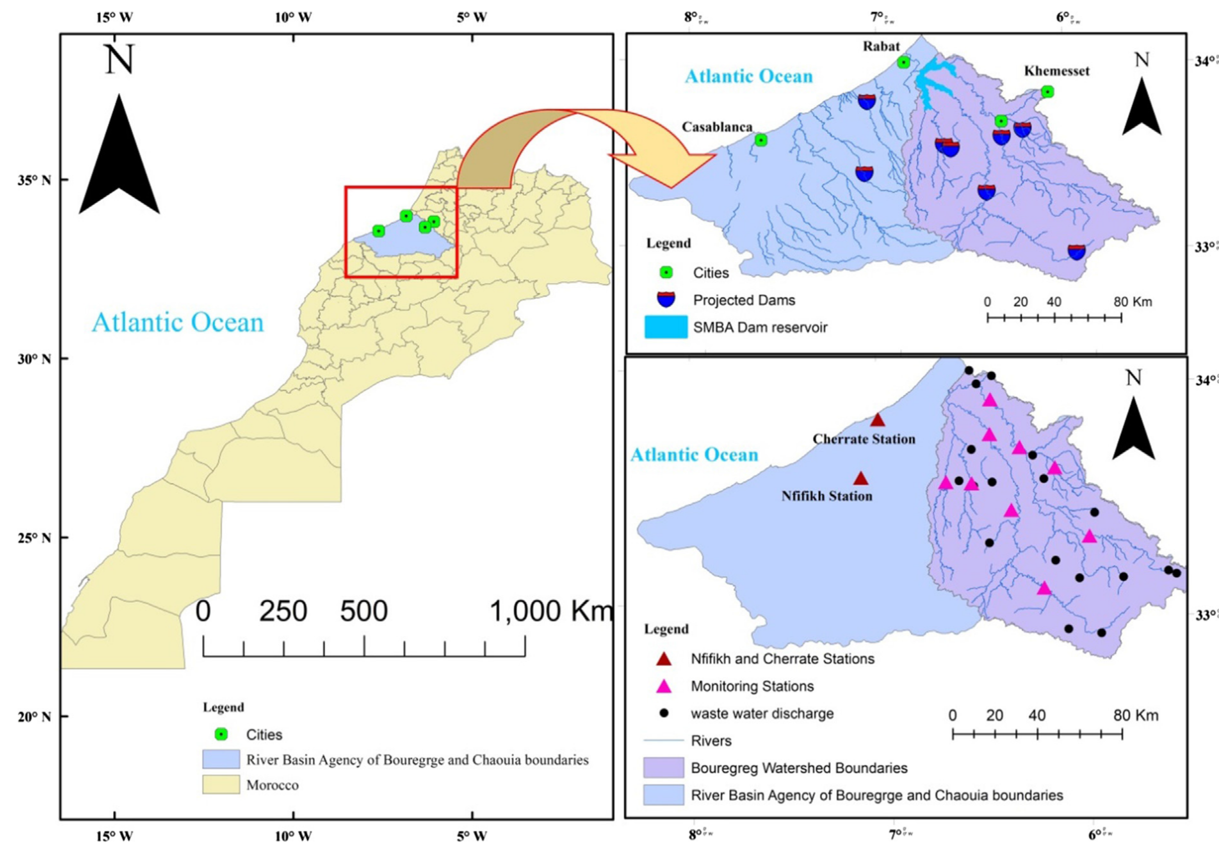

Fig. 1. Study area description, monitoring station locations and waste water discharge locations.

The first step in data processing was controlling the data quality by checking the precision of the analytical procedure using the imbalance between the negative and the positive ions. Among three hundred samples collected at Bouregreg watershed, only 264 samples have imbalance errors within ±5% and therefore were selected for the training and the validation of the models in this study while 29 and 35 samples from 40 were selected for Cherrate and Nfifikh successively. The second step was to calculate the IWQ parameters that include: TDS, SAR, SAR_a, KR, PI, RSC, ESP and %Na and their descriptive statistics using MS-Excel. Table 1 summarizes the statistical characteristics of these parameters.

The third step was the normalization of the input and output features within a range of [0, 1] to improve the prediction performance by reducing the influence of extreme and lower values. All data was recognized in *CSV.file*, and then the study adopted the supervised regression ML models to numerically predict the IWQ parameters.

3.3. Machine learning models

Machine Learning (ML), as one of the AI techniques, is classified into two groups that are the supervised ML and the unsupervised one. In order to develop ML models for predicting IWQ parameters, this study evaluates 8 supervised ML models for numerical prediction of 10 IWQ parameters. These models include: (1) multiple linear regression as conventional method, (2) artificial neural network that shows a high potential accuracy for hydrological modelling in two decades (Govindaraju, 2000; Haykin, 1999). Here, the ANN model was developed on the basis of 15 neurons in a hidden layer and select *Tansig* as a function of activation and *Levenberg* algorithm; (3) Decision Tree (DT); (4) Random Forest (RF) that is bagging technique that operate by constructing an ensemble of DTs in training and the output is the average of the output trees (Ho, 1995; Liaw and Wiener, 2002) (10 trees in parallel in this study); (5) K-Nearest Neighbour (KNN) that is a non-parametric model based on storing all available cases and predicting the outputs by calculating the average of the targets of the nearest neighbours (Burba et al., 2009) (k) (k value is important in determining the k-NN model performance) using the distance function. However, this study adopts 5 as k-value and Euclidian as a distance function; (6) Support Vector Mchine/Regression (SVM/R), that is a discriminative model introduced by Vapnik, 1995. The model constructs a hyper-plane to minimize the error while keeping the error part tolerated. In this study, the v-SVR model with *Sigmoid*

Table 1
Irrigation water quality and input parameters' descriptive statistics.

	EC (µs/cm)	pH	TDS (mg/L)	Cl^-(meq/L)	SAR $(meq/L)^{0.5}$	KR	PI	SAR_a $(meq/L)^{0.5}$	MAR	RSC	ESP	%Na
Max	21,700	9.55	12,290	208.80	46.64	6.38	94.40	53.20	0.82	3.08	86	86
Mean	1403	7.97	949	9.10	3.29	0.78	55.63	3.66	0.45	−3.48	34	37
Min	118	6.30	89	0.42	0.11	0.03	12.66	0.13	0.01	−26.64	3	5
SD	2549	0.48	1504	25.44	6.25	1.00	14.12	7.11	0.12	3.77	18	18
Skewness	6.02	−0.65	6	6.36	5.16	3.23	0.07	5.17	−0.32	−3.34	0.95	0.87
Kurtosis	39.06	1.34	39	42.54	29.87	12.09	0.27	29.82	1.39	15.13	0.38	0.22

Kernel (selected), error penalty factor C = 1 and v, 0.5 parameter are selected for the best regression; (7) Stochastic Gradient Descent (SGD) that is also a discriminative model. It requires a number of hyper parameter including the regularization parameter and iteration. However, this study uses the Ridge (L2) as regularization method, the squared as loss function, and the number of iterations is 1000 with stopping criteria of 0.001 and (8) Adaptive boosting is an enhancement based technique (Freund and Schapire, 1996; Schapire, 1999). All these models can be used for classification and regression and are learned following their own techniques and validated for the unseen dataset during the training process. However, it should be noted that the detailed review of the machine learning algorithm is beyond the aims of this study.

The performance of each model was evaluated according to the traditional statistic performances such as coefficient of correlation (r), Root Mean Square Error (RMSE), and the relative Bias (RBIAS).

$$RMSE = \sqrt{\frac{\sum (Si - Oi)^2}{n}} \tag{9}$$

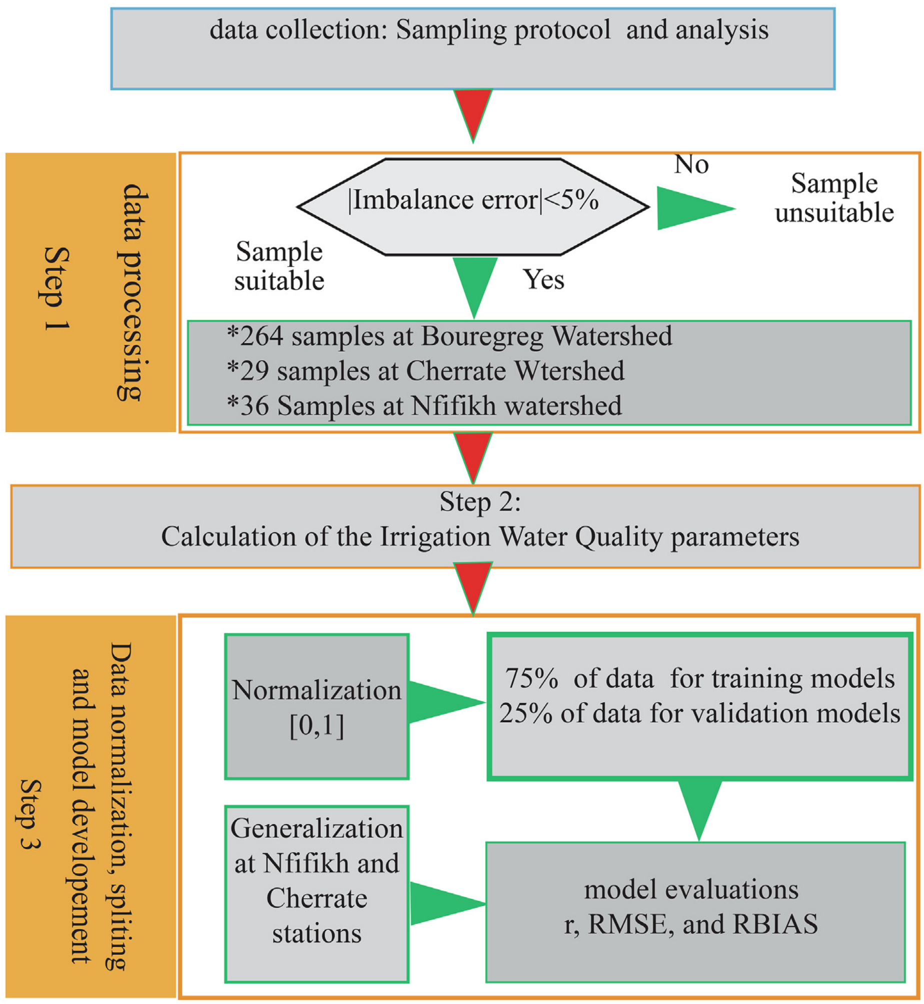

Fig. 2. Flowchart of the developed methodology.

Table 2
Matrix Correlation analyses between input and output variables.

	EC	pH	TDS	Cl^-	SAR	KR	PI	SAR_a	MAR	RSC	ESP	%Na
EC	1.00											
pH	0.04	1.00										
TDS	0.99	0.06	1.00									
Cl-	0.99	0.03	0.99	1.00								
SAR	0.97	0.08	0.97	0.97	1.00							
KR	0.84	0.13	0.84	0.84	0.94	1.00						
PI	0.40	0.07	0.39	0.41	0.54	0.71	1.00					
SAR_a	0.97	0.09	0.97	0.97	1.00	0.93	0.54	1.00				
MAR	0.11	0.07	0.12	0.09	0.08	0.03	−0.11	0.07	1.00			
RSC	−0.83	−0.05	−0.83	−0.81	−0.75	−0.59	0.01	−0.75	−0.21	1.00		
ESP	0.58	0.17	0.58	0.56	0.72	0.88	0.81	0.71	0.08	−0.36	1.00	
%Na	0.57	0.18	0.56	0.55	0.70	0.86	0.84	0.69	0.07	−0.32	0.98	1.00

$$RBIAS = \frac{\sum_{i=1}^{n}(Si - Oi).100}{\sum_{i=1}^{n} Oi} \quad (10)$$

$$r = \left(\frac{\sum_{i=1}^{n}\left(Oi - \text{O}\right)\left(Si - \widehat{\text{S}}\right)}{\left[\sum_{i=1}^{n}\left(Oi - \text{O}\right)^{2}\sum_{i=1}^{n}\left(Si - \widehat{\text{S}}\right)^{2}\right]^{0.5}} \right) \quad (11)$$

where Oi and Si are observed and simulated values sequentially; O represents the mean of observed values and n = number of values considered.

The ML models were built in the ***open-source anaconda distribution*** that is the most popular and easiest platform to perform python package of data science and Machine Learning. These models are developed using 264 samples of surface water quality recorded at 9 monitoring stations in Bouregreg watershed. The used data is classified into two classes: a training–testing dataset and a validation one. Among 264 samples, 198 (75%) are used as a training set while 66 samples are used for model validation phases. As far as the data is concerned, it pertains to the watersheds of Cherrate and Nfifikh that are used to assess the feasibility of model generalization in order to evaluate their applicability in different environmental conditions. Fig. 2 shows the flowchart of the steps of the methodology followed.

4. Results and discussion

In recent decade, AI techniques have been worldwidely used to predict and map various environmental processes. This study investigates the ML models to numerically predict the irrigation water quality parameters that are commonly used for assessing water suitability in agricultural purposes using the easily measurable input variables such as Electrical Conductivity (EC) and pH parameters. This section presents the training of machine learning models, their validation and generalization results. Pearson correlation analysis between all variables (input/output) to investigate their relationships is conducted and the results are presented in Table 2. As this table shows, it is observed that electrical conductivity is highly correlated with the TDS, chloride, SAR, and SAR_a parameters. It has also relatively good correlations with KR and RSC parameters that has low correlations with %Na and ESP parameters, and obtain very low correlations with PI and MAR parameter. For the pH input parameter, it owns poor correlations with all parameters. These results show that the electrical conductivity has a more significant impact than the pH one.

4.1. Training of machine learning models

To examine the possibility of a numerical prediction of IWQ parameters using ML techniques, models of ANN, MLR, Tree, RF, SVM, kNN, SGD and Adaboost are trained and validated and their accuracy is evaluated. The prediction accuracy of these models is determined by three performance criteria namely: coefficient of correlation (r), RMSE, and RBIAS. Table 3 presents the model training results for predicting the TDS, Chloride, SAR, KR, PI, SAR_a, MAR, RSC, ESP and the %Na as main parameters used for IWQ assessment.

As presented in Table 3, generally speaking, all models have a good correlation between the observed and the predicted values of TDS. However, based on the RMSE (in mg/L) and RBIAS of the models with the data of range [90, 12,290] (mg/L), ANN (163) and MLR (163) yielded the best accuracy followed by SGD (171) and Adaboost (358) compared to other models. Moreover, the RBIASs of these models are less than 10% in absolute value showing

Table 3
Training model performance results.

Model		ANN	MLR	DT	RF	SVM	kNN	SGD	AdaBoost
TDS	r	0.99	0.99	0.87	0.97	0.58	0.95	0.99	0.98
	RBIAS	−1%	−1%	−8%	−14%	−35%	−22%	−1%	−10%
	RMSE (mg/L)	162.92	162.92	764.14	564.78	1567.15	1010.27	170.50	357.77
Chloride	r	0.99	0.99	0.87	0.98	0.75	0.96	0.99	0.98
	RBIAS	6%	6%	0%	−15%	−47%	−34%	7%	−7%
	RMSE (meq/L)	3.39	3.39	12.82	7.80	24.08	17.44	3.56	5.28
SAR	r	0.97	0.97	0.90	0.95	0.70	0.94	0.97	0.95
	RBIAS	1%	1%	−7%	−15%	−31%	−26%	1%	−14%
	RMSE $(meq/L)^{0.5}$	1.71	1.71	2.81	2.15	5.16	3.67	1.68	2.20
SAR_a	r	0.97	0.97	0.89	0.94	0.67	0.94	0.97	0.95
	RBIAS	2%	2%	−8%	−19%	−38%	−26%	3%	−14%
	RMSE $(meq/L)^{0.5}$	1.96	1.96	3.24	3.20	6.15	4.09	1.99	2.47
KR	r	0.84	0.84	0.84	0.80	0.10	0.84	0.84	0.83
	RBIAS	−4%	−4%	−11%	−13%	−3%	−19%	−6%	−17%
	RMSE	0.56	0.56	0.57	0.61	1.80	0.60	0.55	0.61
ESP	r	0.58	0.58	0.64	0.68	0.10	0.68	0.58	0.61
	RBIAS	−8%	−8%	−9%	−10%	50%	−10%	−9%	−14%
	RMSE	0.15	0.15	0.15	0.13	1.58	0.14	0.15	0.16
%Na	r	0.57	0.57	0.58	0.63	0.03	0.64	0.57	0.56
	RBIAS	−7%	−7%	−5%	−10%	−77%	−8%	−7%	−11%
	RMSE (%)	15	15	16	14	157	14	15	16
MAR	r	0.08	0.08	0.26	0.24	0.08	0.22	0.08	0.24
	RBIAS	−3%	−3%	2%	0%	−50%	−3%	−4%	4%
	RMSE	0.12	0.12	0.13	0.12	1.57	0.12	0.12	0.13
PI	r	0.40	0.40	0.27	0.32	0.43	0.35	0.40	0.21
	RBIAS	−0.4%	−0.4%	−0.9%	−1.6%	−1.6%	−3.0%	−1.0%	−1.0%
	RMSE (%)	12.87	12.87	16.34	13.89	13.54	13.44	12.86	16.00
RSC	r	0.83	0.83	0.64	0.70	0.59	0.70	0.83	0.69
	RBIAS	−3%	−3%	10%	3%	−12%	−12%	−3%	−4%
	RMSE (meq/L)	2.11	2.11	3.04	2.73	3.35	2.96	2.10	3.05

the high performance of ANN, MLR, SGD and, Adaboost models. Among the other models, the SVM has the worst accuracy for TDS prediction with r, RMSE and RBIAS, 0.58, 1567 mg/L and −35% respectively. However, the RF, kNN and Tree models have a good agreement and correlation between observed and predicted values even if they have high RMSE. Because these models failed to predict some extreme values (more than 4000 mg/l). These results show that the TDS can be predicted using ML algorithms with high accuracy. Except for the SVM and kNN, all other models are generally accurate in terms of magnitude and trend for predict-

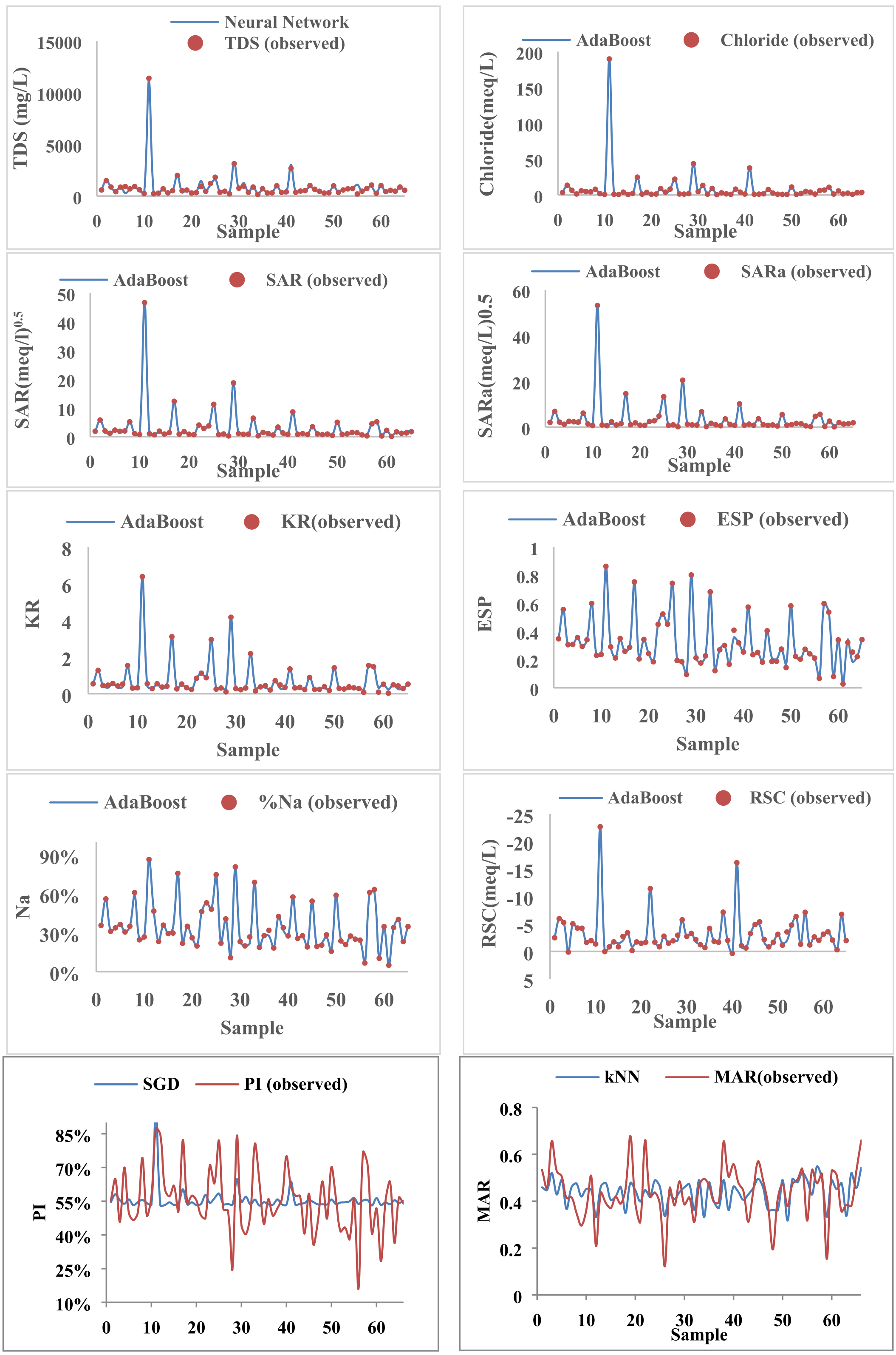

Fig. 3. Graphical plots of both the observed parameter values and the predicted values by the best models for the validation process.

ing the chloride in the water surface during the training. Importantly, the ANN, MLR, SGD and, Adaboost are also the best models for Chloride prediction in Bouregreg watershed with r, 0.99, 0.99, 0.99, 0.98 and RMSE 3, 3, 4, 5 meq/L respectively for a range of [0.42, 208.8] (meq/L). As for the SVM and kNN, they have a high RBIAS up to −47% and −34% indicating the inaccurate prediction of these models.

According to the results presented in Table 3, except SVM, all other ML models that were in the process of training have a coefficient of correlation of more than 0.89 that demonstrates great prediction accuracy for SAR and SAR_a. Moreover, based on the RMSE and RBIAS, the SGD (1.68) followed by ANN (1.7), MLR (1.7), Adaboost (2.47) and DT (2.8) with the RBIAS less than 14% in absolute value are the most accurate models for predicting the SAR with a range of [0.11, 46.64] $(meq/L)^{0.5}$ and the same models are good for SAR_a prediction with RMSE and RBIAS less than 3.2 and 14% in absolute value respectively for data range of [0.13–53.2] $(meq/L)^{0.5}$. For the Kelly Ratio (KR), it is observed, except for SVM; that all other models have acceptable accuracy with a coefficient of correlation (r) between 0.8 and 0.84 and maximal RMSE up to 0.61 for KR with a range of [0.03, 6.38]. For the ESP and %Na parameters, the high coefficient of correlation is 0.68 given both by RF and kNN models for ESP prediction and 0.64 given by kNN for %Na. Such results demonstrate a relatively good correlation between the observed and predicted values. Besides, the smaller values of RBIAS and RMSE of all models, excluding SVM, demonstrate the potential accuracy of ML models for ESP and % Na prediction. For Magnesium Absorption Ration (MAR), all models have poor coefficients of correlations with a high value of 0.26 for DT model. It is observed that the models cannot fit the value of more than 0.6 and less than 0.25. Also, for the Permeability Index (PI) parameter, the coefficients of correlations are poor because the high value of r is about 0.43 provided by SVM models. However, the small values of RBIAS and RMSE indicate relative accurate prediction for all models. It is observed that all models failed to predict the extreme values of MAR and PI parameters. Nevertheless, to further pursue the investigation, the removal of the outlier extremes is carried out and has improved the coefficient of correlation from 0.24 to 0.29 ,but the new RMSE is increased which becomes significant compared to the range of MAR while no improvement of the prediction accuracy is observed for PI. Furthermore, increasing the training dataset to 264 data pairs did not improve the prediction accuracy for all models. Such results show that all the investigated models in this study presented unbalanced prediction accuracy of MAR and PI parameters. The results show that all models demonstrate viable prediction accuracy for the RSC parameter. Notably, ANN, MLR, and SGD presented satisfactory performance with r, 0.83 and RBIAS 3% and RMSE less than 2.11 (meq/L) for a range of [−26.64, 3.08] (meq/L).

4.2. Validation of machine learning models

After training the developed ML models, the study uses 66 samples that are never observed by the models for the validation process. The EC and pH used as inputs of all models even for models that show low accurate prediction during the training phase and the output models are compared to observed values using statistical performances such as r, RMSE and EBIAS. Except for SVM and kNN models, the results show that all other models present high performances for predicting TDS, Chloride, SAR, SAR_a, KR, ESP, Na and RSC. It is observed that the prediction accuracy is high compared to that which observed in the training process because the RMSE is decreased to 100 (mg/L) for TDS given by ANN model and to 0.22 (meq/L) for Chloride given by AdaBoost Model. More importantly, AdaBosst model shows the high accurate prediction of Chloride, SAR, SAR_a, ESP, Na, and RSC with r that is found to be 0.999 and presented excellent agreement between observed and predicted values compared to that in the training process.

Interestingly, the results shows that the lower prediction accuracies are presented by the Tree model for the TDS (with r = 0.85 and RMSE = 422 mg/L), Chloride (with r = 0.85 and RMSE = 7.3 m eq/L), SAR (with r = 0.84 and RMSE = 1.43 $(meq/L)^{0.5}$), and SAR_a (with r = 0.83 and RMSE = 1.6 $(meq/L)^{0.5}$) parameters, Random Forest model for KR parameter (with r = 0.82 and RMSE = 0.25) and the

Table 4
Model generalisation results for Cherrate watershed.

IWQ Parameters	Performance	Models					
		ANN	MLR	DT	RF	SGD	AdaBoost
TDS	r	0.87	0.87	0.61	0.72	0.87	0.50
	RBIAS	−2%	−2%	−5%	−19%	−1%	−25%
	RMSE	200	200	469	361	201	505
Chloride	r	0.13	0.13	0.23	0.28	0.13	0.26
	RBIAS	8%	8%	−13%	−21%	10%	−26%
	RMSE	8.45	8.45	7.40	7.21	8.54	7.34
SAR	r	0.79	0.79	0.56	0.58	0.78	0.59
	RBIAS	35%	35%	4%	5%	35%	3%
	RMSE	1.56	1.56	2.26	2.16	1.54	2.34
SAR_a	r	0.76	0.76	0.50	0.52	0.76	0.50
	RBIAS	28%	28%	−8%	−2%	29%	−10%
	RMSE	1.65	1.65	2.63	2.49	1.68	2.95
KR	r	0.59	0.59	0.40	0.39	0.59	0.43
	RBIAS	36%	36%	16%	26%	34%	18%
	RMSE	0.31	0.31	0.54	0.69	0.30	0.64
ESP	r	0.60	0.60	0.29	0.51	0.58	0.37
	RBIAS	−11%	−11%	0%	8%	−11%	−4%
	RMSE	0.07	0.07	0.15	0.11	0.08	0.15
%Na	r	0.56	0.56	0.34	0.49	0.56	0.32
	RBIAS	−6%	−6%	2%	3%	−7%	1%
	RMSE	0.07	0.07	0.15	0.11	0.07	0.15
RSC	r	0.66	0.66	0.46	0.64	0.66	0.33
	RBIAS	−45%	−45%	−12%	−21%	−45%	−31%
	RMSE	5.1	5.1	4.2	4.0	5.1	5.1

SGD model for ESP (with r = 0.65 and RMSE = 0.08), %Na (with r = 0.64 and RMSE = 0.07) and RSC (with r = 0.80 and RMSE = 1.24 meq/L) parameters. These results, however, are still satisfactory and show better performances of the models for predicting IWQ parameters. Such results may be indicating the undertraining of these models.

Fig. 3 shows graphs of the observed values of the IWQ parameters and the predicted values according to the validated models. As shown in this figure, with the exception of the MAR and PI parameters, the predicted values of the IWQ parameters for the 6 best ML models are closer to the corresponding observed ones, and smoothly follow the observed ones. Generally, it can be observed

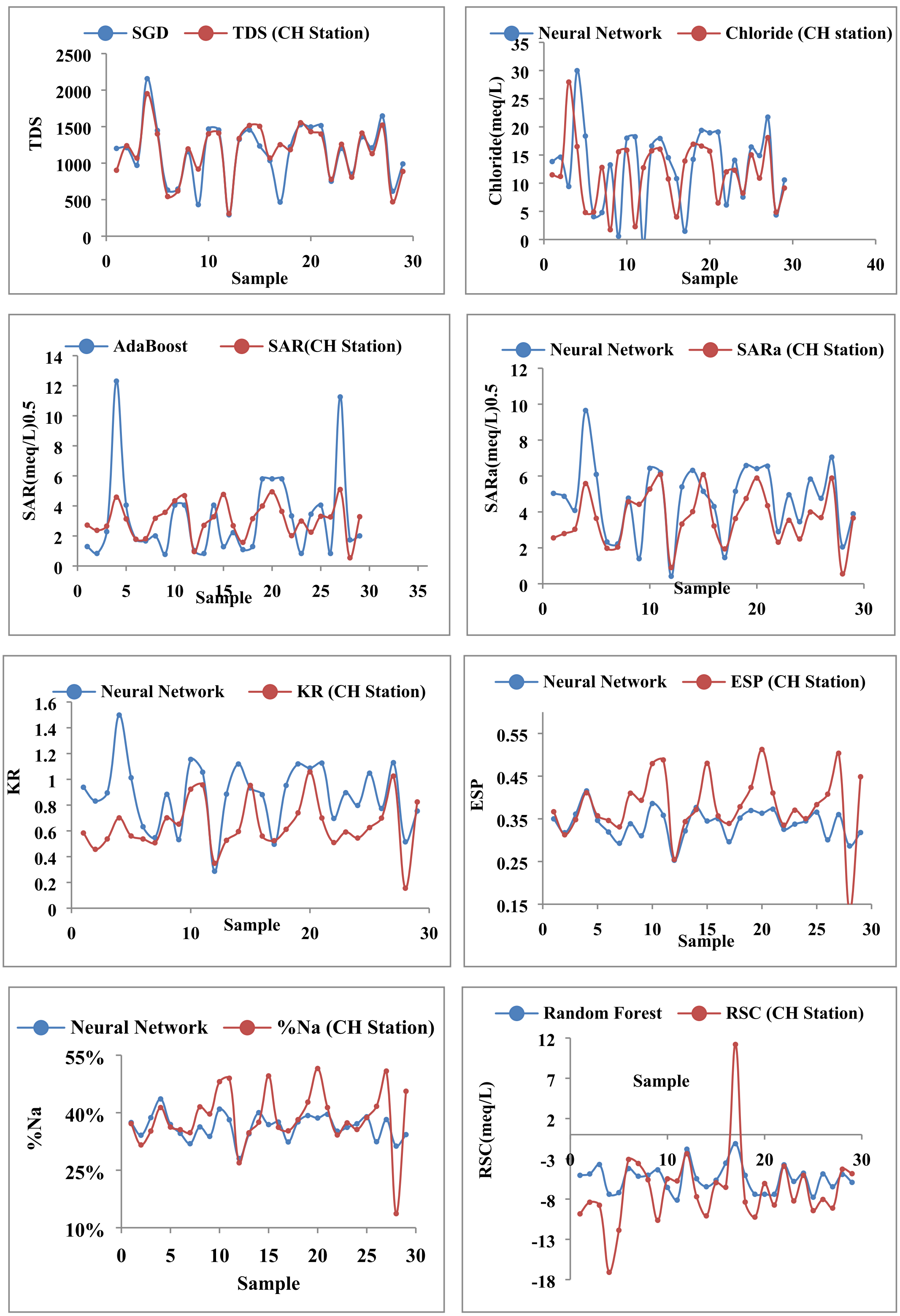

Fig. 4. Graphical plots of both the observed parameter values and the predicted values by the best models for the Cherrate watershed.

that the AdaBoost model provides a reliable and an accurate prediction for seven parameters. With regard to MAR and PI parameters, as Fig. 3 presents, the validation process confirms the failure of all models to predict the extreme values and shows that the models investigated in this study are unable to predict these parameters.

4.3. Machine learning model generalizations

The artificial intelligence techniques often cope easily with the complex systems without explaining their mechanism, but their application for other areas that are different from those used in their development is a great challenge. This study attempts to generalize these models to other areas with 36 and 29 samples collected at Feddane Taba (FT) and Cherrate (CH) monitoring stations respectively (Fig. 1).

4.3.1. Cherrate watershed

Table 4 presents the results of the ML model generalizations for the Cherrate watershed. From these results, it can be seen that the prediction accuracy has decreased compared to the training and the validation phases for Bouregrge watershed. According to this table, it can be seen that the ANN, LR and SGD models perform good results with coefficients of correlations and RMSEs of up to 0.87 and less than 201 mg/l sequentially and, therefore can be implemented to Cherrate watershed for predicting the TDS parameter. The same models have a relative prediction accuracy of SAR and SAR_a parameters. The results also demonstrate that all models are unsuitable for the simulation of other parameters in this watershed.

In addition to that, Fig. 4 shows the graphical plots of the observed parameter values and the model that is best suited for the generalization process to Cherrate watershed. Based on this figure, it is observed that, except for two extreme values, the ANN, MLR, and SGD models simulate the TDS, SAR and SAR_a parameters better. Also, this figure shows an important mismatch between the observed and the predicted values for other parameters.

4.3.2. Nfifikh watershed

The generalization performance results of ML models to Nfifikh watershed are presented in Table 5. These results show that the ANN, MLR, and SGD models accurately predict TDS parameter with a coefficient of correlation and RBIAS, 0.96 and 1%. According to the same table and based on the RBIAS, the Adaboost model successively generalizes chloride parameter in Nfifikh watershed with −4% as a value. Also, the exchangeable sodium percentage and the percentage of sodium are relatively good simulated by the neural network model with a coefficient of correlation and RBIAS 0.66 and 0.65, −13% and 10% respectively. The models have good coefficients of correlations with values of up to 0.87, 0.84 and 0.77 for the SAR, SAR_a and RSC parameters sequentially, indicating that the trends of these parameters are well captured. However, the models failed to predict the magnitude of theses parameter, as the RBIAS is significant with values more than 45%, 40% and 27% for the SAR, SAR_a and RSC parameters respectively. Such results demonstrate that these parameters, in addition to Kelly ratio parameter are poorly simulated, show that all models cannot be generalised for their simulation in this watershed.

Moreover, Fig. 5 illustrates the graphical plots of the observed values of irrigation water quality parameter versus the best suited model values for generalization to Nfifikh watershed. This figure shows the good performance of the SGD and Adaboost models for the simulation of the TDS and chloride parameters respectively, in Nfifikh watershed. It is observed that, except for one extreme value, that the AdaBoost model simulates the chloride with a high agreement between observed and simulated values. The trends of SAR, SAR^a and RSC parameters are fairly captured by the SGD and Tree models respectively. However, these models overestimate the SAR and SAR_a and underestimate the RSC as mentioned in Table 5. Also, the neural network models have a relative agreement between observed and simulated values for the %Na and ESP parameters.

Overall, the results show that the prediction accuracy varies from the models to another, from the parameter to another, and from the watershed to another. This may be due to the model

Table 5
Model generalisation results for Nfifikh watershed.

IWQ Parameter	Performance	Model					
		ANN	MLR	DT	RF	SGD	AdaBoost
TDS	r	0.96	0.96	0.73	0.85	0.96	0.88
	RBIAS	−1%	−1%	−5%	−15%	0%	−13%
	RMSE	248	248	1006	551	250	486
Chloride	r	0.53	0.53	0.19	0.60	0.53	0.64
	RBIAS	21%	21%	16%	−11%	22%	−4%
	RMSE	14.5	14.5	25.2	10.3	14.7	10.5
SAR	r	0.87	0.87	0.73	0.74	0.87	0.69
	RBIAS	45%	45%	72%	73%	44%	58%
	RMSE	3.22	3.22	6.64	5.28	3.17	4.43
SAR_a	r	0.84	0.84	0.71	0.64	0.84	0.58
	RBIAS	40%	40%	69%	64%	40%	45%
	RMSE	3.5	3.5	7.5	6.0	3.5	5.0
KR	r	0.64	0.64	0.53	0.50	0.64	0.52
	RBIAS	53%	53%	121%	165%	49%	138%
	RMSE	0.63	0.63	1.51	1.88	0.59	1.65
ESP	r	0.66	0.66	0.59	0.63	0.64	0.60
	RBIAS	−13%	−13%	29%	23%	−14%	31%
	RMSE	0.09	0.1	0.19	0.15	0.09	0.21
%Na	r	0.65	0.65	0.55	0.58	0.65	0.59
	RBIAS	−10%	−10%	30%	28%	−10%	35%
	RMSE	8%	8%	19%	19%	8%	22%
RSC	r	0.77	0.77	0.71	0.74	0.77	0.55
	RBIAS	−56%	−56%	−27%	−46%	−57%	−37%
	RMSE	8.43	8.43	5.58	7.19	8.51	7.19

structures, to the inputs used for predicting such parameter, and to the conditions of the watershed.

5. Discussion

Assessing the irrigation water quality requires a large number of chemical parameter analyses which are considered cumbersome and very expensive for farmers particularly in developing countries. In this study, various machine learning models (ANN, LR, Tree, RF, SVR, k-NN, SGD and Adaboost) are investigated to predict the TDS, chloride, SAR, SAR_a, KR, PI, MAR, %Na, ESP and RSC parameters of surface water of Bouregreg in Morocco. The data of a total of 264 samples are selected for the training and the validation of the models. Importantly, the adaptive boosting model presents better performance during the validation phase. It is found that the SVR and k-NN models are inaccurate models for predicting the parameters of IWQ assessment. This result is also found by Wang et al. (2020) in predicting anaerobic digestion performance using only 17 samples. However, the significant increase of training dataset or using ensemble models and deep forest improves the classification prediction accuracy of SVM and k-NN models (Chen et al., 2020; Zhou and Feng, 2019). Furthermore, all models fail to predict the MAR and PI parameters. This unsuccessful prediction accuracy of the MAR and PI parameters is possibly due to the low relationship between the EC and the pH used as input

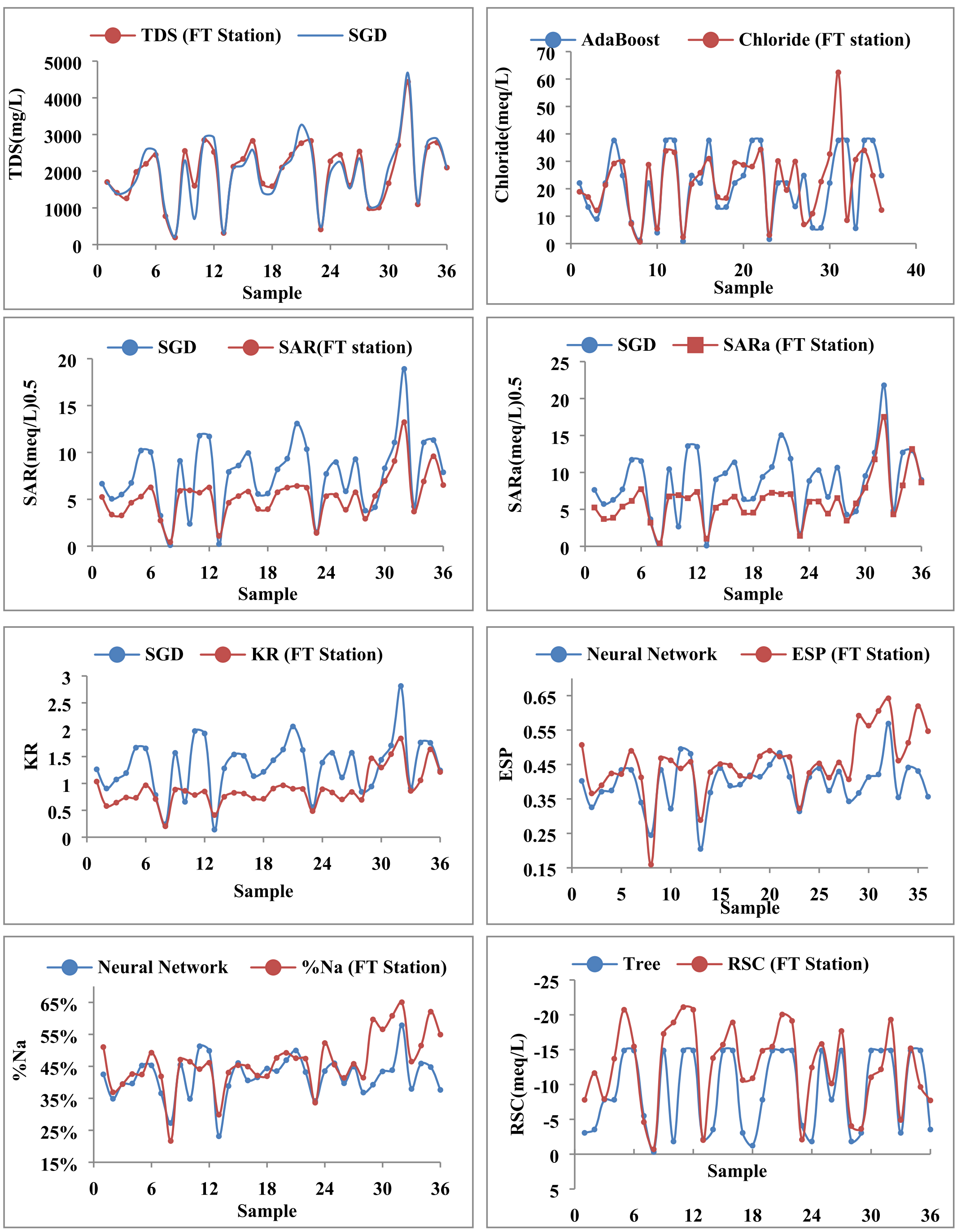

Fig. 5. Graphical plots of both the observed parameter values and the predicted values by the best models for the Nfifikh watershed.

variables and these parameters. Additionally, it is observed that the more the correlation between the input and output variables is significant, the more the performances of the models are high (examples of TDS, chloride, SAR and SAR_a parameters). This indicates that the selection of inputs variables is one of the keystones of machine learning models to have better performances. Also, the investigation of classification models can be efficient in predicting MAR and PI parameters because in this study the outputs are real numbers that may be more difficult to accurately predict them than the class outputs. For example, when a model predicts a value that is different from the one observed, the error can be numerically significant therefore, the model will not be evaluated in a suitable way, but it could be acceptable when the same predicted value includes a class to be predicted that corresponds to the one observed.

However, the results of a high accurate prediction of artificial intelligence techniques for water quality are confirmed by Ahmed et al. (2019), Castrillo and García (2020), Chen et al. (2020), Di et al. (2019), Liu et al. (2019), Lu and Ma (2020) and Wang et al. (2017). The accurate prediction highly depends on the number of input variables and their impact, but the availability of all data with the effective cost is a crucial task. In the present time, there are a few studies that used the parameter which could be measured in situ and in real-time as input variables such as Castrillo and García (2020) study. On the other hand, the generalization of these studies for areas that are different from that used in development must be discussed because there are a lot of variables that significantly impact water quality such as the hydrological regime, land use, geomorphological and geological conditions as well as the anthropogenic activities. These conditions play an important role in the combinations of the input variables used in the predicting process.

The application of ML models to predict 10 IWQ parameters, using only the electrical conductivity and pH as input variables, shows high prediction accuracy for 8 parameters by 6 models in Bouregreg watershed. Also, their generalization shows that the models are fairly suitable to predict the TDS, SAR, SAR_a parameters in the Cherrate watershed and the TDS, chloride, ESP and the %Na parameters in the Nfifikh watershed. However, using the river flow as a predictor may be valuable to improve the generalization of the ML model, and therefore it is suggested for future work.

The results of this study are valuable in that they can improve monitoring the quality of the irrigation water in Bouregreg watershed in real time by the implementation of the automated sensor technologies to measure the electrical conductivity and pH parameters using the best suited models. This is mainly helpful, particularly, for the projected dam reservoirs for agricultural purposes where the evaporation significantly deteriorates the chemical water quality especially during summer periods. Consequently, this work will be helpful for farmers to manage water quality with an effective cost and in a short time. However, given that the evaluation of water for irrigation purposes depends on the type of soil crop according to the water quality class, the classification ML model is suggested for future investigation.

6. Conclusion and recommendations

In this study 8 machine learning models were developed and evaluated to predict 10 IWQ parameters, the generalization of the best models for two other areas was evaluated. The results revealed that the ML models are valuable to overcome some limitations of conventional approaches in evaluating the water suitability for agricultural purposes. The main conclusions that come out of this study are:

(1) Among 8 Machine learning models, 6 models are valuable for accurately predicting 8 from 10 IWQ parameters using the EC and pH as input variables in the surface water of Bouregreg. This finding can help farmers improve monitoring the irrigation water quality by implementing a sensor to measure the input features at water resources that are used for agricultural purposes;
(2) Not all models have been able to numerically predict the magnesium absorption ration (MAR) and the permeability index (PI), so classification models may be able to improve the accuracy of predictions;
(3) Adaptive Boosting model presents better performances compared to other models in Bouregreg watershed;

Conventional methods are widely used for the evaluation of water suitability for irrigation based on a large number of parameters and indexes and are considered efficient tools, but they can be expensive. Rather than relying on these methods for evaluating and predicting the suitability of water for irrigation purposes, the machine learning models (classification and regression models) are suggested for future work and their implementation in the field of engineering application is recommended to improve the control of water quality. Since they do not only present better performances, but they could be valuable to significantly reduce the cost and time for the control of irrigation water quality.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors acknowledge the River basin agency of Bouregrge and Chaouia teams for their help and their support. The authors acknowledge three anonymous reviewers for their constructive suggestions and comments.

Funding sources

The authors have not received any funding.

References

Ahmed, A.N., Binti Othman, F., Abdulmohsin Afan, H., Khaleel Ibrahim, R., Ming Fai, C., Shabbir Hossain, M.D., Ehteram, M., Elshafie, A., 2019. Machine learning methods for better water quality prediction. J. Hydrol. 578, 124084. https://doi.org/10.1016/j.jhydrol.2019.124084.

Ahmed, U., Mumtaz, R., Anwar, H., Shah, A.A., Irfan, R., 2019. Efficient water quality prediction using supervised machine learning. Water (Switzerland), 1–14. https://doi.org/10.3390/w11112210.

Ayers, R.S., Westcot, D.W., 1994. Food, Agriculture Organization of the United Nations (FAO), Water Quality for Agriculture, Irrigation and Drainage, Rome, Paper No. 29. Rev1, M-56.

Bortolini, L., Maucieri, C., Borin, M., 2018. A tool for the evaluation of irrigation water quality in the arid and semi-arid regions. Agronomy 8, 1–15. https://doi.org/10.3390/agronomy8020023.

Burba, F., Ferraty, F., Vieu, P., 2009. K-nearest neighbour method in functional nonparametric regression. J. Nonparametr. Stat. 21, 453–469. https://doi.org/10.1080/10485250802668909.

Busico, G., Kazakis, N., Cuoco, E., Colombani, N., Tedesco, D., Voudouris, K., Mastrocicco, M., 2020. A novel hybrid method of specific vulnerability to anthropogenic pollution using multivariate statistical and regression analyses. Water Res. 171, 115386. https://doi.org/10.1016/j.watres.2019.115386.

Castrillo, M., García, Á.L., 2020. Estimation of high frequency nutrient concentrations from water quality surrogates using machine learning methods. Water Res. 172, 115490. https://doi.org/10.1016/j.watres.2020.115490.

Chaoua, S., Boussaa, S., El Gharmali, A., Boumezzough, A., 2018. Impact of irrigation with wastewater on accumulation of heavy metals in soil and crops in the region of Marrakech in Morocco. J. Saudi Soc. Agric. Sci. 18, 429–436. https://doi.org/10.1016/j.jssas.2018.02.003.

Chen, K., Chen, H., Zhou, C., Huang, Y., Qi, X., Shen, R., Liu, F., Zuo, M., Zou, X., Wang, J., Zhang, Y., Chen, D., Chen, X., Deng, Y., Ren, H., 2020. 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. 171, 115454. https://doi.org/10.1016/j.watres.2019.115454.

Di, Z., Chang, M., Guo, P., 2019. Water quality evaluation of the Yangtze River in China using machine learning techniques and data monitoring on different time scales. Water (Switzerland) 11. https://doi.org/10.3390/w11020339.

Doneen, L.D., 1964. Water Quality for Agriculture. Dep. Irrig. Univ. California, Davis, p. 48.

Droogers, P., Immerzeel, W.W., Terink, W., Hoogeveen, J., Bierkens, M.F.P., Van Beek, L.P.H., Debele, B., 2012. Water resources trends in Middle East and North Africa towards 2050. Hydrol. Earth Syst. Sci. 16, 3101–3114.

Ewaid, S.H., Abed, S.A., Kadhum, S.A., 2018. Predicting the Tigris River water quality within Baghdad, Iraq by using water quality index and regression analysis. Environ. Technol. Innov. 11, 390–398. https://doi.org/10.1016/j.eti.2018.06.013.

Ewaid, S.H., Kadhum, S.A., Abed, S.A., Salih, R.M., 2019. Development and evaluation of irrigation water quality guide using IWQG vol 1 software: A case study of Al-Gharraf Canal, Southern Iraq. Environ. Technol. Innov. 13, 224–232. https://doi.org/10.1016/j.eti.2018.12.001.

Ewaid, S.H., Abed, S.A., 2017. Water quality assessment of Al-Gharraf River, South of Iraq using multivariate statistical techniques. J. Al-Nahrain Univ. 20, 114–122. https://doi.org/10.22401/juns.20.2.16.

FAO, 2002. The Role of Agriculture in the Development of Least-developed Countries and their Integration into the World Economy, 7. FAO.

Fijani, E., Barzegar, R., Deo, R., Tziritis, E., Konstantinos, S., 2019. Science of the Total Environment 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. 648, 839–853. https://doi.org/10.1016/j.scitotenv.2018.08.221.

Fischer, G., Nachtergaele, F., Prieler, S., Van Velthuizen, H.T., Verelst, L., Wiberg, D., 2008. Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008). IIASA, Laxenburg, Austria FAO, Rome, Italy 10.

Freund, Y., Schapire, R.E., 1996. Experiments with a new boosting algorithm. Icml. Citeseer, 148–156.

Gidey, A., 2018. Geospatial distribution modeling and determining suitability of groundwater quality for irrigation purpose using geospatial methods and water quality index (WQI) in Northern Ethiopia. Appl. Water Sci. 8, 1–16. https://doi.org/10.1007/s13201-018-0722-x.

Govindaraju, R.S., 2000. Artificial neural networks in hydrology. II: hydrologic applications. J. Hydrol. Eng. 5, 124–137.

Haykin, S., 1999. Neural Networks, A Comprehensive Foundation, Prentice-Hall Inc. Up. Saddle River, New Jersey 7458, 161–175.

Ho, T.K., 1995. Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition. IEEE, pp. 278–282. https://doi.org/10.1109/ICDAR.1995.598994.

Jahin, H.S., Abuzaid, A.S., Abdellatif, A.D., 2020. Using multivariate analysis to develop irrigation water quality index for surface water in Kafr El-Sheikh Governorate, Egypt. Environ. Technol. Innov. 17. https://doi.org/10.1016/j.eti.2019.100532.

Kelley, W.P., 1963. Use of saline irrigation water. Soil Sci. 95, 385–391.

Kopittke, P.M., So, H.B., Menzies, N.W., 2006. Effect of ionic strength and clay mineralogy on Na–Ca exchange and the SAR–ESP relationship. Eur. J. Soil Sci. 57, 626–633.

Leong, W.C., Bahadori, A., Zhang, J., Ahmad, Z., 2019. Prediction of water quality index (WQI) using support vector machine (SVM) and least square- support vector machine (LS-SVM). Intl. J. River Basin Manag., 1–8 https://doi.org/10.1080/15715124.2019.1628030.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2, 18–22.

Liu, P., Wang, J., Sangaiah, A., Xie, Y., Yin, X., 2019. Analysis and prediction of water quality using LSTM deep neural networks in IoT environment. Sustainability 11, 2058. https://doi.org/10.3390/su11072058.

Lu, H., Ma, X., 2020. Hybrid decision tree-based machine learning models for short-term water quality prediction. Chemosphere 249, 126169. https://doi.org/10.1016/j.chemosphere.2020.126169.

Mbizvo, G.K., Bennett, K., Simpson, C.R., Susan, E., Chin, R.F.M., 2019. ur na l P of. Epilepsy Res., 106192 https://doi.org/10.1016/j.eplepsyres.2019.106192.

Meireles, A.C.M., de Andrade, E.M., Chaves, L.C.G., Frischkorn, H., Crisostomo, L.A., 2010. A new proposal of the classification of irrigation water. Rev. Ciência Agronômica 41, 349–357. https://doi.org/10.1590/s1806-66902010000300005.

Richards, L.A., 1954. Diagnosis and Improvement of. Saline Alkali Soils. Handb. 60. https://www.ars.usda.gov/ARSUserFiles/20360500/hb60_pdf/hb60complete.pdf.

Lesch, S.M., Suarez, D.L., 2009. Technical note: a short note on calculating the adjusted SAR Index. Trans. ASABE 52, 493–496. https://doi.org/10.13031/2013.26842.

Schapire, R.E., 1999. A brief introduction to boosting. IJCAI Int. Jt. Conf. Artif. Intell. 2, 1401–1406.

Singh, S., Ghosh, N.C., Gurjar, S., Krishan, G., Kumar, S., Berwal, P., 2018. Index-based assessment of suitability of water quality for irrigation purpose under Indian conditions. Environ. Monit. Assess. 190. https://doi.org/10.1007/s10661-017-6407-3.

Tyagi, S., Sharma, B., 2014. Water quality assessment in terms of water quality index water quality assessment in terms of water quality index water quality assessment in terms of water quality index. Am. J. Water Resour. 1 (3), 34–38. https://doi.org/10.12691/ajwr-1-3-3.

Vapnik, V.N., 1995. The nature of statistical learning. Theory.

Wagh, V.M., Panaskar, D.B., Muley, A.A., Mukate, S.V., Lolage, Y.P., Aamalawar, M.L., 2016. Prediction of groundwater suitability for irrigation using artificial neural network model: a case study of Nanded tehsil, Maharashtra, India. Model. Earth Syst. Environ. 2, 1–10. https://doi.org/10.1007/s40808-016-0250-3.

Wang, L., Long, F., Liao, W., Liu, H., 2020. Prediction of anaerobic digestion performance and identification of critical operational parameters using machine learning algorithms. Bioresour. Technol. 298, 122495. https://doi.org/10.1016/j.biortech.2019.122495.

Wang, X., Zhang, F., Ding, J., 2017. Evaluation of water quality based on a machine learning algorithm and water quality index for the Ebinur Lake Watershed. China. Sci. Rep. 7, 1–18. https://doi.org/10.1038/s41598-017-12853-y.

Zhou, Z., Feng, J., 2019. Deep forest, 74–86. https://doi.org/10.1093/nsr/nwy108.