

Research Article

Predicting Irrigation Water Quality Indices Based on Data-Driven Algorithms: Case Study in Semiarid Environment

Dimple ¹, Jitendra Rajput ^{2,3}, Nadhir Al-Ansari ⁴, and Ahmed Elbeltagi ⁵

¹SWE Department, CTAE, Maharana Pratap University of Agriculture and Technology, Udaipur 313001, Rajasthan, India

²Division of Agricultural Engineering, IARI, New Delhi, India

³Water Technology Centre, ICAR-Indian Agricultural Research Institute (IARI), New Delhi, India

⁴Environmental and Natural Resources Engineering, Lulea University of Technology, Lulea 97187, Sweden

⁵Agricultural Engineering Dept, Faculty of Agriculture, Mansoura University, Mansoura 35516, Egypt

Correspondence should be addressed to Nadhir Al-Ansari; nadhir.alansari@ltu.se

Received 6 May 2022; Revised 26 July 2022; Accepted 28 July 2022; Published 29 August 2022

Academic Editor: Islam M. Al Akraa

Copyright © 2022 Dimple et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ascertaining water quality for irrigational use by employing conventional methods is often time taking and expensive due to the determination of multiple parameters needed, especially in developing countries. Therefore, constructing precise and adequate models may be beneficial in resolving this problem in agricultural water management to determine the suitable water quality classes for optimal crop yield production. To achieve this objective, five machine learning (ML) models, namely linear regression (LR), random subspace (RSS), additive regression (AR), reduced error pruning tree (REPTree), and support vector machine (SVM), have been developed and tested for predicting of six irrigation water quality (IWQ) indices such as sodium adsorption ratio (SAR), percent sodium (%Na), permeability index (PI), Kelly ratio (KR), soluble sodium percentage (SSP), and magnesium hazards (MH) in groundwater of the Nand Samand catchment of Rajasthan. The accuracy of these models was determined serially using the mean squared error (MSE), correlation coefficients (r), mean absolute error (MAE), and root mean square error (RMSE). The SVM model showed the best-fit model for all irrigation indices during testing, that is, RMSE: 0.0662, 4.0568, 3.0168, 0.1113, 3.7046, and 5.1066; r : 0.9364, 0.9618, 0.9588, 0.9819, 0.9547, and 0.8903; MSE: 0.004381, 16.45781, 9.101218, 0.012383, 13.72447, and 26.078; MAE: 0.042, 3.1999, 2.3584, 0.0726, 2.9603, and 4.0582 for KR, MH, SSP, SAR, %Na, and PI, respectively. The KR and SAR values were predicted accurately by the SVM model in comparison to the observed values. As a result, machine learning algorithms can improve irrigation water quality characteristics, which is critical for farmers and crop management in various irrigation procedures. Additionally, the findings of this research suggest that ML models are effective tools for reliably predicting groundwater quality using general water quality parameters that may be acquired directly on periodical basis. Assessment of water quality indices may also help in deriving optimal strategies to utilise inferior quality water conjunctively with fresh water resources in the water-limited areas.

1. Introduction

Water resources are vital in supplying domestic water, industrial processes, and agriculture. Indeed, improved water quality minimises the cost of water treatment for domestic and industrial purposes and boosts agricultural yield. Demand for water is accelerating as a result of populace growth, intensive agriculture, urbanisation, and industrialisation. Because anthropogenic activities and natural pollution sources endanger water resources in ways that go beyond

their suitability for drinking, irrigation, industrial, and other uses [1, 2], water quality evaluation and prediction are required to establish whether or not a body of water is suitable for a particular purpose and, if it is not, to decide the right remedies or precautions to take. The world's growth and development have resulted in widespread contamination from precipitation outlets such as rivers [3]. Agriculture is the biggest consumer of water, accounting for up to 80% of total consumption, and is also a significant source of water pollution. Thus, water planning and management with an

eye on cost-effective irrigation is necessary to ensure the sustainability of agriculture [4]. Because groundwater is positioned beneath the land surface and is normally not in contact with the atmosphere, it is generally regarded to be safe for ingestion [5]. However, its quality may be harmed as a result of anthropogenic activities such as improper waste disposal and the application of agrochemicals [6]. Naturally, the dissolution of minerals in rocks can have an effect on the quality of groundwater [7].

Artificial intelligence (AI) systems have been researched in recent years and have demonstrated a high capacity for forecasting and monitoring water quality [8–11]. Machine learning (ML), deep learning (DL), and artificial neural networks (ANN) are some of these techniques. For instance, Ahmed et al. [8] explored machine learning models and demonstrated how precise this technique is at predicting water quality for domestic purposes. Leong et al. [12] validated the accuracy of the support vector machine (SVM) model in estimating the water quality index. Nowadays, there are few studies using artificial intelligence to evaluate and forecast the quality of irrigation water. Wagh et al. [13], on the other hand, employed ANN to estimate the suitability of groundwater for irrigation in India, utilising 13 physicochemical properties as input variables. They revealed that the data-driven model outperforms other models in terms of water appropriateness for irrigation use. According to Kouadri et al. [14], eight artificial intelligence algorithms were used to generate WQI predictions in the Illizi region of South East Algeria, including multilinear regression, M5P tree, random subspace, additive regression, random forest, artificial neural network, support vector regression, and locally weighted linear regression. The authors of [15] examined four meta-heuristic algorithms, including the support vector machine, the random tree, the reduced error pruning tree, and the random subspace technique. Similarly, the SVM model was used to predict marine water quality [16] and to monitor wastewater treatment plants [17], with varying degrees of precision. Furthermore, the following studies successfully used machine learning models for the assessment of surface water quality. A study was carried out by the authors of [18] on the Karoun river for the determination of three indices, that is, biochemical oxygen demand (BOD), dissolved oxygen, and chemical oxygen demand (COD), by employing three algorithms and concluded that the EPR model gives the best results during training and testing. The authors of [19] used multivariate adaptive regression spline (MARS) and least square-support vector machine (LS-SVM), as machine learning techniques, to calculate indices of the 5-day biochemical oxygen demand (BOD5) and COD for the Karoun river located in Khuzestan Iran. The results of the research depicted that the LS-SVM-RBF and LS-SVM-Poly methods have given the relatively accurate prediction for BOD5 and COD indices. A study [20] was conducted to determine the water quality index of the surface water resources of the Karoun river watershed for the purpose they used the data-driven models (DDMs) and 12 water quality parameters. Results showed that the FS-M5 MT had the best result for the determination of water quality index classification. The authors of [21] proposed the

multiple-kernel support vector regression (MKSVR) algorithm for the calculation of COD and BOD. Results of the MKSVR were compared with the support vector regression and random forest regression (RFR). The study suggested that the use of MKSVR along with the particle swarm optimisation algorithm could give the superiority of the newly developed support vector machine technique for the water quality parameters determination in natural streams. A study [22] was conducted for the assessment of groundwater quality index in the Rafsanjan plain. The authors used the four robust data-driven techniques. The study reported that the evolutionary polynomial regression (EPR) gave the best results among the rest of the algorithms.

Notably, all published studies have proved that machine learning algorithms are capable of accurately forecasting water quality. The goal of this work is to generate and assess five machine learning (ML) models capable of numerically predicting the quality of water irrigation parameters required to determine its suitability for agricultural usage. Irrigation water just as domestic water quality is dependent on the source of the water, flow path, geology, and processes such as weathering, ion exchange, adsorption, and dissolution. Because the quality of groundwater plays an important role in the sustainability of irrigation, the purpose of this study is to determine the quality and usability of groundwater for irrigational purposes in the Nand Samand catchment.

1.1. Problem Statement. Rajasthan is the largest state in India with just about 1.16% surface water resources of the country. Most of the rivers of the state are rain-fed having no defined drainage basin. Due to the scarcity of surface water resources, about 94% of drinking water delivery schemes and 70% of irrigation schemes rely on groundwater [23]. Deteriorating groundwater quality due to natural and artificial sources is a major concern to meet the water of suitable quality for human consumption and irrigation use. Therefore, the present study aims to assess the irrigation water quality indices employing the machine learning models in the Nand Samand catchment.

2. Materials and Methods

2.1. Study Area Description. The Nand Samand catchment covers an area of Rajasamand and Udaipur. It lies between $72^{\circ}59'$ to $73^{\circ}59'$ E longitude and $24^{\circ}0'$ to $26^{\circ}0'$ N latitude. Figure 1 shows the research area location map. The research area covers a survey of India (SOI) toposheets number of 45G-12, 45H-5, 6, 9, 10, 13, 43NG-9 on a 1:50,000 scale. The total area of the catchment is 865.18 km^2 with the highest elevation of about 1,318 m and the lowest elevation of 570 m above mean sea level (MSL).

2.2. Data Collection. The groundwater samples were taken in 2019 during the pre- and postmonsoon period from open wells (95 sites), which are extensively utilised for drinking and irrigation in the Nand Samand catchment area. The identification of the sampling points was performed using

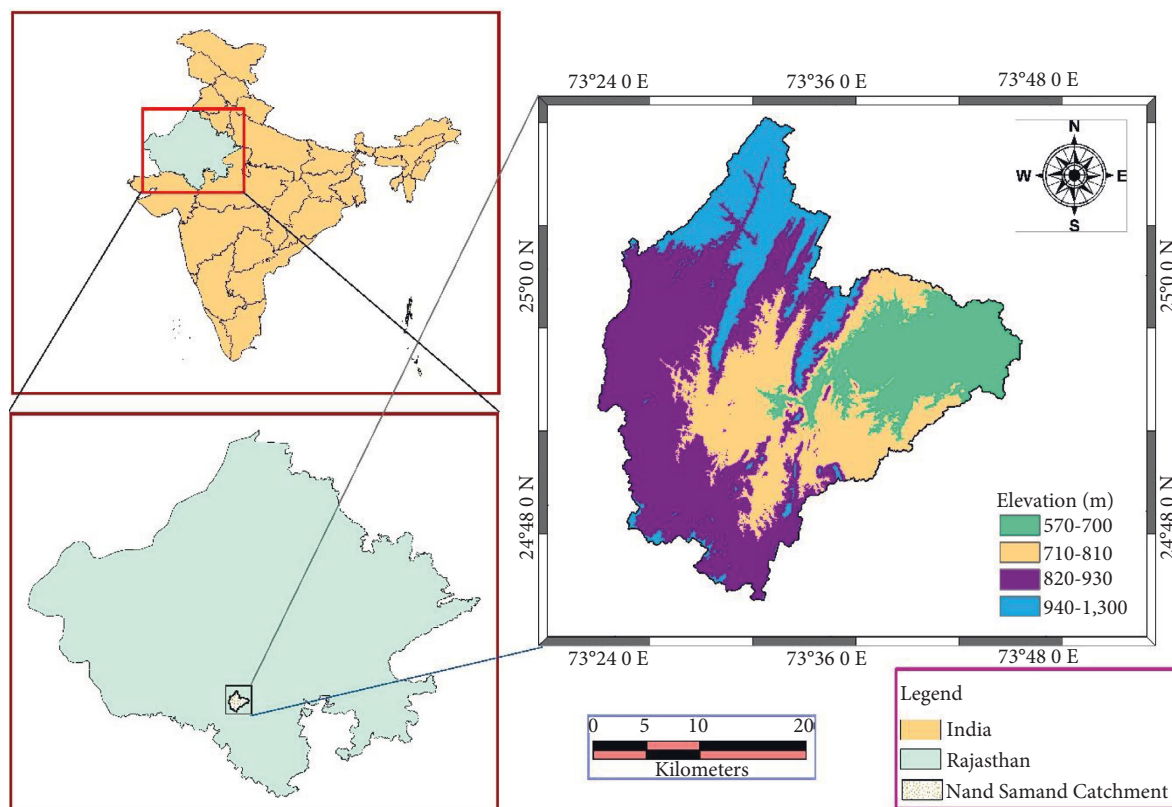


FIGURE 1: Study area map.

Global Positioning System (GPS), and the study area location map was prepared using ArcGIS 10.1 (ESRI, California).

2.3. Methods and Data Preprocessing. Ninety-five water samples were taken at the Nand Samand catchment from 95 monitoring stations. The parameters measured and analysed are: pH, EC, TDS, Ca^{2+} , Mg^{2+} , Na^+ , K^+ , Cl^- , CO_3^{2-} , HCO_3^- , and SO_4^{2-} . The EC, pH, and TDS were evaluated during the sampling protocol using EC, pH, and TDS meter, while the other parameters were analysed in the laboratory that uses the flame photometer for Na^+ and K^+ , titrimetric method for SO_4^{2-} , and titration method for Mg^{2+} , Ca^{2+} , Cl^- , CO_3^{2-} , and HCO_3^- . These samples were used to assess the generalisability of the machine learning model. The irrigation water quality parameters evaluated include SAR, SSP, KR, PI, MH, and % Na and their descriptive statistics using XLSTAT-2021. The statistical characteristics of these parameters are shown in Table 1.

2.4. Machine Learning Models. Machine learning (ML), as one of the methods used in artificial intelligence (Table 2), presented the parameters used for ML models in the present study. At the moment, machine learning models are being utilised to precisely estimate the majority of groundwater quality characteristics and demonstrate their efficiency

[14, 23, 24]. We have utilised 95 water samples for this model. This study uses 76% and 24% data for LR, RSS, AR, REPTree, and SVM models' prediction purposes in training and testing, respectively. Five machine learning models were developed in this study to forecast irrigation water quality indicators, more precisely: LR, RSS, AR, REPTree, and SVM models. Therefore, five machine learning models, LR, RSS, AR, REPTree, and SVM, were selected. The flowchart in Figure 2 illustrates the methodology's steps.

2.4.1. Linear Regression (LR). LR is a conventional statistical technique that describes a target variable Y (referred to as a response variable) as a linear function of a collection of researcher-controlled parameters X_j (called regressors or predictors). The multiple LR model can be represented in the following manner:

$$Y_i = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_k x_{i,k} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (1)$$

where n denotes the sample size. The β_j parameters of the model are estimated using the least squares criteria. The key advantage of the LR technique over other ML methods is the short computing time required to estimate the model's parameters. It enables inferences on regression parameters and predictions under an appropriate theoretical framework. Although the LR technique has demonstrated good performance in a variety of situations and fields, it is

TABLE 1: The descriptive statistics for irrigation water quality and input parameters.

	EC	pH	TDS	Na ⁺	K ⁺	HCO ₃ ⁻	Mg ²⁺	Ca ²⁺	SAR	SSP	KR	PI	MH	%Na
Max	6.11	8.00	4,350	17.30	2.71	6.20	21.67	10.84	8.54	71.75	2.54	82.27	71.62	72.41
Mean	2.38	6.00	1,531	3.91	0.93	4.07	4.61	4.61	1.92	29.67	0.48	46.18	47.59	34.69
Min	1.19	0.99	690	1.22	0.10	0.60	1.16	2.30	0.65	10.49	0.12	16.71	15.79	11.00
SD	0.94	0.59	674	2.51	0.62	1.21	3.08	1.49	1.36	11.54	0.38	12.44	12.55	11.06
Skewness	1.88	-0.09	2.33	3.53	0.89	-0.38	3.10	1.33	3.49	1.41	3.36	0.48	-0.04	1.12
Kurtosis	3.83	-1.09	5.99	14.42	0.67	-0.31	13.18	2.83	13.49	2.63	12.98	0.32	-0.69	1.74

EC (dS/m), TDS (ppm), K, HCO₃, Mg and Ca (meq/l), Max – maximum, Min – minimum, and SD – standard deviation.

TABLE 2: The machine learning algorithm used in the study.

Model name	Description of parameters
Linear regression (LR)	Batch size = 100, bag size percent = 100, attribute selection method = M5 method, eliminate collinear attributes = true
Random subspace (RSS)	Batch size = 100, classifier = random forest, random seed = 1, subspace size = 0.5, numbers of executions slots = 1, number of iteration = 10
Additive regression (AR)	Batch size = 100, classifier = bagging, shrinkage = 1, number of iteration = 30
Reduced error pruning tree (REPTree)	Batch size = 100, initial count = 0, number of folds = 3, random seed = 1, minimum proportion of the variance = 0.001, minimum number = 2, max depth = 1
Support vector machine (SVM)	Kernel = poly kernel; batch size = 100, C = 1, regression optimiser = SMO improved; filter type = normalise training data; cache size = 250,000; omega = 1; sigma = 1

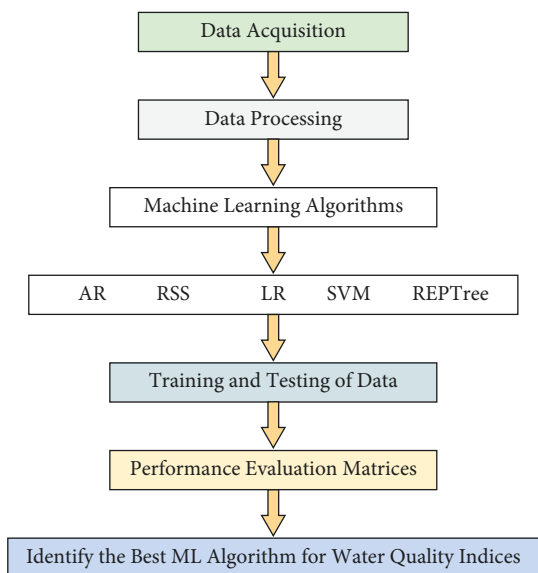


FIGURE 2: Flowchart of the developed methodology.

confined to linear correlations between the response variable and the predictors. However, real-world situations may be nonlinear and complex [25].

2.4.2. Random Subspace (RSS). Ho [26] introduced the RSS model as a revolutionary coupled method for solving natural problems using artificial intelligence. This approach combines and trains numerous classifiers on an updated feature space. The RS inputs are the training set (x), the base classifier (w), and the number of subspaces (L) [27, 28]. This method is strongly recommended by Pham et al. [29] for

avoiding overfitting issues and dealing with the smallest possible data sets. Additionally, when data contains many redundant features, the random subspace gives improved classifiers than the original feature space. The subgroups are randomly chosen from classification training, and the number of subsets is integrated. The first step is to subdivide the initial space into subsets. The results are then calculated using the majority of votes cast by Kushwaha et al. [15] as follows:

$$\beta(x) = \operatorname{argmax}_{y \in \{-1, 1\}} \sum_d \delta_{\text{sng}}(C^b(x)), y, \quad (2)$$

where δ = Kronecker symbol, $y \in \{-1, 1\}$ = decision or class tag of the classifier, and $C^b(x)$ = classification integration ($C = 1, 2, \dots$).

2.4.3. Additive Regression (AR). The generalised additive model (GAM), a variant of the generalised linear model (GLM) [30], has a number of advantages over the latter model. To represent the response, it employs a sum of nonlinear functions, which is based on the theory of non-linear functions, and allows for a more precise representation of the effect of each explanatory variable. This precision makes it a popular technique for modelling the effects of environmental variables, as these effects are frequently nonlinear and difficult to express parametrically [31, 32].

2.4.4. Reduced Error Pruning Tree (REPTree). The REPTree classifier is a rapid decision tree technique that is based on the concept of calculating information acquisition using entropy and reducing variance-induced error [33]. In regression tree modified iterations, the REPTree generates

numerous trees. Then, the best trees produced are chosen. This approach minimises pruning error rates by utilising the linking method. The mistake in the tree's average frame prediction is used to prune the tree. At the start of the modelling process, the values of numerical attributes are sorted. As with the C4.5 Algorithm, this algorithm partitions the corresponding samples and processes the values that are missing [34].

2.4.5. Support Vector Machine (SVM). SVM is a commonly used artificial intelligence technique for pattern recognition, classification, and regression. Because function fitting is a well-known use of SVM, this subset of SVM usage is referred to as support vector regression (SVR). The objective of function fitting with SVM is to minimise error (difference between the model output and observed data). Numerous characteristics of this system make it the perfect choice for use in tackling linear and nonlinear correlation problems [35, 36]. This can be considered an optimisation issue, with the following mathematical expression:

$$\begin{aligned}
 &\text{Minimise} \\
 &R_{\text{svm}}(\omega, \xi^*) = \frac{1}{2}\omega^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*), \\
 &\text{Subject} \\
 &di - \omega\varphi(xi) + bi \leq \varepsilon + \xi_i, \\
 &\omega\varphi(xi) + bi - di \leq \varepsilon + \xi_i, \\
 &\xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, l.
 \end{aligned} \tag{3}$$

where ω is a normal vector, $1/2\omega^2$ is the regularisation factor, C is the error penalty factor, b is a bias, ε is the error function, xi is the input vector, di is the target value, l is the number of elements in the training data set, $\varphi(xi)$ is a feature space, and ξ_i and ξ_i^* are upper and lower excess deviation, respectively [37].

2.5. Agriculture Water Quality Parameters. Six irrigation water quality indices were selected and calculated in this study: KI/KR, MH, PI, percent Na, SAR, and SSP. These indices were used to forecast the value of water quality variables and to determine the acceptable groundwater quality for agricultural applications in the study area.

2.5.1. Kelly Ratio (KR). The Kelly ratio is a key indicator for determining the appropriateness of groundwater for irrigation purposes. Groundwater classified as having a KR value of 0 to 1 is suitable for irrigation purposes; groundwater classified as having a KR value greater than 1 is not suitable for irrigation purposes. The KR of groundwater was estimated using the following formula:

$$KR = \frac{Na^+}{(Ca^{2+} + Mg^{2+})} \tag{4}$$

2.5.2. Magnesium Hazards (MHs). Magnesium hazard is another important indicator for determining the irrigation quality of groundwater. Magnesium concentrations in groundwater are critical for crop productivity and growth. Calcium and magnesium, in general, sustain the state of groundwater equilibrium. Excessive magnesium (Mg^{2+}) degrades soil structure, increasing the soil's alkaline character and inhibiting plant growth. Groundwater is classified as acceptable for irrigation if the MH value is less than 50 and unsuitable if the MH value is greater than 50.

The MH value of groundwater was determined using the following formula:

$$MH = \frac{Mg^{2+}}{(Ca^{2+} + Mg^{2+})} \times 100. \tag{5}$$

2.5.3. Permeability Index (PI). Permeability of the soil is important for crop yield and water circulation in the field. For an extended period of time, soil permeability is impacted by elevated Na^+ , Ca^{2+} , Mg^{2+} , and bicarbonate concentrations in groundwater. Groundwater is classified into three classes based on PI: class I (more than 75%), class II (25–75%), and class III (less than 25%). Prolonged long-period groundwater utilisation to irrigate crops influences the groundwater permeability index [38]. PI is defined by

$$PI = \frac{Na^+ + \sqrt{HCO_3^-}}{Ca^{2+} + Mg^{2+} + Na^+} \times 100. \tag{6}$$

2.5.4. Percent Sodium (%Na). Wilcox [39] introduced a classification scheme for rating irrigation water based on percent sodium (%Na). Sodium percent or the proportion of sodium among all the anions is usually expressed in terms of percent sodium. Using the following formula, it was estimated:

$$\%Na = \frac{(Na^+ + K^+)}{(Ca^{2+} + Mg^{2+} + K^+ + Na^+)} \times 100. \tag{7}$$

2.5.5. Sodium Adsorption Ratio (SAR). The SAR is an important irrigation water quality parameter for determining the appropriateness of groundwater for agricultural use. The SAR value determined the relative Na^+ , Ca^{2+} , and Mg^{2+} concentrations in groundwater. Excess amount of sodium in groundwater degrades soil quality and the groundwater equilibrium structure. The ratio of sodium concentration to the sum of Ca^{2+} and Mg^{2+} concentrations yields the SAR value for groundwater.

$$\text{SAR} = \frac{\text{Na}^+}{\sqrt{(\text{Ca}^{2+} + \text{Mg}^{2+})/2}} \quad (8)$$

2.5.6. *Soluble Sodium Percentage (SSP)*. Ca^{2+} , Mg^{2+} , and Na^+ concentrations all have a significant effect in the quality of groundwater used for irrigation. Groundwater with an SSP of less than 50% is good for irrigation purposes, while groundwater with an SSP of more than 50% is unsuitable.

$$\text{SSP} = \frac{\text{Na}}{\text{Ca} + \text{Mg} + \text{Na}} \times 100. \quad (9)$$

2.6. *Model's Performance Criteria*. The following statistical performance criteria measurements were utilised in equations (6)–(9) to evaluate the performance of the developed machine learning models.

2.6.1. *Mean Absolute Error (MAE)*. Mean absolute error is a fairly common metric to calculate the error. It measures the performance of models that are applied to continuous variables, such as the water quality index. Error is defined as the prediction error of a model (actual value–predicted value). This value is determined for each row of data. Then the absolute value of each difference is averaged. It is also a linear score, meaning that every individual error is weighted equally in the average [40].

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |P_i - M_i|. \quad (10)$$

2.6.2. *Mean Squared Error (MSE)*. Mean squared error is almost identical to mean absolute error; however, the difference between actual and predicted is squared instead of put into an absolute value. Squaring the differences emphasises the differences as larger differences become amplified by the squaring. This better accounts for the larger errors and provides a more accurate measure of error. The formula of mean squared error is as follows:

$$\text{MSE} = \frac{\sum_{i=1}^n (Z_i - Z^i)^2}{n}. \quad (11)$$

2.6.3. *Root Mean Squared Error*. The root mean squared error technique is exactly the same as the mean squared error but is squared as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (Q_A^i - Q_P^i)^2}. \quad (12)$$

The square root function makes this a good indicator of the standard deviation of the errors, which can demonstrate whether the model is consistent in its accuracy or varies based on input. An important note about this metric is that

having a small root mean squared error is not necessarily adequate proof that the model is sufficient. If the error is too small, this could signify that the model is suffering from overfitting, meaning that it will only perform with high accuracy on the data set it was given [41].

2.6.4. *Coefficient of Correlation (R)*. The correlation coefficient (r) is a statistic that indicates how well the model fits the experimental data. The coefficient of correlation (r) was calculated by

$$r = \sqrt{1 - \frac{\sum_{i=1}^n (Z_i - Z^i)^2}{\sum_{i=1}^n (Z_i)^2}}, \quad (13)$$

where Z_i/P_i and Z^i/M_i are the measures and calculated value and n is the model's number of values. Two models can be beneficial for regression and sorting; each model is learned using a unique method and evaluated for concealed data during the training process.

3. Results

3.1. *Correlation Analysis*. Correlation analysis is a statistical approach commonly exercised to determine the strength of a linear relationship between two dependent parameters. The variables are not chosen due to their independence or dependence. The correlation analysis method was used in the majority of the studies to determine the linear relationship between two variables. The correlation matrix was produced by computing the correlation coefficients for several parameters. p values were used to determine the correlation's significance. If p is smaller than 0.05 and 0.01 ($p < 0.05$ and $p < 0.01$), the variation is significant (Tables 3 and 4). The change is not significant when $p > 0.05$. The level of significance is set between 0.01 and 0.05 [14, 42, 43]. Pearson correlation analysis is used to analyse the correlations between all variables (input/output), and the results are included in Tables 3 and 4. The two optimal input combinations are determined mostly by nonlinear subset regression and sensitivity analysis. Numerous researches have documented the benefit of employing a nonlinear sensitivity input variable selection strategy to carefully identify the most significant components [10, 32, 44, 45]. The optimal subset regression analysis is used to determine the optimal input combinations for all six irrigation water quality indices presented in Table 5. We found that the best combinations were Ca/Mg/Na , $\text{EC/TDS/SO}_4/\text{Ca/Mg}$, $\text{pH/EC/TDS/CO}_3/\text{SO}_4/\text{Cl/Ca/Mg/Na/K}$, $\text{pH/CO}_3/\text{HCO}_3/\text{SO}_4/\text{Cl/Ca/Mg/Na/K}$, $\text{pH/EC/CO}_3/\text{HCO}_3/\text{SO}_4/\text{Cl/Ca/Mg/Na/K}$, $\text{EC/TDS/SO}_4/\text{Cl/Ca/Mg/Na/K}$ and achieved high correlation and less statistical errors for SSP, MH, KR, PI, %Na, SAR, respectively. Furthermore, all of the combinations that were identified showed positive results.

The Box–Whisker plots of irrigation water quality parameters for pre- and postmonsoon seasons are shown in Figures 3 and 4, respectively.

TABLE 3: Premonsoon input-output correlation matrix.

Variables	HCO ₃ ⁻	CO ₃ ²⁻ (meq/l)	Ca ²⁺	MH	Mg ²⁺	Cl ⁻	TDS	EC (dS/m)	Na ⁺	SAR	KR	SSP	%Na	PI	K ⁺	SO ₄ ²⁻	pH
HCO ₃ ⁻	1																
CO ₃ ²⁻ (meq/l)	0.043	1															
Ca ²⁺	0.032	-0.168	1														
MH	0.050	0.040	-0.239	1													
Mg ²⁺	0.058	-0.083	0.465	0.682	1												
Cl ⁻	0.061	-0.028	0.453	0.296	0.683	1											
TDS	-0.067	0.206	0.264	0.023	0.309	0.334	1										
EC (dS/m)	-0.090	0.166	0.135	0.054	0.222	0.240	0.887	1									
Na ⁺	-0.055	0.211	-0.087	0.046	-0.013	0.100	0.522	0.586	1								
SAR	-0.047	0.223	-0.233	-0.059	-0.188	-0.016	0.474	0.550	0.971	1							
KR	-0.036	0.224	-0.320	-0.132	-0.286	-0.082	0.433	0.510	0.910	0.982	1						
SSP	-0.011	0.257	-0.422	-0.190	-0.419	-0.190	0.280	0.401	0.820	0.908	0.937	1					
%Na	-0.069	0.295	-0.472	-0.228	-0.490	-0.233	0.203	0.295	0.729	0.837	0.884	0.948	1				
PI	0.210	0.281	-0.574	-0.307	-0.615	-0.332	0.100	0.213	0.584	0.729	0.804	0.912	0.893	1			
K ⁺	-0.216	0.094	0.015	0.002	-0.014	-0.014	-0.181	-0.277	-0.207	-0.197	-0.178	-0.172	0.126	-0.175	1		
SO ₄ ²⁻	-0.020	0.060	0.043	0.051	0.081	0.083	0.255	0.235	0.154	0.139	0.125	0.110	0.070	0.048	-0.107	1	
pH	0.002	0.134	-0.227	0.099	-0.022	0.085	0.166	0.149	0.153	0.182	0.202	0.197	0.167	0.225	-0.125	0.247	1

TABLE 4: Postmonsoon input-output correlation matrix.

Variables	CO ₃ ²⁻ (meq/l)	Na ⁺	SAR	KR	SSP	%Na	PI	Mg ²⁺	MH	SO ₄ ²⁻	EC (dS/m)	TDS	Ca ²⁺	Cl ⁻	K ⁺	HCO ₃ ⁻	pH
CO ₃ ²⁻ (meq/l)	1																
Na ⁺	0.043	1															
SAR	0.114	0.924	1														
KR	0.166	0.794	0.963	1													
SSP	0.161	0.784	0.944	0.974	1												
%Na	0.154	0.756	0.921	0.955	0.970	1											
PI	0.155	0.271	0.544	0.699	0.718	0.707	1										
Mg ²⁺	-0.102	0.175	-0.120	-0.305	-0.293	-0.315	-0.593	1									
MH	-0.009	0.007	-0.042	-0.093	-0.080	-0.082	-0.118	0.669	1								
SO ₄ ²⁻	0.002	0.246	0.027	-0.099	-0.087	-0.080	-0.355	0.420	0.020	1							
EC (dS/m)	-0.023	0.463	0.174	-0.004	-0.013	-0.049	-0.394	0.528	-0.073	0.541	1						
TDS	-0.119	0.442	0.160	-0.014	-0.025	-0.039	-0.404	0.489	-0.099	0.545	0.894	1					
Ca ²⁺	-0.096	0.207	-0.103	-0.267	-0.292	-0.314	-0.607	0.403	-0.345	0.486	0.802	0.790	1				
Cl ⁻	-0.132	0.141	-0.025	-0.113	-0.159	-0.079	-0.366	0.265	-0.074	0.452	0.447	0.505	0.515	1			
K ⁺	-0.035	0.126	0.070	0.023	-0.005	0.211	-0.095	0.082	0.025	0.125	0.001	0.075	0.067	0.330	1		
HCO ₃ ⁻	-0.044	0.014	0.015	0.001	0.014	-0.039	0.236	0.102	0.178	-0.206	-0.016	-0.036	-0.098	-0.227	-0.146	1	
pH	0.045	0.023	0.082	0.121	0.108	0.053	0.164	-0.047	0.027	-0.141	0.013	-0.075	-0.092	-0.212	-0.196	0.224	1

TABLE 5: Best subset regression analysis for irrigation water quality indices input combinations.

Indices	Variables	MSE	R ²	Adjusted R ²	Mallows' Cp	Akaike's AIC	Schwarz's SBC	Amemiya's PC
SSP	Ca/Mg/Na	32.768	0.781	0.778	4.609	666.951	679.939	0.226
MH	EC (dS/m)/TDS/SO ₄ /Ca/Mg	49.492	0.755	0.748	6.966	747.248	766.730	0.258
SAR	EC (dS/m)/TDS/SO ₄ /Cl/Ca/Mg/Na/K	0.041	0.971	0.970	8.863	-599.143	-569.920	0.031
PI	pH/CO ₃ (meq/l)/HCO ₃ /SO ₄ /Cl/Ca/Mg/Na/K	52.228	0.793	0.782	9.957	761.294	793.764	0.228
KR	pH/EC (dS/m)/TDS/CO ₃ (meq/l)/SO ₄ /Cl/Ca/Mg/Na/K	0.015	0.883	0.876	10.426	-781.725	-746.008	0.130
%Na	pH/EC (dS/m)/CO ₃ (meq/l)/HCO ₃ /SO ₄ /Cl/Ca/Mg/Na/K	28.660	0.801	0.790	10.004	648.215	683.933	0.221

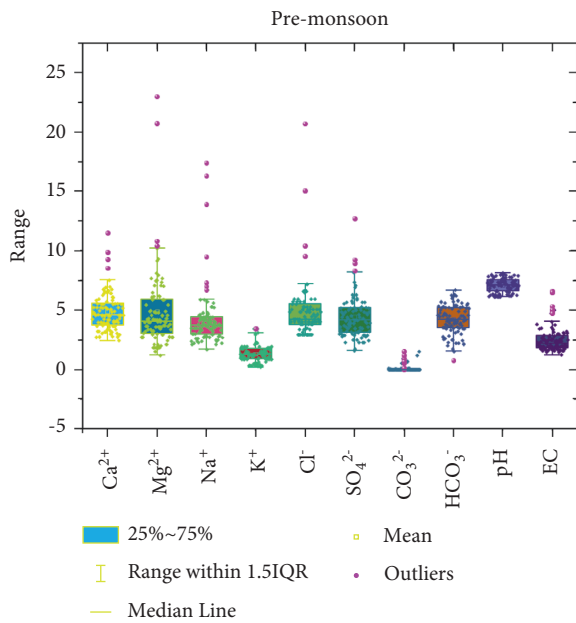


FIGURE 3: Box-Whisker plot of premonsoon season.

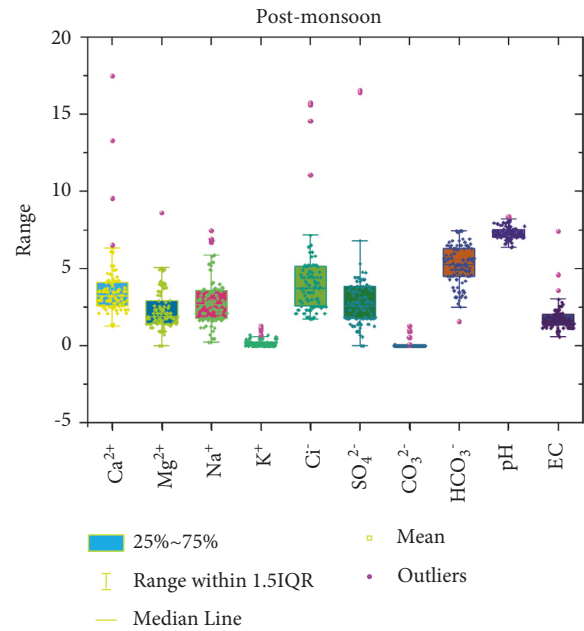


FIGURE 4: Box-Whisker plot of postmonsoon season.

3.2. Evaluation of Results. Machine learning models have been widely applied in a variety of fields in recent years. They can aid in the forecast of future conservation and natural process scenarios. In this study, we have studied LR, RSS, AR, REPTree, and SVM models with two seasons (pre- and postmonsoon) findings that are commonly used to determine the viability of groundwater for irrigation drives when simple quantifiable input variables such as pH, EC, TDS, Ca²⁺, Mg²⁺, Na⁺, K⁺, Cl⁻, CO₃²⁻, HCO₃⁻, and SO₄²⁻ for two seasons machine learning models. The finding of the study shows that machine learning models are quite good at predicting water quality levels [14, 35]. This article discusses the training and testing of LR, RSS, AR, REPTree, and SVM models. The comparison between LR, RSS, AR, REPTree, and SVM models' performance is presented in Table 6. Subsections below provide a more extensive description.

3.2.1. Data Set Comparisons. We included all of the irrigation water quality variable data sets that were required for training and testing the generated model in our analysis. The

training and testing results obtained by LR, RSS, AR, REPTree, and SVM are presented in Table 6.

As depicted in Table 6, the LR model has shown the maximum $r=0.9853$ for the SAR value of training data and $r=0.9811$ and 0.9811 for SSP and %Na for testing data, respectively. The SVM model has shown the maximum $r=0.9819$ for SAR value of testing data. According to available literature, the r value above 0.8 is reasonable [46, 47]. Other variables of LR models such as KR, MH, SSP, SAR, %Na, and PI value are above $r=0.8$, and the other values are RMSE = 0.1347, 7.5467, 6.1867, 0.2183, 5.7977, and 7.5174 and RMSE = 0.0754, 4.7873, 3.8575, 0.1222, 3.2215, and 6.1774 in training and testing, respectively; MSE = 0.01814, 56.9535, 38.2754, 0.04762, 33.6145, and 56.5104 in training and MSE = 0.00571, 22.9183, 14.8804, 0.01495, 10.3783, and 38.1609 in testing; and MAE = 0.0916, 5.3615, 4.4615, 0.1444, 4.2076, and 5.6192 in training and MAE = 0.0602, 4.0303, 3.2127, 0.0952, 2.4304, and 5.2347 in testing.

Similarly, RSS-model developed based on the parameters such as Kelly's ratio, magnesium hazard, sodium adsorption

TABLE 6: Comparison between different models' performance.

	LR					RSS					SVM					AR					REPTree				
	RMSE	r	MSE	MAE	RMSE	r	MSE	MAE	RMSE	r	MSE	MAE	RMSE	r	MSE	MAE	RMSE	r	MSE	MAE	RMSE	r	MSE	MAE	
Train	KR	0.1347	0.936	0.018142	0.0916	0.1072	0.9636	0.01149	0.0655	0.1165	0.956	0.013571	0.071	KR	0.1031	0.9661	0.010633	0.0621	31.6945	0.9556	0.01473	0.068			
	MH	7.5467	0.8454	56.95347	5.3615	3.909	0.9616	15.27995	2.8828	4.2215	0.9546	17.82021	2.814	MH	3.82	0.9634	14.59281	2.7928	4.1625	0.9561	17.32617	2.8226			
	SSP	6.1867	0.8749	38.27535	4.4615	2.7788	0.9761	7.722282	1.9971	3.0118	0.972	9.070936	2.0718	SSP	2.6868	0.9778	7.219327	1.9196	2.8648	0.9746	8.207105	2.0187			
	SAR	0.2183	0.9853	0.047618	0.1444	0.2861	0.979	0.081844	0.145	0.2767	0.9807	0.076575	0.1308	SAR	0.3004	0.9744	0.090245	0.1488	0.4003	0.961	0.160246	0.1479			
	%Na	5.7977	0.8825	33.61445	4.2076	3.5897	0.9567	12.88595	2.7885	3.2326	0.9652	10.44961	2.2881	%Na	3.4375	0.9611	11.81618	2.6419	3.7704	0.9534	14.21579	2.6201			
	PI	7.5174	0.877	56.51042	5.6192	4.3149	0.9622	18.61812	3.2678	4.592	0.956	21.08588	3.3637	PI	4.5108	0.9577	20.34772	3.5802	4.6661	0.9551	21.77212	3.6209			
Test	KR	0.0754	0.9016	0.005705	0.0602	0.0837	0.8888	0.006988	0.0533	0.0662	0.9364	0.004381	0.042	KR	0.0712	0.9223	0.005083	0.0553	0.0781	0.9083	0.006108	0.0559			
	MH	4.7873	0.942	22.9183	4.0303	3.9627	0.9584	15.70277	3.2096	4.0568	0.9618	16.45781	3.1999	MH	4.1999	0.9587	17.64038	3.5776	3.9804	0.9561	15.84309	3.3622			
	SSP	3.8575	0.9811	14.88043	3.2127	2.5604	0.965	6.555419	1.9217	3.0168	0.9588	9.101218	2.3584	SSP	3.6843	0.9335	13.57334	2.795	2.6323	0.9679	6.928808	2.0269			
	SAR	0.1222	0.9733	0.014953	0.0952	0.1411	0.9666	0.019914	0.1022	0.1113	0.9819	0.012383	0.0726	SAR	0.1401	0.9717	0.019609	0.1052	0.1379	0.9659	0.019049	0.0982			
	%Na	3.2215	0.9811	10.3783	2.4304	4.026	0.9168	16.20922	2.9814	3.7046	0.9547	13.72447	2.9603	%Na	3.4726	0.9395	12.05931	2.7483	4.0133	0.9252	16.10663	2.838			
	PI	6.1774	0.8321	38.16091	5.2347	6.0438	0.8253	36.53067	4.6719	5.1066	0.8903	26.078	4.0582	PI	6.0103	0.8323	36.12515	4.589	5.4483	0.8575	29.68431	4.2895			

ration, permeability index, and soluble sodium percent. Statistical indices, RMSE, r , MSE, and MAE have been applied, and the values of these indices revealed that KR, MH, SSP, SAR, %Na, and PI coefficient correlation values for all irrigation indices in training and testing shows reasonable ' r ' value except PI in testing ($r = 0.8253$), and RMSE values of these irrigation indices are 0.1072, 3.909, 2.7788, 0.2861, 3.5897, and 4.3149 and RMSE are 0.0837, 3.9627, 2.5604, 0.1411, 4.026, 6.0438 in training and testing, respectively. The MAE values are 0.0655, 2.8828, 1.9971, 0.145, 2.7885, and 3.2678, and 0.0533, 3.2096, 1.9217, 0.1022, 2.9814, and 4.6719 in training and testing, respectively.

Furthermore, the SVM training and testing models of KR, MH, SSP, SAR, %Na, and PI have shown a good coefficient correlation, and RMSE values range between 0.1165 (KR) and 4.592 (PI) in training and 0.0662 (KR) and 5.1066 (PI) in testing. Similarly, training and testing results of SVM model of KR, MH, SSP, SAR, %Na, and PI have shown MSE and MAE values of 0.01357, 17.8202, 9.07094, 0.07658, 10.4496, and 21.0859 and 0.00438, 16.4578, 9.10122, 0.01238, 13.7245, and 26.078 and 0.071, 2.814, 2.0718, 0.1308, 2.2881, and 3.3637 and 0.042, 3.1999, 2.3584, 0.0726, 2.9603, and 4.0582, respectively.

Similarly, for AR-model, training and testing models were developed based on the parameters such as Kelly's ratio, magnesium hazard, sodium adsorption ration, permeability index, and soluble sodium percent, which have shown RMSE, r , MSE, and MAE values of KR, MH, SSP, SAR, %Na, and PI coefficient correlation values for all irrigation indices in training and testing shows reasonable r value except PI in testing ($r = 0.8323$), RMSE = 0.1031 for KR in training and RMSE = 0.0712 for KR in testing, respectively.

Furthermore, the REPTree training and testing models of KR, MH, SSP, SAR, %Na, and PI have shown a good coefficient correlation, and RMSE value ranges between 0.4003 (SAR) and 31.6945 (KR) in training and 0.0781 (KR) and 5.4483 (PI) in testing. Similarly, training and testing results of the REPTree model of KR, MH, SSP, SAR, %Na, and PI have shown MSE and MAE values ranging from 0.01473 (KR) to 21.7721 (PI) and 0.00611 (KR) to 29.6843 (PI) and 0.068 (KR) to 3.6209 (PI) and 0.0559 (KR) to 4.2895 (PI), respectively.

All four statistical indicators show the LR-model best fit for KR data in training and testing, followed by SAR. Similarly, AR-model and SVM-model are best-fit for KR and SAR data during training and testing. Furthermore, for the RSS-model, all four statistical indicators show the best results for KR in training. For testing, this model shows the best results for RMSE, MSE, and MAE for KR, and correlation coefficient for SAR data. For REPTree machine learning model for RMSE, r statistical indicators show best results for SAR irrigation indices, and MSE and MAE show best-fit for KR in training and testing. KR shows best results for all four statistical indicators followed by SAR. As a result, it can be concluded that SVM is the best model for all irrigation water quality indices for all four statistical indicators. This observation was similar to the findings obtained in their study [15, 48].

The scatter plots in Figure 5 depict the observed and simulated values generated by the models during the validation process. The models' accuracy is sufficient when the values are distributed uniformly over or across both sides of the line XY, indicating that the errors follow the Gaussian distribution. This figure demonstrates that the ensemble models (SVM and REPTree) predict values that are more evenly distributed along the XY axis than the LR, RSS, and AR models.

Along with the models discussed previously, a radar map of performance indicators was used to evaluate the effectiveness of the applied models. Figure 6 depicts the values of performance indicators to aid in diagnosing the efficiency of all models. As illustrated in the figures, the SVM model has a lower MAE, MSE, root relative squared error (RRSE), relative absolute error (RAE), and RMSE value, as well as a higher Pearson's r value than the other models for the SAR, PI, KR, and MH irrigation indices.

During testing, the Taylor diagram was used to execute a more in-depth comparative examination of the models (Figure 7) for all six irrigation indices. Based on the standard deviation and correlation coefficient, the SVM model was the most closely related to the observed location. In contrast, the LR model was shown to be the most far from the observed site for most irrigation indexes. LR was found to be the worst model, and SVM was found to be the best model among the models used for the current study.

4. Discussion

The current study has examined irrigation water quality utilising a variety of machine learning models (LR, RSS, AR, REPTree, and SVM) in predicting the irrigation water quality indices such as SAR, KR, PI, percent Na, KR, and MH of groundwater in the Nand Samand catchment, Rajasthan, India. A sum of 95 (pre-and postmonsoon season) water samples were used to train and test the selected models. Agrawal et al. [49] conducted a study on artificial intelligence approaches for groundwater quality evaluation in the Pindrawan tank command region in Chhattisgarh's upper Mahanadi River valley (southeastern section), Raipur district. Groundwater samples were acquired from 37 sites. They evaluated the efficacy of artificial intelligence strategies for determining the water quality index, comprising particle swarm optimisation (PSO), naïve Bayes classifier (NBC), and SVM. The results indicated that the PSO-SVM accuracy for WQI indices was 77.60%. Additionally, they concluded that higher the significant correlation among the input and output variables, the better the model's performance. These findings demonstrate that selection of input variables is a critical step in determining the performance of machine learning models. Groundwater quality analysis could be assessed by the four robust artificial intelligence techniques for the data of long period [22]. Bilali and Taleb [4] mentioned identical findings by utilising the machine learning models to predict TDS, chloride, SAR, and SARA parameters. Aldhyani et al. [50] predicted water quality by employing artificial intelligence algorithms. They used the SVM model for predicting water quality classification (WQC) and

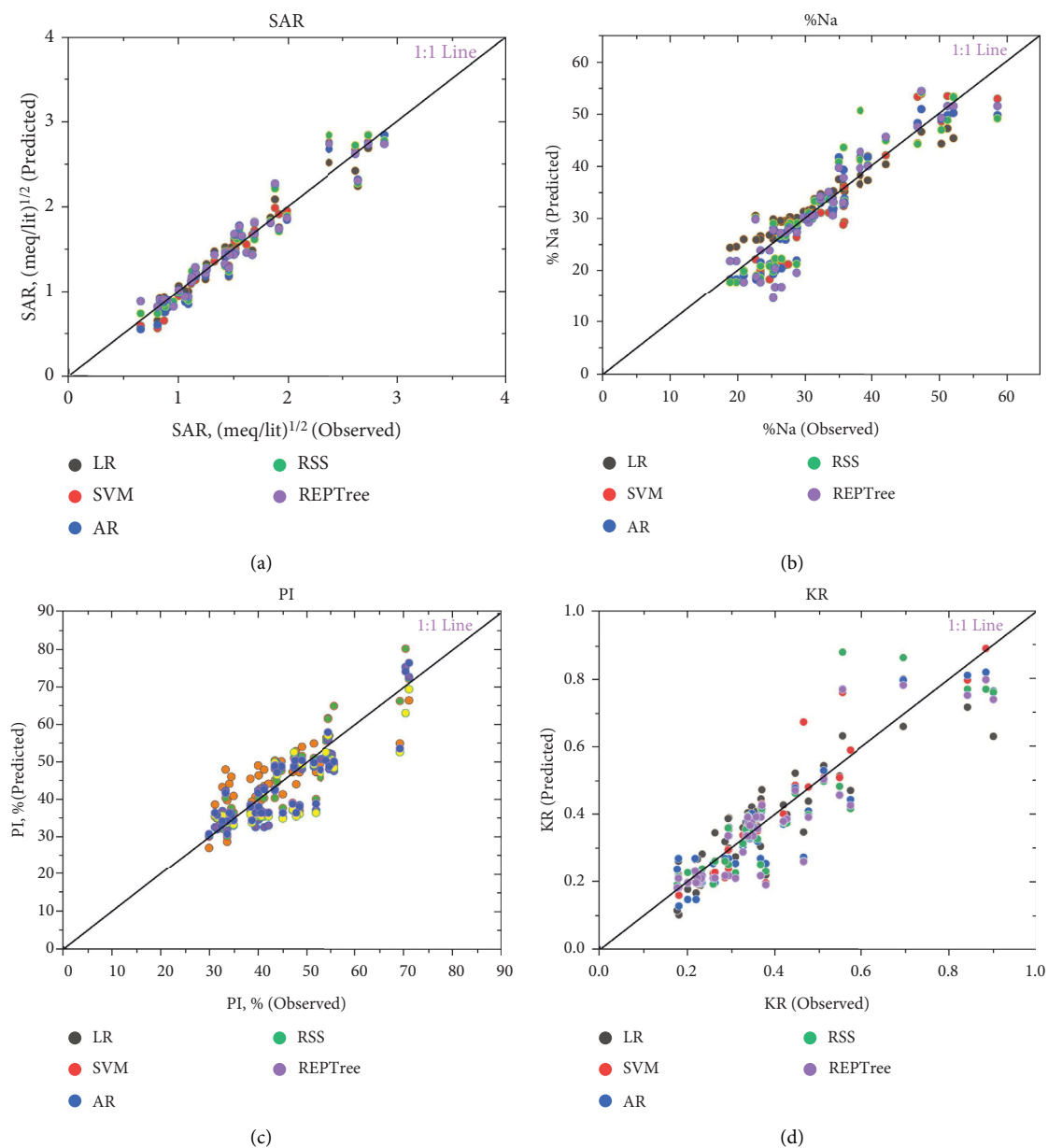


FIGURE 5: Continued.

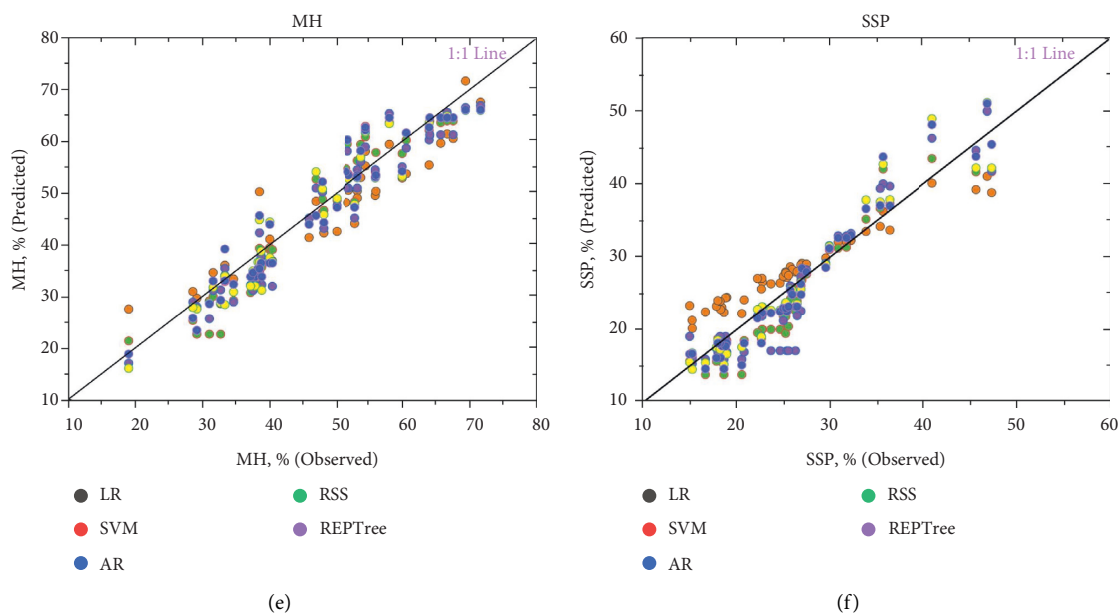


FIGURE 5: Scatter plots of the predicted and observed values for predicting the LR, SVM, AR, RSS, and REPTree model performance during the testing phase.

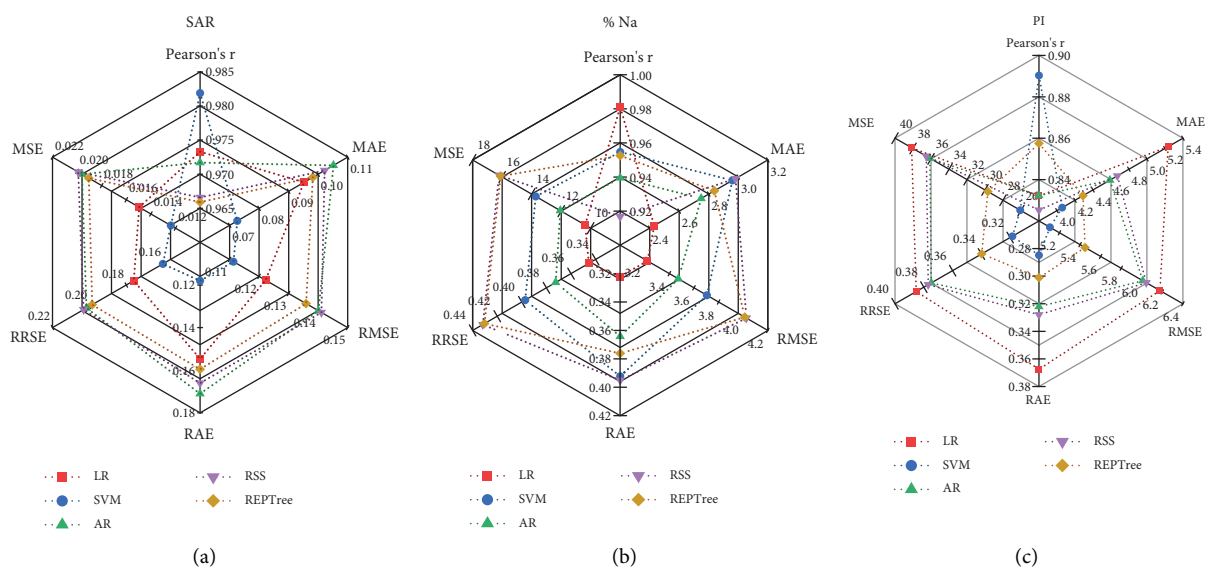


FIGURE 6: Continued.

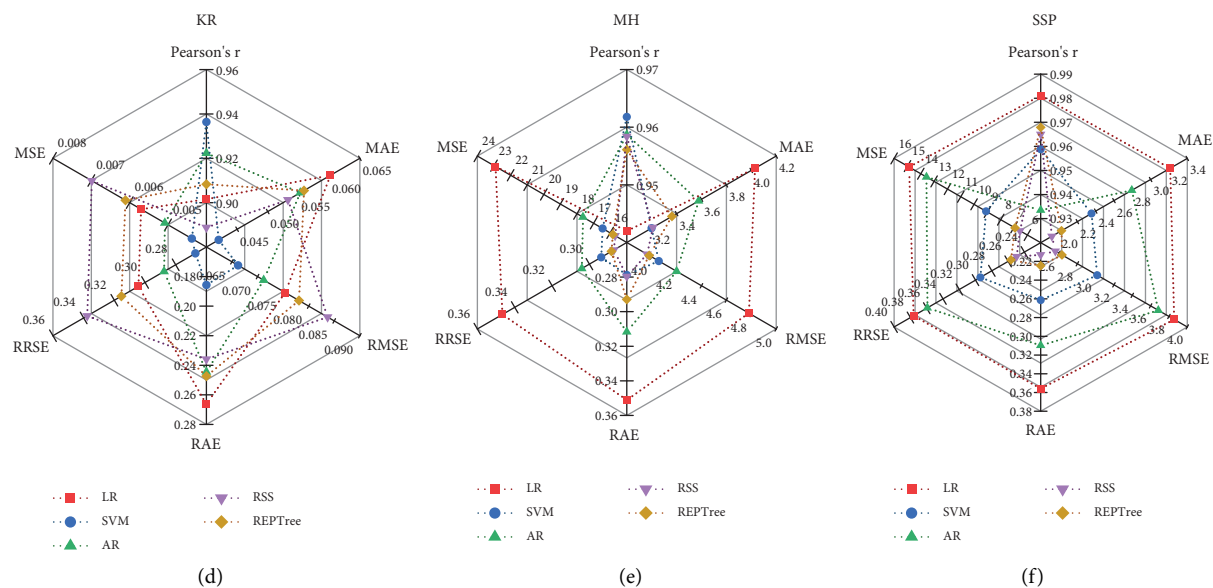


FIGURE 6: Indicators for all six irrigation indices are shown in a radar graphic (a–f).

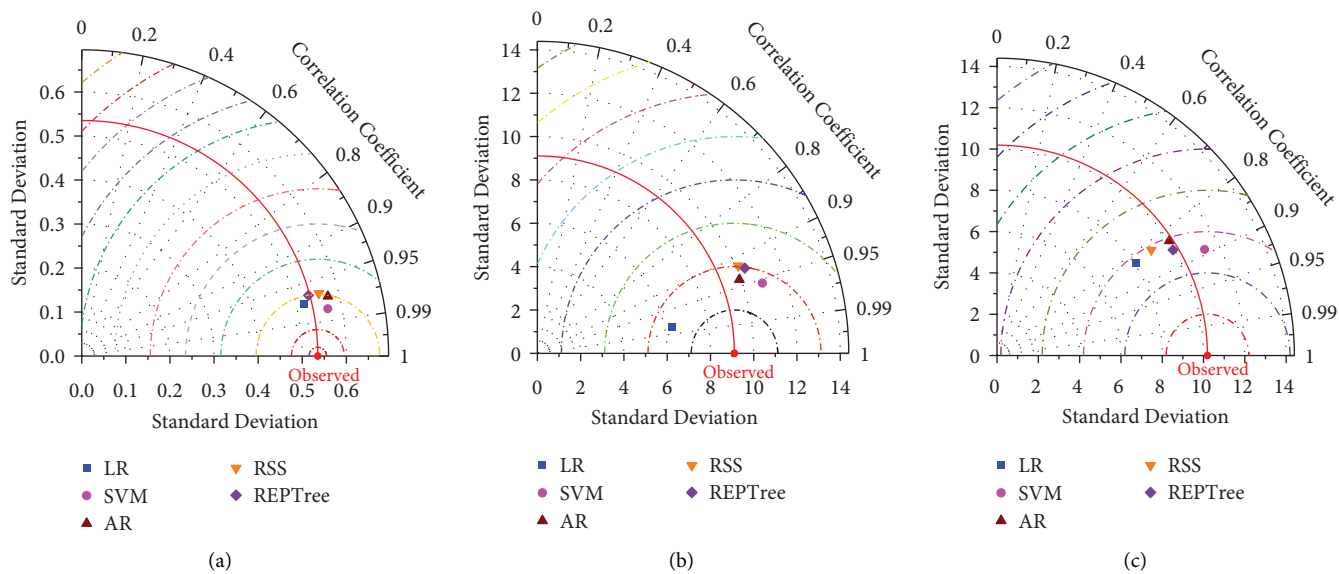


FIGURE 7: Continued.

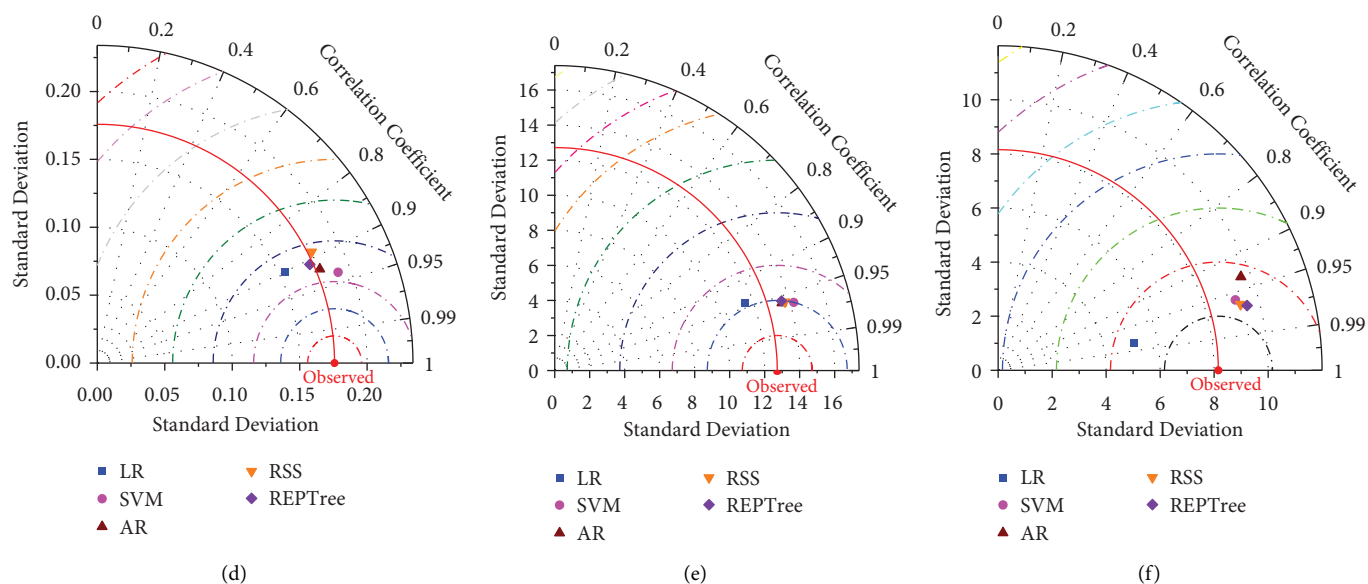


FIGURE 7: Comparison of model's performances using the Taylor diagram for irrigation indices (a–f).

reported the SVM model's highest accuracy (97.01%) for WQC prediction over other algorithms. The authors [21] used a newly developed version of the support vector machine on the basis of kernel learning for the water quality index. They concluded that using other water quality metrics and the support vector regression algorithm, biochemical oxygen demand, and chemical oxygen demand was computed with acceptable accuracy. Haghiabi et al. [37] concluded similar findings for southwest Iran's Tیره River Basin. Machine learning models quickly assess water quality index and indices, and therefore, these are useful for quick evaluation. Bilali and Taleb [4] modeled an irrigation water quality index in a semiarid climate in Morocco's Bouregreg watershed employing machine learning algorithms. Eight ML models were developed and tested in predicting 10 characteristics of irrigation water quality (IWQ). Three hundred samples were analysed, processed, and chosen to train and test the models at nine monitoring sites. Evaluation of magnesium absorption ratio (MAR) and permeability index (PI) indicated that all other models are strongly efficient in predicting the other parameters, with coefficients of correlation (r) ranging between [0.56, 0.99] and [0.64, 0.99] for the training and validation phases, respectively. The SVM model outperforms than ANN model [51] except for support vector regression (SVR) and k -nearest neighbour (k -NN) models, as well as. The findings of this study indicated that SVM is an effective technique for resolving a variety of environmental concerns [21, 52].

Furthermore, all validated results of a high-accuracy prediction using artificial intelligence algorithms for water quality have been confirmed by the authors [8, 10, 53–55]. Although decent prediction is largely dependent on the variety of input factors and their effect, ensuring that all data are available at an affordable cost is vital. Besides this, generalisation of these research findings to regions other than those used in advancement must be explored since

numerous variables substantially influence water quality, including hydrological system, land use and land cover, landform, morphological, geological conditions, and anthropogenic factors. Such criteria significantly impact the way the input variables utilised in the prediction process are integrated. The findings of this study are significant as these can be used to boost real-time supervision of the irrigation water quality in the Nand Samand catchment. However, automatic sensing technologies to assess electrical conductivity and pH parameters employ the most appropriate models that need to be applied. These results can be extremely beneficial to plan dam reservoirs for agriculture, where evaporation greatly affects the chemical water quality, notably during the warmer seasons. As a result, our effort would enable farmers to manage water quality effectively and efficiently. Therefore, since the appraisal of water quality for irrigation purposes is largely reliant on soil type, crop, and water quality class, the categorisation of the machine learning model is proposed for further research.

5. Conclusions

Five ML models were created and evaluated to predict six irrigation indices. The results indicated that machine learning models could help overcome some constraints associated with conventional approaches for assessing water appropriateness for agricultural applications. The study's major conclusions are as follows: conventional methods are extensively used to evaluate irrigation water appropriateness based on a huge variety of characteristics and indexes. They are considered to be effective instruments, yet they might be expensive. It is recommended that rather than depending on these methods for assessing and forecasting the suitability of water for agricultural use, future research on ML models (classification and regression models) can be carried out, as

well as their application in the field of engineering to improve water quality control. Because they function better, they also have the potential to considerably cut the cost and time associated with irrigation water quality control. In conclusion, the SVM model has demonstrated its efficiency and usefulness in predicting the water quality of the Nand Samand catchment.

Data Availability

The data used to support the findings of this study are available from the first author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The first author greatly acknowledges the Department of Science and Technology, INSPIRE-Fellowship, the Government of India, for fellowship.

References

- [1] G. Busico, N. Kazakis, E. Cuoco et al., "A novel hybrid method of specific vulnerability to anthropogenic pollution using multivariate statistical and regression analyses," *Water Research*, vol. 171, Article ID 115386, 2020.
- [2] G. K. Mbizvo, K. Bennett, C. R. Simpson, S. E. Duncan, and R. F. M. Chin, "Epilepsy-related and other causes of mortality in people with epilepsy: a systematic review of systematic reviews," *Epilepsy Research*, vol. 157, Article ID 106192, 2019.
- [3] A. Ringler, I. Budzbon, H. Kremer, N. Fernandez, P. Mmayi, and K. Alverson, *A Snapshot of the World's Water Quality: Towards a Global Assessment*, United Nations Environment Programme, Nairobi, 2016.
- [4] A. El Bilali and A. Taleb, "Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment," *Journal of the Saudi Society of Agricultural Sciences*, vol. 19, no. 7, pp. 439–451, 2020.
- [5] A. Asante-Annor, P. N. Bewil, and D. Boateng, "Evaluation of groundwater suitability for irrigation in the Lambussie-Karni district of Ghana," *Ghana Mining Journal*, vol. 18, no. 1, pp. 9–19, 2018.
- [6] M. Salifu, F. Aidoo, M. S. Hayford, D. Adomako, and E. Asare, "Evaluating the suitability of groundwater for irrigational purposes in some selected districts of the upper West Region of Ghana," *Applied Water Science*, vol. 7, pp. 1–10, 2015.
- [7] N. Aghazadeh and A. A. Mogaddam, "Assessment of groundwater quality and its suitability for drinking and agricultural uses in the Oshnavieh area, Northwest of Iran," *Journal of Environmental Protection*, vol. 01, no. 01, pp. 30–40, 2010.
- [8] A. Najah Ahmed, F. Binti Othman, H. Abdulmohsin Afan et al., "Machine learning methods for better water quality prediction," *Journal of Hydrology*, vol. 578, Article ID 124084, 2019.
- [9] E. Fijani, R. Barzegar, R. Deo, E. Tziritis, and K. Skordas, "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," *The Science of the Total Environment*, vol. 648, pp. 839–853, 2019.
- [10] P. Liu, J. Wang, A. Sangaiah, Y. Xie, and X. Yin, "Analysis and prediction of water quality using LSTM deep neural networks in IoT environment," *Sustainability*, vol. 11, no. 7, Article ID 2058, 2019.
- [11] H. Lu and X. Ma, "Hybrid decision tree-based machine learning models for shortterm water quality prediction," *Chemosphere*, vol. 249, Article ID 126169, 2020.
- [12] W. C. Leong, A. Bahadori, J. Zhang, and Z. Ahmad, "Prediction of water quality index (WQI) using support vector machine (SVM) and least square support vector machine (LS-SVM)," *International Journal of River Basin Management*, vol. 19, no. 2, pp. 149–156, 2021.
- [13] V. M. Wagh, D. B. Panaskar, A. A. Muley, S. V. Mukate, Y. P. Lolage, and M. L. Aamalawar, "Prediction of groundwater suitability for irrigation using artificial neural network model: a case study of Nanded tehsil, Maharashtra, India," *Modeling Earth Systems and Environment*, vol. 2, no. 4, pp. 1–10, 2016.
- [14] S. Kouadri, C. B. Pande, B. Panneerselvam, K. N. Moharir, and A. Ahmed Elbeltagi, "Prediction of irrigation groundwater quality parameters using ANN, LSTM, and MLR models," *Environmental Science and Pollution Research*, vol. 29, 2021.
- [15] N. L. Kushwaha, J. Rajput, A. Elbeltagi et al., "Data intelligence model and meta-heuristic algorithms-based Pan evaporation modelling in two different agro-climatic zones: a case study from northern India," *Atmosphere*, vol. 12, Article ID 1654, 2021.
- [16] T. Deng, K. W. Chau, and H. F. Duan, "Machine learning based marine water quality prediction for coastal hydro-environment management," *Journal of Environmental Management*, vol. 284, Article ID 112051, 2021.
- [17] V. Nourani, G. Elkiran, and S. I. Abba, "Wastewater treatment plant performance analysis using artificial intelligence—an ensemble approach," *Water Science and Technology*, vol. 78, no. 10, pp. 2064–2076, 2018.
- [18] M. Najafzadeh, A. Ghaemi, and S. Emamgholizadeh, "Prediction of water quality parameters using evolutionary computing-based formulations," *International journal of Environmental Science and Technology*, vol. 16, no. 10, pp. 6377–6396, 2018.
- [19] M. Najafzadeh and A. Ghaemi, "Prediction of the five day biochemical oxygen demand and chemical oxygen demand in natural streams using machine learning methods," *Environmental Monitoring and Assessment*, vol. 191, no. 6, p. 380, 2019.
- [20] M. Najafzadeh, F. Homaei, and H. Farhadi, "Reliability assessment of water quality index based on guidelines of national sanitation foundation in natural streams: integration of remote sensing and data-driven models," *Artificial Intelligence Review*, vol. 54, no. 6, pp. 4619–4651, 2021a.
- [21] M. Najafzadeh and S. Niazmardi, "A novel multiple-kernel support vector regression algorithm for estimation of water quality parameters," *Natural Resources Research*, vol. 30, no. 5, pp. 3761–3775, 2021b.
- [22] M. Najafzadeh, F. Homaei, and S. Mohamadi, "Reliability evaluation of groundwater quality index using data-driven models," *Environmental Science and Pollution Research*, vol. 29, no. 6, pp. 8174–8190, 2022.
- [23] P. K. Singh, K. K. Yadav, and M. Singh, *Water Budgeting of Rajasthan*, p. 154, Maharana Pratap University of Agriculture and Technology, Udaipur, 2021.
- [24] J. Y. Ho, H. A. Afan, A. H. El-Shafie et al., "Towards a time and cost effective approach to water quality index class prediction," *Journal of Hydrology*, vol. 575, pp. 148–165, 2019.

- [25] R. Torres-Sanchez, H. Navarro-Hellin, A. Guillamon-Frutos, R. San-Segundo, M. C. Ruiz-Abellón, and R. Domingo-Miguel, "A decision support system for irrigation management: analysis and implementation of different learning techniques," *Water*, vol. 12, no. 2, p. 548, 2020.
- [26] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, 1998.
- [27] X. Luo, F. Lin, Y. Chen et al., "Coupling logistic model tree and random subspace to predict the landslide susceptibility areas with considering the uncertainty of environmental features," *Scientific Reports*, vol. 9, no. 1, pp. 15369–15413, 2019.
- [28] Q. Wang, W. Xu, and H. Zheng, "Combining the wisdom of crowds and technical analysis for financial market prediction using deep random subspace ensembles," *Neurocomputing*, vol. 299, pp. 51–61, 2018.
- [29] B. T. Pham, D. Tien Bui, I. Prakash, and M. B. Dholakia, "Hybrid integration of multilayer perceptron neural networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS," *Catena*, vol. 149, pp. 52–63, 2017.
- [30] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, CRC Press, Boca Raton, FL, USA, 1989.
- [31] B. Bruneau and F. Grégoire, "Étude de la distribution spatiale des données d'abondance de maquereau bleu (*Scomber scombrus*) et de capelan (*Mallotus villosus*) des relevés d'hiver aux Poissons de fond des Divisions 4VW de l'OPANO à l'aide de modèles additifs généralisés," *Rapport Technique Canadien des Sciences Halieutiques Et Aquatiques*, vol. 2930, 2011.
- [32] S. Kouadri, A. Elbeltagi, A. R. M. T. Islam, and S. Kateb, "Performance of machine learning methods in predicting water quality index based on irregular data set: application on Illizi region (Algerian southeast)," *Applied Water Science*, vol. 11, no. 12, p. 190, 2021b.
- [33] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Series in Data Management Systems, Washington, DC, USA, 2005.
- [34] J. Quinlan, "Simplifying decision trees," *International Journal of Man-Machine Studies*, vol. 27, no. 3, pp. 221–234, 1987.
- [35] A. Elbeltagi, C. B. Pande, S. Kouadri, and A. R. M. T. Islam, "Applications of various data-driven models for the prediction of groundwater quality index in the Akot basin, Maharashtra, India," *Environmental Science and Pollution Research*, vol. 29, no. 12, pp. 17591–17605, 2021a.
- [36] A. Elbeltagi, N. Kumari, J. K. Dharpure et al., "Prediction of combined terrestrial evapotranspiration index (Cte_i) over large river basin based on machine learning approaches," *Water (Switzerland)*, vol. 13, no. 4, pp. 547–618, 2021b.
- [37] A. H. Haghiabi, A. H. Nasrolahi, and A. Parsaie, "Water quality prediction using machine learning methods," *Water Quality Research Journal*, vol. 53, pp. 3–13, 2018.
- [38] A. Roy, T. Keesari, H. Mohokar, U. K. Sinha, and S. Bitra, "Assessment of groundwater quality in hard rock aquifer of central Telangana state for drinking and agriculture purposes," *Applied Water Science*, vol. 8, no. 5, p. 124, 2018.
- [39] L. V. Wilcox, *Classification and Use of Irrigation Waters*, USDA, Washington, DC, 1955.
- [40] S. Garmsiri, *Art of Choosing Metrics in Supervised Models Part 1*, 2018, <https://towardsdatascience.com/art-of-choosing-metrics-in-supervised-models-part-1-f960ae46902e>.
- [41] J. Moody, *What Does RMSE Really Mean?*, 2019, <https://towardsdatascience.com/what-does-rmse-really-mean-806b65f2e48e>.
- [42] S. K. Sar, M. Sahu, S. Singh, V. Diwan, M. Jindal, and A. Arora, "Assessment of uranium in ground water from durg District of Chhattisgarh state and its correlation with other quality parameters," *Journal of Radioanalytical and Nuclear Chemistry*, vol. 314, no. 3, pp. 2339–2348, 2017.
- [43] R. K. Tiwary, B. Kumari, and D. B. Singh, "Water quality assessment and correlation study of physico-chemical parameters of Sukinda chromite mining area, Odisha, India," *Environmental Pollution*, vol. 77, pp. 357–370, 2018.
- [44] D. T. Bui, K. Khosravi, J. Tiefenbacher, H. Nguyen, and N. Kazakis, "Improving prediction of water quality indices using novel hybrid machine-learning algorithms," *The Science of the Total Environment*, vol. 721, 2020.
- [45] O. Kisi, A. Azad, H. Kashi et al., "Modeling groundwater quality parameters using hybrid neuro-fuzzy methods," *Water Resources Management*, vol. 33, no. 2, pp. 847–861, 2018.
- [46] A. H. Gandomi, A. H. Alavi, M. R. MirzaHosseini, and F. M. Nejad, "Nonlinear genetic-based models for prediction of flow number of asphalt mixtures," *Journal of Materials in Civil Engineering*, vol. 23, no. 3, pp. 248–263, 2011.
- [47] M. I. Shah, W. S. Alaloul, A. Alqahtani, A. Aldrees, M. A. Musarat, and M. F. Javed, "Predictive modeling approach for surface water quality: development and comparison of machine learning models," *Sustainability*, vol. 13, no. 14, Article ID 7515, 2021.
- [48] A. S. Abobakr Yahya, A. N. Ahmed, F. Binti Othman et al., "Water quality prediction model based support vector machine model for ungauged river catchment under dual scenarios," *Water*, vol. 11, no. 6, Article ID 1231, 2019.
- [49] P. Agrawal, A. Sinha, S. Kumar et al., "Exploring artificial intelligence techniques for groundwater quality assessment," *Water*, vol. 13, no. 9, Article ID 1172, 2021.
- [50] T. H. H. Aldhyani, M. Al-Yaari, H. Alkahtani, and M. Maashi, "Water quality prediction using artificial intelligence algorithms," *Applied Bionics and Biomechanics*, vol. 2020, Article ID 6659314, 2020.
- [51] R. M. Adnan, X. Yuan, O. Kisi, and Y. Yuan, "Stream flow forecasting using artificial neural network and support vector machine models," *American Scientific Research Journal for Engineering, Technology, and Sciences*, vol. 29, no. 1, pp. 286–294, 2017.
- [52] M. Koranga, P. Pant, D. Pant et al., "SVM model to predict the water quality based on physicochemical parameters," *International Journal of Mathematical, Engineering and Management Sciences*, vol. 6, no. 2, pp. 645–659, 2021.
- [53] M. Castrillo and Á. L. García, "Estimation of high frequency nutrient concentrations from water quality surrogates using machine learning methods," *Water Research*, vol. 172, Article ID 115490, 2020.
- [54] K. Chen, H. Chen, C. Zhou et al., "Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data," *Water Research*, vol. 171, Article ID 115454, 2020.
- [55] Z. Di, M. Chang, and P. Guo, "Water quality evaluation of the Yangtze River in China using machine learning techniques and data monitoring on different time scales," *Water*, vol. 11, 2019.