

## Article

# Comparative Assessment of Individual and Ensemble Machine Learning Models for Efficient Analysis of River Water Quality

Abdulaziz Alqahtani <sup>1</sup>, Muhammad Izhar Shah <sup>2,\*</sup>, Ali Aldrees <sup>1</sup> and Muhammad Faisal Javed <sup>2</sup>

<sup>1</sup> Department of Civil Engineering, College of Engineering in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj 16273, Saudi Arabia; a.qahtani@psau.edu.sa (A.A.); a.aldrees@psau.edu.sa (A.A.)

<sup>2</sup> Department of Civil Engineering, COMSATS University Islamabad, Abbottabad Campus, Abbottabad 22060, Pakistan; arabfaisal@cuiatd.edu.pk

\* Correspondence: mizharshah@gmail.com; Tel.: +92-301-8040796

**Abstract:** The prediction accuracies of machine learning (ML) models may not only be dependent on the input parameters and training dataset, but also on whether an ensemble or individual learning model is selected. The present study is based on the comparison of individual supervised ML models, such as gene expression programming (GEP) and artificial neural network (ANN), with that of an ensemble learning model, i.e., random forest (RF), for predicting river water salinity in terms of electrical conductivity (EC) and dissolved solids (TDS) in the Upper Indus River basin, Pakistan. The projected models were trained and tested by using a dataset of seven input parameters chosen on the basis of significant correlation. Optimization of the ensemble RF model was achieved by producing 20 sub-models in order to choose the accurate one. The goodness-of-fit of the models was assessed through well-known statistical indicators, such as the coefficient of determination ( $R^2$ ), mean absolute error (MAE), root mean squared error (RMSE), and Nash–Sutcliffe efficiency (NSE). The results demonstrated a strong association between inputs and modeling outputs, where  $R^2$  value was found to be 0.96, 0.98, and 0.92 for the GEP, RF, and ANN models, respectively. The comparative performance of the proposed methods showed the relative superiority of the RF compared to GEP and ANN. Among the 20 RF sub-models, the most accurate model yielded the  $R^2$  equal to 0.941 and 0.938, with 70 and 160 numbers of corresponding estimators. The lowest RMSE values of 1.37 and 3.1 were yielded by the ensemble RF model on training and testing data, respectively. The results of the sensitivity analysis demonstrated that  $\text{HCO}_3^-$  is the most effective variable followed by  $\text{Cl}^-$  and  $\text{SO}_4^{2-}$  for both the EC and TDS. The assessment of the models on external criteria ensured the generalized results of all the aforementioned techniques. Conclusively, the outcome of the present research indicated that the RF model with selected key parameters could be prioritized for water quality assessment and management.



**Citation:** Alqahtani, A.; Shah, M.I.; Aldrees, A.; Javed, M.F. Comparative Assessment of Individual and Ensemble Machine Learning Models for Efficient Analysis of River Water Quality. *Sustainability* **2022**, *14*, 1183. <https://doi.org/10.3390/su14031183>

Academic Editor: Jae Kwang (Jim) Park

Received: 7 September 2021

Accepted: 17 January 2022

Published: 21 January 2022

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



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** environmental sustainability; machine learning; ensemble learners; water quality modeling; comparative analysis; sensitivity analysis

## 1. Introduction

Rivers are one of the essential components of surface water, which is needed for industrial processes, agricultural production, and hydroelectricity generation. With the economic development and growing use of water resources, the surface water gets contaminated, lowering water quality and consequently posing serious threats to human health. The streams and rivers carry most of the waste load due to their dynamic natures [1]. Some of the main responsible factors for water pollution are human-induced activities, resulting in sewage, industrial discharge, and wastewater from urban areas [2–5]. The present study considered the total dissolved solids (TDS) and electrical conductivity (EC) as water quality indicators. Both the TDS and EC are well-accepted parameters for measuring water quality, examining salt content and organic matter in water [6,7]. The evaluation of water

quality variables such as TDS and EC is a tedious and labor-intensive procedure which requires expertise and specialized equipment [8]. On the contrary, artificial intelligence-based modeling techniques can provide a reasonable alternative to laboratory testing. The most readily used modeling methods include numerical, statistical, deterministic, and stochastic models. However, these models have some limitations, such as inadequate competencies and complex structures, and require exhaustive details about the model development [9–12]. Moreover, these traditional models showed relatively low prediction accuracies and unbalanced forecasts for various levels of water quality. The statistical-based techniques for water quality modeling assume a linear association between several predictors and predicting variables. The available literature demonstrated that the conventional models often produced inaccurate results due to the complicated hydrological processes. Therefore, more suitable, reliable, and robust modeling methods are required for water quality assessment [6,13].

Recently, an emerging class of machine learning (ML) models, such as artificial neural networks (ANNs), random forest (RF), adaptive neuro-fuzzy inference-based system (ANFIS), gene expression programming (GEP), group method of data handling (GMDH), support vector machine (SVM), and ensemble ML models were proposed and successfully applied in the literature for surface water and groundwater quality prediction [14–31]. The ANNs are the computational network models based on the biological neural network that forms the structure of human brain. Similar to the human brain in which neurons are connected to each other, the ANNs also have neurons that are interconnected to each other in different layers of the network [32,33]. These neurons are called nodes, which try to mimic the network of neurons similar to the human brain so that the computers will have an option to understand things and make decisions in a manner like the human brain. The nodes in the ANN take the input data and perform some operations. Afterward, the output of each operation is transferred to other neurons [16,34,35]. The main drawback of the ANN is that it is considered to be a black box model due to its unexplainable behavior and lack of information about the adopted process for predicting the output. Similarly, GEP is another computer-based evolutionary algorithm. The GEP has a complex tree-like structure which learns and adapts by changing shape, size, and composition just like a living organism [36]. All the GEP programs are encoded in simple and linear structures—the chromosomes. In GEP, the linear chromosomes work as the genotype and the parse tree as the phenotype, creating a multigenic system encoding multiple parse trees in each chromosome. In GEP, the parse trees are termed as expression trees (ETs), which are the result of gene expression. The ETs are used to represent different expressions in the modeling process. Despite the obvious advantages, one of the main drawbacks of the GEP is that it cannot handle exceptions, such as invalid expressions, divided by zero, and infinity [37,38]. Another ML technique that can solve classification and regression problems is known as random forest (RF). This method is referred to as the ensemble learning method, which combines various classifiers to provide the best solution to complex problems [39]. The forest created in the RF algorithm is a group of decision trees, which is commonly trained on the bagging method. Instead of relying on a single decision tree, the RF considers the prediction from each tree and, based on the majority votes of the prediction, it forecasts the output [40]. The RF creates a large number of trees and combines their outputs, making the model more complex, time consuming, and ineffective for real-time forecasting. A small change in the dataset can change the RF algorithm and reduce the model capacity for accurate predictions outside of the training data [19].

Various researchers used different models to estimate water quality parameters. Raheli et al. (2017) [41] predicted dissolved oxygen (DO) and biochemical oxygen demand (BOD) in the Langat River basin, Malaysia, using the multilayer perceptron method combined with a firefly algorithm (MLP-FFA). The prediction of the MLP-FFA was compared with the MLP model. The authors reported that the results of MLP-FFA outperformed the MLP in term of modeling accuracy. Palani et al. (2008) [42] employed the ANN in forecasting the chlorophyll-a, salinity, water temperature, and DO for coastal water. The

results revealed a good approximation, with NS and  $R^2$  both ranging from 0.8 to 0.9. Bozorg Haddad et al. (2017) [43] modeled the water quality, including TDS and EC, using support vector regression and the GP method. Good results were obtained, with  $R^2$  and NSE above 0.9 for all the cases. Najafzadeh et al. (2019) [38] used GEP, the model tree (MT), and evolutionary polynomial regression (EPR) to forecast DO, BOD, and chemical oxygen demand (COD) in the surface water. The authors reported the superior performance of the EPR compared to other methods. Nemati et al. (2015) [44] used the ANFIS, MLR, and ANN models to predict DO concentration in the Tai Po River, Hong Kong. Various parameters were used as inputs for model development. The results demonstrated the better performance of the ANN, as compared to the MLR and ANFIS. The chloride and water temperature turned out to be the most sensitive parameters in the DO prediction. Shah et al. (2021) [45] used ML- and regression-based techniques in forecasting the TDS and EC level in river water. The authors reported the accurate results of GEP as compared to other ML and regression techniques. Mosavi et al. (2021) [46] predicted the salinity of groundwater by mean of six different machine learning techniques. The authors reported the superior performance of the support vector machine compared to the other techniques. Kadam et al. (2019) [47] applied the ANN and regression modeling approach for the prediction of the groundwater quality index. Various physicochemical water quality parameters were considered for water quality index forecasting. The results of the study revealed a satisfactory range for the parameters. The predictions of the ANN were acceptable and showed satisfactory results for both the seasons, as compared with the regression model.

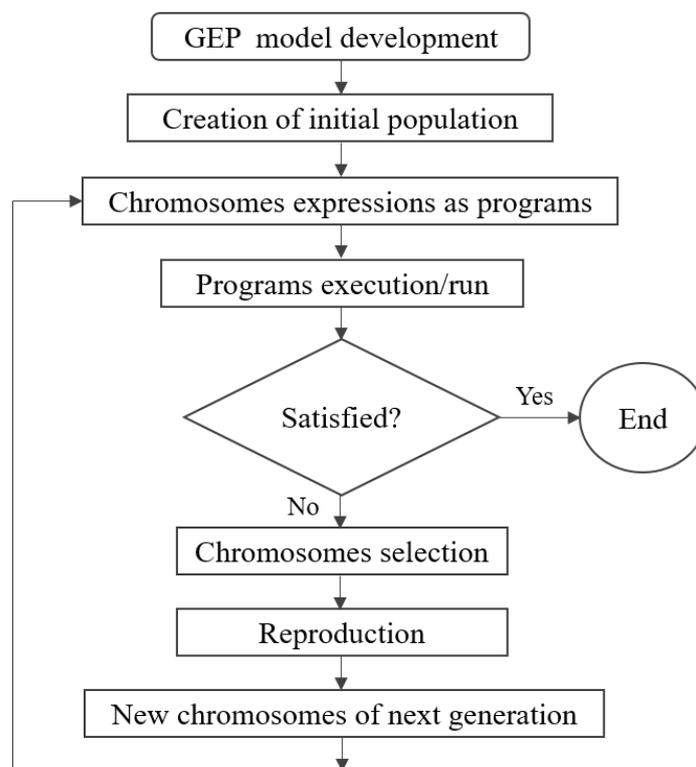
In the aforementioned literature survey, it is observed that the individual ML techniques and ensemble learners behave differently on a given dataset. Therefore, this leads the author to estimate the performance of individual and ensemble ML methods for the efficient forecast of river water quality parameters. The objective was accomplished by applying two supervised models (i.e., GEP and ANN) and an ensemble learning technique (i.e., RF). The training and testing of the models were completed based on a thirty-year dataset measured on a monthly timescale at the Upper Indus River basin (UIB) located in Pakistan. The UIB is part of Indus basin situated upstream of the Tarbela reservoir, with a 1150 km length and 165,400 km<sup>2</sup> drainage area. The optimization of the RF model was done by producing 20 sub-models in order to select the accurate one. Thereafter, the performance of the proposed models was computed on the basis of well-known statistical indicators. Percent relative error (RE%) graphs were prepared for the comparative analysis of the aforementioned individual and ensemble models. Moreover, sensitivity analysis was implemented to observe the significance of each variable in predicting the TDS and EC variables. At last, the external validation criteria were applied to judge the behavior of the developed models.

## 2. Material and Methods

### 2.1. Gene Expression Programming (GEP)

GEP is a computer-based program that imitates the biological system to model some phenomena. This efficient expression–mutation system was first developed by Candida Ferreira back in 2001 [48], and allows the encoding of expressions for the rapid application of a wide range of cross-breeding and mutation techniques. GEP is an enhanced form of the genetic algorithm (GA), which is known as evolutionary computing. These evolutionary algorithms are based on Darwin's theory: "survival of the fittest" [49]. In GEP, the chromosome (also known as genome) is comprised of one or more genes in a linear string of fixed length. The individual gene is also of fixed length and composed of primitives, which may be a terminal or a function. In the GEP process, a function can accept an argument and return a result after evaluation, while a terminal represents a variable or a constant in a given program. In the GEP process, a gene can be distributed into two parts. The first one is called the "head", which is formed by terminals and functions, while the second part is known as the "tail", which is formed by terminals only. Moreover, GEP uses the arbitrary population of individuals, adopts the fitness criteria for selection, and introduces a variation in genes based on genetic operators. GEP has the ability to quickly

mutate expressions in the way it encodes various symbols in genes. In the present research, the GEP method was applied using GeneXpro 5.0 software tools developed by Gepsoft Limited located in Portugal. It is a freely available software package which can be used for genetic-based algorithms. The flowchart of the GEP process is illustrated in Figure 1.



**Figure 1.** Flowchart of the GEP model development.

### 2.2. Random Forest (RF)

The RF is a machine learning (ML) technique that is mainly used to solve regression and classification-based problems. The RF method is based on an ensemble approach, which combines many classifiers to solve complex problems. The bagging method is usually used to train the RF algorithm, which is an ensemble technique for improving the accuracy of ML models. The RF algorithm is comprised of many decision trees, which is a basic building block of this algorithm [19]. A typical decision tree is comprised of three basic nodes: root node, leaf node, and decision node. The algorithm of the decision tree distributed a dataset into various branches, which further divided into various branches. The process of segregation into sub-branches continues until a leaf node is reached, after which no further distribution is possible. Figure 2 shows the different types of nodes in a decision tree.

### 2.3. Artificial Neural Network (ANN)

The ANNs are computing systems, and a branch of deep learning algorithms. The structure and name of ANNs are inspired by the human brain, imitating it in the way in which neurons signal to one another [50]. The artificial neuron, or node, is the basic building block for simulating the microstructures of the biological nervous system. The ANNs are composed of different node layers, encompassing input layers, hidden layers, and output layers. An individual node in the system links to an alternate one which has a specific weight and threshold. If the output of a node is beyond the indicated limit, that particular node is activated, which transfers it to the subsequent layer of the network. The neural networks solely depend on the training data to improve the network accuracy. The ANNs have been used in many fields for a variety of applications [7,51,52]. Figure 3 shows the architecture of a typical ANN.

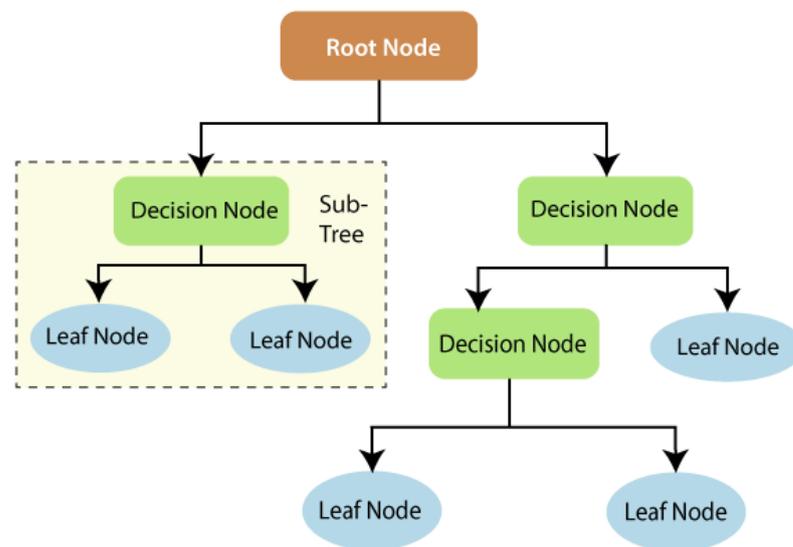


Figure 2. Types of nodes in the decision tree of the random forest (RF) algorithm.

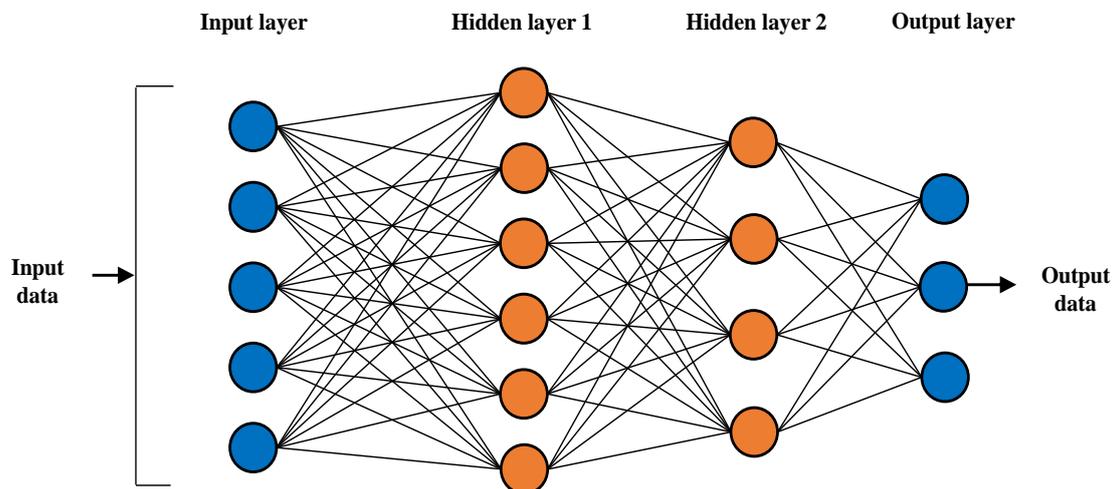


Figure 3. Structure of an ANN with different node layers.

### 3. Case Study and Modeling Database

#### 3.1. Depiction of the Study Area—The Upper Indus River Basin (UIB)

The Indus River is selected as a case study which is situated in different parts of China, Afghanistan, India, and Pakistan. With a total length of 2880 km and a drainage area of nearly 912,000 km<sup>2</sup>, the Indus River is one the major rivers in Asia. The Upper Indus Basin (UIB) is a part of the Indus basin situated upstream of the Tarbela Dam, and is 1150 km in length, 165,400 km<sup>2</sup> in drainage area, and has ice reserves of 2174 km<sup>3</sup> [53]. The altitude in the UIB varies from 455 m to 8611 m because the climate inside the basin varies significantly [45,54]. The mean annual precipitation in the UIB ranges from 100 to 200 mm, which is mainly due to the turbulences in the western mid-latitude [55,56]. The detailed land-use characteristics and soil classes with percent area is tabulated in Tables 1 and 2, respectively. All the dataset collected for this study belong to the Bisham Qilla point, which is the final gauging station before the Tarbela Reservoir. The mean annual discharge at the Bisham Qilla outlet is 2425 m<sup>3</sup>/s, as per our findings. Details of the UIB are given in Figure 4.

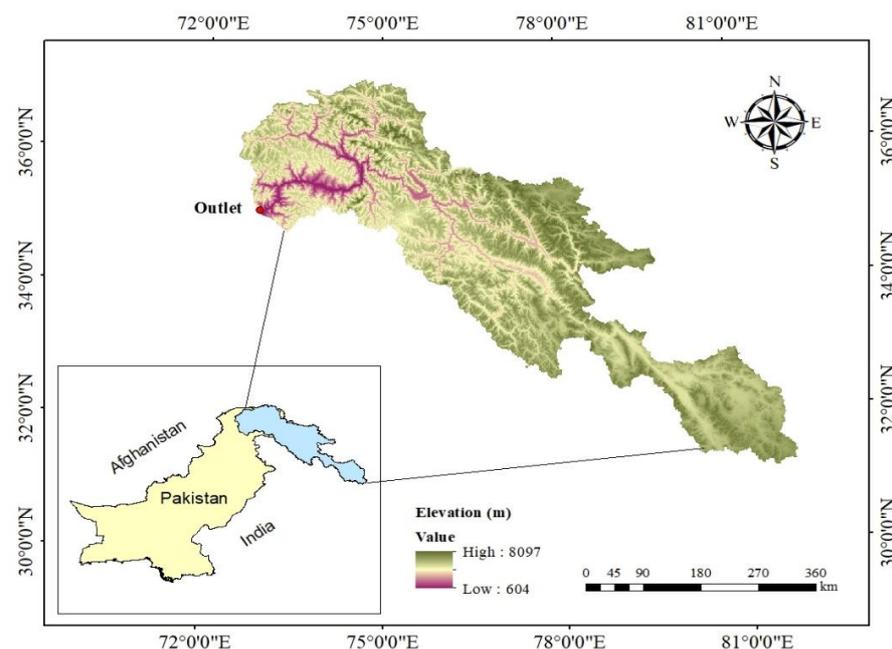
**Table 1.** Land-use classes in the study region [57].

S. No	Land-Use/Land-Cover Classes	SWAT * Classes	% Area Covered
1	Water	WATR	0.24
2	Agricultural Land—Generic	AGRL	0.70
3	Agricultural Land—Row Crops	AGRR	0.10
4	Agricultural Land—Close-grown	AGRC	0.02
5	Hay	HAY	0.11
6	Forest—Mixed	FRST	0.18
7	Forest—Deciduous	FRSD	2.82
8	Forest—Evergreen	FRSE	0.27
9	Wetlands—Mixed	WETL	0.01
10	Wetlands—Forested	WETF	20.08
11	Wetlands—Non-Forested	WETN	0.01
12	Pasture	PAST	0.37
13	Summer Pasture	SPAS	0.01
14	Winter Pasture	WPAS	0.01
15	Range—Grasses	RNGE	15.45
16	Range—Brush	RNGB	59.62

\* Soil and Water Assessment Tool.

**Table 2.** Soil classes in the study region [57].

S. No	FAO Soil Type	% Area Covered	Texture	Clay%	Silt%	Sand%
1	Ao72-2b-3644	4.14	Sandy—Loam	16	19	65
2	Be72-2a-3669	4.91	Loam	22	36	42
3	Be72-2c-3671	2.59	Loam	22	36	42
4	Be78-2c-3679	6.15	Loam	23	37	40
5	GLACIER-6998	19.67	UWB	5	25	70
6	I-B-U-2c-3503	9.15	Loam	26	30	44
7	I-B-U-2c-3713	10.55	Loam	26	30	44
8	I-B-U-3712	15.98	Loam	26	30	44
9	I-Gx-2c-3720	0.01	Loam	19	34	48
10	I-K-U-2c-3723	0.04	Loam	26	28	46
11	I-X-2c-3731	12.51	Loam	22	33	45
12	I-Y-2c-3733	14.30	Loam	23	39	38



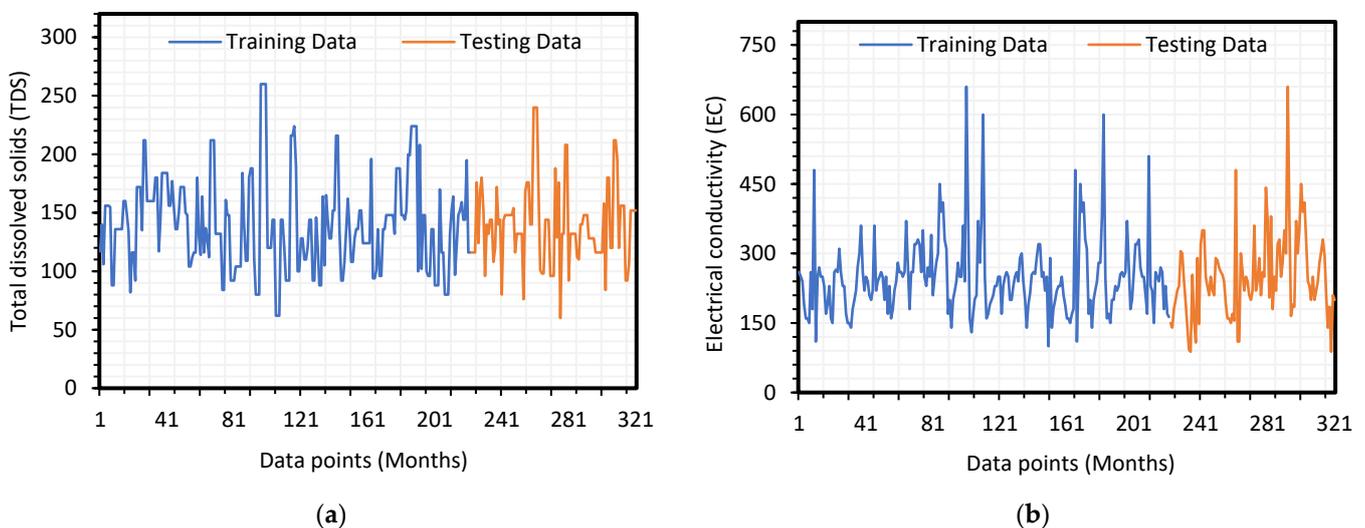
**Figure 4.** Details of the study area.

### 3.2. Modeling Dataset

For the present research, the water quality dataset measured at the Bisham Qilla point was obtained from WAPDA, Pakistan. The final dataset contained 321 data points collected monthly between 1975 and 2005. The dataset included the information of a total of nine parameters, namely; bicarbonates ( $\text{HCO}_3^-$ ), calcium ( $\text{Ca}^{2+}$ ), sulphate ( $\text{SO}_4^{2-}$ ), hydrogen ion concentration (pH), magnesium ( $\text{Mg}^{2+}$ ), chloride ( $\text{Cl}^-$ ), sodium ( $\text{Na}^+$ ), total dissolved solids (TDS), and electrical conductivity (EC). The datasets were statistically analyzed to figure out the relationship of the inputs and outputs. The basic statistical variables, including mean, minimum value, maximum value, and standard deviation of the modeling dataset, are illustrated in Table 3. Figure 5 graphically demonstrates a variation in TDS and EC with time. Considering a substantial correlation of TDS and EC with other water quality data, the seven most correlated parameters ( $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Na}^+$ ,  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{HCO}_3^-$ , and pH) were used as modeling inputs for GEP, RF, and ANN models. The correlation matrix of TDS and EC data is tabulated in Table 4. The final dataset was divided into 70% and 30% for models training and testing, respectively.

**Table 3.** Summary of statistics of the water quality parameters.

Variable	Unit	Minimum	Maximum	Mean Value	Range	Standard Deviation
$\text{Ca}^{2+}$	meq/L	0.65	2.45	1.46	1.80	0.32
$\text{Mg}^{2+}$	meq/L	0.04	2.64	0.63	2.61	0.33
$\text{Na}^+$	meq/L	0.05	9.0	0.53	8.95	0.69
$\text{Cl}^-$	meq/L	0.05	4.2	0.28	4.15	0.28
$\text{SO}_4^{2-}$	meq/L	0.1	3.2	0.55	3.10	0.37
$\text{HCO}_3^-$	meq/L	0.3	7.4	1.73	7.10	0.63
TDS	ppm	60	260	139.87	200	38.64
EC	$\mu\text{S}/\text{cm}$	92	650	242.65	558	67.49
pH	-	7.08	8.3	7.83	1.22	0.65



**Figure 5.** Observed data used for models training and testing: (a) Total Dissolved Solids (TDS); (b) Electrical Conductivity (EC) of the Bisham Qilla station on Upper Indus River.

The water quality modeling using the GEP, RF, and ANN models was carried out using four main steps: (1) data preparation, (2) model development, (3) model assessment and validation, and (4) model robustness analysis. In the first phase, the data were analyzed to remove the outliers and statistically confirm the association of modeling outputs with the input dataset. In the second phase, the GEP, RF, and ANN models were developed with 70% data used for training and 30% for model testing. In the third phase, the models were assessed by employing four well-known statistical assessment indicators. On the basis of

statistical values, the most accurate and reliable model was selected. At last, the accuracy of the models was verified using a validation criterion adopted from the literature.

**Table 4.** Correlation of the output parameters with modeling inputs.

Parameters	Ca	Mg	Na	HCO <sub>3</sub> <sup>−</sup>	Cl	SO <sub>4</sub>	PH	TDS
Ca <sup>2+</sup>	1							
Mg <sup>2+</sup>	0.0194	1						
Na <sup>+</sup>	−0.0037	0.4712	1					
HCO <sub>3</sub> <sup>−</sup>	0.0363	0.5324	0.7414	1				
Cl <sup>−</sup>	0.0239	0.5035	0.7041	0.5296	1			
SO <sub>4</sub> <sup>2−</sup>	0.0212	0.5415	0.4853	0.2749	0.3698	1		
pH	0.0025	0.0737	0.0415	0.0545	0.0561	−0.0445	1	
TDS	0.7452	0.7001	0.8629	0.8176	0.7411	0.6297	0.5210	1
Ca <sup>2+</sup>	1							
Mg <sup>2+</sup>	0.0194	1						
Na <sup>+</sup>	−0.0137	0.4712	1					
HCO <sub>3</sub> <sup>−</sup>	0.0675	0.5324	−0.5756	1				
Cl <sup>−</sup>	0.0239	0.5086	0.7638	0.5296	1			
SO <sub>4</sub> <sup>2−</sup>	0.0674	0.4671	−0.3460	0.2749	0.2698	1		
pH	0.0455	0.3005	0.0627	0.2177	0.3417	0.4297	1	
EC	0.6539	0.8632	0.5672	0.8545	0.8951	0.7954	0.6202	1

### 3.3. Models Performance Evaluation and Validation

The statistical indicators were used to evaluate the performance and relative position or change in the model-predicted data. The selected indicators for this study are well-known, and included the coefficient of determination ( $R^2$ ), root mean squared error (RMSE), Nash–Sutcliffe efficiency (NSE), and mean absolute error (MAE) [58]. The aforementioned indicators are frequently used in modeling studies to assess the accuracy of the model's predicted and actual data. These indicators are respectively defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - M_i)^2}{N}} \quad (1)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (M_i - P_i)^2}{\sum_{i=1}^n (M_i - \bar{M}_i)^2} \quad (2)$$

$$R^2 = \frac{\sum_{i=1}^n (M_i - \bar{M}_i) (P_i - \bar{P}_i)}{\sqrt{\sum_{i=1}^n (M_i - \bar{M}_i)^2 \sum_{i=1}^n (P_i - \bar{P}_i)^2}} \quad (3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - M_i| \quad (4)$$

where  $n$  is the total number of data points,  $M_i$  and  $P_i$  are the respective actual and predicted datasets, and  $\bar{M}_i$  and  $\bar{P}_i$  respectively represent the average of actual and predicted data.

### 3.4. Sensitivity Analysis

Sensitivity analysis often involves varying the inputs of a system to assess the impact of each input on the desired output, which ultimately gives information on the various effects of the individually tested variable. This method is used in a variety of disciplines to

optimize the efficacy of a given system [59]. The sensitivity analysis could be a powerful tool in providing additional insights for a system that would have otherwise been missed. This method has the potential to describe the substantial input parameters that have a direct effect on the output. It is indispensable in ML-based modeling to figure out the most sensitive input parameters which have been proposed in literature [45,60]. In the present study, the technique proposed by Gandomi et al. (2013) [61] was adopted for sensitivity analysis. The authors developed Equations (5) and (6) to demonstrate the impact of inputs.

$$N_i = f_{max}(x_i) - f_{min}(x_i) \quad (5)$$

$$S_i = \frac{N_i}{\sum_{j=1}^n N_j} \times 100 \quad (6)$$

where  $f_{max}(x_i)$ . and  $-f_{min}(x_i)$  indicated the highest and lowest values over the  $i$ th output.

## 4. Results and Discussion

### 4.1. GEP Model Development and Output

The GEP model developed for water quality forecasts was chosen after completing a set of iterations with basic function sets and the smallest head size. The final model, trained on monthly TDS and EC datasets, was used to develop the below-mentioned Equations (7) and (8) in order to estimate TDS and EC concentrations with some input variables, i.e., Ca, Mg, Na, Cl,  $\text{SO}_4$ ,  $\text{HCO}_3^-$ , and pH.

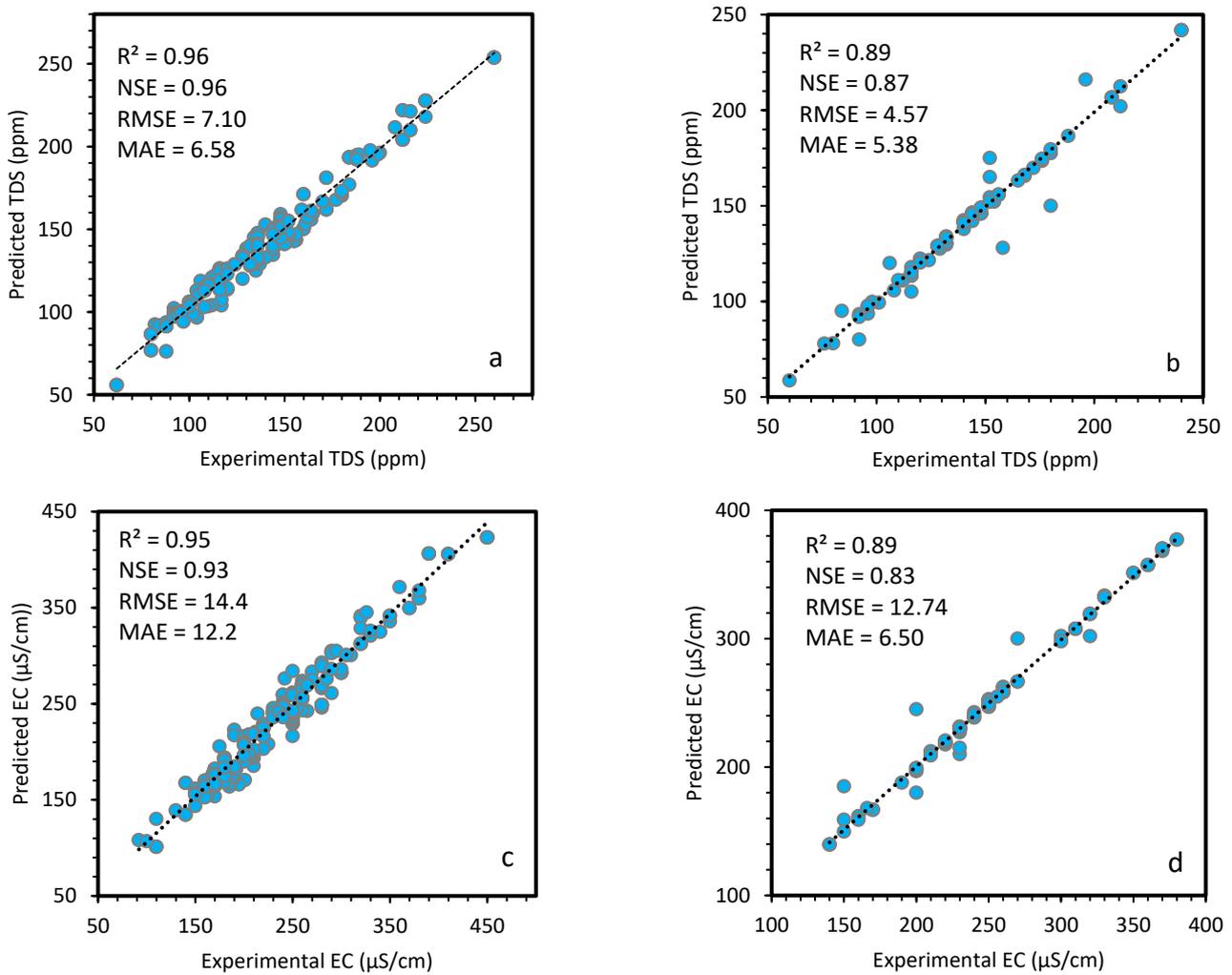
$$\begin{aligned} \text{TDS} = & \left\{ \left( \frac{20390}{\text{Ca}} \right)^{\frac{1}{3}} - 22\text{HCO}_3 \right\} \times (\text{SO}_4 - \text{HCO}_3)^{\frac{1}{3}} + \frac{1}{\text{HCO}_3^{\frac{1}{3}} \times \ln(8.14\text{Cl} - 1.11)^2} + (4.15 + \text{Na}) \times 25 \\ & - \text{Na} \times \text{HCO}_3 - \frac{28}{\text{Ca}} \left( \text{Mg} \times \text{Cl} \times 1.17 + \frac{\text{SO}_4}{1.03} \right) \times (51 - 7.33\text{Cl}) \end{aligned} \quad (7)$$

$$\begin{aligned} \text{EC} = & (9.6\text{Cl} + 5.1\text{SO}_4 - \ln \text{HCO}_3) \times (4.9 - \text{HCO}_3)^2 + \left\{ (\text{SO}_4 \times \text{PH} \times 2.6 - \frac{32.7}{\text{Ca}}) - 5.8 \right\} \times \text{HO}_3 \\ & + (\text{Na} + \text{HCO}_3 \times 10.4 - \text{SO}_4 \times \text{Cl}) \times 12.12 \end{aligned} \quad (8)$$

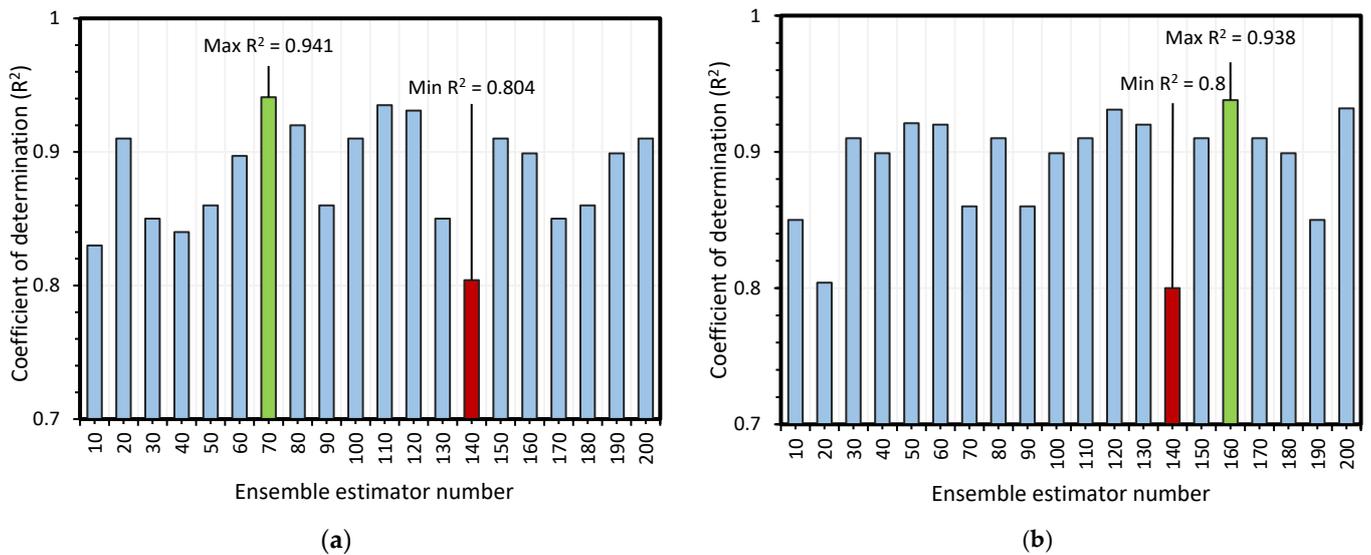
Figure 6a–d show the GEP model predictions for the TDS and EC data, which demonstrated that the proposed model was successfully trained and tested on the given input data. The output of the GEP model estimation and the actual TDS data depict the NSE,  $R^2$ , MAE, and RMSE values as 0.96, 0.96, 6.58, and 7.10 on the training set, and 0.87, 0.89, 5.38, and 4.57 on the testing set, respectively. Likewise, the aforementioned indicators between the GEP-simulated results and the actual EC data are 0.93, 0.95, 12.2, and 14.4 for the training set, and the goodness-of-fit are found to be 0.83, 0.89, 6.50, and 12.74 for the EC testing dataset, respectively. The literature survey demonstrated that the coefficient of determination ( $R^2$ ) above 0.8 is considered a reasonable value [62]. Moreover, a small RMSE value and higher  $R^2$  and NSE values indicated the acceptable estimation made by the model when compared with actual data [63]. The results forecasted by the GEP model showed the  $R^2$  value as being above 0.85 for the TDS and EC training, as well as the testing data.

### 4.2. Random Forest Model for TDS and EC

The RF is an ensemble machine learning approach based on weak learners which gives the finest model on the basis of  $R^2$  value. The RF model applied in the present study randomly divided the whole dataset into 20 sub-models based on ensemble  $n$ -estimators. Among all the sub-models, the optimized model provided the best  $R^2$  value, as depicted in Figure 7a,b for the TDS and EC models, respectively. All the sub-models for the TDS have  $R^2$  values above 0.75, with an average value of 0.9221, indicating that each of the sub-models correlated with the TDS-predicted and actual data. Among all the 20 sub-models, the maximum and minimum  $R^2$  for the TDS was found to be 0.941 and 0.805, with 70 and 140 estimators, respectively. Furthermore, among the sub-classes for the EC model, the mean and maximum  $R^2$  value turned out to be 0.928 and 0.938, respectively.

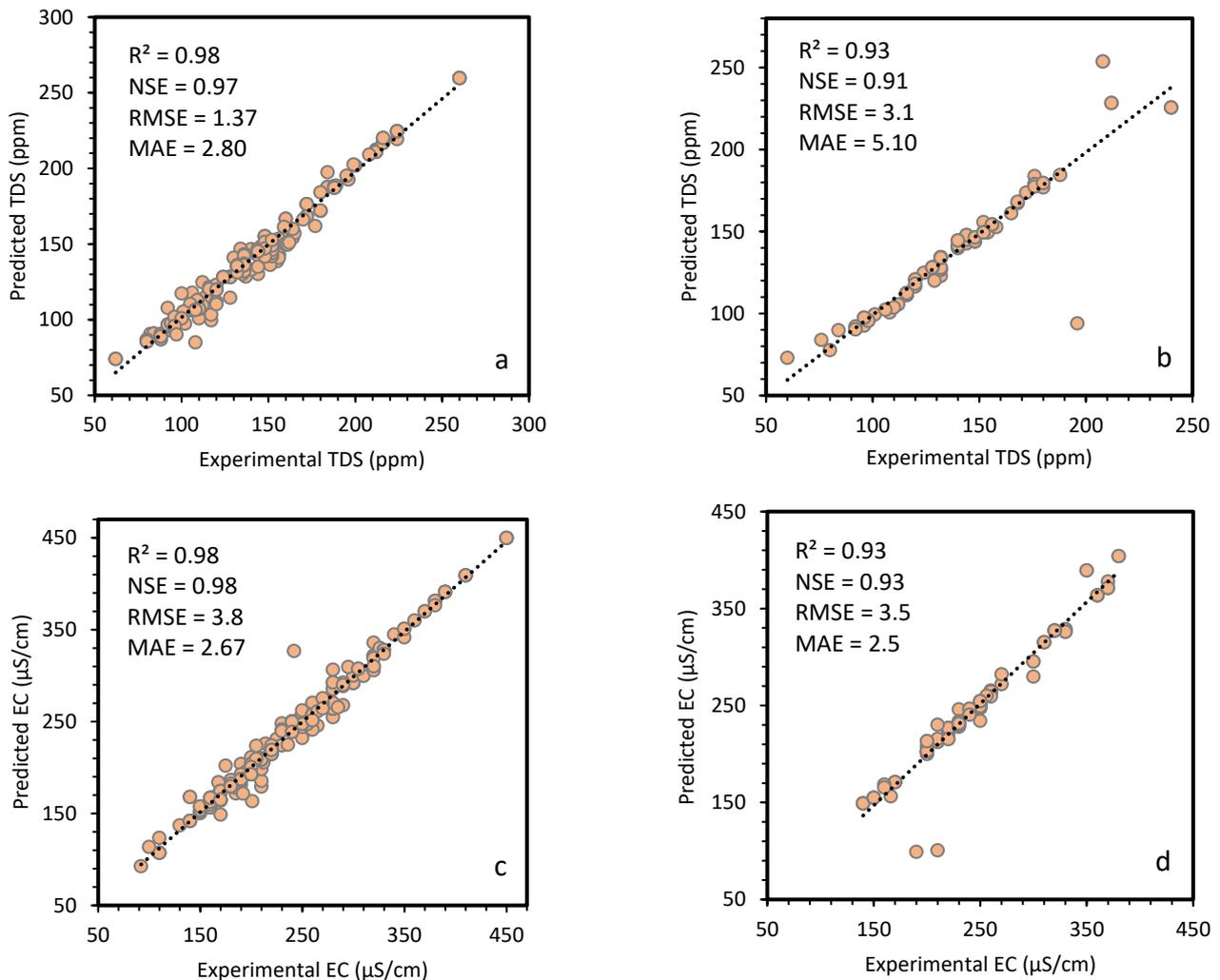


**Figure 6.** Analysis of the GEP-predicted data versus the actual data: (a) TDS training; (b) TDS testing; (c) EC training; (d) EC testing.



**Figure 7.** A total of 20 RF sub-models with varying numbers of estimators with the highest  $R^2$  (green) selected for: (a) TDS; (b) EC.

The RF model output is given in Figure 8a–d against the measured TDS and EC data. The assessment indicators, i.e., NSE,  $R^2$ , and RMSE, were found to be 0.97, 0.98, and 1.37, respectively, for training the RF model on TDS data, and 0.91, 0.93, and 3.10, respectively, for the TDS testing data. Furthermore, the RF model-simulated outcome versus the actual EC data is given in Figure 8c,d for the training and testing datasets, respectively. The values of NSE,  $R^2$ , MAE, and RMSE were equal to 0.98, 0.98, 2.67, and 3.81 for the EC training set, and 0.93, 0.93, 2.5, and 3.5 for the testing of the RF model for the EC estimation. The results predicted by the RF model demonstrated that the NSE and  $R^2$  values were above 0.9 for all the cases, therefore highlighting the accuracy and supremacy of the ensemble-based modeling.

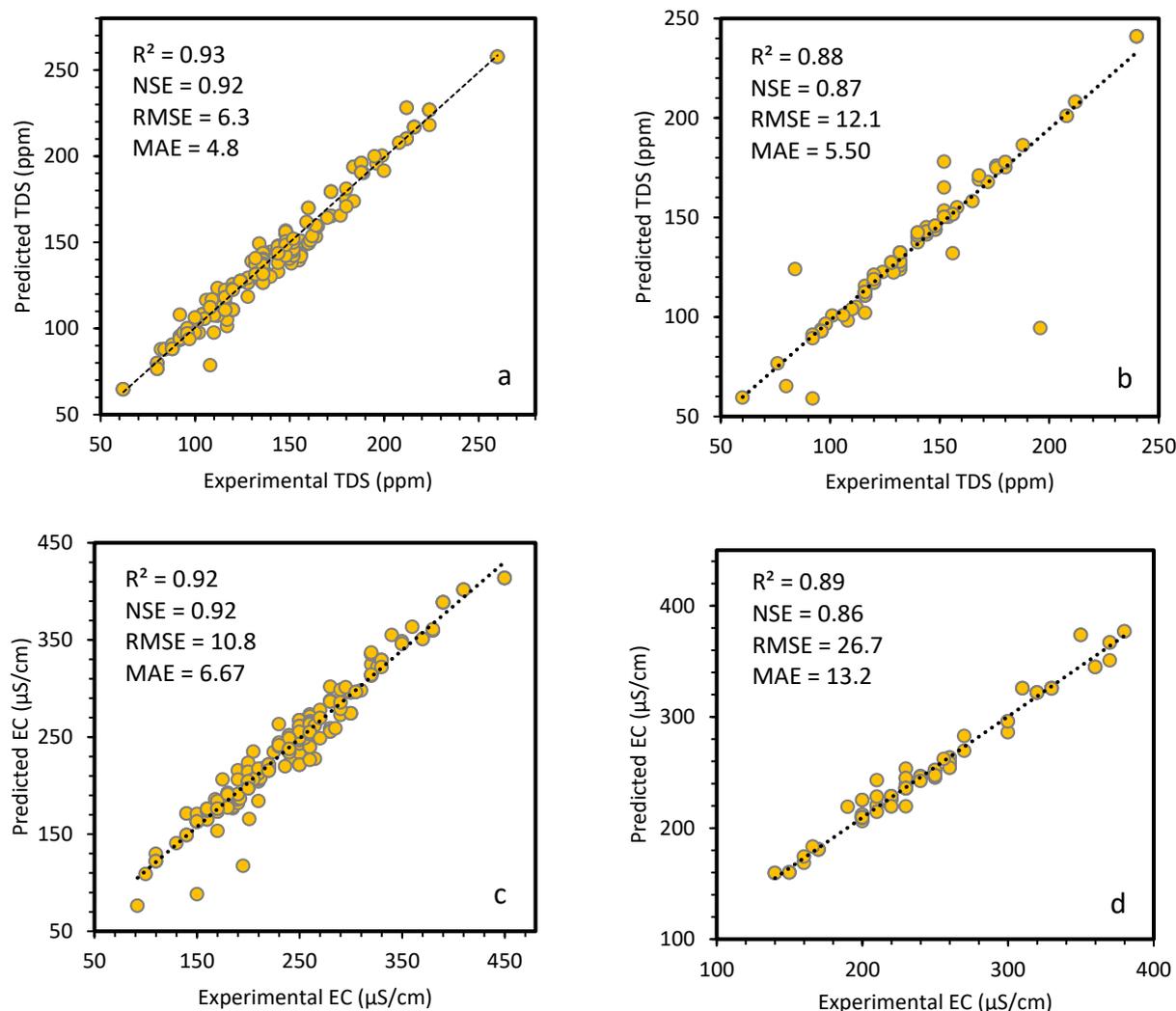


**Figure 8.** Analysis of the RF prediction versus actual data: (a) TDS training; (b) TDS testing; (c) EC training; (d) EC testing.

#### 4.3. Neural Network Model for TDS and EC

The output of the ANN model is illustrated in Figure 9a,b for the TDS, and Figure 9c,d for the EC data, respectively. The ANN-forecasted outcome shows a good agreement against the actual TDS and EC data. The evaluation parameters, NSE,  $R^2$ , MAE, and RMSE, for the TDS training set were observed to be 0.92, 0.93, 4.8, and 6.3. The aforementioned indicators were 0.87, 0.88, 5.50, and 12.1, respectively, for the TDS testing set. Similarly, the ANN modeling outcome for forecasting the EC data is illustrated in Figure 9c,d for the training as well as testing dataset. The RMSE was found to be 10.8 and 26.7 for the EC training and testing sets, respectively. Moreover, the NSE and  $R^2$  were found to be 0.92 each (training set) and 0.86 and 0.89 (testing set) in EC modeling. The results of the ANN

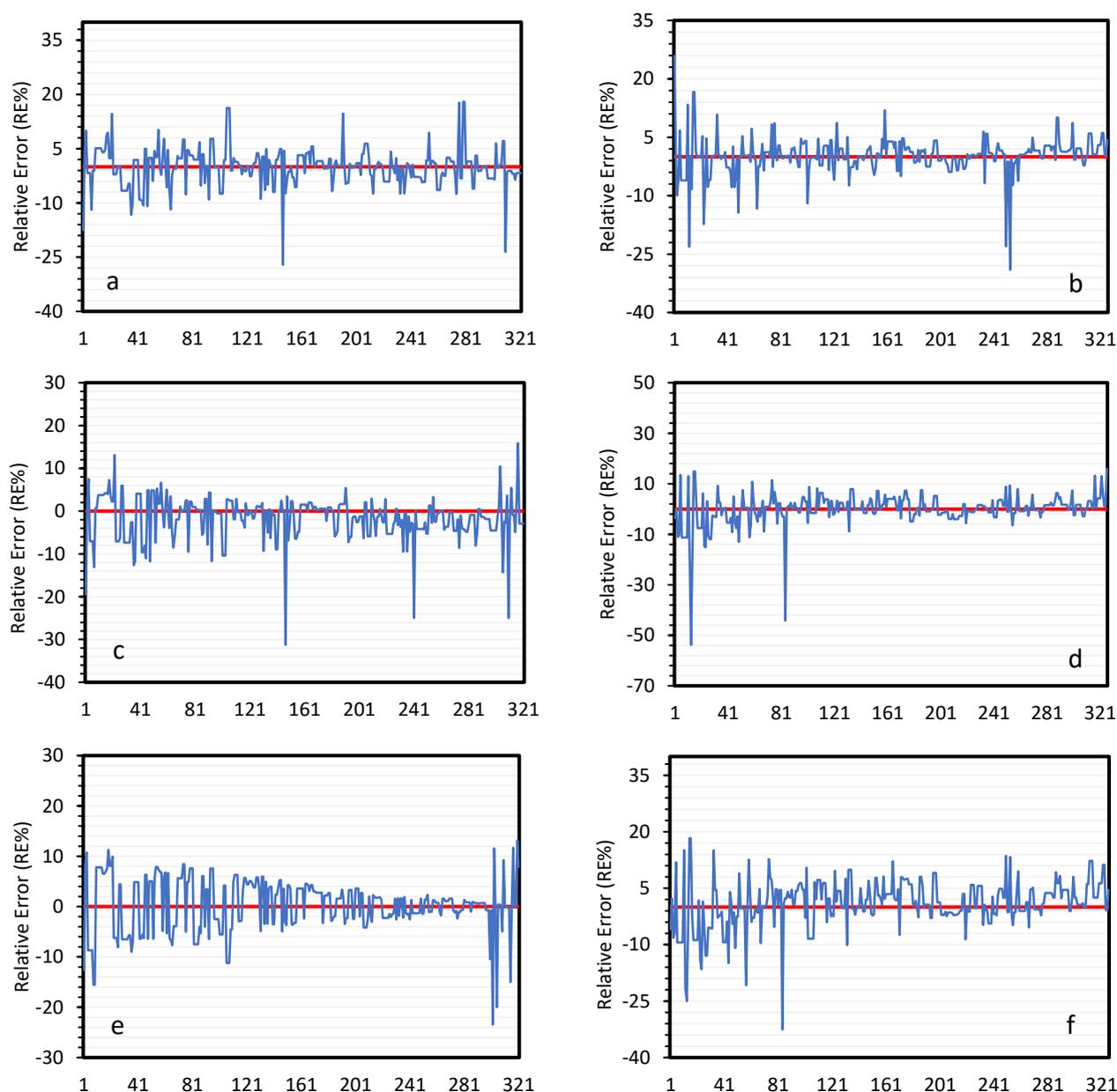
modeling demonstrated the reduced efficacy of the model on the testing set compared to the training data. The values of the statistical indicators, i.e., NSE and  $R^2$ , were reduced on the TDS and EC testing datasets, which may be attributed to the complex structure, a reduced trust in the network, and the unexplainable behavior of the ANN models.



**Figure 9.** Analysis of the ANN prediction versus actual data: (a) TDS training; (b) TDS testing; (c) EC training; (d) EC testing.

#### 4.4. Comparative Analysis of GEP, RF, and ANN Models

Figure 10a–f demonstrates the percent relative error (RE%) graphs of the GEP, RF, and ANN models. The results of the error demonstrated that most of the GEP model results range from  $-25\%$  to  $+25\%$  for both of the datasets. The error yielded by the GEP model equals to 6.1 and 7.01 for the TDS and EC predictions. The maximum error yielded in the RF model was 5.3 and 6.9 for the TDS and EC data, respectively. The results of the RF model depicted that almost 80% of the TDS and EC data lies from  $-20\%$  to  $+20\%$ , which highlighted the accuracy and predictability of the RF-based ensemble learning method. The average error in the TDS and EC data yielded by the ANN model was 8.9 and 10.5, respectively.



**Figure 10.** Percent relative error (RE%) yielded: (a,b) TDS and EC predicted by GEP; (c,d) TDS and EC predicted by RF; (e,f) TDS and EC predicted by ANN.

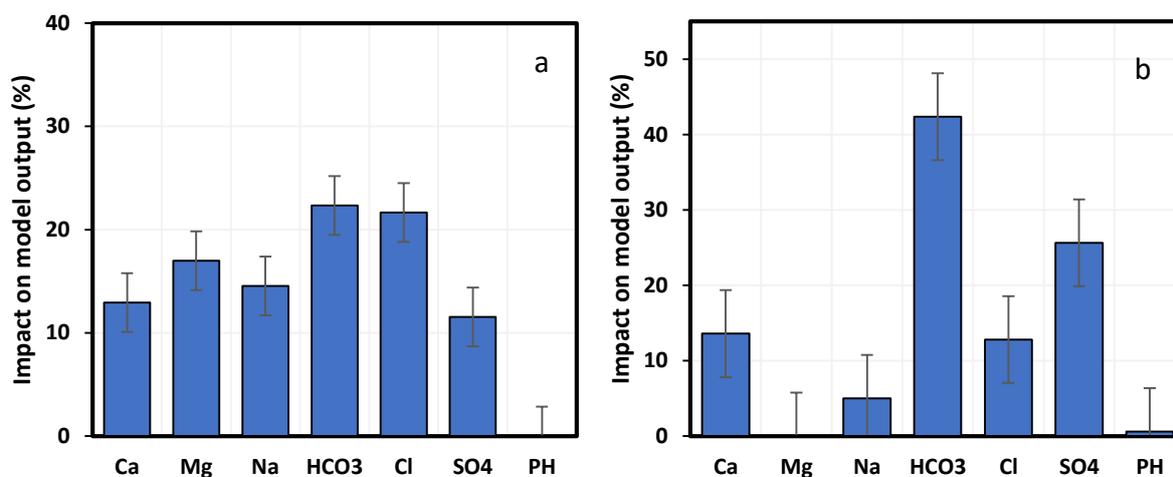
The performance-measured statistical indicators for all the models, i.e., GEP, RF, and ANN, are listed in Table 5. In term of accuracy, the RF model outclasses other models on the training dataset as well as the testing dataset, followed by the GEP and ANN models. The highest  $R^2$  value (0.98) and lowest RMSE value (3.1) were both attained by the RF model. Moreover, the lowest %RE was also observed in the RF model estimation, which highlights the overall efficiency of the ensemble RF method. The RF model was followed by the GEP, which also achieved the  $R^2$  above 0.9 for all the datasets. The performance of the ANN was found to be the least accurate when compared to the RF and GEP models. However, the ANN provided acceptable results, with the  $R^2$  being more than 0.85 for the TDS and EC data. The ANN model provided good results on training data but showed poor performance on testing data, which can be related to inexplicable behavior and difficult network structure [64,65].

**Table 5.** Results for GEP, RF, and ANN models.

		Training Set				Testing Set			
		R <sup>2</sup>	RMSE	NSE	MAE	R <sup>2</sup>	RMSE	NSE	MAE
TDS	GEP	0.96	7.10	0.96	6.58	0.89	4.57	0.87	5.38
	RF	0.98	1.37	0.97	2.80	0.93	3.1	0.91	5.10
	ANN	0.92	6.37	0.93	4.80	0.88	13.1	0.87	5.50
EC	GEP	0.95	14.4	0.93	12.2	0.89	12.74	0.83	6.51
	RF	0.98	3.8	0.98	2.67	0.93	3.52	0.93	2.5
	ANN	0.92	10.81	0.92	6.67	0.89	26.7	0.86	13.2

#### 4.5. Sensitivity Analysis Output

In the present study, the sensitivity analysis was carried out by holding all the parameters at their mean values while changing one of the inputs to observe the effects of changing one variable. Through this procedure, the input parameters were ranked based on their sensitivity and effect on the output, as illustrated in Figure 11. The results demonstrated that the input parameters, i.e., Ca, Mg, Na, HCO<sub>3</sub><sup>-</sup>, Cl, SO<sub>4</sub>, and pH, contributed 12.92%, 16.98, 14.55%, 22.33%, 21.66%, 11.55%, and 0% in the modeling for the TDS, which shows that the HCO<sub>3</sub><sup>-</sup> is the most important and sensitive variable for the TDS. Similarly, the result depicts the influence of Ca, Mg, Na, HCO<sub>3</sub><sup>-</sup>, Cl, SO<sub>4</sub>, and pH as 13.59%, 0%, 5.01%, 42.36%, 12.8%, 25.63%, and 0.61%, respectively, on EC output. Overall, the results showed that HCO<sub>3</sub><sup>-</sup> is the most sensitive parameter, while pH is the least influential parameter for both TDS and EC.

**Figure 11.** Variable importance and contribution of inputs to targeted output: (a) TDS; (b) EC.

#### 4.6. External Validation (EV) of the Models

Evaluating the generalizability of machine learning (ML) models is essential as it has various implications. Therefore, external validation (EV) criteria have been recommended in the literature. EV often involves the utilization of independent datasets to validate the performance of a model that was initially trained on an input dataset [66]. The EV is considered as being significant evidence to judge the generalizability of ML models. As the EV comes from an external validation source, any feature selected peculiarly in the input dataset would likely fail in this process. Therefore, a better performance of the models during EV is considered as being proof of the generalizability. The present study adopted the same EV criteria for model assessment through the measures recommended in the literature studies. Gholampour et al. (2017) [67] argued that the precision of a model was mainly dependent on the number of data points in the training set. Frank and Todeschini [68] put forward that the ratio among the number of data points and input variables should be above five for the better approximation capability of the model. In

this study, the suggested ratio is 45.8 (321/7), which completely fulfills the EV criteria. Another criterion put forward by Golbraikh et al. (2002) [69] proposed that the slope of the particular line which is passing from the origin must be nearly equal to one. A new assessment indicator, i.e.,  $R_m$  for EV, was developed by Roy et al. (2008) [70], who suggested that for a good model, the  $R_m$  must be higher than 0.5. Moreover, the two new indicators, i.e.,  $R_0^2$  and  $R_0'^2$ , were proposed by Alavi et al. (2011) [71], and these criteria are satisfied when both of the indicators yield approximately one. In this study, the predicted performance of all the developed models has been evaluated through the abovementioned EV criteria. The result shows that all the models successfully passed the EV criteria and the outcome is given in Table 6.

**Table 6.** Results of external validation criteria.

S. No.	Equation	Condition	Model	Value	Suggested by
1.	$R = \frac{\sum_{i=1}^n (M_i - \bar{M}_i)(P_i - \bar{P}_i)}{\sqrt{\sum_{i=1}^n (M_i - \bar{M}_i)^2 \sum_{i=1}^n (P_i - \bar{P}_i)^2}}$	$R > 0.8$	GEP	0.96	[68]
			RF	0.98	
			ANN	0.97	
2.	$k = \frac{\sum_{i=1}^n (M_i - P_i)}{M_i^2}$	$0.85 < k < 1.15$	GEP	1.004	[69]
			RF	0.997	
			ANN	0.992	
3.	$k' = \frac{\sum_{i=1}^n (M_i - P_i)}{P_i^2}$	$0.85 < k' < 1.15$	GEP	0.995	[69]
			RF	1.002	
			ANN	1.007	
4.	$R_m = R^2 \times \left(1 - \sqrt{ R^2 - R_0^2 }\right)$	$R_m > 0.5$	GEP	0.799	[70]
			RF	0.820	
			ANN	0.811	
5.	$R_0^2 = \frac{\sum_{i=1}^n (P_i - M_i^0)^2}{\sum_{i=1}^n (P_i - \bar{P}_i^0)^2}, M_i^0 = k \times P_i$	$R_0^2 \cong 1$	GEP	0.999	[71]
			RF	0.999	
	ANN	0.999			
	$R_0'^2 = \frac{\sum_{i=1}^n (M_i - P_i^0)^2}{\sum_{i=1}^n (M_i - \bar{M}_i^0)^2}, P_i^0 = k' \times M_i$	$R_0'^2 \cong 1$	GEP	0.999	
	RF		0.999		
		ANN	0.999		

### 5. Conclusions

The present research aims to develop and compare the forecasting precision of the traditional and ensemble machine learning models for predicting surface water quality, and to identify the best model with essential parameters to serve the water quality assessment in the best and most precise manner. The predictive performances of two individual supervised models (GEP and ANN) and one ensemble learning model (RF) were comprehensively developed and compared based on a consistent water quality dataset. The key findings of this study are:

- An excellent prediction capability has shown by the RF model compared to other methods, which highlighted the overall supremacy of the ensemble learning techniques;
- Two mathematical expressions were established for TDS and EC prediction, highlighting the uniqueness of the GEP method. These expressions can be easily used for water quality monitoring and assessment with some known input parameters;
- The reduced performance of the ANN was observed in comparison with the RF and GEP, which can be attributed to the inexplicable behavior and difficult structure of the neural network;

- Important variables for TDS and EC modeling were identified through sensitivity analysis, where the  $\text{HCO}_3^-$  remained the most sensitive input for both the outputs;
- The modeling outcome verified by external criteria ensured the generalized modeling capability of the aforementioned techniques;
- The research conducted in this paper can be reckoned as a data mining-based study for water quality monitoring and assessment. Eventually, the authors recommend to conduct a heavy study and establish a widespread database for other water quality parameters considering a number of explanatory variables.

It is recommended to study the new database using other advanced AI modelling techniques, for instance, gradient-boosted trees (GBT), multivariate adaptive regression spline (MARS), support vector regression (SVR), recurrent neural network (RNN), multilayer perceptron neural network (MLPNN), among others. The machine learning techniques are still encountering problems due to flaws like knowledge extraction, interpretability, and the uncertainty of the model. To have a better understanding of the learning process, extra emphasis must be devoted to expanding on previous knowledge about the underlying physical phenomena, such as engineering judgement or human skill.

**Author Contributions:** Supervision, investigation, and review, A.A. (Abdulaziz Alqahtani); conceptualization, data analysis, methodology, writing—original draft preparation, M.I.S.; data curation, investigation, and review, A.A. (Ali Aldrees); formal analysis, modeling, and validation, M.F.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This project was supported by the Deanship of Scientific Research at Prince Sattam Bin Abdulaziz University under the research project No. 2020/01/16488.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data used in this study was collected from Water & Power Development Authority (WAPDA), Pakistan.

**Conflicts of Interest:** The authors declare that they have no conflict of interest.

## References

1. Najafzadeh, M.; Ghaemi, A. Prediction of the five-day biochemical oxygen demand and chemical oxygen demand in natural streams using machine learning methods. *Environ. Monit. Assess.* **2019**, *191*, 1–21. [\[CrossRef\]](#)
2. Al-Mukhtar, M.; Al-Yaseen, F. Modeling water quality parameters using data-driven models, a case study Abu-Ziriq marsh in south of Iraq. *Hydrology* **2019**, *6*, 24. [\[CrossRef\]](#)
3. Li, K.; Wang, L.; Li, Z.; Xie, Y.; Wang, X.; Fang, Q. Exploring the spatial-seasonal dynamics of water quality, submerged aquatic plants and their influencing factors in different areas of a lake. *Water* **2017**, *9*, 707. [\[CrossRef\]](#)
4. Singh, K.P.; Malik, A.; Mohan, D.; Sinha, S. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—A case study. *Water Res.* **2004**, *38*, 3980–3992. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Shah, M.I.; Javed, M.F.; Alqahtani, A.; Aldrees, A. Environmental assessment based surface water quality prediction using hyper-parameter optimized machine learning models based on consistent big data. *Process Saf. Environ. Prot.* **2021**, *151*, 324–340. [\[CrossRef\]](#)
6. Jamei, M.; Ahmadianfar, I.; Chu, X.; Yaseen, Z.M. Prediction of surface water total dissolved solids using hybridized wavelet-multigene genetic programming: New approach. *J. Hydrol.* **2020**, *589*, 125335. [\[CrossRef\]](#)
7. Najah, A.; El-Shafie, A.; Karim, O.A.; El-Shafie, A.H. Application of artificial neural networks for water quality prediction. *Neural Comput. Appl.* **2013**, *22*, 187–201. [\[CrossRef\]](#)
8. Sattari, M.T.; Joudi, A.R.; Kusiak, A. Estimation of Water Quality Parameters with Data—Driven Model. *J. Am. Water Work. Assoc.* **2016**, *108*, E232–E239. [\[CrossRef\]](#)
9. Vats, S.; Sagar, B.B.; Singh, K.; Ahmadian, A.; Pansera, B.A. Performance evaluation of an independent time optimized infrastructure for big data analytics that maintains symmetry. *Symmetry* **2020**, *12*, 1274. [\[CrossRef\]](#)
10. Pakdaman, M.; Falamarzi, Y.; Yazdi, H.S.; Ahmadian, A.; Salahshour, S.; Ferrara, F. A kernel least mean square algorithm for fuzzy differential equations and its application in earth's energy balance model and climate. *Alex. Eng. J.* **2020**, *59*, 2803–2810. [\[CrossRef\]](#)
11. Mosavi, A.; Hosseini, F.S.; Choubin, B.; Goodarzi, M.; Dineva, A.A. Groundwater salinity susceptibility mapping using classifier ensemble and Bayesian machine learning models. *IEEE Access* **2020**, *8*, 145564–145576. [\[CrossRef\]](#)

12. Molekoa, M.D.; Avtar, R.; Kumar, P.; Minh, H.V.T.; Kurniawan, T.A. Hydrogeochemical assessment of groundwater quality of Mokopane area, Limpopo, South Africa using statistical approach. *Water* **2019**, *11*, 1891. [[CrossRef](#)]
13. Shah, M.I.; Abunama, T.; Javed, M.F.; Bux, F.; Aldrees, A.; Tariq, M.A.U.R.; Mosavi, A. Modeling Surface Water Quality Using the Adaptive Neuro-Fuzzy Inference System Aided by Input Optimization. *Sustainability* **2021**, *13*, 4576. [[CrossRef](#)]
14. Firat, M.; Güngör, M. Monthly total sediment forecasting using adaptive neuro fuzzy inference system. *Stoch. Environ. Res. Risk Assess.* **2010**, *24*, 259–270. [[CrossRef](#)]
15. Chen, L.; Jamal, M.; Tan, C.; Alabbadi, B. A Study of Applying Genetic Algorithm to Predict Reservoir Water Quality. *Int. J. Model. Opt.* **2017**, *7*, 98. [[CrossRef](#)]
16. Martí, P.; Shiri, J.; Duran-Ros, M.; Arbat, G.; De Cartagena, F.R.; Puig-Bargués, J. Artificial neural networks vs. gene expression programming for estimating outlet dissolved oxygen in micro-irrigation sand filters fed with effluents. *Comp. Elect. Agricul.* **2013**, *99*, 176–185. [[CrossRef](#)]
17. Basant, N.; Gupta, S.; Malik, A.; Singh, K.P. Linear and nonlinear modeling for simultaneous prediction of dissolved oxygen and biochemical oxygen demand of the surface water—A case study. *Chemomet. Intel. Lab. Sys.* **2010**, *104*, 172–180. [[CrossRef](#)]
18. Amin, R.; Shah, K.; Khan, I.; Asif, M.; Salimi, M.; Ahmadian, A. Efficient Numerical Scheme for the Solution of Tenth Order Boundary Value Problems by the Haar Wavelet Method. *Mathematics* **2020**, *8*, 1874. [[CrossRef](#)]
19. Farooq, F.; Nasir Amin, M.; Khan, K.; Rehan Sadiq, M.; Faisal Javed, M.; Aslam, F.; Alyousef, R.A. Comparative Study of Random Forest and Genetic Engineering Programming for the Prediction of Compressive Strength of High Strength Concrete (HSC). *Appl. Sci.* **2020**, *10*, 7330. [[CrossRef](#)]
20. Aslam, F.; Farooq, F.; Amin, M.N.; Khan, K.; Waheed, A.; Akbar, A.; Alabduljabbar, H. Applications of Gene Expression Programming for Estimating Compressive Strength of High-Strength Concrete. *Adv. Civ. Eng.* **2020**, *2020*, 8850535. [[CrossRef](#)]
21. Najafzadeh, M.; Tafarjoruz, A. Evaluation of neuro-fuzzy GMDH-based particle swarm optimization to predict longitudinal dispersion coefficient in rivers. *Environ. Earth Sci.* **2016**, *75*, 157. [[CrossRef](#)]
22. Saberi-Movahed, F.; Najafzadeh, M.; Mehrpooya, A. Receiving more accurate predictions for longitudinal dispersion coefficients in water pipelines: Training group method of data handling using extreme learning machine conceptions. *Water Resour. Manag.* **2020**, *34*, 529–561. [[CrossRef](#)]
23. Najafzadeh, M.; Sattar, A.M. Neuro-fuzzy GMDH approach to predict longitudinal dispersion in water networks. *Water Resour. Manag.* **2015**, *29*, 2205–2219. [[CrossRef](#)]
24. Choubin, B.; Borji, M.; Hosseini, F.S.; Mosavi, A.; Dineva, A.A. Mass wasting susceptibility assessment of snow avalanches using machine learning models. *Sci. Rep.* **2020**, *10*, 18363. [[CrossRef](#)]
25. Mosavi, A.; Hosseini, F.S.; Choubin, B.; Abdolshahnejad, M.; Gharechae, H.; Lahijanzadeh, A.; Dineva, A.A. Susceptibility Prediction of Groundwater Hardness Using Ensemble Machine Learning Models. *Water* **2020**, *12*, 2770. [[CrossRef](#)]
26. Mosavi, A.; Golshan, M.; Janizadeh, S.; Choubin, B.; Melesse, A.M.; Dineva, A.A. Ensemble models of GLM, FDA, MARS, and RF for flood and erosion susceptibility mapping: A priority assessment of sub-basins. *Geocarto Int.* **2020**, 1–20. [[CrossRef](#)]
27. Mosavi, A.; Shirzadi, A.; Choubin, B.; Taromideh, F.; Hosseini, F.S.; Borji, M.; Dineva, A.A. Towards an ensemble machine learning model of random subspace based functional tree classifier for snow avalanche susceptibility mapping. *IEEE Access* **2020**, *8*, 145968–145983. [[CrossRef](#)]
28. Wagh, V.M.; Panaskar, D.B.; Muley, A.A.; Mukate, S.V.; Lolage, Y.P.; Aamalawar, M.L. Prediction of groundwater suitability for irrigation using artificial neural network model: A case study of Nanded tehsil, Maharashtra, India. *Model. Earth Syst. Environ.* **2016**, *2*, 1–10. [[CrossRef](#)]
29. Panahi, F.; Ehteram, M.; Ahmed, A.N.; Huang, Y.F.; Mosavi, A.; El-Shafie, A. Streamflow prediction with large climate indices using several hybrid multilayer perceptrons and copula Bayesian model averaging. *Ecol. Indic.* **2021**, *133*, 108285. [[CrossRef](#)]
30. Seifi, A.; Ehteram, M.; Singh, V.P.; Mosavi, A. Modeling and uncertainty analysis of groundwater level using six evolutionary optimization algorithms hybridized with ANFIS, SVM, and ANN. *Sustainability* **2020**, *12*, 4023. [[CrossRef](#)]
31. Asadi, E.; Isazadeh, M.; Samadianfard, S.; Ramli, M.F.; Mosavi, A.; Nabipour, N.; Chau, K.W. Groundwater quality assessment for sustainable drinking and irrigation. *Sustainability* **2020**, *12*, 177. [[CrossRef](#)]
32. Haykin, S. *Neural Networks: A Comprehensive Foundation*; Prentice-Hall, Inc.: Upper Saddle River, NJ, USA, 1999; Volume 7458, pp. 161–175.
33. Taghizadeh-Mehrdardi, R.; Emadi, M.; Cherati, A.; Heung, B.; Mosavi, A.; Scholten, T. Bio-inspired hybridization of artificial neural networks: An application for mapping the spatial distribution of soil texture fractions. *Remote Sens.* **2021**, *13*, 1025. [[CrossRef](#)]
34. Zounemat-Kermani, M.; Seo, Y.; Kim, S.; Ghorbani, M.A.; Samadianfard, S.; Naghshara, S.; Singh, V.P. Can decomposition approaches always enhance soft computing models? Predicting the dissolved oxygen concentration in the St. Johns River, Florida. *Appl. Sci.* **2019**, *9*, 2534. [[CrossRef](#)]
35. Wagh, V.; Panaskar, D.; Muley, A.; Mukate, S.; Gaikwad, S. Neural network modelling for nitrate concentration in groundwater of Kadava River basin, Nashik, Maharashtra, India. *Groundw. Sustain. Dev.* **2018**, *7*, 436–445. [[CrossRef](#)]
36. Mohammadzadeh, S.D.; Kazemi, S.F.; Mosavi, A.; Nasseralshariati, E.; Tah, J.H. Prediction of compression index of fine-grained soils using a gene expression programming model. *Infrastructures* **2019**, *4*, 26. [[CrossRef](#)]
37. Javed, M.F.; Farooq, F.; Memon, S.A.; Akbar, A.; Khan, M.A.; Aslam, F.; Rehman, S.K.U. New prediction model for the ultimate axial capacity of concrete-filled steel tubes: An evolutionary approach. *Crystals* **2020**, *10*, 741. [[CrossRef](#)]

38. Najafzadeh, M.; Ghaemi, A.; Emamgholizadeh, S. Prediction of water quality parameters using evolutionary computing-based formulations. *Int. J. Environ. Sci. Technol.* **2019**, *16*, 6377–6396. [[CrossRef](#)]
39. Tyrallis, H.; Papacharalampous, G.; Langousis, A. A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water* **2019**, *11*, 910. [[CrossRef](#)]
40. Ahmed, U.; Mumtaz, R.; Anwar, H.; Shah, A.A.; Irfan, R.; García-Nieto, J. Efficient water quality prediction using supervised machine learning. *Water* **2019**, *11*, 2210. [[CrossRef](#)]
41. Raheli, B.; Aalami, M.T.; El-Shafie, A.; Ghorbani, M.A.; Deo, R.C. Uncertainty assessment of the multilayer perceptron (MLP) neural network model with implementation of the novel hybrid MLP-FFA method for prediction of biochemical oxygen demand and dissolved oxygen: A case study of Langat River. *Environ. Earth Sci.* **2017**, *76*, 1–16. [[CrossRef](#)]
42. Palani, S.; Liong, S.-Y.; Tkalich, P. An ANN application for water quality forecasting. *Mar. Pollut. Bull.* **2008**, *56*, 1586–1597. [[CrossRef](#)]
43. Bozorg-Haddad, O.; Soleimani, S.; Loáiciga, H.A. Modeling water-quality parameters using genetic algorithm–least squares support vector regression and genetic programming. *J. Environ. Eng.* **2017**, *143*, 04017021. [[CrossRef](#)]
44. Nemati, S.; Fazelifard, M.H.; Terzi, Ö.; Ghorbani, M.A. Estimation of dissolved oxygen using data-driven techniques in the Tai Po River, Hong Kong. *Environ. Earth Sci.* **2015**, *74*, 4065–4073. [[CrossRef](#)]
45. Shah, M.I.; Javed, M.F.; Abunama, T. Proposed formulation of surface water quality and modelling using gene expression, machine learning, and regression techniques. *Environ. Sci. Pollut. Res.* **2021**, *28*, 13202–13220. [[CrossRef](#)]
46. Mosavi, A.; Hosseini, F.S.; Choubin, B.; Taromideh, F.; Ghodsi, M.; Nazari, B.; Dineva, A.A. Susceptibility mapping of groundwater salinity using machine learning models. *Environ. Sci. Pollut. Res.* **2021**, *28*, 10804–10817. [[CrossRef](#)] [[PubMed](#)]
47. Kadam, A.K.; Wagh, V.M.; Muley, A.A.; Umrikar, B.N.; Sankhua, R.N. Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India. *Model. Earth Syst. Environ.* **2019**, *5*, 951–962. [[CrossRef](#)]
48. Ferreira, C. Gene expression programming: A new adaptive algorithm for solving problems. *arXiv* **2001**, arXiv:cs/0102027.
49. Faradonbeh, R.S.; Armaghani, D.J.; Monjezi, M.; Mohamad, E.T. Genetic programming and gene expression programming for flyrock assessment due to mine blasting. *Int. J. Rock Mech. Min. Sci.* **2016**, *88*, 254–264. [[CrossRef](#)]
50. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]
51. Chebud, Y.; Naja, G.M.; Rivero, R.G.; Melesse, A.M. Water quality monitoring using remote sensing and an artificial neural network. *Water Air Soil Pollut.* **2012**, *223*, 4875–4887. [[CrossRef](#)]
52. Azamathulla, H.M.; Rathnayake, U.; Shatnawi, A. Gene expression programming and artificial neural network to estimate atmospheric temperature in Tabuk, Saudi Arabia. *Appl. Water Sci.* **2018**, *8*, 184. [[CrossRef](#)]
53. Bajracharya, S.R.; Shrestha, B.R. *The Status of Glaciers in the Hindu Kush-Himalayan Region*; International Centre for Integrated Mountain Development (ICIMOD): Kathmandu, Nepal, 2011.
54. Tahir, A.A.; Chevallier, P.; Arnaud, Y.; Neppel, L.; Ahmad, B. Modeling snowmelt-runoff under climate scenarios in the Hunza River basin, Karakoram Range, Northern Pakistan. *J. Hydrol.* **2011**, *409*, 104–117. [[CrossRef](#)]
55. ul Hasson, S. Future water availability from Hindukush-Karakoram-Himalaya Upper Indus Basin under conflicting climate change scenarios. *Climate* **2016**, *4*, 40. [[CrossRef](#)]
56. Ali, S.; Li, D.; Congbin, F.; Khan, F. Twenty first century climatic and hydrological changes over Upper Indus Basin of Himalayan region of Pakistan. *Environ. Res. Lett.* **2015**, *10*, 014007. [[CrossRef](#)]
57. Shah, M.I.; Khan, A.; Akbar, T.A.; Hassan, Q.K.; Khan, A.J.; Dewan, A. Predicting hydrologic responses to climate changes in highly glacierized and mountainous region Upper Indus Basin. *R. Soc. Open Sci.* **2020**, *7*, 191957. [[CrossRef](#)] [[PubMed](#)]
58. Montaseri, M.; Ghavidel, S.Z.Z.; Sanikhani, H. Water quality variations in different climates of Iran: Toward modeling total dissolved solid using soft computing techniques. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 2253–2273. [[CrossRef](#)]
59. Shah, M.I.; Alaloul, W.S.; Alqahtani, A.; Aldrees, A.; Musarat, M.A.; Javed, M.F. Predictive Modeling Approach for Surface Water Quality: Development and Comparison of Machine Learning Models. *Sustainability* **2021**, *13*, 7515. [[CrossRef](#)]
60. Iqbal, M.F.; Liu, Q.F.; Azim, I.; Zhu, X.; Yang, J.; Javed, M.F.; Rauf, M. Prediction of mechanical properties of green concrete incorporating waste foundry sand based on gene expression programming. *J. Hazard. Mater.* **2020**, *384*, 121322. [[CrossRef](#)]
61. Gandomi, A.H.; Yun, G.J.; Alavi, A.H. An evolutionary approach for modeling of shear strength of RC deep beams. *Mater. Struct.* **2013**, *46*, 2109–2119. [[CrossRef](#)]
62. Gandomi, A.H.; Alavi, A.H.; Mirzahosseini, M.R.; Nejad, F.M. Nonlinear genetic-based models for prediction of flow number of asphalt mixtures. *J. Mater. Civ. Eng.* **2011**, *23*, 248–263. [[CrossRef](#)]
63. Azim, I.; Yang, J.; Javed, M.F.; Iqbal, M.F.; Mahmood, Z.; Wang, F.; Liu, Q.F. Prediction model for compressive arch action capacity of RC frame structures under column removal scenario using gene expression programming. *Structures* **2020**, *25*, 212–228. [[CrossRef](#)]
64. Ferrero Bermejo, J.; Gomez Fernandez, J.F.; Olivencia Polo, F.; Crespo Marquez, A. A Review of the Use of Artificial Neural Network Models for Energy and Reliability Prediction. A Study of the Solar PV, Hydraulic and Wind Energy Sources. *Appl. Sci.* **2019**, *9*, 1844. [[CrossRef](#)]
65. Tung, T.M.; Yaseen, Z.M. A survey on river water quality modelling using artificial intelligence models: 2000–2020. *J. Hydrol* **2020**, *585*, 124670.

66. Ho, S.Y.; Phua, K.; Wong, L.; Goh, W.W.B. Extensions of the External Validation for Checking Learned Model Interpretability and Generalizability. *Patterns* **2020**, *1*, 100129. [[CrossRef](#)] [[PubMed](#)]
67. Gholampour, A.; Gandomi, A.H.; Ozbakkaloglu, T. New formulations for mechanical properties of recycled aggregate concrete using gene expression programming. *Con. Build. Mat.* **2017**, *130*, 122–145. [[CrossRef](#)]
68. Frank, I.E.; Todeschini, R. *The Data Analysis Handbook*; Elsevier: Amsterdam, The Netherlands, 1994.
69. Golbraikh, A.; Tropsha, A. Beware of  $q^2$ ! *J. Mol. Graph. Model.* **2002**, *20*, 269–276. [[CrossRef](#)]
70. Roy, P.P.; Roy, K. On some aspects of variable selection for partial least squares regression models. *QSAR Comb. Sci.* **2008**, *27*, 302–313. [[CrossRef](#)]
71. Alavi, A.H.; Ameri, M.; Gandomi, A.H.; Mirzahosseini, M.R. Formulation of flow number of asphalt mixes using a hybrid computational method. *Constr. Build. Mater.* **2011**, *25*, 1338–1355. [[CrossRef](#)]