Research Article

Data-Driven Model for the Prediction of Total Dissolved Gas: Robust Artificial Intelligence Approach

Mohamed Khalid AlOmar [ID],[1] **Mohammed Majeed Hameed** [ID],[1] **Nadhir Al-Ansari** [ID],[2] **and Mohammed Abdulhakim AlSaadi**[3,4]

[1]*Department of Civil Engineering, Al-Maaref University College, Ramadi, Iraq*
[2]*Civil Engineering Department, Environmental and Natural Resources Engineering, Lulea University of Technology,, 97187 Lulea, Sweden*
[3]*National Chair of Materials Science and Metallurgy, University of Nizwa, Nizwa, Oman*
[4]*Nanotechnology & Catalysis Research Centre (NANOCAT), IPS Building, University of Malaya, 50603 Kuala Lumpur, Malaysia*

Correspondence should be addressed to Mohamed Khalid AlOmar; mohd.alomar@yahoo.com

Saturated total dissolved gas (TDG) is recently considered as a serious issue in the environmental engineering field since it stands behind the reasons for increasing the mortality rates of fish and aquatic organisms. The accurate and more reliable prediction of TDG has a very significant role in preserving the diversity of aquatic organisms and reducing the phenomenon of fish deaths. Herein, two machine learning approaches called support vector regression (SVR) and extreme learning machine (ELM) have been applied to predict the saturated TDG% at USGS 14150000 and USGS 14181500 stations which are located in the USA. For the USGS 14150000 station, the recorded samples from 13 October 2016 to 14 March 2019 (75%) were used for training set, and the rest from 15 March 2019 to 13 October 2019 (25%) were used for testing requirements. Similarly, for USGS 14181500 station, the hourly data samples which covered the period from 9 June 2017 till 11 March 2019 were used for calibrating the models and from 12 March 2019 until 9 October 2019 were used for testing the predictive models. Eight input combinations based on different parameters have been established as well as nine statistical performance measures have been used for evaluating the accuracy of adopted models, for instance, not limited, correlation of determination (R^2), mean absolute relative error (MAE), and uncertainty at 95% (U_{95}). The obtained results of the study for both stations revealed that the ELM managed efficiently to estimate the TDG in comparison to SVR technique. For USGS 14181500 station, the statistical measures for ELM (SVR) were, respectively, reported as R^2 of 0.986 (0.986), MAE of 0.316 (0.441), and U_{95} of 3.592 (3.869). Lastly, for USGS 14181500 station, the statistical measures for ELM (SVR) were, respectively, reported as R^2 of 0.991 (0.991), MAE of 0.338 (0.396), and U_{95} of 0.832 (0.837). In addition, ELM's training process computational time is stated to be much shorter than that of SVM. The results also showed that the temperature parameter was the most significant variable that influenced TDG relative to the other parameters. Overall, the proposed model (ELM) proved to be an appropriate and efficient computer-assisted technology for saturated TDG modeling that will contribute to the basic knowledge of environmental considerations.

1. Introduction

Water encounters substantial volumes of air and bubbles during the flood discharge and is transferred down the watershed to the deep-water basin. Since the pressure in the quenching basin not only increases with increasing depth of water but also with kinetic pressure, and subsequently, the air and bubbles are under much greater pressure than the surface atmosphere. Consequently, a significant amount of air dissolves in the water and the total dissolved gas (TDG) is supersaturated [1]. The average dissolved gas content in water is often controlled by two parameters, the water temperature and the barometric pressure. Many essential gases, such as oxygen, nitrogen, argon, and carbon dioxide,

are known to contribute significantly to TDG formation [2]. The formation of the TDG is highly complex and depends on several variables in which it may be triggered by a mechanism affected by human or natural conditions. Subsequently, it can be divided into the following: first, physical and chemical processes where the air bubbles are produced and transferred from the dam to the spillway; and second, mixing and interaction with the involvement of mass transfer equations between water and bubbles [3].

Saturated TDG is recently considered as a serious issue in the environmental engineering field since it could cause increased mortality rates in fish and aquatic organisms [4]. The phenomena take place when fish consumes water with a high level of saturated TDG, and the dissolved gases flow into the bloodstream and balance with the external pressure of water. The problem begins to be drastically worse once fish sink in depths of the river; at this moment, the difference in pressures can be clearly observed resulting in bubble construction in the tissues of fish and bloodstream which lead to gas bubble trauma. In addition, the potential harmful environmental impacts of saturated TDG% beside its effects on fish and other marine species, the concentration of TDG% level in water may have an impact on the water quality and dissolved oxygen.

Based on the forgoing, prediction of TDG is vital due to its effects on water quality, sediments, hydrology, and economy [5–7]. Early attempts to model TDGs downstream are based on laboratory, field, and data fitting studies [8]. However, one downside of this technique is that the derived TDG analytical results are limited to the geometry and range of measures required to attain the model parameters. Weber et al., who solved tailrace hydrodynamics using Reynolds-averaged Navier–Stokes (RANS), made the first attempt to use a computational fluid mechanics (CFD) model to both forecast the hydrodynamics and TDG [9]. A scalar transport equation was used for the TDG with the gas volume fraction and source-term function as model parameters. For evaluating the two model parameters, the TDG field data measured in the river of Columbia were used. Feng et al. recently employed an averaged 2D model in a deep reservoir to simulate TDG. They used a scalar transport equation to model TDG with a bubble dissolution and mass transfer in the free surface as source term. The dissolution of the bubble was calculated using an interfacial area of the bubble dissipation coefficient [10]. Polydisperse dual phase flows and unstable 2-phase 3D flow approaches have been used by Politano et al. to develop TDG models [3, 11]. Later, in order to estimate the concentration of TDG, Fu et al. developed a two-phase 3D flow model, using feedback on the velocity, pressure, and volume of air involved in the supersaturation of gas under uncontrolled release conditions for the Gezhouba project in China [12]. Several algorithms have therefore been developed using various numerical, fluid mechanical, and hydrodynamic equations, which have shown that the TDG mechanism is precisely simulated [2, 13–18]. Previous research has typically focused on the application of data-driven approaches to solve a variety of environmental challenges. However, less attention is attributed to the modeling of TDG utilizing data-driven models [19].

Recently, the revolution of Artificial Intelligence (AI) has conquered almost all fields of science and engineering [20] including environmental applications [21–27]. Lately, several environmental problems have been solved using robust AI modeling approaches including Extreme Learning Machine (ELM) and Support Vector Machine (SVM) [28–30]. The ELM modeling approach is considered one of the most beneficial AI tools due to its ability to avoid problems such as overfitting, which can be seen in other forms of iterative learning algorithms, in addition to the slow learning and local minimization issues. Compared to standard neural network learning algorithms, the ELM model can complete the training of a given dataset reasonably quickly. The ELM model requires only one iteration of the learning process. The ELM model can also be used for kernel SLFNs such as Radial Basis Functions (RBF) in which the kernel function of the ELM model is a nonlinear, integrable function [24]. On the contrary, the SVM, a mathematical learning tool that can be used to solve classification and regression issues, is also considered as a highly robust AI system due to its generalization capability, highly scalable, global optimization, and statistical analysis skill. Therefore, a fast, precise, and powerful model can describe the SVM model. Due to the ability of the kernel model to gain expert knowledge, this approach can better describe complex nonlinear relationships than other models [31, 32]. SVM is often used as a kernel model and selection is a priority for maximum efficiency. The kernel can be formed using space and nonlinear boundaries. The radial basis function (RBF) is extensively used along with others, such as linear, polynomial, or sigmoid, due to its small or no error advantage in testing and validation [33].

As regards, using AI modeling techniques to predict TDG, Heddam utilized generalized regression neural network (GRNN) for predicting TDG concentration based on several variables including, temperature of water, barometric pressure, dam spill, sensor depth, and average flow. The GRNN model outperforms the multiple linear regression (MLR) model [2]. Later, the same researcher used Adaptive Neuro-Fuzzy Inference System (ANFIS) and Dynamic Evolving Neural-Fuzzy Inference System (DENFIS) to predict the DTG based on data generated by the dam from spell (SFD) [19]. Keshtegar et al. developed four models to predict TDG including high-order response surface methodology (HRSM), least squares support vector machine (LSSVM), M5 model tree (M5Tree), and multivariate adaptive regression splines (MARS). The data used in their study was collected from four United States Geological Survey (SGS) stations at Columbia River. It was reported that the HRSM with five variables demonstrated the best performance with regard to predicting TDG among the other models used in their study, where the HRSM recorded 0.911 coefficient of correlation [34].

Establishing modern and robust AI models is very effective in developing early warning systems [21, 32, 35] that can track saturated TDG% anomalies. Predicting one hour ahead of saturated TDG% may give more information about TDG% concentrations in water to the decision maker. This information is very important and may help to maintain the

safer level of TDG% concentration by switching off the hydropower station for several minutes. To the best of our knowledge, up to date, the studies involving AI technologies in predicting TDG are still limited. Based on the foregoing, this study attempts to develop two robust modeling approaches to predict the DTG utilizing data derived from historical dataset of Willamette River and North Santiam River. Extreme learning machine (ELM) and support vector regression (SVR) were employed for the first time in building a prediction model for TDG.

2. Case Studies and Data Collection

The historical dataset used for constituting and developing models was collected from the official website of the United States Geological Survey (USGS) [36]. The current study contains two stations, namely, USGS 14150000 at middle fork Willamette River, near Dexter, Lan County, OR (latitude 43° 56′45″; longitude 122° 50′10″) and USGS 14181500 North Santiam River, at Niagara, Marion County, OR (latitude 44°45′13.6″, longitude 122°17′50.8″). The location of each reservoir site is illustrated in Figure 1. The obtained parameters are statistically summarized in Table 1, where X_{min}, X_{max}, X_{mean}, X_{std}, C_v, and S_x denote the minimum, maximum, average, standard deviation, variation coefficient, and skewness coefficient, respectively. Five variables measured at hourly time step were used in this study including, discharge (D), barometric pressure (BP), water temperature (T), gage height (GH), and percent of saturated total dissolved gas (TDS%). These variables were used for establishing an AI model to predict one hour ahead TDG. The measured samples covered a long period of time from 13 October 2016 to 13 October 2019, where 25,667 samples were countered for station USGS 14150000 while for station USGS 14181500 the data obtained was 18,210 samples to cover the period from 9 June 2017 to 10 October 2019.

3. Methodology

3.1. Extreme Learning Machine Forecasting Model. Extreme learning machine (ELM) is a novel learning algorithm that generally has a simple structure consisting of three layers, namely, input layer, hidden layer, and output layer. The hidden layer is one of the most important layers in the structure of ELM including numerous numbers of nonlinear hidden nodes. ELM can be primarily characterized by the fact which the model's internal parameters such as hidden neurons do not require to be tuned. Additionally, ELM is considered an updated version of traditional ANN due to its ability to solve regression issues with minimum time consumption [37–39]. The reason behind is that the weights linking the input layer with the hidden layer and bias values in the hidden layer are randomly assigned where the output weights are optimally calculated using the Moore–Penrose approach [40]. This can lead to improved results compared to other forecasting models that can be

established using the ANN technique [41–43]. ELM is also presented as an efficient and alternative approach for conventional modeling techniques such as ANN which commonly suffers from several issues such as overfitting, slow convergence ability, local minimum problems, poorer generalization, and long time execution, as well as the essential for iterative tuning. Based on the fundamental structure of ELM, randomly assigned hidden neurons are tuned, so ELM is powerfully robust to achieve a global minimum solution, resulting in universal approximation abilities [44]. Based on the fundamental structure of the ELM, the randomly assigned hidden neurons are tuned in such a way that the ELM is powerfully resilient to achieve a global minimum solution, resulting in universal approximation capabilities [44]. Figure 2 shows and visualizes the basic structure of ELM.

ELM model can be mathematically expressed as shown in the following equation:

$$\sum_{k=1}^{L} B_k g_k \left(\alpha_k \cdot x_k + \beta_k \right) = z_t, \quad k = 1, \ldots, N, \qquad (1)$$

where L (number of hidden nodes), $g_k (\alpha_k \cdot x_k + \beta_k)$ (hidden layer output function), (α_k and β_k) (the parameters of hidden nodes which are randomly initialized), B_k (the weight values linking the kth hidden node(s) with the output node), and z_t (the ELM target).

The number of hidden nodes is determined by trial and error which belong to the range from 1 to 25. This current study used the hybrid tangent sigmoid transfer function to activate hidden nodes, while the forecasting values of the ELM model were obtained from the output layer based on linear activation function [45].

The selection of hidden node parameters in the ELM forecasting model can be randomly determined, where this process neither requires any detailed information about training data nor needs to iteratively tune the hidden layer neurons according to the lowest sum square error. Thus, for any randomly assigned sequence $\left\{ (\alpha_k, \beta_k)_{k=1}^{L} \right\}$ and any continuous target function $f(x)$, equation (2) is employed to approximate and calculate a set of N training sample as follows [46]:

$$\lim_{L \longrightarrow \infty} \left\| f(x) - f_L(x) \right\| = \lim_{L \longrightarrow \infty} \left\| f(x) - \sum_{k=1}^{L} B_k g_k \left(\alpha_k \cdot x_k + \beta_k \right) \right\| = 0.$$
$$(2)$$

The main merits of the nontuned ELM forecasted model are that the hidden layer weight values are randomly attained. This can reach a zero error, providing the opportunity to the network's target weight values (B) analytically for the training dataset. It is very significant to mention that the value of internal transfer function factors (α_k and β_k) are assigned according to a probability distribution. Finally, $Y = GB$ is considered an equivalent to equation (2) which can be linearly expressed as explained by [40]
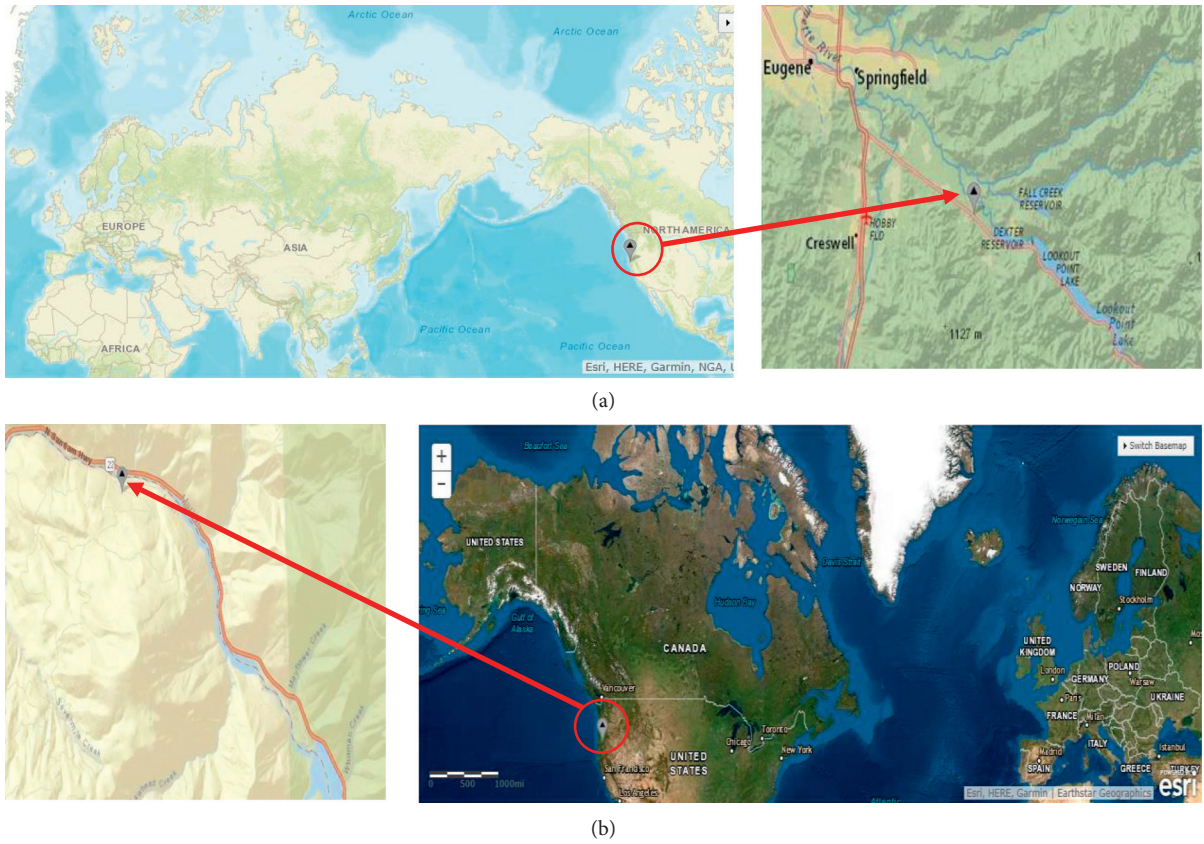
(a)

(b)

Figure 1: The location site of case studies. (a) USGS 14150000. (b) USGS 14181500.

Table 1: Statistical variables of adopted dataset for both stations.

Station	Dataset	Unit	X_{min}	X_{max}	X_{mean}	X_{std}	C_v	S_x
	D	kcfs	3.50	20.30	11.26	4.34	0.39	0.25
	BP	mmHg	1160.00	14400.00	2784.69	1933.69	0.69	2.58
USGS 14150000	T	˚C	2.76	9.41	3.91	1.16	0.30	1.71
	GH	feet	723.00	762.00	747.31	4.80	0.01	−0.49
	TDS%	—	96.00	119.00	102.81	5.11	0.05	1.12
	D	kcfs	957.00	10200.00	2056.95	1329.40	0.65	2.13
	BP	mmHg	710.00	747.00	733.67	4.44	0.01	−0.32
USGS 14181500	T	˚C	3.70	17.10	8.73	2.98	0.34	0.11
	GH	feet	2.82	7.61	3.71	0.86	0.23	1.39
	TDS%	—	96.00	132.00	104.55	6.37	0.06	1.54

kcfs: cubic feet per second; mmHg: millimeters of mercury; ˚C: degrees in Celsius.
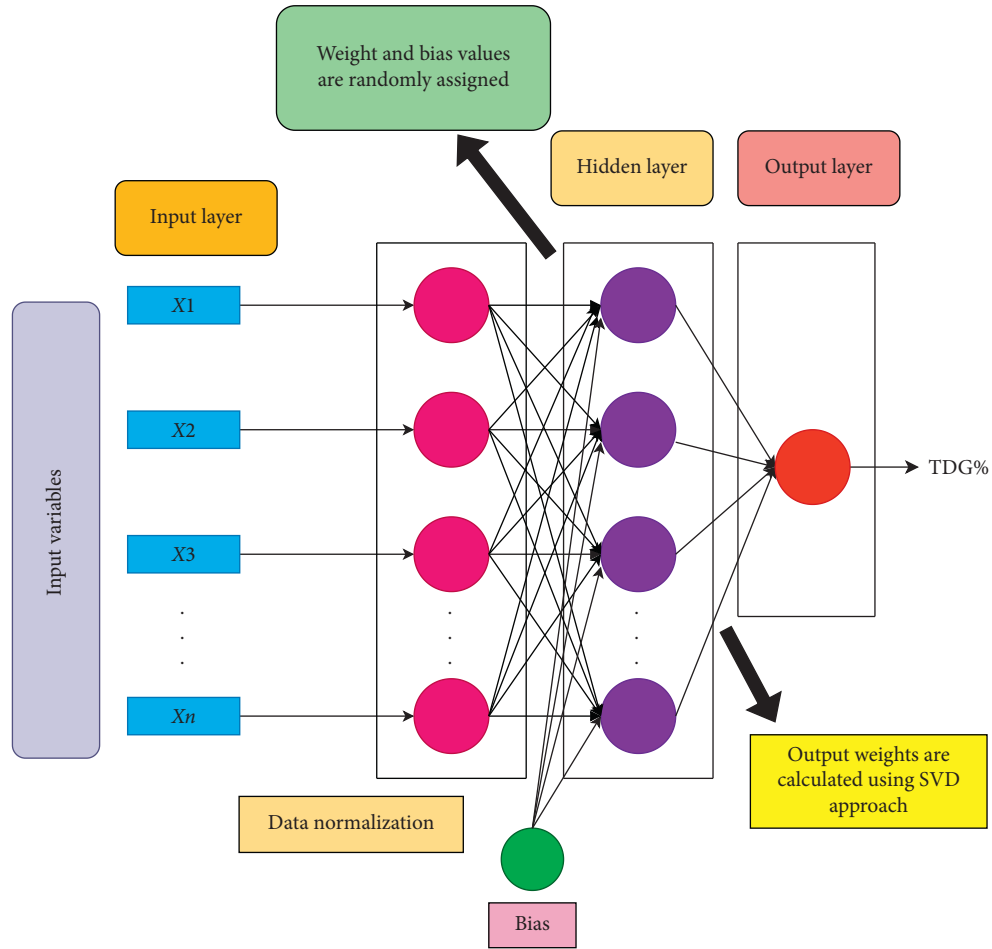
FIGURE 2: Extreme learning machine (ELM) structure.

$$G(\alpha, \beta, x) = \begin{bmatrix} g(x_k) \\ . \\ . \\ . \\ g(x_n) \end{bmatrix} = \begin{bmatrix} g_k(a_1 \cdot x_1 + \beta_1), \ldots, g_L(a_L, b_L, x_1) \\ . \\ . \\ . \\ g_k(a_N \cdot x_N + \beta_1), \ldots, g_L(a_L, b_L, x_1) \end{bmatrix},$$

(3)

$$B = \begin{bmatrix} B_1^T \\ . \\ . \\ . \\ . \\ B_L^T \end{bmatrix},$$

$$Y = \begin{bmatrix} Y_1^T \\ . \\ . \\ . \\ . \\ Y_N^T \end{bmatrix},$$

(4)

where G is the output matrix of the hidden layer and T is the transpose matrix, and equation (3) can be summarized as

$$HB = Y.$$

(5)

The lowest norm square of equation (4) can be calculated as

$$\widehat{B} = \breve{H}Y,$$

(6)

where $\breve{H}$ represents the Moore–Penrose generalized inverse of Hussain matrix which is employed to calculate the output weights of the ELM model. Singular Value Decomposition (SVD) method is mainly used as an efficient approach for the ELM learning process.

3.2. Support Vector Machine Regression Model. Support Vector Machine (SVM) is a sort of AI technique introduced by Cortes and Vapnik in 1995 [47], dealing with classification issues based on Structural Risk Minimization (SRM) and Statistical Learning Theory (SLT). This approach has been increasingly applied in different kinds of sectors for solving issues related to prediction and regression. The design of SVM density approximation uses the principle of SRM which has illustrated much efficient performance and accuracy compared to classical Empirical Risk Minimization (ERM) principle which mainly utilizes traditional learning algorithms such as neural network systems. SRM aims at

minimizing the upper and lower bounds on the generalization error, while ERM employs to minimize the total error on the training dataset. For that reason, SVM is more efficient in several statistical applications especially when it comes to constituting a predictive model [48]. Recently, SVM has been applied to carry out many tasks related to machine learning in numerous areas of research since it is a reliable and effective tool [49–53].

Given dataset points, $D = \{(x_i, y_i)\} \in R^d * R, i = 1: n$.

Here, the main principle is to identify and discover a function f in a Hilbert space based on SRM, constituting a certain relationship between variable x and the grandeur to obtain the model y, where $y = f(x)$ based on the measurement data D:

$$f(x) = (wx) + b, \quad b \in R, \tag{7}$$

$$f(x) = (w\varnothing(x)) + b. \tag{8}$$

Both equations (7) and (8) can define the function $f(x)$ for linear and nonlinear regression issues, respectively. Suppose the nature of the issue or the data does not belong to the linear relationship in its input space; in that case, that data can be derived to a higher dimension feature by applying a specific kernel function. The aim is to calculate the optimal weight values (w) and bias (b) and define the criteria to determine the best set of weight values. This task can be carried out using two stages. The first stage is to apply the Euclidean norm method (i.e., minimize w^2) for smoothing the weight values. The second stage is to minimize the empirical risk function by reducing the generated error values to the lowest level as possible. Finally, it can be summarized that the regularized risk function $R_{reg}(f)$, as illustrated below, should be minimized by

$$R_{reg}(f) = R_{emp}(f) + \frac{1}{2}\|w\|^2. \tag{9}$$

The empirical error is mathematically expressed as follows:

$$R_{emp}(f) = C\frac{1}{N}\sum_{1}^{N} L(x_i, y_i, f(x_i, W)), \tag{10}$$

where $L(\cdot)$ is the cost function to be derived. There are two common cost function which can be utilized: the first one is the ε-insensitive loss function presented by Vapnik, as shown in Figure 3, and the second is called the quadratic loss function which is usually related to least squares support vector machine (LSSVM) [54].

"C" indicates regularization constant which calculates the balance between the regularization term and the empirical risk. Additionally, ε is the size of the tube, denoting the accuracy of the function should be approximated. Accepted errors within a certain range made the problem more feasible. To consider the errors, the slack variables, ξ_i and ξ_i^*, are commonly presented. The main formulation of the optimization problem is as shown in the following equations:

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} (\xi_i + \xi_i^*), \tag{11}$$

$$\text{under the constraints} \begin{cases} y_i - w\varnothing(x) - b \leq \varepsilon + \xi_i, \\ y_i - w\varnothing(x) - b \geq -\varepsilon - \xi_i^*, \quad \forall_i \in \{1, \ldots, n\}, \\ \xi_i, \quad \xi_i^* \geq 0. \end{cases} \tag{12}$$

For optimally minimizing the regularized risk and efficiently calculating the optimal weight values, the quadratic programming problem is applied (utilizing the ε-insensitive loss function) based on Lagrange multipliers and using optimality constraints (further details can be seen in [55]); then, Lagrange multiplier α_i and $\alpha_i^*, i = 1$ to n, can be determined by minimizing the following equation:
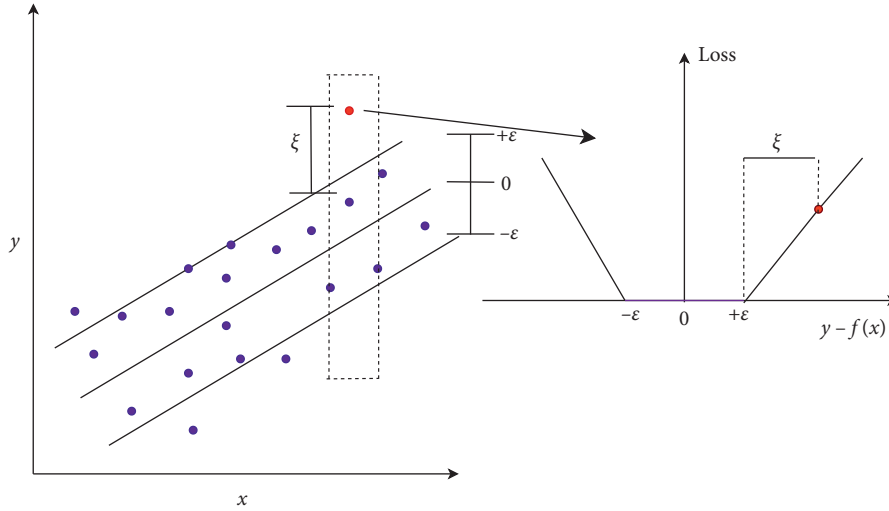
$$\min L(\alpha_i, \alpha_i^*) = -\sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) + \varepsilon\sum_{i=1}^{n} y_i(\alpha_i + \alpha_i^*) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(X_i, X_j), \tag{13}$$

with the constraints

$$0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \ldots, n,$$

$$\sum_{i=1}^{n} (\alpha_i - \alpha_i^*) = 0. \tag{14}$$

The regression function can be mathematically introduced by

$$f(x, \widehat{\alpha_i}, \widehat{\alpha_j^*}) = \sum_{i=1}^{n} (\widehat{\alpha_i} - t\widehat{\alpha_j^*})K(X, X_i) + b^*, \tag{15}$$

FIGURE 3: SVM regression with ε-insensitive loss function.

where $K(x_i, x_j) = \varnothing(x_i) * \varnothing(x_i)$ is known as the kernel function, and its value is represented as the scalar product of both vectors x_i and x_j in the feature space $\varnothing(x_i)$ and $\varnothing(x_j)$.

The selection of the proper kernel function is a significant task and mainly depends on Mercer's conditions; therefore, any function that satisfies these conditions can be applied as a kernel function for SVM approach. This current study adopted Radial Basis Function (RBF) which is mathematically expressed below since it can handle and map the nonlinear relationships between labels and features [56, 57]:

$$K\left(x_i, x_j\right) = \exp\left(-\frac{\left\|X_i - X_j^2\right\|}{2\sigma^2}\right), \quad (16)$$

where σ is the bandwidth of the RBF kernel.

It is worth mentioning that the most important parameters of the SVM model such as C, ε, and σ have been optimized by using a sequential minimal optimization (SMO) algorithm. Figure 4 shows the stage of prediction saturated TDG using the SVM model.

3.3. Preprocessing Dataset.

The preprocessing stage is one of the most significant stages in developing a predictive model due to its great effect on the accuracy of the model. This step includes two stages which are selecting the input combinations and data normalization and choosing the proper input variables which play an important role to obtain reliable and efficient predictive modes. Artificial Intelligent (AI) models generally considered robust techniques usually employing nonlinear functions for mapping input to their responses. These sophisticated methods have recently achieved great successes in many fields and outperformed the traditional approaches. Therefore, the selection of best input variables for a certain AI model is a difficult task and probably cannot be carried out by using common approaches such as linear relationships between predictors and their response. Additionally, each AI modeling technique

has a specific structure and methodology, consequently, selecting the best input parameters for a target might importantly vary from one model to others. This paper adopted different kinds of input combinations for both stations, as shown in Table 2, and introduced each combination to AI models (support vector machine and extreme learning machine) to predict hourly saturated TGG based on previously measured data points. Moreover, the employed procedure is very crucial to highlight the relative importance of the five variables, and we carried out different scenarios including several variables combinations for getting detailed information about affecting each of these factors to the saturated TDG concentration as well as conducting further cooperation for the responses of each input combination.

The data normalization stage is a very essential process in developing the AI models during training and testing phases because it maintains the stability of AI model performances [58] and reduces the time required to learn the model [59, 60]. Generally, there are two reasons for normalizing data before presented to the modeling approach. First, the process of normalization data ensures all available variables during the learning phase which take equal attention. Second, preprocessing is very essential for increasing the accuracy of models by improving the efficiency of the training algorithm. In this study, all data variables were rescaled with range from −1 to 1 based on

$$X_{\text{new}} = -1 + \frac{X - X_{\min}}{X_{\max} - X_{\min}} * 2, \quad (17)$$

where X_{new} is the scaled value, X is the current data point, and $X_{\max}$ and $X_{\min}$ are the maximum and minimum records in the dataset, respectively. Besides, the normalization process also ensures to control the values of each recorded sample within a certain range where the minimum values were kept to −1, whereas the highest values became +1. The normalization pattern was selected because this type of scaling data is centered on zero which can enhance the quality of predictive models.
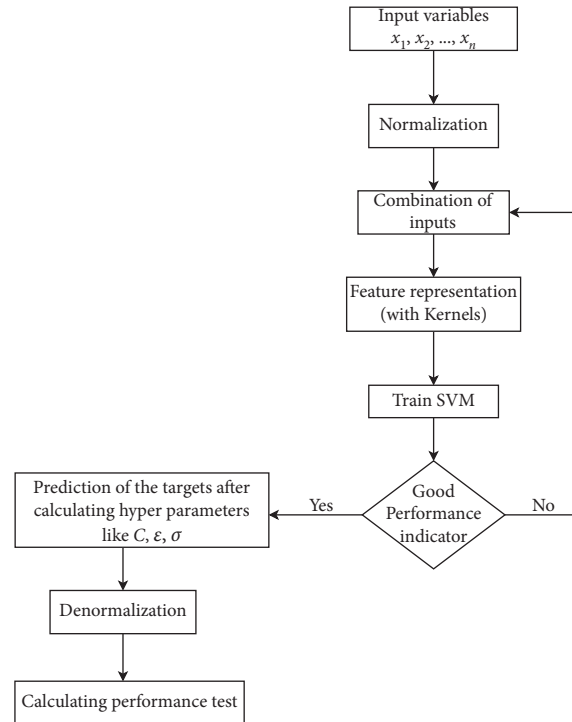
Figure 4: Flowchart of establishing the SVM model.

Table 2: Input combination set of several forecasting models.

No. of combinations	ELM	SVM	Input variables	Target
Comb1	ELM1	SVM1	TDS_{-1}	TDS_{+1}
Comb2	ELM2	SVM2	$[TDS_{-1}, D_{-1}]$	TDS_{+1}
Comb3	ELM3	SVM3	$[TDS_{-1}, BP_{-1}]$	TDS_{+1}
Comb4	ELMz4	SVM4	$[TDS_{-1}, T_{-1}]$	TDS_{+1}
Comb5	ELM5	SVM5	$[TDS_{-1}, GH_{-1}]$	TDS_{+1}
Comb6	ELM6	SVM6	$[TDS_{-1}, D_{-1}, BP_{-1}]$	TDS_{+1}
Comb7	ELM7	SVM7	$[TDS_{-1}, D_{-1}, BP_{-1}, T_{-1}]$	TDS_{+1}
Comb8	ELM8	SVM8	$[TDS_{-1}, D_{-1}, BP_{-1}, T_{-1}, GH_{-1}]$	TDS_{+1}

It is a general practice to split the raw dataset into two-phases: training and testing phase; then, these data separately recalled according to equation (17) before introducing to machine learning models. In this study, for both stations, 75% of the available data were used for the model constitution, and the reset was used for the testing phase. Figure 5 illustrates the methodology of the TDG prediction model using ELM and SVR models. For the station USGS 14150000, the record samples from 13 October 2016 to 14 March 2019 (822 days) were used for training set and the rest from 15 March 2019 to 13 October 2019 (273 days) were used for testing the accuracy of AI models. It can be pointed out that, about 19,251 hourly measured sample points were employed for the learning stage and 6416 hourly records used for the testing set. The first set of data that was utilized to train the AI model for the station USGS 14181500 including 640 days (13,658 samples measured hourly) covered the period from 9 June 2017 till 11 March 2019, while 213 days (4552 sample points measured hourly) were used for testing the accuracy of predictive models.

3.4. Model Performance Measures. Generally, the accuracy of the predictive modeling approach is evaluated by carrying out a comparison between observed responses and computed output. In this study, the forecasting of each model performance is assessed using ten statistical criteria including, root mean square error (RMSE), correlation of determination (R^2), mean absolute error (MAE), mean absolute relative error (MARE), root mean square relative error (RMSRE), relative root mean square error (RRMSE), maximum absolute relative error (erMAX), relative error (RE%), and uncertainty at 95% (U_{95}).

In environmental modeling, the RMSE criterion is frequently used to measure the performance of forecasting models, while the MAE index is considered as a vital indicator to evaluate the error in time series analysis. Furthermore, the value of MAE is very important to determine how well the model's output matches the actual values. However, the other statistical measures such as RE% try to fill the gaps left because it provides additional and detailed information about the capabilities of forecasting models. The
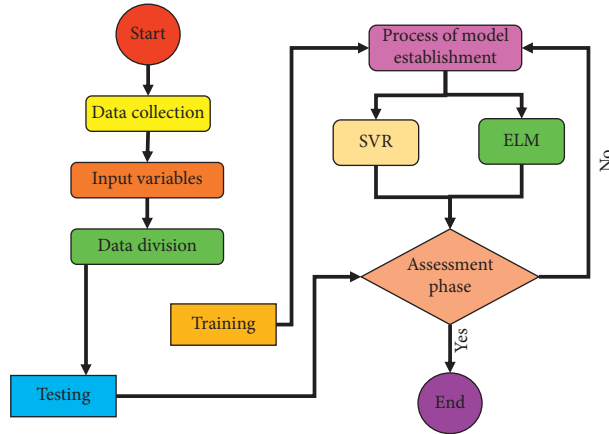
FIGURE 5: TDG prediction model methodology.

mean absolute relative error (MARE) is an absolute mathematical error for the code (the difference between the actual points and the predicted points). The MARE parameter is called the mean absolute percentage relative error (MARE) when the percentage is defined. The mean square error for relative root (RRMSE) is possible to measure the mean of actual data points by splitting RMSE criteria. This parameter is very critical for a model's accurate evaluation. The model is called outstanding: if RRMSE <10%, good, if

RRMSE ranged between 10 and 20%, fair, and if ranged between 20 and 30%, the model can be considered unacceptable with RRMSE >30% [61, 62]. Finally, in the selection of an effective prediction model among different models, the uncertainty is very effective criteria at 95 percent (U_{95}), whereas U_{95} contains very valuable knowledge on a model deviation. Formulae for determining (R^2), RMSE, MAE, RE, MARE, RMSRE, RRMSRE, erMAX, and U_{95} are expressed as follows:

$$R^2 = 1 - \frac{\sum_{t=1}^{n} \left(TDG(obser)_t - TDG(sim)_t \right)^2}{\sum_{t=1}^{n} \left(TDG(obser)_t - \overline{TDG(sim)_t} \right)^2},$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left(TDG(obser)_t - TDG(sim)_t \right)^2},$$

$$MAE = \frac{1}{n} \sum_{t=1}^{n} \left| TDG(obser)_t - TDG(sim)_t \right|,$$

$$RE = 100 * \left[\frac{TDG(obser)_t - TDG(sim)_t}{TDG(obser)_t} \right],$$

$$MARE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{TDG(obser)_t - TDG(sim)_t}{TDG(obser)_t} \right|,$$

$$RMSRE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left(\frac{TDG(obser)_t - TDG(sim)_t}{TDG(obser)_t} \right)^2},$$

$$RRMSRE = \sqrt{\frac{RMSE}{(1/n) \sum_{t=1}^{n} TDG_{(obser)_t}}},$$

$$erMAX = \left(\left| \frac{TDG(obser)_t - TDG(sim)_t}{TDG(obser)_t} \right| \right),$$

$$U_{95} = 1.96 \left(SD^2 + RMSE^2 \right)^{(1/2)},$$

(18)

Table 3: Statistical criteria of each predicted models for USGS 14181500 station during the training set.

Model	MAE	RMSE	MARE	RMSRE	RRMSE	R^2	erMAX	U_{95}
ELM-M1	1.4311	1.9436	0.0132	0.0173	1.8625	0.9827	1.2733	63.0982
SVR-M1	8.3559	10.4287	0.0770	0.0926	9.9934	0.2	1.0000	42756.9442
ELM-M2	0.6033	1.0258	0.0056	0.0092	0.9830	0.9824	1.2391	6.8153
SVR-M2	**0.4244**	**0.8826**	**0.0040**	**0.0079**	**0.8458**	**0.9810**	**1.2207**	**4.7401**
ELM-M3	0.4764	0.9483	0.0044	0.0084	0.9087	0.9825	1.2349	5.5517
SVR-M3	1.1037	1.5475	0.0102	0.0138	1.4829	0.9828	1.2610	27.1510
ELM-M4	0.3272	0.8571	0.0030	0.0076	0.8213	0.9822	1.2265	4.2314
SVR-M4	1.3372	1.5625	0.0128	0.0147	1.4972	0.9806	1.2292	20.2709
ELM-M5	0.4599	0.9717	0.0042	0.0086	0.9311	0.9825	1.2366	6.2246
SVR-M5	0.6284	0.9966	0.0059	0.0090	0.9550	0.9824	1.2333	5.8313
ELM-M6	**0.3230**	**0.8462**	**0.0030**	**0.0075**	**0.8109**	**0.9820**	**1.2258**	**4.0191**
SVR-M6	2.3503	2.8870	0.0221	0.0265	2.7665	0.9515	1.2317	261.4542
ELM-M7	0.7394	1.2223	0.0068	0.0109	1.1713	0.9815	1.2468	13.1359
SVR-M7	1.1364	1.6076	0.0107	0.0147	1.5405	0.9469	1.1783	45.4376
ELM-M8	0.5822	0.9770	0.0054	0.0087	0.9363	0.9813	1.2184	5.6848
SVR-M8	0.5510	1.0121	0.0051	0.0090	0.9698	0.9826	1.2406	6.8033

where $\mathrm{TDG(obser)}_t$ and $\mathrm{TDG(sim)}_t$ are the actual and simulated values of saturated TDG, respectively, $\overline{\mathrm{TDG(obser)}_t}$ and $\overline{\mathrm{TDG(sim)}_t}$ are the mean actual and predicted values of saturated TDG, and n is the total number of samples.

4. Result and Discussion

In this part of the study, results of SVR and ELM models for different proposed combinations are presented. The obtained results for both USGS 14150000 and USGS 14181500 stations are also further discussed in this section in order to select the best predictive model which can provide more accurate results related to saturated TDG. In general, the quantitative and visualized analyses show that the trend ELM models are more stable than SVR models. However, the SVR approaches sometimes provide an acceptable prediction of TDG.

Table 3 presented the evaluation performance of each predicted model for USGS 14181500 station during the training set. The results pointed out that the performances of both SVR and ELM approaches depended mainly on the input combinations. For SVR technique, the best accuracy model was SVR-M2, reporting lowest forecasted error (MAE = 0.4244, RMSE = 0.8826, MARE = 0.0040, RMSRE = 0.0079, RRMSE = 0.8458, R^2 = 0.9810, erMAX = 1.2207, and U_{95} = 4.7401). On the contrary, the ELM-M6 generated the best performance in comparison with other comparable ELM models. The ELM-M6 model reported less forecasted error based on statistical measures indexes (MAE = 0.3230, RMSE = 0.8462, MARE = 0.0030, RMSRE = 0.0075, RRMSE = 0.8109, R^2 = 0.9820, erMAX = 1.2258, and U_{95} = 4.0191). The given results intensively indicated that the SVR models required fewer input parameters in comparison with ELM models. Moreover, the most important feature can be seen that there was a unique superiority in terms of prediction TDG for the ELM model over SVR during the training phase.

Table 4 provided detailed information on the performances of both predictive models (SVR and ELM) with different input combinations at USGS 14150000 station during the training step. In general, the majority of comparable models reported excellent predictions of TDG concentration. In accordance with the presented result, the most accurate SVR model based on eight different input combinations was SVR-M6. The proposed model achieved higher accuracy indexes (MAE = 0.296, RMSE = 0.503, MARE = 0.003, RMSRE = 0.005, RRMSE = 0.495, R^2 = 0.985, erMAX = 1.085, and U_{95} = 0.492), while the ELM models provided in general a good accurate predictions, and ELM-M4 generated the best simulated results of TDG compared with actual values (MAE = 0.309, RMSE 0.566, MARE = 0.003, RMSRE = 0.006, RRMSE = 0.557, R^2 = 0.981, erMAX = 1.086, and U_{95} = 0.802). Based on given results, the SVR technique tended to give greater accuracy with a relatively higher number of input combinations.

For the selection of the most accurate models, the testing set is considered the most efficient step since the models in the training set relying on given input parameters and their responses target, while in the testing set, the actual performance of each model is easily recognized due to the fact that only input variable is introduced to the predictive model [63]. Moreover, in the testing phase, better evaluation of the accuracy of the model as well as generalization capabilities can be efficiently revealed. For adequate evaluation, it is necessary to check out the performances of the models which provided the most accurate estimations throughout the training set.

The performances of each model showed in Table 5 during the testing set at USGS 14181500 station. Generally, the observable note that both models (SVR-M2 and ELM-M6), which gave the most accurate predictions during the calibration step, produced relatively lower performances during the testing set. The SVR-M2 model provide relatively higher forecasted errors (MAE = 4.005, RMSE = 4.548, MARE = 0.037, RMSRE = 0.041, RRMSE = 4.327, R^2 = 0.925, erMAX = 1.092, and U_{95} = 1271.496). With respect to the ELM-M6 model, this model did not consider the ideal one during the testing phase; however, it produced an acceptable level of accuracy (MAE = 0.538, RMSE = 0.904, MARE = 0.005, RMSRE = 0.008,

TABLE 4: Statistical criteria of each predicted models for USGS 14150000 station during the training set.

Model	MAE	RMSE	MARE	RMSRE	RRMSE	R^2	erMAX	U_{95}
ELM-M1	0.374	0.583	0.004	0.006	0.574	0.981	1.086	0.862
SVR-M1	0.293	0.569	0.003	0.006	0.560	0.981	1.086	0.819
ELM-M2	0.328	0.575	0.003	0.006	0.567	0.981	1.087	0.852
SVR-M2	0.319	0.546	0.003	0.005	0.538	0.982	1.086	0.698
ELM-M3	0.326	0.569	0.003	0.006	0.561	0.981	1.086	0.819
SVR-M3	0.296	0.564	0.003	0.006	0.556	0.981	1.086	0.795
ELM-M4	**0.309**	**0.566**	**0.003**	**0.006**	**0.557**	**0.981**	**1.086**	**0.802**
SVR-M4	0.280	0.556	0.003	0.005	0.548	0.981	1.086	0.750
ELM-M5	0.344	0.572	0.003	0.006	0.563	0.981	1.085	0.820
SVR-M5	0.271	0.530	0.003	0.005	0.522	0.983	1.086	0.618
ELM-M6	0.369	0.578	0.004	0.006	0.570	0.981	1.084	0.832
SVR-M6	**0.296**	**0.503**	**0.003**	**0.005**	**0.495**	**0.985**	**1.085**	**0.492**
ELM-M7	0.367	0.585	0.004	0.006	0.576	0.980	1.089	0.918
SVR-M7	0.288	0.568	0.003	0.006	0.559	0.981	1.086	0.817
ELM-M8	0.439	0.624	0.004	0.006	0.614	0.979	1.092	1.163
SVR-M8	2.810	3.103	0.027	0.030	3.056	0.661	1.073	465.864

TABLE 5: Statistical indicators of each predicted models for USGS 14181500 station during the testing phase.

Model	MAE	RMSE	MARE	RMSRE	RRMSE	R^2	erMAX	U_{95}
ELM-M1	1.308	1.804	0.012	0.016	1.716	0.979	1.214	48.218
SVR-M1	8.116	10.540	0.074	0.091	10.027	0.008	1.000	47895.737
ELM-M2	0.645	1.054	0.006	0.010	1.002	0.985	1.193	7.309
SVR-M2	4.005	4.548	0.037	0.041	4.327	0.925	1.092	1271.496
ELM-M3	0.593	1.027	0.005	0.009	0.977	0.985	1.191	6.948
SVR-M3	19.363	20.324	0.188	0.199	19.335	0.009	1.283	417056.225
ELM-M4	**0.316**	**0.823**	**0.003**	**0.008**	**0.783**	**0.986**	**1.181**	**3.592**
SVR-M4	5.229	6.330	0.048	0.056	6.022	0.760	1.048	5579.844
ELM-M5	0.442	0.918	0.004	0.008	0.874	0.985	1.188	5.021
SVR-M5	**0.441**	**0.862**	**0.004**	**0.008**	**0.820**	**0.986**	**1.183**	**3.869**
ELM-M6	0.538	0.904	0.005	0.008	0.860	0.986	1.170	4.229
SVR-M6	1.416	1.896	0.013	0.017	1.804	0.945	1.152	84.101
ELM-M7	1.119	1.540	0.010	0.014	1.465	0.981	1.207	26.109
SVR-M7	0.904	1.777	0.008	0.015	1.691	0.932	1.128	76.123
ELM-M8	0.469	0.877	0.004	0.008	0.835	0.985	1.172	4.111
SVR-M8	0.719	1.171	0.007	0.011	1.114	0.984	1.198	10.954

RRMSE = 0.860, R^2 = 0.986, erMAX = 1.170, and U_{95} = 4.229). The given results clearly showed that the optimal models of SVR during the training set were suffering from overfitting issues (highest accuracy in the training set and lowest accuracy in the testing set). However, after reviewing the performances of both models, it is vital to present the most two efficient models for the same station. In accordance with Table 5, the ELM-M4 model is identified as the best model which can more effectively predict one step ahead of saturated TDG. The obtained results from the model showed there was a perfect similarity with the actual values (MAE = 0.316, RMSE = 0.823, MARE = 0.003, RMSRE = 0.008, RRMSE = 0.783, R^2 = 0.986, erMAX = 1.181, and U_{95} = 3.592). The second best model during the testing set was SVR-M5 which also produced fewer errors (MAE = 0.441, RMSE = 0.862, MARE = 0.004, RMSRE = 0.008, RRMSE = 0.820, R^2 = 0.986, erMAX = 1.183, and U_{95} = 3.869). The given evaluation disclosed that the performance of ELM-M4 generated more accurate estimations,

and it also provided reasonable and adequate estimations in the training set (see Table 3). On the contrary, the predictability of SVR-M5 generated higher forecasted errors during the training set. Consequently, the ELM-M4 was the most efficient model in estimation of TDG for USGS 14181500 station.

While the ELM approach outperformed SVR techniques for the estimation of TDG at USGS 14181500 station, it is important to see the capability of the proposed approach in the prediction of TDG in USGS 14150000 station. Table 4 exhibited the predicted results of each predictive model which was established based on different input combinations during the training set. Statistical parameters indicated that the SVR-M6 was the most excellent performance during the calibration step (MAE = 0.296, RMSE = 0.503, MARE = 0.003, RMSRE = 0.005, RRMSE = 0.495, R^2 = 0.985, erMAX = 1.085, and U_{95} = 0.8492). In addition, the performance of the ELM-M4 was relatively less accurate compared to SVR-M6 as a result of generating higher computed errors (MAE = 0.309,

TABLE 6: Statistical indicators of each predicted models for USGS 14150000 station during the testing phase.

Model	MAE	RMSE	MARE	RMSRE	RRMSE	R^2	erMAX	U_{95}
ELM-M1	0.379	0.586	0.004	0.006	0.549	0.991	1.047	0.898
SVR-M1	0.317	0.572	0.003	0.005	0.537	0.991	1.047	0.842
ELM-M2	0.339	0.576	0.003	0.005	0.540	0.991	1.047	0.864
SVR-M2	3.822	4.273	0.037	0.041	4.008	0.723	1.079	2053.114
ELM-M3	0.373	0.577	0.004	0.005	0.541	0.991	1.046	0.851
SVR-M3	8.619	10.380	0.078	0.093	9.736	0.018	1.021	39877.590
ELM-M4	**0.338**	**0.571**	**0.003**	**0.005**	**0.536**	**0.991**	**1.047**	**0.832**
SVR-M4	8.799	10.463	0.080	0.094	9.814	0.086	1.018	39559.262
ELM-M5	0.463	0.618	0.004	0.006	0.580	0.991	1.044	0.990
SVR-M5	8.875	10.572	0.081	0.095	9.916	0.054	1.017	41293.085
ELM-M6	0.495	0.639	0.005	0.006	0.599	0.991	1.044	1.080
SVR-M6	3.282	4.321	0.030	0.038	4.053	0.550	1.102	2508.932
ELM-M7	0.517	0.688	0.005	0.006	0.646	0.988	1.055	1.754
SVR-M7	**0.396**	**0.572**	**0.003**	**0.005**	**0.536**	**0.991**	**1.047**	**0.837**
ELM-M8	0.513	0.680	0.005	0.006	0.637	0.988	1.052	1.672
SVR-M8	5.034	5.942	0.048	0.057	5.573	0.009	1.116	9694.100

RMSE = 0.566, MARE = 0.003, RMSRE = 0.006, RRMSE = 0.557, R^2 = 0.981, erMAX = 1.086, and U_{95} = 0.802).

Based on the forgoing, the reliable models should produce better performance in the vital step (testing) as well as in the training step. Herein, it is crucial to review the performance of both models SVR-M6 and ELM-M4, which yielded the best estimated results through the training set, during the testing set. As shown in Table 6, the performance of the SVR-M6 was awful and its estimations were extremely inaccurate and unacceptable (MAE = 3.282, RMSE = 4.321, MARE = 0.030, RMSRE = 0.038, RRMSE = 4.053, R^2 = 0.550, erMAX = 1.102, and U_{95} = 2508.932). On the other hand, the performance of ELM-M4 was excellent and can be considered the most reliable model during the testing set by generating the lowest forecasted errors (MAE = 0.338, RMSE = 0.571, MARE = 0.003, RMSRE = 0.005, RRMSE = 0.536, R^2 = 0.991, erMAX = 1.047, and U_{95} = 0.832). Similarly, SVR-M7 achieved a good performance prediction with lower error measures (MAE = 0.396, RMSE = 0.572, MARE = 0.003, RMSRE = 0.005, RRMSE = 0.536, R^2 = 0.991, erMAX = 1.047, and U_{95} = 0.837).

According to the presented outcomes, the ELM provided much more valid and efficient estimations than SVR techniques. It is also obvious from the quantitative analyses that the SVR suffered from overfitting issues, thereby reducing its ability to provide reliable prediction of saturated TDG. After presenting the quantitative assessment, it is necessary to carry out a visualization assessment to ideally select the best predictive models. For visually evaluating the predictive models against actual values of TDG during the testing set, boxplot and scatterplot diagrams were established.

The capacity of each adopted model has been graphically compared with actual values as illustrated in the boxplot diagram at USGS 14181500 station (see Figure 6). The useful information was that all ELM models were found to predict TDG more precisely. Moreover, ELM models managed to efficiently estimate the peak values of TDG which were considered the most important values due

to its dangers impact on the ecosystem. However, most SVR models produced less predictability precision compared to the observed TDG. ELM models found to have median and interquartile range (IQR) closer to the observed median and IQR. Dissimilarly, several SVR models' characteristics (median and IQR) were found farther to the actual ones. The figures also showed that the efficiency of SVR models effectively relied upon input combinations. In general, SVR approaches could give a good prediction in limited input combinations; however, the ELM has relatively stable performances in all adopted input groups as well as has less sensitivity when input parameters changed. Figure 7 represents the boxplot for USGS 14150000 station. In accordance with Figures 6 and 7, the distribution of saturated TDG% obtained from the SVR modeling approach in several cases was very poor and gave poor accuracy prediction. For instance, with respect to USGS 14181500 station, the SVR-M1 and SVR-M3 models generated the worst estimations. Similarly, in USGS 14150000 station, the four models, SVR-M3, SVR-M4, SVR-M5, and SVR-M8, gave the worst accuracy compared with others. The reason behind this might be due to the nature of the SVR modeling approach which tends to give lower-accurate estimations when using a relatively large number of input variables. However, it can be observed that the SVR techniques may encounter some difficulties in developing a univariate model. Moreover, the structure of the SVR approach required a larger number of coefficients in simulating data including higher numbers of observations; hence, fluctuating in accuracy of predictions took place occasionally.

Scatterplots were also prepared for both aforementioned stations to measure how well-estimated values of TDG relate to the actual points. The scatterplots, as shown in Figures 8–11, provided significant visualization information on the diversion between predicted and actual TDG, as well as correlation of determination magnitude (R^2). In addition, the line equation can be also presented in the graphs ($y = ax + b$) where a and b are representing the slope of the line and intercept point, respectively. A closer value of a to 1
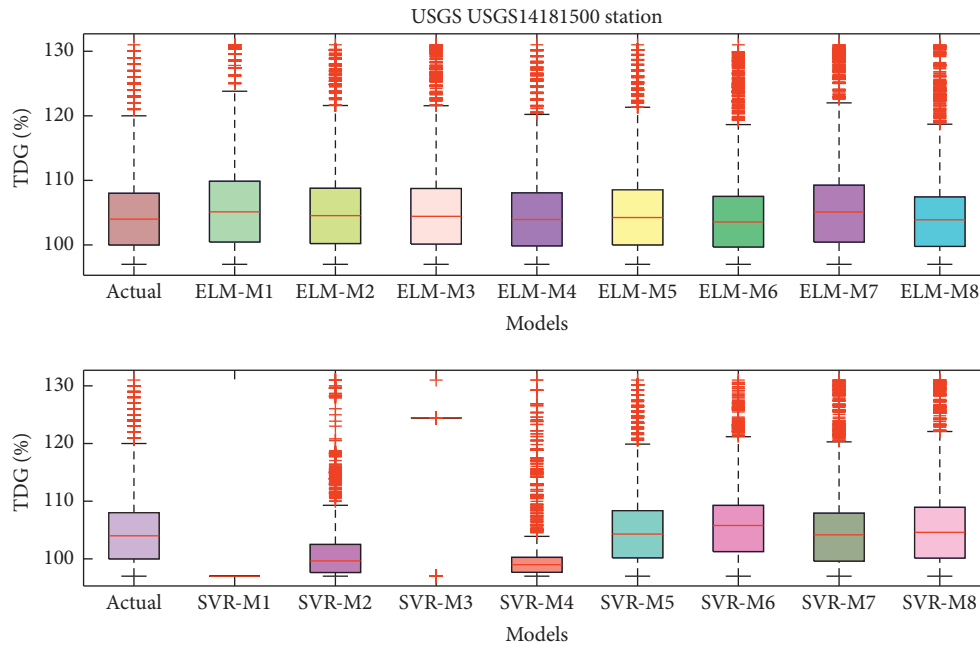
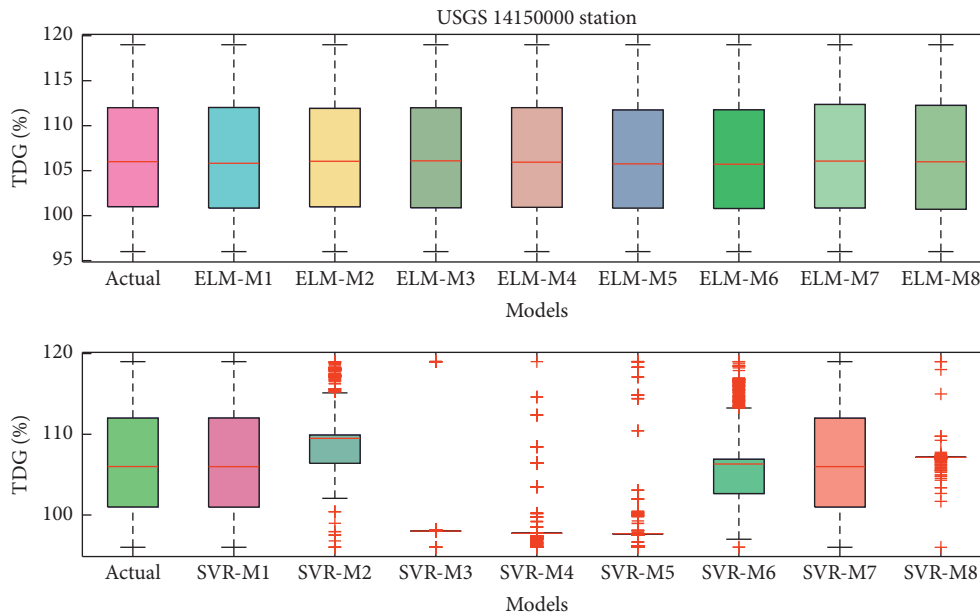FIGURE 6: Boxplot presentation for USGS 14181500 station: (a) ELM and (b) SVR models.

FIGURE 7: Boxplot presentation for USGS 14150000 station: (a) ELM and (b) SVR models.

and b to zero refer to the best predictive model achieved. In accordance with Figures 8–11, the ELM-M4 was the best predictive model for USGS 14150000 and USGS 14181500 stations.

Finally, the most unique observation from the analytical results of ELM for both stations is that the combination M4 has a more significant influence on TDG than other adopted combinations. That means the temperature of the water has a vital effect on the concentrations of the saturated TGD in the water. Moreover, the ELM algorithm has a perfect advantage in terms of low computational cost as well as ease of

implementations. However, the SVR technique exhibited a very slow learning process compared with ELM approaches. Based on Table 7, the ELM approaches required less time (seconds) to complete the training process on average of 0.018s and 0.021s for USGS 14181500 and USGS 14150000, respectively. On the contrary, the algorithm of SVR required a huge time to complete the calibration process, in average of 1223.991s and 2759.280s for USGS14181500 and USGS14150000, respectively. Lastly, according to the given results from the best predictive model (ELM), the input parameters such as discharge, gage height, and aerometric
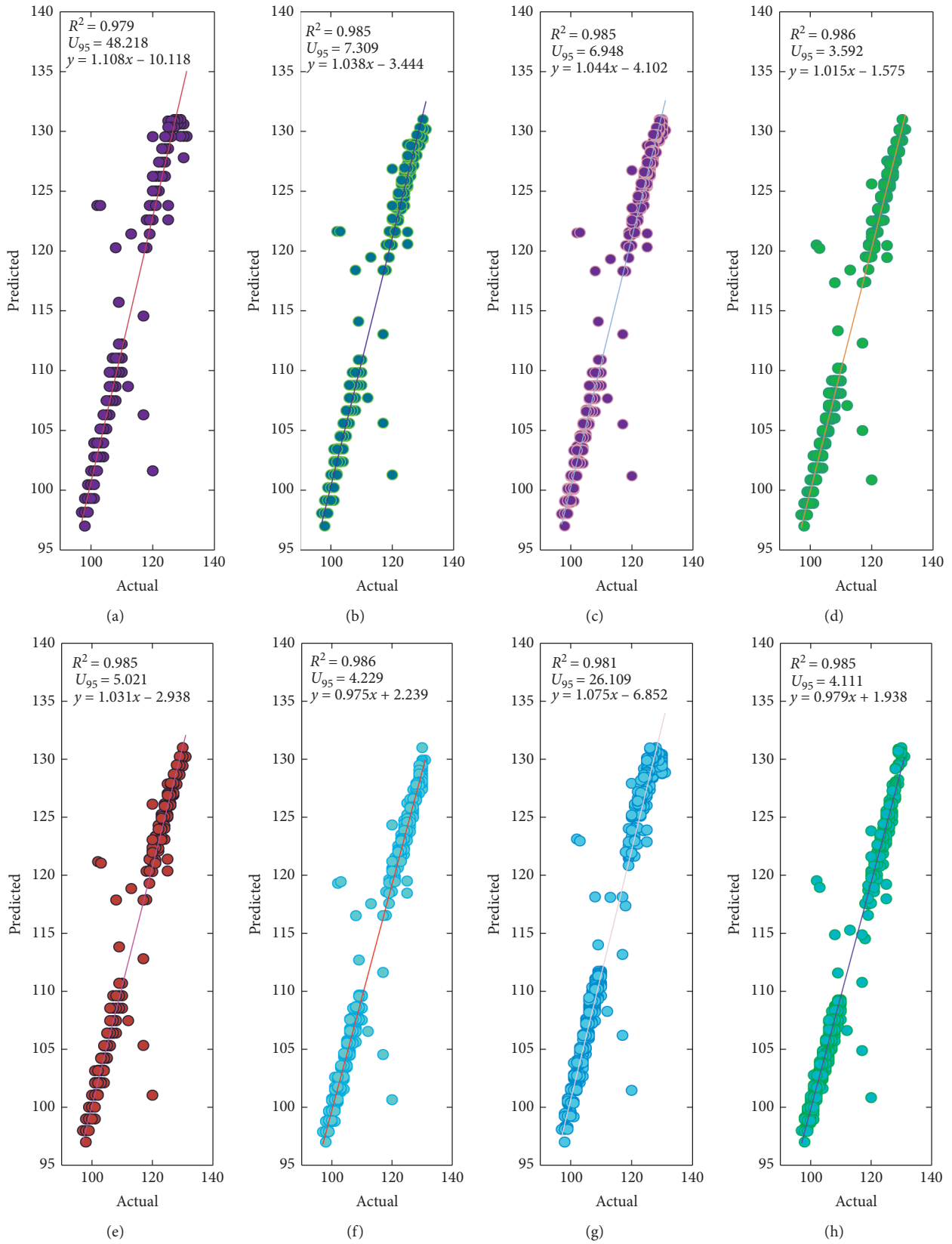
FIGURE 8: ELM models for USGS 14181500 station: coefficient of determination. (a) ELM-M1. (b) ELM-M2. (c) ELM-M3. (d) ELM-M4. (e) ELM-M5. (f) ELM-M6. (g) ELM-M7. (h) ELM-M8.
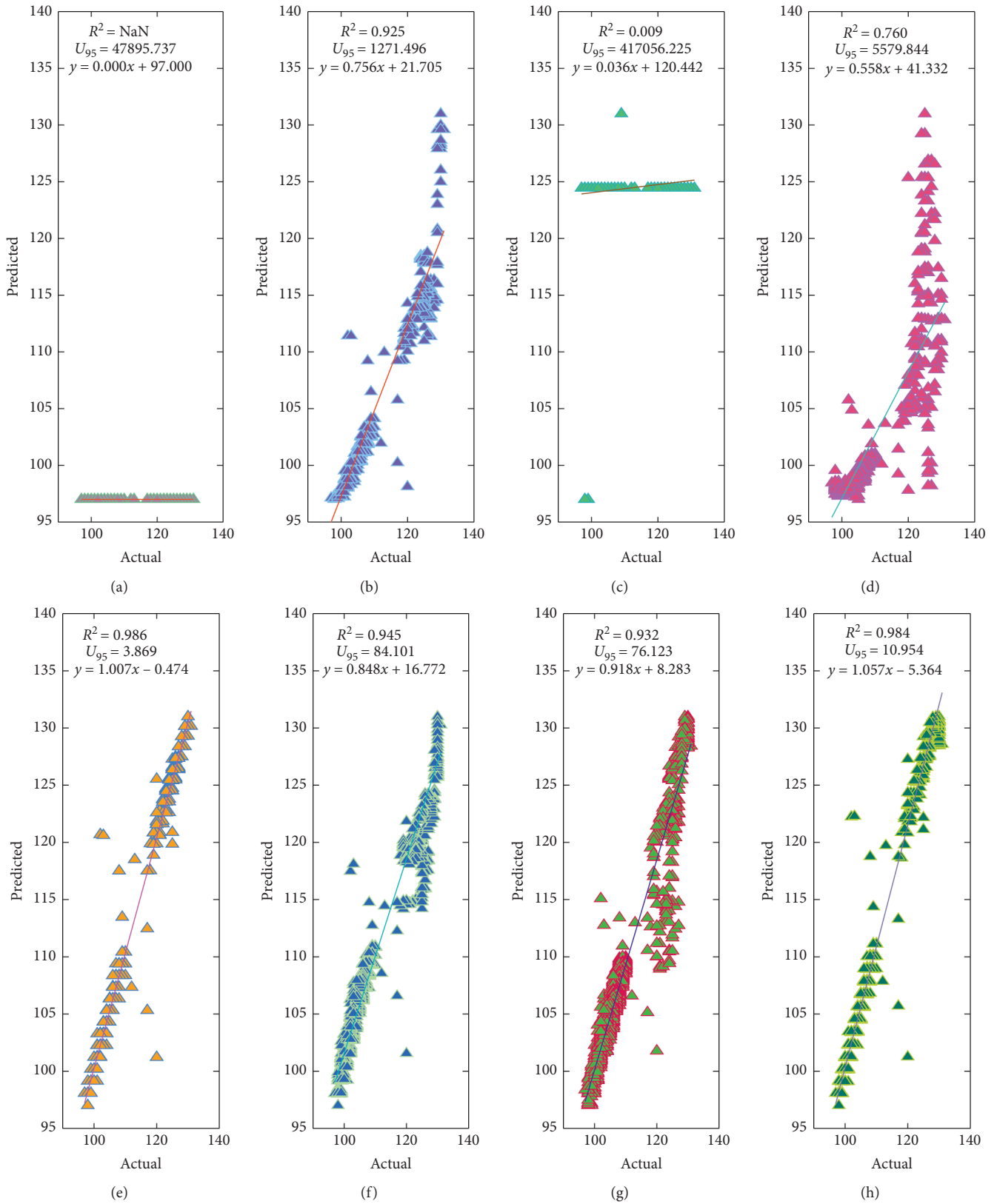
Figure 9: SVR models for USGS 14181500 station: coefficient of determination. (a) SVR-M1. (b) SVR-M2. (c) SVR-M3. (d) SVR-M4. (e) SVR-M5. (f) SVR-M6. (g) SVR-M7. (h) SVR-M8.
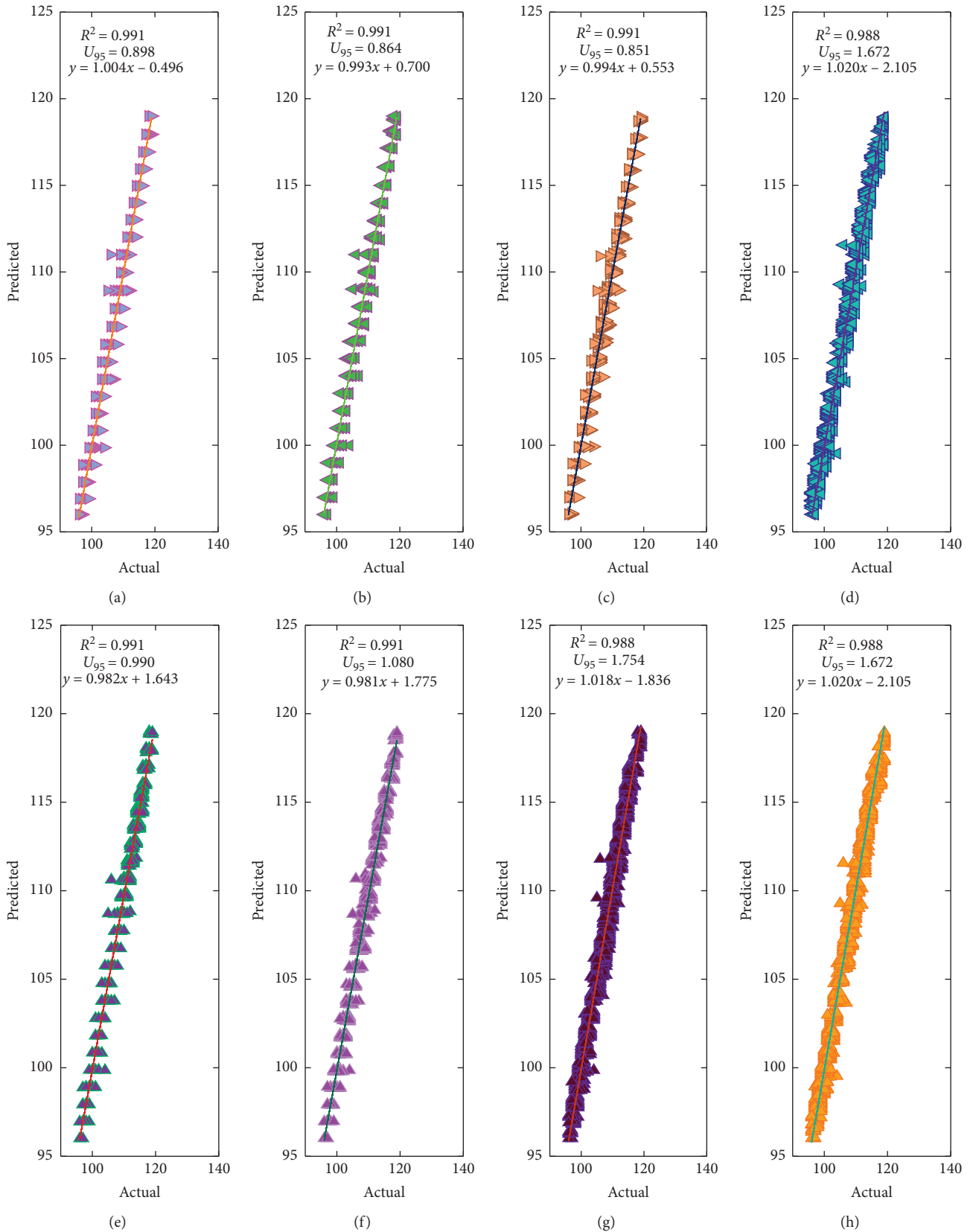
Figure 10: ELM models for USGS 14150000 station: coefficient of determination. (a) ELM-M1. (b) ELM-M2. (c) ELM-M3. (d) ELM-M4. (e) ELM-M5. (f) ELM-M6. (g) ELM-M7. (h) ELM-M8.
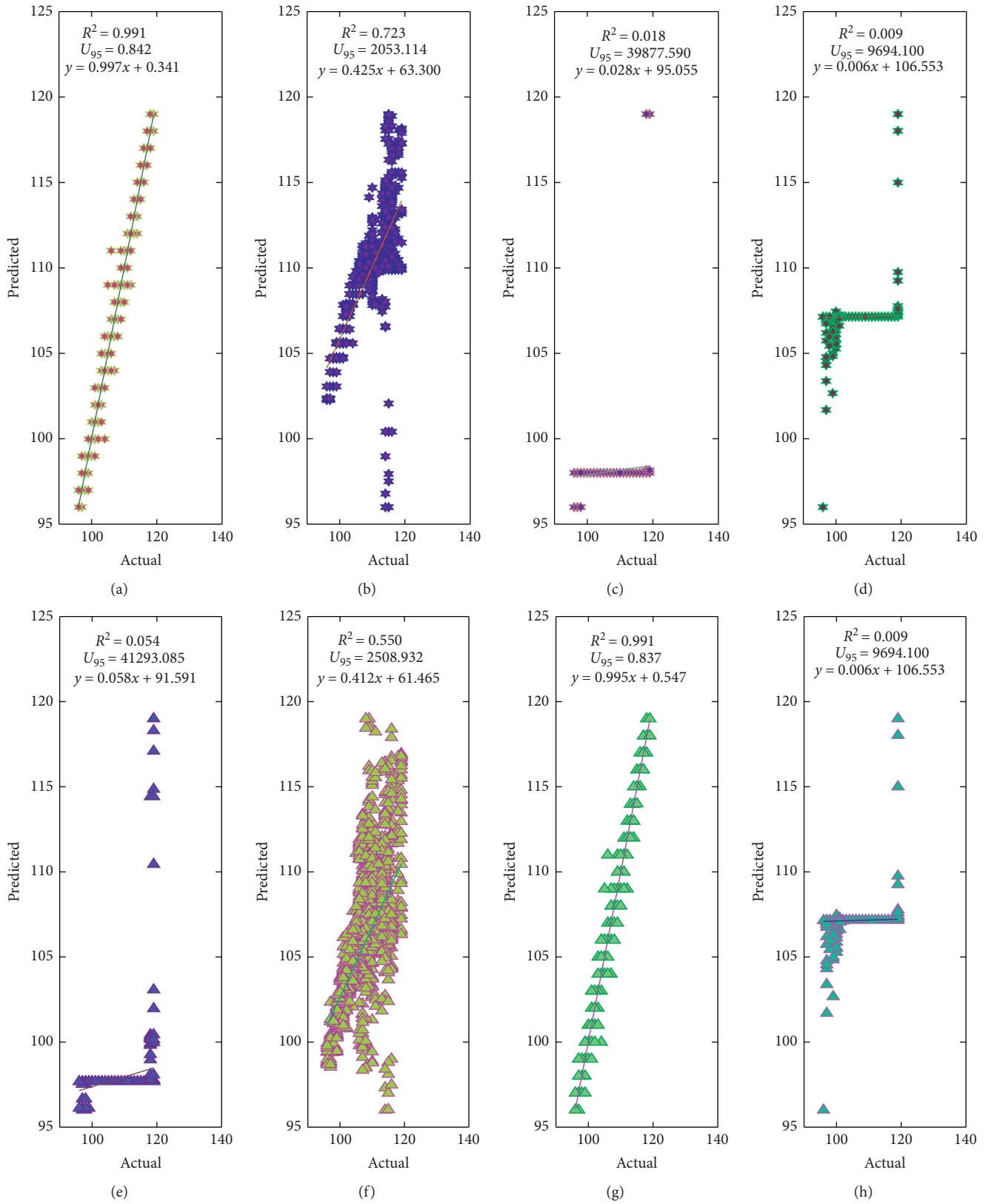
FIGURE 11: SVR models for USGS 14150000 station: coefficient of determination. (a) SVR-M1. (b) SVR-M2. (c) SVR-M3. (d) SVR-M4. (e) SVR-M5. (f) SVR-M6. (g) SVR-M7. (h) SVR-M8.

TABLE 7: The computational time required for the training phase from each model.

Model	Time (s)	Model	Time (s)
USGS 14181500 station			
ELM-M1	0.007	SVR-M1	5486.311
ELM-M2	0.019	SVR-M2	961.291
ELM-M3	0.015	SVR-M3	75.623
ELM-M4	0.004	SVR-M4	1757.157
ELM-M5	0.018	SVR-M5	869.159
ELM-M6	0.024	SVR-M6	397.216
ELM-M7	0.023	SVR-M7	160.172
ELM-M8	0.033	SVR-M8	85.000
Mean	0.018		1223.991
USGS 14150000 station			
ELM-M1	0.012	SVR-M1	3472.969
ELM-M2	0.003	SVR-M2	8305.326
ELM-M3	0.008	SVR-M3	267.823
ELM-M4	0.015	SVR-M4	645.349
ELM-M5	0.033	SVR-M5	345.758
ELM-M6	0.049	SVR-M6	7281.479
ELM-M7	0.017	SVR-M7	1446.514
ELM-M8	0.031	SVR-M8	309.020
Mean	0.021		2759.280

pressure has less effect on the saturated TDG for both stations.

5. Conclusion

Total dissolved gas (TDG) is considered as one of the most problematic phenomena associated with the expansion of dams and reservoirs infrastructure which affects the ecological system. Herein, the potential for producing a robust predictive model utilizing two artificial intelligence methodologies to estimate one hour ahead TDG based on environmental variables. Eight input combinations were used as inputs for both types of machine learning, i.e., ELM, and SVM. The ELM model outperformed the SVM models in all the statistical measures at the testing phase. The given results also showed that the temperature has the most significant influence on the TDG and played a substantial role in increasing the accuracy of prediction. Moreover, several SVR models provided very low performances with higher forecasted errors. In general, the reason behind achieving low prediction accuracies is that the SVR suffers from overfitting issues, thereby reducing its ability of generalization when it comes to deal with huge data sets. Besides, the computation time of training the SVM algorithm was very huge in comparison with the ELM, where the average time required for the training process ranged 0.018s–0.021s and 1223.991s–2759.280s for ELM and SVM, respectively. It is worth mentioning that the use of U_{95} gave a great advantage in specifying the best model especially when other indicators recorded a close point. Finally, this study successfully produced a data-driven model to predict TDG based on machine learning approaches. The current study recommends to increasingly apply the ELM approach to cope with the environmental issues which contain huge data samples due to its ability to provide excellent outcomes as well as

requiring less time to complete the training processes. For future research, (a) prior to learning method, feature selection approach could be used to select the best input variables; (b) exploring the effectiveness of environmental and hydro-environmental variables such as PH, DO, evaporation, turbidity, and suspended sediment load on the prediction of saturated TDG%; (c) estimation of multihour ahead TDG% using AI models approaches.

Data Availability

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] Q. Ma, R. Li, J. Feng, J. Lu, and Q. Zhou, "Cumulative effects of cascade hydropower stations on total dissolved gas supersaturation," *Environmental Science and Pollution Research*, vol. 25, no. 14, pp. 13536–13547, 2018.

[2] S. Heddam, "Generalized regression neural network based approach as a new tool for predicting total dissolved gas (TDG) downstream of spillways of dams: a case study of Columbia river basin dams, USA," *Environmental Processes*, vol. 4, no. 1, pp. 235–253, 2017.

[3] M. Politano, P. Carrica, and L. Weber, "A multiphase model for the hydrodynamics and total dissolved gas in tailraces," *International Journal of Multiphase Flow*, vol. 35, no. 11, pp. 1036–1050, 2009.

[4] Y. Wang, Y. Li, R. An, and K. Li, "Effects of total dissolved gas supersaturation on the swimming performance of two endemic fish species in the upper Yangtze river," *Scientific Reports*, vol. 8, no. 1, Article ID 10063, 2018.

[5] M. Mumba and J. R. Thompson, "Hydrological and ecological impacts of dams on the Kafue Flats floodplain system, southern Zambia," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 30, no. 6-7, pp. 442–447, 2005.

[6] G. Wang, Q. Fang, L. Zhang, W. Chen, Z. Chen, and H. Hong, "Valuing the effects of hydropower development on watershed ecosystem services: case studies in the Jiulong River Watershed, Fujian Province, China," *Estuarine, Coastal and Shelf Science*, vol. 86, no. 3, pp. 363–368, 2010.

[7] D. Anderson, H. Moggridge, P. Warren, and J. Shucksmith, "The impacts of "run-of-river" hydropower on the physical and ecological condition of rivers," *Water and Environment Journal*, vol. 29, no. 2, pp. 268–276, 2015.

[8] D. A. Geldert, J. S. Gulliver, and S. C. Wilhelms, "Modeling dissolved gas supersaturation below spillway plunge pools," *Journal of Hydraulic Engineering*, vol. 124, no. 5, pp. 513–521, 1998.

[9] L. Weber, H. Huang, Y. Lai, and A. McCoy, "Modeling total dissolved gas production and transport downstream of spillways: three-dimensional development and applications,"

International Journal of River Basin Management, vol. 2, no. 3, pp. 157–167, 2004.

[10] J.-j. Feng, R. Li, H.-x. Yang, and J. Li, "A laterally averaged two-dimensional simulation of unsteady supersaturated total dissolved gas in deep reservoir," *Journal of Hydrodynamics*, vol. 25, no. 3, pp. 396–403, 2013.

[11] M. S. Politano, P. M. Carrica, C. Turan, and L. Weber, "A multidimensional two-phase flow model for the total dissolved gas downstream of spillways," *Journal of Hydraulic Research*, vol. 45, no. 2, pp. 165–177, 2007.

[12] X.-l. Fu, D. Li, and X.-f. Zhang, "Simulations of the three-dimensional total dissolved gas saturation downstream of spillways under unsteady conditions," *Journal of Hydrodynamics*, vol. 22, no. 4, pp. 598–604, 2010.

[13] J. Feng, R. Li, R. Liang, and X. Shen, "Eco-environmentally friendly operational regulation: an effective strategy to diminish the TDG supersaturation of reservoirs," *Hydrology and Earth System Sciences*, vol. 18, no. 3, pp. 1213–1223, 2014.

[14] Y. Ou, R. Li, Y. Tuo, J. Niu, J. Feng, and X. Pu, "The promotion effect of aeration on the dissipation of supersaturated total dissolved gas," *Ecological Engineering*, vol. 95, pp. 245–251, 2016.

[15] M. Politano, A. Arenas Amado, S. Bickford, J. Murauskas, and D. Hay, "Evaluation of operational strategies to minimize gas supersaturation downstream of a dam," *Computers & Fluids*, vol. 68, pp. 168–185, 2012.

[16] M. Politano, A. Castro, and B. Hadjerioua, "Modeling total dissolved gas for optimal operation of," *Multireservoir Systems*, vol. 143, no. 6, Article ID 04017007, 2017.

[17] A. Witt, K. Stewart, and B. Hadjerioua, "Predicting total dissolved gas travel time in hydropower reservoirs," *Journal of Environmental Engineering*, vol. 143, no. 12, Article ID 06017011, 2017.

[18] Y. Wang, M. Politano, and L. Weber, "Spillway jet regime and total dissolved gas prediction with a multiphase flow model," *Journal of Hydraulic Research*, vol. 57, no. 1, pp. 26–38, 2019.

[19] S. Heddam and O. Kisi, "Evolving connectionist systems versus neuro-fuzzy system for estimating total dissolved gas at forebay and tailwater of dams reservoirs," in *Intelligent Data Analytics for Decision-Support Systems in Hazard Mitigation: Theory and Practice of Hazard Mitigation*, R. C. Deo, P. Samui, O. Kisi, and Z. M. Yaseen, Eds., pp. 109–126, Springer Singapore, Singapore, 2021.

[20] M. M. Hameed and M. K. AlOmar, *Prediction of Compressive Strength of High-Performance Concrete: Hybrid Artificial Intelligence Technique*, Springer International Publishing, Cham, Switzerland, 2020.

[21] M. K. AlOmar, M. M. Hameed, and M. A. AlSaadi, "Multi hours ahead prediction of surface ozone gas concentration: robust artificial intelligence approach," *Atmospheric Pollution Research*, vol. 11, no. 9, pp. 1572–1587, 2020.

[22] S. S. Fiyadh, M. A. AlSaadi, M. K. AlOmar et al., "The modelling of lead removal from water by deep eutectic solvents functionalized CNTs: artificial neural network (ANN) approach," *Water Science and Technology*, vol. 76, no. 9, pp. 2413–2426, 2017.

[23] S. S. Fiyadh, M. A. AlSaadi, M. K. AlOmar, S. S. Fayaed, F. S. Mjalli, and A. El-Shafie, "BTPC-based DES-functionalized CNTs for As^{3+} removal from water," *NARX Neural Network Approach*, vol. 144, no. 8, Article ID 04018070, 2018.

[24] H. Sanikhani, R. C. Deo, P. Samui et al., "Survey of different data-intelligent modeling strategies for forecasting air temperature using geographic information as model predictors,"

Computers and Electronics in Agriculture, vol. 152, pp. 242–260, 2018.

[25] S. Naganna, P. Deka, M. Ghorbani, S. Biazar, N. Al-Ansari, and Z. Yaseen, "Dew point temperature estimation: application of artificial intelligence model integrated with nature-inspired optimization algorithms," *Water*, vol. 11, no. 4, p. 742, 2019.

[26] W. H. M. Wan Mohtar, K. N. Abdul Maulud, N. S. Muhammad, S. Sharil, and Z. M. Yaseen, "Spatial and temporal risk quotient based river assessment for water resources management," *Environmental Pollution*, vol. 248, pp. 133–144, 2019.

[27] H. Tao, A. A. Ewees, A. O. Al-Sulttani et al., "Global solar radiation prediction over North Dakota using air temperature: development of novel hybrid intelligence model," *Energy Reports*, vol. 7, pp. 136–157, 2021.

[28] M. H. Kashani, M. A. Ghorbani, M. Shahabi, S. Raghavendra Naganna, and L. Diop, "Multiple AI model integration strategy—application to saturated hydraulic conductivity prediction from easily available soil properties," *Soil and Tillage Research*, vol. 196, Article ID 104449, 2020.

[29] S. Shamshirband, F. Esmaeilbeiki, D. Zarehaghi et al., "Comparative analysis of hybrid models of firefly optimization algorithm with support vector machines and multilayer perceptron for predicting soil temperature at different depths," *Engineering Applications of Computational Fluid Mechanics*, vol. 14, no. 1, pp. 939–953, 2020.

[30] A. Ashrafzadeh, A. Malik, V. Jothiprakash, M. A. Ghorbani, and S. M. Biazar, "Estimation of daily pan evaporation using neural networks and meta-heuristic approaches," *ISH Journal of Hydraulic Engineering*, vol. 26, no. 4, pp. 421–429, 2020.

[31] N. S. Raghavendra and P. C. Deka, "Support vector machine applications in the field of hydrology: a review," *Applied Soft Computing*, vol. 19, pp. 372–386, 2014.

[32] T. M. T. Tiyasha, T. M. Tung, and Z. M. Yaseen, "A survey on river water quality modelling using artificial intelligence models: 2000–2020," *Journal of Hydrology*, vol. 585, Article ID 124670, 2020.

[33] A. H. Haghiabi, A. H. Nasrolahi, and A. Parsaie, "Water quality prediction using machine learning methods," *Water Quality Research Journal*, vol. 53, no. 1, pp. 3–13, 2018.

[34] B. Keshtegar, S. Heddam, O. Kisi, and S. P. Zhu, "Modeling total dissolved gas (TDG) concentration at Columbia river basin dams: high-order response surface method (H-RSM) vs. M5Tree, LSSVM, and MARS," *Arabian Journal of Geosciences*, vol. 12, no. 17, p. 544, 2019.

[35] T. Jin, S. Cai, D. Jiang, and J. Liu, "A data-driven model for real-time water quality prediction and early warning by an integration method," *Environmental Science and Pollution Research*, vol. 26, no. 29, pp. 30374–30385, 2019.

[36] U.S. Department of the Interior, USGS, *Title: NWIS Site Information for USA: Site Inventory*, United States Geological Survey (USGS), Reston, VA, USA, 2020.

[37] R. C. Deo, N. Downs, A. V. Parisi, J. F. Adamowski, and J. M. Quilty, "Very short-term reactive forecasting of the solar ultraviolet index using an extreme learning machine integrated with the solar zenith angle," *Environmental Research*, vol. 155, pp. 141–166, 2017.

[38] R. C. Deo and M. Şahin, "Application of the extreme learning machine algorithm for the prediction of monthly effective drought index in eastern Australia," *Atmospheric Research*, vol. 153, pp. 512–525, 2015.

[39] J. Kim, J. Kim, G.-J. Jang, and M. Lee, "Fast learning method for convolutional neural networks using extreme learning

machine and its application to lane detection," *Neural Networks*, vol. 87, pp. 109–121, 2017.

[40] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.

[41] M. Atiquzzaman and J. Kandasamy, "Prediction of hydrological time-series using extreme learning machine," *Journal of Hydroinformatics*, vol. 18, no. 2, pp. 345–353, 2015.

[42] I. Ebtehaj, H. Bonakdari, F. Moradi, B. Gharabaghi, and Z. S. Khozani, "An integrated framework of extreme learning machines for predicting scour at pile groups in clear water condition," *Coastal Engineering*, vol. 135, pp. 1–15, 2018.

[43] J. Shiri, S. Shamshirband, O. Kisi et al., "Prediction of water-level in the Urmia lake using the extreme learning machine approach," *Water Resources Management*, vol. 30, no. 14, pp. 5217–5229, 2016.

[44] G.-B. Huang and L. Chen, "Convex incremental extreme learning machine," *Neurocomputing*, vol. 70, no. 16–18, pp. 3056–3062, 2007.

[45] R. C. Deo and M. Şahin, "An extreme learning machine model for the simulation of monthly mean streamflow water level in eastern Queensland," *Environmental Monitoring and Assessment*, vol. 188, no. 2, p. 90, 2016.

[46] Z. M. Yaseen, S. O. Sulaiman, R. C. Deo, and K.-W. Chau, "An enhanced extreme learning machine model for river flow forecasting: state-of-the-art, practical applications in water resource engineering area and future research direction," *Journal of Hydrology*, vol. 569, pp. 387–408, 2019.

[47] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[48] A. A. Farag and R. M. Mohamed, "Classification of multi-spectral data using support vector machines approach for density estimation," in *Proceedings of the 7th INES 2003*, Assiut, Egypt, March 2003.

[49] J. Fan, L. Wu, F. Zhang et al., "Evaluating the effect of air pollution on global and diffuse solar radiation prediction using support vector machine modeling based on sunshine duration and air temperature," *Renewable and Sustainable Energy Reviews*, vol. 94, pp. 732–747, 2018.

[50] X. Ji, X. Shang, R. A. Dahlgren, and M. Zhang, "Prediction of dissolved oxygen concentration in hypoxic river systems using support vector machine: a case study of Wen-Rui Tang River, China," *Environmental Science and Pollution Research*, vol. 24, no. 19, pp. 16062–16076, 2017.

[51] K. Mohammadi, S. Shamshirband, M. H. Anisi, K. A. Alam, and D. Petković, "Support vector regression based prediction of global solar radiation on a horizontal surface," *Energy Conversion and Management*, vol. 91, pp. 433–441, 2015.

[52] K. Roushangar, M. T. Alami, J. Shiri, and M. M. Asl, "Determining discharge coefficient of labyrinth and arced labyrinth weirs using support vector machine," *Hydrology Research*, vol. 49, no. 3, pp. 924–938, 2017.

[53] 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.

[54] N. Sapankevych and R. Sankar, "Time series prediction using support vector machines: a survey," *IEEE Computational Intelligence Magazine*, vol. 4, no. 2, pp. 24–38, 2009.

[55] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer Science & Business Media, Berlin, Germany, 2013.

[56] S. Belaid and A. Mellit, "Prediction of daily and mean monthly global solar radiation using support vector machine in an arid climate," *Energy Conversion and Management*, vol. 118, pp. 105–118, 2016.

[57] O. Kisi, S. Heddam, and Z. M. Yaseen, "The implementation of univariable scheme-based air temperature for solar radiation prediction: new development of dynamic evolving neural-fuzzy inference system model," *Applied Energy*, vol. 241, pp. 184–195, 2019.

[58] M. M. Hameed, S. S. Sharqi, Z. M. Yaseen, H. A. Afan, A. Hussain, and A. Elshafie., "Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in tropical region, Malaysia," *Neural Computing and Applications*, vol. 28, no. 1, pp. 893–905, 2017.

[59] H. A. Afan, A. El-Shafie, Z. M. Yaseen, M. M. Hameed, W. H. M. Wan Mohtar, and A. Hussain, "ANN based sediment prediction model utilizing different input scenarios," *Water Resources Management*, vol. 29, no. 4, pp. 1231–1245, 2015.

[60] Z. M. Yaseen, A. El-Shafie, H. A. Afan, M. Hameed, W. H. M. W. Mohtar, and A. Hussain, "RBFNN versus FFNN for daily river flow forecasting at Johor River, Malaysia," *Neural Computing and Applications*, vol. 27, no. 6, pp. 1533–1542, 2016.

[61] M.-F. Li, X.-P. Tang, W. Wu, and H.-B. Liu, "General models for estimating daily global solar radiation for different solar radiation zones in mainland China," *Energy Conversion and Management*, vol. 70, pp. 139–148, 2013.

[62] K. Khosravi, P. Daggupati, M. T. Alami et al., "Meteorological data mining and hybrid data-intelligence models for reference evaporation simulation: a case study in Iraq," *Computers and Electronics in Agriculture*, vol. 167, Article ID 105041, 2019.

[63] G. Zhang, Z. H. Ali, M. S. Aldlemy et al., "Reinforced concrete deep beam shear strength capacity modelling using an integrative bio-inspired algorithm with an artificial intelligence model," *Engineering with Computers*, pp. 1–14, 2020.