Prediction of aeration efficiency of parshall and modified venturi flumes: application of soft computing versus regression models

Parveen Sihag, Omer Faruk Dursun, Saad Shauket Sammen, Anurag Malik, and Anita Chauhan

Department of Civil Engineering, Shoolini University, Solan, India
Civil Engineering Department, Faculty of Engineering, Inonu University Malatya, Turkey
Department of Civil Engineering, College of Engineering, Diyala University, Diyala 15 Governorate, Iraq
Punjab Agricultural University, Regional Research Station, Bathinda 151001, Punjab, India
Department of Law, Shoolini University, Solan, India

Corresponding author. E-mail: amalik19@pau.edu

ABSTRACT

In this study, the potential of soft computing techniques namely Random Forest (RF), M5P, Multivariate Adaptive Regression Splines (MARS), and Group Method of Data Handling (GMDH) was evaluated to predict the aeration efficiency (AE20) at Parshall and Modified Venturi flumes. Experiments were conducted for 26 various Modified Venturi flumes and one Parshall flume. A total of 99 observations were obtained from experiments. The results of soft computing models were compared with regression-based models (i.e., MLR: multiple linear regression, and MNLR: multiple nonlinear regression). Results of the analysis revealed that the MARS model outperformed other soft computing and regression-based models for predicting the AE20 at Parshall and Modified Venturi flumes with Pearson’s correlation coefficient (CC) = 0.9997, and 0.9992, and root mean square error (RMSE) = 0.0015, and 0.0045 during calibration and validation periods. Sensitivity analysis was also carried out by using the best executing MARS model to assess the effect of individual input variables on AE20 of both flumes. Obtained results on sensitivity examination indicate that the oxygen deficit ratio (r) was the most effective input variable in predicting the AE20 at Parshall and Modified Venturi flumes.

Key words: aeration efficiency, regression-based models, soft computing models

HIGHLIGHTS

- Aeration Efficiency of Parshall and Modified Venturi Flumes was predicted by using soft computing techniques.
- M5P, RF, MARS, and GMDH models were employed to predict the aeration efficiency at the first time.
- Outcomes of soft computing models were compared against regression-based models.
- Effectiveness of applied models was evaluated using performance evaluation indicators.
- MARS-based model outperformed other models.

INTRODUCTION

Several main sources cause water pollution such as poisonous substances, detergents, repellents, and products of mining, organic, and industrial wastes (Schwarzenbach et al. 2010). The quality of water in nature such as rivers and artificial systems such as canals or reservoirs are related to the existence of dissolved oxygen (DO). Therefore, the DO is considered necessary for aquatic life. The ideal value of DO in natural water bodies is ranged between 5 and 6 mg/L (Sánchez et al. 2007). Therefore, it is required to keep the amount of DO in this range because if the level of DO is drop than 5 mg/L, this will affect the water quality and the fish will die within hours if the level of DO is a drop below than 1–2 mg/L (Baylar et al. 2009b). On other hand, if the level of DO should be not greater than 110%, otherwise, it may be harmful to aquatic life. The physical processes that include the transmission of DO from the atmosphere to the water will help to replenish the existing oxygen. The term given for this process is aeration and considered a significant process in wastewater treatment plants (Sangeeta & Tiwari 2019).

The hydraulic structure could be considered as the best choice to improve the water quality in the river system by aeration where it helps to raise the amount of DO. Only one hydraulic structure can provide the same amount of DO that may occur along several kilometers of river. This speedy transfer of DO happens due to a large number of bubbles that form because of entrained air with water flow, which further enhances the surface area thereby leading to mass transfer. Different types of spillways such as stepped or overflow spillways could
be used for aeration purposes in the river system. On the other hand, these structures have not been considered the best solution in the straight flow canal. Therefore, other types of hydraulic structures such as drop structures and weir are preferred. For the prismatic channels, the small Parshall flume is represented as the best solution to provide the required aeration.

Several studies reviewed the experimental works which deal with the evaluation of aeration performance of hydraulic structures such as Gulliver et al. (1990), Wilhelms et al. (1993), Chanson (1995), and Ervine (1998), and. In last two decades, Baylar & Bagatur (2000, 2006), Baylar et al. (2006, 2008, 2009a, 2009c), Hanbay et al. (2009), and Baylar & Emiroglu (2003) have examined sharp-crested barriers with diverse cross-sectional geometry. Their findings revealed that the rate of air entrainment rate and the aeration efficiency of barriers varied depending on the barrier. Parshall flume was designed and used for the first time by Ralph Parshall. He starts his study based on work that was achieved by Cone in 1917 and improved in 1928. After that, this cascade was used in several irrigation projects (USBR 2001; Kim et al. 2010; Dursun 2016). Recently, the Parshall flume was utilized for several projects around the world to measure the discharge in channels and river systems.

In the last few years, Artificial Intelligent (AI) techniques have been widely applied in different field of water resources as they are capable of resolving the intricate problems that did not have the tractable solution (Aghelpour et al. 2019; Jahani & Mohammadi 2019; Sihag et al. 2019b, 2019c; Mohammadi et al. 2020; Thakur et al. 2021). AI-based models/algorithms have been employed extensively in many water resources applications; including hydrology (Banadkooki et al. 2020; Malik et al. 2020b; Tikhamarine et al. 2020c; Ghasempour et al. 2021; Sihag et al. 2021), hydraulics (Parsaie 2016; Parsaie et al. 2016, 2018, 2019; Ebtehaj et al. 2017; Najafzadeh et al. 2017; Sihag et al. 2019a), and water flow/quality (Heddam & Kisi 2017; Parsaie & Haghiabi 2017a, 2019; Haghiabi et al. 2018; Singh et al. 2019; Esmaeilbeiki et al. 2020; Pandhian et al. 2020). Nevertheless, limited numbers of studies have considered the application of AI for the evaluation of Parshall flume aeration performance. Therefore, in this study, the MSP, RF, MARS, and GMDH techniques were used to developed the AI-based models to predict the aeration efficiency at Parshall and Modified Venturi flumes and compared with regression-based models.

**MATERIALS AND METHODS**

**Mechanism of aeration**

Aeration is the most dominant factor in wastewater treatment. In aeration, oxygen is introduced into the water to increase the DO level for the survival of aquatic life. Aeration in the Parshall flume is different from the weirs. A Parshall flume consists of three portions. The first portion is made up of the upstream narrowing approach portion followed by a short and sloping throat portion from where the flow continues to the diverging downstream portion as shown in Figure 1. The upstream portion is the largest one with no slope, the short narrow throat portion has a downstream sloping portion and the floor again rises in the downstream portion (Parshall 1926). In weirs, aeration is done by creating the hydraulic jump (Baylar & Emiroglu 2003). Figure 2 demonstrates an explanation of the aeration process for various stages of the cascade from the weir (Tsang 1987). But aeration process of the Parshall flumes is that it accelerates the flow velocity through contracting sidewalls in the converging portion. At the throat portion, the flow velocity got accelerated from a subcritical to supercritical due to contraction and drop however at diverging section flow changes from fast, supercritical to slow, subcritical following a just at outlet of the flume resulting in aeration (Figure 1).

**Aeration transfer efficiency**

The process of oxygen dispersion in the flowing water bodies is ruled by turbulent mixing (hydraulic jump), molecular diffusion (bubble formation), or both. It has been submitted the concept of the existence of two laminar layers remains on each side of the water-air interface. Both of these layers control the movement of oxygen in the water. The oxygen transfer rate from one side to the other is directly proportional to the concentration gradient and expressed as:

\[
\frac{dC}{dt} = K_L \frac{A}{V} (C_S - C)
\]  

where \( C \) is the concentration of dissolved oxygen, \( K_L \) is the coefficient of the liquid layer for oxygen, \( A \) is the surface area, \( V \) is the volume over which oxygen transfer occurs, \( C_S \) represents saturation concentration, and \( t \) indicates time.
Equation (1) excludes the source of oxygen transfer as its rate is very slow compared to the rate of oxygen transfer by most fluidic structures. Gulliver et al. (1990) has given a predictive association between $C_S$ which is constant concerning time and oxygen aeration efficiency (AE) as shown below:

$$AE = \frac{C_D - C_U}{C_S - C_U} = 1 - \frac{1}{r}$$

Figures 1 and 2 | Definition sketch of Parshall flume.

Figure 2 | Different stages of aeration process by free fall jet from weir (Tsang 1987).

Equation (1) excludes the source of oxygen transfer as its rate is very slow compared to the rate of oxygen transfer by most fluidic structures. Gulliver et al. (1990) has given a predictive association between $C_S$ which is constant concerning time and oxygen aeration efficiency (AE) as shown below:

$$AE = \frac{C_D - C_U}{C_S - C_U} = 1 - \frac{1}{r}$$

where subscript $U$, and $D$ represents upstream and downstream locations, while $r$ represents oxygen deficit ratio.
For full saturation of the water body, the value of aeration efficiency should be unity which represents the full transfer of oxygen to water. Oxygen would not be transferred if the value of aeration efficiency is zero. For the uniformity in measured experiment results, the temperature is generally normalized at 20 °C as a standard. The equation showing the relation of temperature and aeration efficiency was given by Gulliver et al. (1990) as:

$$1 - AE_{20} = (1 - AE_{T})^{1/f}$$

In which, $AE$ is oxygen aeration transfer efficiency at the water temperature of measurement, $AE_{20}$ denotes oxygen transfer efficiency at 20 °C. $f$ denotes aeration exponent and calculated as:

$$f = 1 + 0.02103 (T - 20) + 8.261 \times 10^{-5}(T - 20)^2$$

In Equation (4), $T$ is water temperature, °C.

**Multiple linear regression (MLR)**

MLR method is based on linear regression used for developing the association between dependent and independent variables. The commonly used equation for MLR is as follows (Malik et al. 2021a):

$$AE_{20} = c_0 + c_1x_1 + c_2x_2 + c_3x_3 + \ldots + c_nx_n$$

In Equation (5), $AE_{20}$ represents the dependent (output) variable, $c_0, c_1, c_2, c_3, \ldots, c_n$ denotes the regression coefficients, and $x_1, x_2, x_3, \ldots, x_n$ are the independent variables.

**Multiple nonlinear regression (MNLR)**

The application of MNLR requires the observed/experimental data set to establish the relationship between input and output variables. For nonlinear and complex issues, MNLR is commonly chosen. This study uses XLSTAT tools for the creation of the MNLR-based model. Equation (6) is classified as the most common form of the MNLR (Malik et al. 2017):

$$AE = c_0x_1^{c_1}x_2^{c_2}x_3^{c_3}x_4^{c_4} \ldots \ldots x_n^{c_n}$$

**Review on MARS**

MARS is a non-parametric method that considers no assumptions regarding the association between independent and dependent datasets. The space of input characteristics is split into sub-domains in the MARS system, and a linear regression equation is installed for each sub-domain. A boundary value between knots is called sub-domains. The fitted linear regression is defined as the base functions (BFs). The following forms are given by the BFs:

$$\max(0, x - k) \text{or} \max(0, k - x)$$

where, $x$ denotes a separate parameter, and $k$ denotes boundary value. The following is the deduced mathematical formula using the MARS design for the desired phenomenon:

$$y = f(x) = S_0 + \sum_{n=1}^{N} S_n \beta_n(x)$$

where $y$ denotes the dependent parameter predicted by the function $f(x)$, $S_0$ denotes the value of a constant and $n$ denotes the number of BFs. $S_n$ symbolizes the coefficient multiplied in BFs. $\beta_n$ indicates BFs. Arithmetic design is sufficient to complete two phases using the MARS process. The first stage is the creation of the model, where the function input space is split into several subdomains. This stage of the MARS design is the growth of names that feed-forward algorithms. The pruning stage is the second step. The BFs established in the previous stage have no major impact on enhancing the accuracy of the model at this stage are therefore pruned based on a criterion called Generalized Cross-Validation (GCV). Thus, the structure of the derived MARS model type system was
adopted by GCV. As below, the GCV is defined:

$$GCV = \frac{1}{N} \sum_{i=1}^{N} [y_i - f(x_i)]^2, \quad C(H) = (h + 1) + dH$$

where \(N\) signifies the data number, and \(C(H)\) is the penalty for complexity that escalates by the number of BFs. The above equation shows that for each BF, \(d\): penalty number and \(H\): number of BFs obtained from the MARS process (Parsaie & Haghiabi 2017b, 2017c; Sihag et al. 2019d).

**Review on GMDH**

GMDH is an auto-adjusting design that eventually models in any complicated system the dynamic relationship between inputs and output. Ivakhnenko (1971) suggested GMDH approach. Separate pair of inputs are added to a neuron in the GMDH technique. As seen below, the governing equation is a quadratic polynomial:

$$\tilde{y} = G(x_i, x_j) = w_0 + w_1x_i + w_2x_j + w_3x_i^2 + w_4x_j^2 + w_5x_ix_j$$

This is the discrete form of the Volterra series as presented by Kolmogorov-Gabor polynomial (Najafzadeh & Lim 2015; Alfaifi et al. 2020).

**Review on M5P tree**

Basically, model trees depend on the idea of regression trees having linear functions on their leaves (Pal & Deswal 2009). In piece-wise type, they are similar to linear equations. M5P model tree is a binary decision tree at the terminal (leaf) nodes with linear equations that can approximate continuous mathematical characteristics (Quinlan 1992). The chance of overfitting is reduced through pruning in this algorithm. A separation method is introduced for each node to deduce enhanced knowledge with minimal deviation within the intra-subset class values in each branch. There are three main steps in preparing the M5P model: tree growth, pruning, and smoothing. Using separation standards, which estimate the standard deviations of the class values extending to nodes, the basic tree model is generated. Linear functions in each node are generated by this approach. It utilizes the standard deviation approach by measuring the predicted error at the terminal node. The standard reduction is given as follows:

$$SDR = sd(Z) - \sum_{i}^{n} \left( \frac{|Z_i|}{|Z|} sd(Z_i) \right)$$

In Equation (12), \(Z\) is a list of examples at the \(Z_i\) node showing the outcome of the subset of possible set examples, \(sd\) is the standard equation. A considerable tree structure with high prediction precision is generated by this form of technique.

**Review on RF**

The random forest (RF) procedure was primarily introduced by Breiman (1996). RF is a versatile approach that has been chosen to solve various nonlinear or intricate engineering issues (Mohammadi & Mehdizadeh 2020). In this technique, a substantial quantity of trees is created having different bootstrap (bagging) samples of the original data set at the root node. The division is performed at each node using a randomly chosen subset of the estimation parameters. The random forest algorithm is comparatively insensitive to features of the training set and can achieve high prediction accuracy (Breiman 2001). It entails the usage of two user-defined parameters: the number of trees
cultivated \((k)\) and the number of input characteristics \((m)\). For model development, a trial-and-error process is used. The WEKA 3.9 software was used to develop the random forest-based model in this current investigation.

**Performance evaluation indicators**

For the assessment of the accuracy of the implemented AI and regression based-models, six different types of performance/statistical indicators namely Pearson’s coefficient of correlation (CC), scatter index (SI), mean absolute error (MAE), root mean square error (RMSE), Bias, and of the Nash-Sutcliffe efficiency (NSE) were considered and defined as:

Pearson’s correlation coefficient: The CC shows the interrelation between observed and predicted values and ranges from \(-1\) to \(+1\). It can be calculated as (Malik et al. 2019b, 2020a):

\[
CC = \frac{\sum_{i=1}^{N} (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{N} (O_i - \bar{O})^2 \sum_{i=1}^{N} (P_i - \bar{P})^2}}
\] (13)

Root mean square error: The RMSE is one of the most used errors to assess the model fitness. As the name indicates, it is the square root of the mean square error. A zero value indicates the best prediction and expressed as (Elbeltagi et al. 2020; Malik et al. 2021c):

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (O_i - P_i)^2}
\] (14)

Scatter index: It is the ratio of RMSE and average of actual observations, and written as (Tao et al. 2018; Malik et al. 2019a):

\[
SI = \frac{RMSE}{O}
\] (15)

Bias: It is the average difference among observed and predicted values. It ranges from \(-\infty \) to \(\infty \), and calculated as (Sihag et al. 2017):

\[
Bias = \frac{\sum_{i=1}^{N} (O - P)}{N}
\] (16)

Mean absolute error: The MAE is the mean of the absolute difference between observed and predicted values. The range of MAE is \(0 \) to \(\infty \). The formula for MAE (Tikhamarine et al. 2020a; Malik et al. 2021b; Mohammadi et al. 2021):

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |P_i - O_i|
\] (17)

Nash-Sutcliffe efficiency: Nash & Sutcliffe (1970) introduced the NSE which was used to evaluate the efficacy of soft computing models. The range of NSE lies between \(-\infty \) to \(1\). If NSE is equal to 1, it shows a perfect model, and computed as (Tikhamarine et al. 2020b):

\[
NSE = 1 - \frac{\sum_{i=1}^{N} (O_i - P_i)^2}{\sum_{i=1}^{N} (O_i - \bar{O})^2}
\] (18)

In Equations (13)–(18), \(O_i\) and \(P_i\) are the observed (actual) and the predicted aeration efficiency, \(\bar{O}\) and \(\bar{P}\) are the average of observed and predicted aeration efficiency, and \(N\) denotes the number of observations.
Dataset
For model establishment and validation, a total of 99 experimental observations of aeration efficiency at 20 °C with Parshall and Modified Venturi flumes were used. Two separate classes of the total data set were separated. The division method was subjective. The model production training dataset included 69 observations while the remaining 30 observations were considered in the model validation test data set. Six independent variables were considered to be inputs, flow rate (Q), throat widths (W), throat lengths (L), sill heights (N), oxygen deficit ratio (r), and exponent (f), while aeration efficiency at 20 °C (AE$_{20}$) was observed as the target for the model establishment and validation goal. Table 1 outlines the features of the data used for model development (training) and validation.

RESULTS AND DISCUSSION
For the prediction of aeration efficiency at Parshall and modified Venturi flumes, soft computing and regression-based models were used in this investigation. Six standard statistical parameters, CC, MAE, Bias, SI, RMSE, and NSE, were chosen to test the working of all implemented models. Lower RMSE, Bias, SI, and MAE values show higher model accuracy, and the higher CC and NSE values show higher model accuracy. Model preparation is a method of trial-and-error. After several tests, the optimum values of the user-defined parameters were achieved. There are well-defined statistical criteria for selecting and defining user-defined parameters that are unique to the model.

Results of linear and nonlinear regression-based models
In this investigation, linear and nonlinear regression-based designs for the prediction of aeration efficiency at Parshall and Modified Venturi flumes have also been developed. XLSTAT software employing the least square method was used to develop these equations. For all developed equations, the performance measurement parameter values are listed in Table 2. Linear and nonlinear equations based on regression-based models are as follows:

**MLR Equation:**

\[
AE_{20} = 0.33066 + 0.00019Q + 0.00008W + 0.00122N + 0.62605r - 1.02413f 
\]

**MNLR Equation:**

\[
AE_{20} = 0.00396Q^{0.22}W^{0.1648}L^{0.036}N^{0.0854}r^{3.499}f^{-18.026} 
\]

Table 1 | Characteristics of the data used for model development and validation

| Range         | Q       | W       | L       | N       | r       | f       | AE$_{20}$ | Data set |
|---------------|---------|---------|---------|---------|---------|---------|-----------|----------|
| Mean          | 59.0909 | 7.7273  | 15.0000 | 5.0000  | 1.1283  | 0.9170  | 0.1177    | Total    |
|               | 61.5942 | 8.0797  | 14.7101 | 5.0000  | 1.1250  | 0.9171  | 0.1157    | Training |
| Minimum       | 25.0000 | 5.0000  | 10.0000 | 2.5000  | 1.0284  | 0.9070  | 0.0300    | Total    |
|               | 25.0000 | 5.0000  | 10.0000 | 2.5000  | 1.0284  | 0.9070  | 0.0300    | Training |
|               | 25.0000 | 5.0000  | 10.0000 | 2.5000  | 1.0371  | 0.9070  | 0.0389    | Testing  |
| Maximum       | 100.0000| 10.0000 | 20.0000 | 7.5000  | 1.7440  | 0.9213  | 0.4547    | Total    |
|               | 100.0000| 10.0000 | 20.0000 | 7.5000  | 1.7440  | 0.9213  | 0.4547    | Training |
|               | 100.0000| 10.0000 | 20.0000 | 7.5000  | 1.6641  | 0.9213  | 0.4261    | Testing  |
| Standard deviation | 26.8344     | 1.9914   | 4.1033  | 2.0516  | 0.0980  | 0.0030  | 0.0641    | Total    |
|               | 26.6305     | 1.8760   | 4.0114  | 2.1004  | 0.0910  | 0.0030  | 0.0599    | Training |
|               | 26.8564     | 2.0430   | 4.3018  | 1.9696  | 0.1139  | 0.0031  | 0.0736    | Testing  |
| Confidence level (95.0%) | 5.3520     | 0.3972   | 0.8184  | 0.4092  | 0.0195  | 0.0006  | 0.0128    | Total    |
|               | 6.3974     | 0.4507   | 0.9636  | 0.5046  | 0.0219  | 0.0007  | 0.0144    | Training |
|               | 10.0283    | 0.7629   | 1.6063  | 0.7355  | 0.0425  | 0.0011  | 0.0275    | Testing  |
Comparison of the performance of the MLR and MNLR models was done based on six statistical performance measures viz., CC, RMSE, Bias, SI, MAE, and NSE (Table 2). As per Table 2, the higher CC, NSE (0.9840 and 0.9670), and lower RMSE, Bias, SI, and MAE (0.0131, −0.0019, 0.1073, and 0.0111) in the testing phase of the MLR-based model, and concluded that the MLR-based model had better performance compared to MNLR-based model. Figure 3 displays agreement plots using training and testing datasets, separately, between observed and predicted aeration efficiency at Parshall and Modified Venturi flumes by the MLR and MNLR based models.

As depicted in the graph, values predicted using MLR based model are close to the line of perfect agreement. Results of RF-based model

Figure 4 provides plots of agreement between observed and predicted aeration efficiency at Parshall and modified Venturi flumes by RF-based model for training and testing stages. Predicted values from the RF-based model closely follow the observed values. Table 2 summarizes the results of the training and testing datasets which reflects that the performance of the RF design was appropriate for the prediction of aeration efficiency at Parshall and modified Venturi flumes with CC, RMSE, Bias, SI, MAE, and NSE values are 0.9849, 0.0134, −0.0014, 0.0111, 0.9961.

| Models | CC      | RMSE    | Bias    | SI      | MAE    | NSE  |
|--------|---------|---------|---------|---------|--------|------|
| Training dataset | | | | | | |
| MLR    | 0.9813  | 0.0115  | 0.0000  | 0.0991  | 0.0091 | 0.9629 |
| MNLR   | 0.9302  | 0.0220  | 0.0111  | 0.1903  | 0.0187 | 0.8630 |
| MSP_pruned | 0.9960  | 0.0053  | −0.0002 | 0.0461  | 0.0037 | 0.9920 |
| MSP_unpruned | 0.9971  | 0.0047  | −0.0003 | 0.0405  | 0.0035 | 0.9938 |
| RF     | 0.9849  | 0.0134  | −0.0014 | 0.1162  | 0.0037 | 0.9489 |
| GMDH   | 0.9999  | 0.0009  | 0.0000  | 0.0080  | 0.0005 | 0.9999 |
| MARS   | 0.9997  | 0.0015  | 0.0000  | 0.0134  | 0.0010 | 0.9993 |
| Testing dataset | | | | | | |
| MLR    | 0.9840  | 0.0131  | −0.0019 | 0.1073  | 0.0111 | 0.9670 |
| MNLR   | 0.9077  | 0.0341  | −0.0090 | 0.2787  | 0.0278 | 0.7776 |
| MSP_pruned | 0.9971  | 0.0060  | −0.0014 | 0.0494  | 0.0041 | 0.9930 |
| MSP_unpruned | 0.9983  | 0.0032  | −0.0014 | 0.0426  | 0.0038 | 0.9948 |
| RF     | 0.9760  | 0.0259  | −0.0067 | 0.2111  | 0.0110 | 0.8725 |
| GMDH   | 0.9958  | 0.0189  | 0.0029  | 0.1540  | 0.0033 | 0.9649 |
| MARS   | 0.9992  | 0.0045  | −0.0018 | 0.0368  | 0.0021 | 0.9961 |

Figure 3 | Agreement plot of observed and predicted aeration efficiency at Parshall and modified Venturi flumes by the MLR and MNLR-based models using training and testing datasets, respectively.
0.1162, 0.0037 and 0.9489 for training stage and 0.9760, 0.0259, −0.0067, 0.2111, 0.0110 and 0.8725 for testing stage respectively.

Results of M5P tree-based model

The development of M5P based model for predicting the values of the aeration efficiency at Parshall and Modified Venturi flumes is identical to the RF-based model development process. M5 model tree algorithm entails the usage of linear regression models for defining the input-output relationship, based on the splitting of the parameter space of the data set into several sub-spaces. In this study pruned and unpruned both type M5P models were developed and their structures are shown in Figures 5 and 6. Developed linear equations using pruned and unpruned M5P based models are enlisted in Tables 3 and 4. Figure 7 shows the results obtained on the M5P model to predict the aeration efficiency at Parshall and Modified Venturi flumes for training and testing stages. Results of performance evaluation parameters suggests that unpruned M5P based model is more accurate than pruned M5P based model for predicting the aeration efficiency at Parshall and Modified Venturi flumes with CC = 0.9971, 0.9983, RMSE = 0.0047, 0.0052, Bias = −0.0003, −0.0014, SI = 0.0405, 0.0426, MAE = 0.0035, 0.0038 and NSE = 0.9938, 0.9948 for model training and testing periods. The overall evaluation

![Figure 4](http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2021.161/895027/ws2021161.pdf)

**Figure 4** | Agreement plot of observed and predicted aeration efficiency at Parshall and Modified Venturi flumes by the RF model using training and testing datasets, respectively.

![Figure 5](http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2021.161/895027/ws2021161.pdf)

**Figure 5** | Structure of M5P_pruned model.
suggested (Figure 7 and Table 2) that both pruned and unpruned M5P-based models are suitable for predicting the aeration efficiency at Parshall and modified Venturi flumes.

**Results of MARS based model**

The dataset was chosen for the development (training) of MARS based model for the prediction of aeration efficiency at Parshall and modified Venturi flumes are given in Table 1. Thirty primary functions were chosen in the initial stage of development of the MARS model while in the next step ten primary functions were pruned. Finally, 7 basic functions and 16 total effective parameters were found to be the optimal for MARS model. Equation (21) shows the basic form of the MARS model. Table 5 enlists a comprehensive form of the MARS model. The amount of pruning as presented with GVC constraint in the standardization of the MARS model was found to be 0.0000040565. Figure 8 represents the results obtained on the MARS model for training and testing periods. It is evident from these figures that the MARS model’s competence level is high for predicting the aeration efficiency at Parshall and Modified Venturi flumes. Assessment parameters of performance indicators propose that the MARS model was appropriate for the prediction of aeration efficiency at Parshall and modified Venturi flumes with CC = 0.9997, 0.9992, RMSE = 0.0015, 0.0045, Bias = 0.0000, -0.0018, SI = 0.0134, 0.0368, MAE = 0.0010, 0.0021 and NSE = 0.9995, 0.9961 for model development (training) and validation (testing) periods.

$$AE_{20} = 0.1781 + \sum_{M=1}^{8} \beta_m BF_i(x)$$  \hspace{1cm} (21)
Table 4 | Linear equations developed using M5P_unpruned model

| LM num | Equations |
|--------|-----------|
| 1      | $AE_{20} = 0.0001 * Q + 0.0004 * N + 0.8293 * r - 0.8198$ |
| 2      | $AE_{20} = 0.0001 * Q + 0.0004 * N + 0.8293 * r - 0.8198$ |
| 3      | $AE_{20} = 0.0001 * Q + 0.0004 * N + 0.8328 * r - 0.8441$ |
| 4      | $AE_{20} = 0.0001 * Q + 0.0004 * N + 0.8351 * r - 0.8464$ |
| 5      | $AE_{20} = 0.0001 * Q + 0.0004 * N + 0.8364 * r - 0.8261$ |
| 6      | $AE_{20} = 0.0001 * Q + 0.0004 * N + 0.7994 * r - 0.7641$ |
| 7      | $AE_{20} = 0.0001 * Q + 0.0004 * N + 0.8190 * r - 0.7819$ |
| 8      | $AE_{20} = 0.0001 * Q + 0.0004 * N + 0.8190 * r - 0.7819$ |
| 9      | $AE_{20} = 0.0001 * Q + 0.0004 * N + 0.8253 * r - 0.5957$ |
| 10     | $AE_{20} = 0.0001 * Q + 0.0004 * N + 0.7915 * r - 0.5765$ |
| 11     | $AE_{20} = 0.0001 * Q + 0.0004 * N + 0.7915 * r - 0.5765$ |
| 12     | $AE_{20} = 0.0001 * Q + 0.0004 * N + 0.7915 * r - 0.5765$ |
| 13     | $AE_{20} = 0.0001 * Q + 0.0004 * N + 0.7915 * r - 0.5765$ |
| 14     | $AE_{20} = 0.0001 * Q + 0.0004 * N + 0.7915 * r - 0.5765$ |
| 15     | $AE_{20} = 0.0001 * Q + 0.0004 * N + 0.7915 * r - 0.5765$ |
| 16     | $AE_{20} = 0.0001 * Q + 0.0004 * N + 0.7915 * r - 0.5765$ |
| 17     | $AE_{20} = 0.0001 * Q + 0.0004 * N + 0.7915 * r - 0.5765$ |
| 18     | $AE_{20} = 0.0001 * Q + 0.0004 * N + 0.7915 * r - 0.5765$ |
| 19     | $AE_{20} = 0.0001 * Q + 0.0004 * N + 0.7915 * r - 0.5765$ |
| 20     | $AE_{20} = 0.0001 * Q + 0.0004 * N + 0.7915 * r - 0.5765$ |

Figure 7 | Agreement plot of observed and predicted aeration efficiency at Parshall and modified Venturi flumes by the M5P_pruned and unpruned models using training and testing datasets, respectively.

Table 5 | Basic functions and related coefficients of the MARS model

| Basic Function | $\beta_m$ |
|----------------|-----------|
| BF1 = max(0, $r - 1.1776$) | 0.678484169449607 |
| BF2 = max(0, 1.1776 -$r$) | -0.99086139783535 |
| BF3 = max(0, 1.0783) | -0.145667947631348 |
| BF4 = BF3*max(0, W-.5) | 0.0103736565914279 |
| BF5 = BF3*max(0, 7.5 -W) | -0.0113773669860508 |
| BF6 = BF3*max(0, 0.9192-f) | 0.06266243416642 |
Results of GMDH based model

The development of GMDH design entails a trial-and-error process that is similar to RF, M5P, and MARS-based models. Table 2 outlines the outcomes of the developed GMDH model. Three hidden layers are included in GMDH developed model. For the development of the GMDH model, the same dataset (training and testing) was used for the development of other soft computing-based models. Table 6 enlists outcomes of constants

### Table 6 | Results of adjusting the parameters of the transfer function of the GMDH model

| Layers | neurons | a0     | a1     | a2     | a3     | a4     | a5     |
|--------|---------|--------|--------|--------|--------|--------|--------|
| 1      | 1       | -1.5473| 0.0024 | 2.0871 | 0.0000 | -0.5345| -0.0029|
| 2      | 13.6203 | -0.4349| -29.9403| -0.5146| 14.6721| 2.6730 |
| 3      | -1.5211 | 0.0006 | 2.0254 | 0.0000 | -0.5023| -0.0006|
| 4      | -1.4979 | 0.0007 | 2.0146 | 0.0000 | -0.5132| -0.0005|
| 5      | -1.5142 | 0.0020 | 2.0188 | 0.0000 | -0.5042| -0.0015|
| 2      | -0.0002 | 0.3328 | 0.6698 | -7.6643| -8.7161| 16.3742|
| 3      | 0.0001  | -0.3447| 1.3467 | -439.9819| -446.5554| 886.5322|
| 4      | -0.0004 | -0.0378| 1.0426 | 49.8484 | 42.5080| -92.4401|
| 5      | -0.0004 | 0.0463 | 0.9615 | -835.4749| -841.8852| 1677.3435|
| 3      | 0.0001  | 1.1741 | -0.1725| -186.4012| -183.7293| 370.1263|

### Figure 8 | Agreement plot of observed and predicted aeration efficiency at Parshall and Modified Venturi flumes by the MARS model using training and testing datasets, respectively.

### Figure 9 | Agreement plot of observed and predicted aeration efficiency at Parshall and Modified Venturi flumes by the GMDH model using training and testing datasets, respectively.
and coefficients of transfer functions. Figure 9 illustrates the results of the GMDH model during the training and testing phases. Assessment parameters of performance indicates that the GMDH was appropriate for the prediction of aeration efficacy at Parshall and modified Venturi flumes with CC = 0.9999, 0.9958, RMSE = 0.0009, 0.0189, Bias = 0.0000, 0.0029, SI = 0.0080, 0.1540, MAE = 0.0005, 0.0033 and NSE = 0.9999, 0.9649 for model development and validation periods, individually.

Figure 10 | Performance of various applied models for the prediction of aeration efficiency at Parshall and Modified Venturi flumes during the testing stage.
Intercomparison of regression and soft computing-based models

Contrasting soft computing-based designs (Table 2 and Figure 10) indicates that the MARS-based model works better than other regression and soft computing-based models. The potential of regression and soft computing-based models for predicting the aeration efficiency at Parshall and Modified Venturi flumes was assessed through performance indicators, agreement plots, and error as plotted in Figure 10 for the testing stage. The predicted values produced by the MARS-based model were found to be near the observed aeration efficiency at Parshall and Modified Venturi flumes as indicated in the plots. Also, the predicted aeration efficiency at Parshall and modified Venturi flumes values are noted to follow the analogous pattern as observed in actual aeration efficiency at Parshall and modified Venturi flumes. Unpruned M5P based model outperforms pruned M5P based model. Overall, the MARS-based model performance was more accurate than M5P, GMDH, MLR, MNLR, and RF-based models. MLR based model is also suitable and better than MNLR for predicting the aeration efficiency at Parshall and modified Venturi flumes using this dataset. Also, the box plot is plotted in Figure 11 for the comparison of observed (actual) and predicted values using various applied designs for the testing stage. Descriptive statistics of observed and applied models during the testing stage are enlisted in Table 7. Figure 11 and Table 7 suggest that the MARS model was outperforming in comparison to other applied models. Minimum and maximum values of actual and predicted values using the MARS model are very close. The widths of the lower and upper Quartile are almost the same in Figure 11.

Taylor diagram is a graphical illustration of the performance of developed models in terms of CC, RMSE, and the standard deviation is shown in Figure 12, which indicates that the MARS was the best performing model and the performance of the MNLR model was least for the prediction of aeration efficiency at Parshall and Modified Venturi flumes using this dataset.

Sensitivity analysis

Analysis of sensitivity was carried out to find the most influential input variable in AE20 at Parshall and modified Venturi flumes. The analysis was done by utilizing the superlative performing model (MARS). The set of training

![Box plot of observed and applied models during the testing stage.](http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2021.161/895027/ws2021161.pdf)

**Table 7** | Descriptive statistics of observed and applied models during testing stage

| Statistic | Observed | MLR | MNLR | RF | M5P_pruned | M5P_unpruned | MARS | GMDH |
|-----------|----------|-----|------|----|-------------|--------------|------|------|
| Minimum   | 0.0389   | 0.0586 | 0.0678 | 0.0470 | 0.0450 | 0.0450 | 0.0388 | 0.0401 |
| Maximum   | 0.4261   | 0.4461 | 0.3779 | 0.3020 | 0.4300 | 0.4250 | 0.4146 | 0.4997 |
| 1st Quartile | 0.0701   | 0.0816 | 0.0871 | 0.0753 | 0.0733 | 0.0733 | 0.0705 | 0.0703 |
| Median    | 0.1070   | 0.1054 | 0.1043 | 0.1065 | 0.1060 | 0.1055 | 0.1065 | 0.1065 |
| Mean      | 0.1225   | 0.1206 | 0.1136 | 0.1159 | 0.1211 | 0.1212 | 0.1208 | 0.1254 |
| 3rd Quartile | 0.1450   | 0.1378 | 0.1211 | 0.1475 | 0.1430 | 0.1430 | 0.1459 | 0.1452 |
data was used, fluctuated as it was developed after eliminating one input variable at a time. Results were listed in terms of coefficient of correlation, RMSE, MAE, and NSE in Table 8 which indicates that the oxygen deficit ratio (r) is the most effective input variable in predicting the AE20 at Parshall and Modified Venturi flumes using this dataset.

CONCLUSION

In this research, the M5P, Random Forest (RF), Multivariate Adaptive Regression Splines (MARS), and Group method of data handling (GMDH) have been developed to predict aeration efficiency (AE20) at Parshall and Modified Venturi flumes and compared with multiple linear regression (MLR) and multiple nonlinear regression (MNLR) models. This was done by conducting experiments for 26 different Modified Venturi flumes and one Parshall flume. The comparison analysis using performance evaluation indices concludes that the MARS approach outperformed the rest of the models (i.e., M5P, RF, GMDH, MLR, and MLR) during development (training) and validation (testing) periods, separately. Other major outcomes from this study are that the MLR based model performs better than MNLR based model. In M5P based models, the unpruned model works better than the pruned model. Overall, the M5P based models outperform GMDH, MLR, RF, and MNLR based models. By utilizing the anticipated best performing MARS design, sensitivity analysis was carried out to evaluate each input variable effect on AE20. The finding of the sensitivity analysis indicates that the oxygen deficit ratio (r) was the most effective input variable for estimating the AE20 at Parshall and Modified Venturi flumes using this dataset.

CONFLICT OF INTEREST

None.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.
REFERENCES

Aghelpour, P., Mohammadi, B. & Biazar, S. M. 2019 Long-term monthly average temperature forecasting in some climate types of Iran, using the models SARIMA, SVR, and SVR-FA. *Theor. Appl. Climatol.* https://doi.org/10.1007/s00704-019-02905-w.

Aliafi, H., Mohammadian, A. & Bonakdari, H. 2020 Experimental investigation and model development of geometric characteristics of negatively buoyant jets inclined at 15° and 32° using GMDH method. *J. Coast. Res.* https://doi.org/10.2112/COASTRES-D-19-00105.1.

Banadkooki, F. B., Ehteram, M., Panahi, F. et al. 2020 Estimation of total dissolved solids (TDS) using new hybrid machine learning models. *J. Hydrol.* 587, 124989. https://doi.org/10.1016/j.jhydrol.2020.124989.

Baylar, A. & Bagatur, T. 2000 Aeration performance of weirs. *Water SA.*

Baylar, A. & Bagatur, T. 2006 Experimental studies on air entrainment and oxygen content downstream of sharp-crested weirs. *Water Environ. J.* https://doi.org/10.1111/j.1747-6595.2005.00002.x.

Baylar, A. & Emiroglu, M. E. 2003 Study of aeration efficiency at stepped channels. *Proc. Inst. Civ. Eng. Water Marit. Eng.* https://doi.org/10.1680/wame.2003.156.3.257.

Baylar, A., Emiroglu, M. E. & Bagatur, T. 2006 An experimental investigation of aeration performance in stepped spillways. *Water Environ. J.* https://doi.org/10.1111/j.1747-6595.2005.00099.x.

Baylar, A., Hanbay, D. & Ozpolat, E. 2008 An expert system for predicting aeration performance of weirs by using ANFIS. *Expert Syst. Appl.* https://doi.org/10.1016/j.eswa.2007.08.019.

Baylar, A., Aydin, M. C., Unsal, M. & Ozkan, F. 2009a Numerical modeling of venturi flows for determining air injection rates using fluent V6.2. *Math. Comput. Appl.* https://doi.org/10.3390/mca14020099.

Baylar, A., Hanbay, D. & Batan, M. 2009b Application of least square support vector machines in the prediction of aeration performance of plunging overfall jets from weirs. *Expert Syst. Appl.* https://doi.org/10.1016/j.eswa.2008.10.061.

Baylar, A., Kisi, O. & Emiroglu, M. E. 2009c Modeling air entrainment rate and aeration efficiency of weirs using ANN approach. *Gazi. Univ. J. Sci.*

Breiman, L. 1996 Bagging predictors. *Mach. Learn.* 24, 123–140.

Breiman, L. 2001 Random forests. *Mach. Learn.* https://doi.org/10.2307/2424575.

Chanson, H. 1995 *Air Bubble Entrainment in Free-Surface Turbulent Flows: Experimental Investigations.* CH46/95 Dep Civ Eng Univ Queensl.

Dursun, O. F. 2016 An experimental investigation of the aeration performance of parshall flume and venturi flumes. *KSCE J. Civ. Eng.* https://doi.org/10.1007/s12205-015-0645-0.

Ebrahimi, I., Sattar Ahmed, M. A., Bonakdari, H. & Zají, A. H. 2017 Prediction of scour depth around bridge piers using self-adaptive extreme learning machine. *J. Hydroinform.* https://doi.org/10.1080/14635144.2016.1211465.

Elbeltagi, A., Deng, J., Wang, K. et al. 2020 Modeling long-term dynamics of crop evapotranspiration using deep learning in a semi-arid environment. *Agric. Water Manag.* 241, 106334. https://doi.org/10.1016/j.agwat.2020.106334.

Ervine, D. A. 1998 Air entrainment in hydraulic structures: a review. *Proc. Inst. Civ. Eng. – Water Marit. Energy* 130, 142–153. https://doi.org/10.1080/14635144.1998.10581617.

Esmaeilbeiki, F., Nikpour, M. R., Singh, V. K. et al. 2020 Exploring the application of soft computing techniques for spatial evaluation of groundwater quality variables. *J. Clean. Prod.* https://doi.org/10.1016/j.jclepro.2020.124206.

Ghasempour, R., Roushangar, K. & Sihag, P. 2021 Suspended sediment load prediction in consecutive stations of river based on ensemble pre-post-processing kernel based approaches. *Water Supply* https://doi.org/10.2166/wss.2021.094.

Gulliver, J. S., Thene, J. R. & Rindels, A. J. 1990 Indexing gas transfer in self-aerated flows. *J. Environ. Eng. https://doi.org/10.1061/(ASCE)0733-9372(1990)116:3(503).*

Haghiabi, A. H., Nasrolahi, A. H. & Parsaie, A. 2018 Water quality prediction using machine learning methods. *Water Qual. Res. J. Canada.* https://doi.org/10.2166/wqrj.2018.025.

Hanbay, D., Baylar, A. & Batan, M. 2009 Prediction of aeration efficiency on stepped cascades by using least square support vector machines. *Expert Syst. Appl.* https://doi.org/10.1016/j.eswa.2008.03.003.

Heddam, S. & Kisi, O. 2017 Extreme learning machines: a new approach for modeling dissolved oxygen (DO) concentration with and without water quality variables as predictors. *Environ. Sci. Pollut. Res.* https://doi.org/10.1007/s11356-017-9283-z.

Ivakhnenko, A. G. 1971 Polynomial theory of complex systems. *IEEE Trans. Syst. Man. Cybern.* https://doi.org/10.1109/TSMC.1971.1430830.

Jahani, B. & Mohammadi, B. 2019 A comparison between the application of empirical and ANN methods for estimation of daily global solar radiation in Iran. *Theor. Appl. Climatol.* https://doi.org/10.1007/s00704-018-2666-3.

Kim, S. D., Lee, H. J. & Oh, B. D. 2010 Investigation on application of parshall flume for flow measurement of low-flow season in Korea. *Meas. Sci. Rev.* https://doi.org/10.2478/v10048-010-0019-1.

Malik, A., Kumar, A. & Piri, J. 2017 Daily suspended sediment concentration simulation using hydrological data of Pranhita River Basin, India. *Comput. Electron. Agric.* 138, 20–28. https://doi.org/10.1016/j.compag.2017.04.005.

Malik, A., Kumar, A., Ghorbani, M. A. et al. 2019a The viability of co-active fuzzy inference system model for monthly reference evapotranspiration estimation: case study of Uttarakhand State. *Hydrol. Res.* 50, 1623–1644. https://doi.org/10.2166/hnr.2019.059.

Malik, A., Kumar, A., Kisi, O. & Shiri, J. 2019b Evaluating the performance of four different heuristic approaches with Gamma test for daily suspended sediment concentration modeling. *Environ. Sci. Pollut. Res.* 26, 22670–22687. https://doi.org/10.1007/s11356-019-05553-9.
Malik, A., Rai, P., Heddam, S. et al. 2020a Pan evaporation estimation in Uttarakhand and Uttar Pradesh States, India: validity of an integrative data intelligence model. Atmosphere (Basel) 11, 1–26. https://doi.org/10.3390/atmos11060535.

Malik, A., Tikhamarine, Y., Souag-Gamane, D. et al. 2020b Support vector regression optimized by meta-heuristic algorithms for daily streamflow prediction. Stoch. Environ. Res. Risk Assess. 34, 1755–1773. https://doi.org/10.1007/s00477-020-02187-1.

Malik, A., Kumar, A., Rai, P. & Kurigi, A. 2021a Prediction of multi-scalar standardized precipitation index by using artificial intelligence and regression models. Climate 9, 28. https://doi.org/10.3390/cli9020028.

Malik, A., Tikhamarine, Y., Sammen, S. S. et al. 2021b Prediction of meteorological drought by using hybrid support vector regression optimized with HHO versus PSO algorithms. Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-021-13445-0.

Malik, A., Tikhamarine, Y., Souag-Gamane, D. et al. 2021c Support vector regression integrated with novel meta-heuristic algorithms for meteorological drought prediction. Meteorol. Atmos. Phys. https://doi.org/10.1007/s00703-021-00787-0.

Mohammadi, B. & Mehdizadeh, S. 2020 Modeling daily reference evapotranspiration via a novel approach based on support vector regression coupled with whale optimization algorithm. Agric. Water Manag. 237, 106145. https://doi.org/10.1016/j.agwat.2020.106145.

Mohammadi, B., Guan, Y., Aghelpour, P. et al. 2020 Simulation of Titicaca lake water level fluctuations using hybrid machine learning technique integrated with grey wolf optimizer algorithm. Water 12, 3015. https://doi.org/10.3390/w12113015.

Mohammadi, B., Guan, Y., Moazenzadeh, R. & Safari, M. J. S. 2021 Implementation of hybrid particle swarm optimization-differential evolution algorithms coupled with multi-layer perceptron for suspended sediment load estimation. Catena 198, 105024. https://doi.org/10.1016/j.catena.2020.105024.

Najafzadeh, M. & Lim, S. Y. 2015 Application of improved neuro-fuzzy GMDH to predict scour depth at sluice gates. Earth Sci. Inform. https://doi.org/10.1007/s12145-014-0144-8.

Najafzadeh, M., Tafarorjnoruz, A. & Lim, S. Y. 2017 Prediction of local scour depth downstream of sluice gates using data-driven models. ISH J. Hydraul. Eng. 23, 195–202. https://doi.org/10.1080/09715010.2017.1286614.

Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models part I – A discussion of principles. J. Hydrol. 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.

Pal, M. & Deswal, S. 2009 M5 model tree based modelling of reference evapotranspiration. Hydrol. Process. https://doi.org/10.1002/hyp.7266.

Pandhiani, S. M., Sihag, P., Shabri, A. B. et al. 2020 Time-series prediction of streamflows of Malaysian rivers using data-driven techniques. J. Irrig. Drain. Eng. https://doi.org/10.1061/(asce)ir.1943-4774.0001463.

Parsaie, A. 2016 Predictive modeling of the side weir discharge coefficient using neural network. Model. Earth Syst. Environ. https://doi.org/10.1007/s40808-016-0125-9.

Parsaie, A. & Haghiaabi, A. H. 2017a Numerical routing of tracer concentrations in rivers with stagnant zones. Water Sci. Technol. Water Supply. https://doi.org/10.21666/ws.2016.175.

Parsaie, A. & Haghiaabi, A. H. 2017b Mathematical expression of discharge capacity of compound open channels using MARS technique. J. Earth Syst. Sci. https://doi.org/10.1007/s12040-017-0807-1.

Parsaie, A. & Haghiaabi, A. H. 2017c Improving modelling of discharge coefficient of triangular labyrinth lateral weirs using SVM, GMDH and MARS techniques. Irrig. Drain. https://doi.org/10.1002/ird.2125.

Parsaie, A. & Haghiaabi, A. H. 2019 Inception point of flow aeration on quarter-circular crested stepped spillway. Flow Meas. Instrum. https://doi.org/10.1016/j.flowmeasinst.2019.101618.

Parsaie, A., Najaﬁan, S. & Shamsi, Z. 2016 Predictive modeling of discharge of flow in compound open channel using radial basis neural network. Model. Earth Syst. Environ. https://doi.org/10.1007/s40808-016-0207-6.

Parsaie, A., Haghiaabi, A. H., Saniec, M. & Torabi, H. 2018 Prediction of energy dissipation of flow over stepped spillways using data-driven models.Iran. J. Sci. Technol. Trans. Civ. Eng. 42, 39–53. https://doi.org/10.1007/s40996-017-0060-5.

Parsaie, A., Haghiaabi, A. H. & Moradinejad, A. 2019 Prediction of scour depth below river pipeline using support vector machine. KSCE J. Civ. Eng. https://doi.org/10.12205/19-1327.

Parshall, R. L. 1926 The improved venturi. Trans. Am. Soc. Civ. Eng. https://doi.org/10.1061/taceat.0003626.

Quinlan, J. R. 1992 Learning with continuous classes. In: Machine Learning, pp. 343–348.

Sánchez, E., Colmenarejo, M. F., Vicente, J. et al. 2007 Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution. Ecol. Indic. https://doi.org/10.1016/j.ecolind.2006.02.005.

Sangeeta, S. R. & Tiwari, N. K. 2019 Aeration efficiency evaluation of modified small Parshall flume using M5P and adaptive neuro-fuzzy. Sustain. Eng. Proc. EGRWSE 30, 243–252.

Schwarzenbach, R. P., Egli, T., Hofstetter, T. B. et al. 2010 Global water pollution and human health. Annu. Rev. Environ. Resour. https://doi.org/10.1146/annurev-environ-100809-125342.

Sihag, P., Tiwari, N. K. & Ranjan, S. 2017 Estimation and inter-comparison of infiltration models. Water Sci. https://doi.org/10.1016/j.wsj.2017.03.001.

Sihag, P., Esmaeilbeiki, F., Singh, B. et al. 2019a Modeling unsaturated hydraulic conductivity by hybrid soft computing techniques. Soft Comput. https://doi.org/10.1007/s00500-019-03847-1.

Sihag, P., Esmaeilbeiki, F., Singh, B. & Pandhiani, S. M. 2019b Model-based soil temperature estimation using climatic parameters: the case of Azerbaijan Province, Iran. Geol. Ecol. Landsc. https://doi.org/10.1080/24749508.2019.1610841.
Sihag, P., Singh, V. P., Angelaki, A. et al. 2019c Modelling of infiltration using artificial intelligence techniques in semi-arid Iran. *Hydrol. Sci. J.* https://doi.org/10.1080/02626667.2019.1659965.

Sihag, P., Suthar, M. & Mohanty, S. 2019d Estimation of UCS-FT of dispersive soil stabilized with fly ash, cement clinker and GGBS by artificial intelligence. *Iran. J. Sci. Technol. – Trans. Civ. Eng.* https://doi.org/10.1007/s40996-019-00329-0.

Sihag, P., Kumar, M. & Sammen, S. S. 2021 Predicting the infiltration characteristics for semi-arid regions using regression trees. *Water Supply.* https://doi.org/10.2166/ws.2021.047.

Singh, B., Sihag, P., Pandhiani, S. M. et al. 2019 Estimation of permeability of soil using easy measured soil parameters: assessing the artificial intelligence-based models. *ISH J. Hydraul. Eng.* 1–11. https://doi.org/10.1080/09715010.2019.1574615.

Tao, H., Diop, L., Bodian, A. et al. 2018 Reference evapotranspiration prediction using hybridized fuzzy model with firefly algorithm: Regional case study in Burkina Faso. *Agric. Water Manag.* 208, 140–151. https://doi.org/10.1016/j.agwat.2018.06.018.

Thakur, M. S., Pandhiani, S. M., Kashyap, V. et al. 2021 Predicting bond strength of FRP bars in concrete using soft computing techniques. *Arab. J. Sci. Eng.* https://doi.org/10.1007/s13369-020-05314-8.

Tikhamarine, Y., Malik, A., Pandey, K. et al. 2020a Monthly evapotranspiration estimation using optimal climatic parameters: efficacy of hybrid support vector regression integrated with whale optimization algorithm. *Environ. Monit. Assess.* https://doi.org/10.1007/s10661-020-08659-7.

Tikhamarine, Y., Malik, A., Souag-Gamane, D. & Kisi, O. 2020b Artificial intelligence models versus empirical equations for modeling monthly reference evapotranspiration. *Environ. Sci. Pollut. Res.* 27, 30011–30019. https://doi.org/10.1007/s11356-020-08792-3.

Tikhamarine, Y., Souag-Gamane, D., Najah Ahmed, A. et al. 2020c Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey Wolf optimization (GWO) algorithm. *J. Hydrol.* 582, 124435. https://doi.org/10.1016/j.jhydrol.2019.124435.

Tsang, C. 1987 *Hydraulic and Aeration Performance of Labyrinth Weirs.* Dr Diss Univ London.

USBR 2001 *Water Measurement Manual: A Guide to Effective Water Measurement Practices for Better Water Management,* 3rd edn. rev Repr US Dep Inter Bur Reclam, Washington, DC.

Wilhelms, S. C., Gulliver, J. S. & Parkhill, K. 1993 Reaeration at low-head hydraulic structures. *Hydraul. Lab US Army Corps. Eng. Water Exp. Sta.* Vicksbg, MS Techn.

Received 9 April 2021; accepted in revised form 17 May 2021. Available online 3 June 2021