scientific reports

Check for updates

OPEN

LXGB: a machine learning algorithm for estimating the discharge coefficient of pseudo-cosine labyrinth weir

Somayeh Emami[1]✉, Hojjat Emami[2] & Javad Parsa[3]

One of the practical and financial solutions to increase the efficiency of weirs is to modify the geometry of the plan and increase the length of the weir to a specific width. This increases the discharge coefficient (C_d) of the weir. In this study, a new weir referred to pseudo-cosine labyrinth weir (PCLW) was introduced. A hybrid machine learning LXGB algorithm was introduced to estimate the C_d of the PCLW. The LXGB is a combination of the linear population size reduction history-based adaptive differential evolution (LSHADE) and extreme gradient boosting (XGB) algorithm. Seven different input scenarios were presented to estimate the discharge coefficient of the PCLW weir. To train and test the proposed method, 132 data series, including geometric and hydraulic parameters from PCLW1 and PCLW2 models were used. The root mean square error (*RMSE*), relative root mean square error (*RRMSE*), and Nash–Sutcliffe model efficiency coefficient (*NSE*) indices were used to evaluate the proposed approach. The results showed that the input variables were the ratio of the radius to the weir height (*R/W*), the ratio of the length of the weir to the weir height (*L/W*), and the ratio of the hydraulic head to the weir height (*H/W*), with the average values of *RMSE* = 0.009, *RRMSE* = 0.010, and *NSE* = 0.977 provided better results in estimating the C_d of PCLW1 and PCLW2 models. The improvement compared to SAELM, ANFIS-FFA, GEP, and ANN in terms of R^2 is 2.06%, 3.09%, 1.03%, and 5.15%. In general, intelligent hybrid approaches can be introduced as the most suitable method for estimating the C_d of PCLW weirs.

One of the main concerns of hydraulic engineers is the optimal management of limited water resources, in Iran. The ever-increasing growth of national investment in water projects leads to the optimization of water control and management projects in order to save national capital[1–3]. In recent years, hydraulic engineers have tried to measure the discharge with proper accuracy by building and installing measuring structures in the channels. One of the common structures in many dams and water transfer channels are labyrinth weirs, which are used for draining, measuring, and controlling the water level[4,5]. These types of weirs are among the most practical surface structures, which have recently attracted the attention of various researchers. The pseudo-cosine labyrinth weirs (PCLW) with a long crown have a suitable performance for regulating the water level compared to other weirs. Numerous parameters are effective in determining the C_d in labyrinth weir with different plans. These parameters are related to several factors, including the upstream total hydraulic head (H_u), downstream hydraulic head (H_d), weir height (W), radius (R), number of cycles (N), shape of the weir crest (CR), collision of nape (Na), the approach flow conditions (AF), etc.[4]. Nowadays, several issues, including the increase in costs, time-consuming, and the occurrence of human error, have led to the use of 3D and computer models[6,7]. Since manual calculations may involve human error, it is necessary to use novel intelligent methods such as meta-heuristic algorithms, artificial neural networks, fuzzy logic, etc. Several studies have been carried out by researchers in the investigation of the C_d of labyrinth weirs[8–15]. Considering some structural limitations (such as structure dimensions and weir angle) and using classical calculation methods such as linear and non-linear regression methods, the researchers have determined the C_d of weirs.

Azamathulla and Wu[16] used the support vector machine (SVM) to accurately estimate the longitudinal dispersion coefficients in natural rivers. With a test on real-world datasets, the SVM algorithm is proven to generate encouraging results. In another work, Azamathulla et al.[17] proposed SVM to estimate the C_d in side weirs. The

[1]Department of Water Engineering, University of Tabriz, Tabriz 5971982284, Iran. [2]Department of Computer Engineering, University of Bonab, Bonab, Iran. [3]Department of Water Engineering, University of Tabriz, Tabriz, Iran. ✉email: somayehemami70@gmail.com

experimental results proved the superiority of the SVM compared with counterpart adaptive neuro-fuzzy inference systems (ANFIS) and artificial neural networks (ANNs). Bilhan et al.[18] estimate the C_d of labyrinth weirs using support vector regression (SVR) and an outlier robust extreme learning machine. The results showed that machine learning methods estimated the C_d values more accurately. Safarrazavizadeh et al.[19] performed a laboratory investigation of the flow on labyrinth weirs with a semicircular and sinusoidal plan. Observations showed that the discharge coefficient in labyrinth weirs with a semicircular and sinusoidal plan, unlike linear weirs, has an upward trend in low water loads ($H_T/P < 0.35$) and decreases after reaching its maximum value. Bonakdari et al.[20] investigated the effectiveness of the gene expression programming (GEP) method for estimating C_d. Results show that the GEP method provides better results in predicting C_d. Shafiei et al.[21] used the ANFIS-firefly algorithm (ANFIS-FFA) method to estimate the C_d of triangular labyrinth weirs. Results showed that the ANFIS-FFA model is more accurate in predicting the C_d of triangular labyrinth weirs. Emami et al.[8] estimated the C_d of W-planform labyrinth weirs using the improved self-adaptive differential evolutionary algorithm and support vector regression (ISaDE-SVR) method. ISaDE-SVR is highly effective in estimating the C_d of W-planform weirs. Norouzi et al.[22] simulated C_d using a self-adaptive robust learning machine (SAELM) model. The results showed that the SAELM model estimated the C_d with high accuracy. Wang et al.[23], investigated the application of genetic algorithm (GA), particle swarm optimization (PSO), and traditional BP neural network in predicting the C_d of triangular labyrinth weir. The results showed that GA-BPNN and PSO-BPNN methods have high efficiency in predicting C_d. Chen et al.[24] used SVM, random forest (RF), linear regression, SVM, k-nearest neighbor (KNN), and decision tree (DT) in predicting the C_d of streamlined weirs. Ahmad et al.[25] used the ANN model to predict the C_d of an arced labyrinth side weir. The results indicated that C_d calculated by ANN is more accurate. Emami et al.[26] used the Walnut algorithm and SVR method to predict the C_d of triangular labyrinth weirs. Safari et al.[27] evaluated ANN, GEP, and regression models to estimate the C_d of the broad-crested weir. The results showed that ANN estimates the C_d better than GEP models and regression models.

In the previous studies, according to the many geometrical models that have been investigated by different researchers, the C_d of PCLW has not been investigated. Therefore, in the present study, by using the intelligent model of the differential evolution (LSHADE) and extreme gradient boosting (XGB) approach, the C_d of the PCLW was estimated. The proposed approach was investigated with different combinations of features to identify the high-performance combination of features.

The contributions of this paper are as follows:

(a) Introducing the LXGB algorithm, which integrates the LSHADE with XGB to tune the XGB parameters and further enhance its estimation performance.
(b) Using the LXGB algorithm to estimate the C_d of PCLW. The proposed algorithm models the
(c) Evaluating the proposed model with a real-world dataset and compared with state-of-the art algorithms. The experimental results show the superiority of the proposed method compared with counterparts in terms of performance measures.

The remaining sections of this study are organized as follows. Section "Material and methods" illustrates the experimental materials and the presented hybrid approach. Section "Results and Discussion" presents the results and discussions. Section "Conclusion" summarizes the paper and supplies recommendations for coming work.

Material and methods

Dimensional analysis. The 1-dimensional equation of the flow on the PCLW is as follows[28]:

$$Q = \frac{2}{3} C_d L \sqrt{2g} H_T^{1.5} \tag{1}$$

where Q is the discharge, g shows the acceleration of gravity, L is the length of the weir, and H_T is the hydraulic height ($h + V_2/2g$). The C_d of labyrinth weirs in free flow conditions depends on geometric and hydraulic parameters as follows:

$$C_d = f_1(B, L, H_T, H_d, V, W, R, S, t, \alpha, N, g, \rho, \mu, \sigma, CS, JS, SW) \tag{2}$$

where B is the channel width, H_d is the total hydraulic height (downstream of the weir), V shows the flow velocity, W indicates the height of the weir, R is the radius of weir curvature, S is the length of the straight part between the curves of the weir, t is the thickness of the weir, α represents the angle of the straight section between the weir curves with the direction of the channel, N indicates the number of cycles, ρ indicates the fluid density, μ the dynamic viscosity, σ shows the surface tension, CS means the shape of the weir crest, JS denotes the shape of the flowing blade, and SW represents the approaching flow and the sidewall effect.

Equation (2) can be written as follows:

$$C_d = f_2\left(\mathrm{Re}, We, Fr, \frac{H_T}{W}, \frac{H_d}{W}, \frac{L}{W}, \frac{B}{W}, \frac{R}{W}, \frac{S}{W}, \frac{t}{W}, \alpha, N, CS, JS, SW\right) \tag{3}$$

where Re is the Reynolds number, We mean the Weber number, and Fr is the Froude number. Henderson[29] concluded that if $Re < 2000$, the effect of viscosity can be neglected. Novak et al.[30] concluded that if the water height on the weir is more than 3 to 4 cm, the effect of surface tension is ignored. Due to the turbulent flow and minimum water height of 5 cm on the weir, the impacts of the Re and We numbers were removed. The shape of the edge of all used weirs was selected as a sharp-crested, and the effect of CS was ignored. Due to the installation

of weirs perpendicular to the main flow and the absence of local contraction at their installation location, the conditions of the approaching *SW* flow were considered the same for all experiments.

Equation (3). is simplified as the following equation:

$$C_d = f_3\left(\frac{H_T}{W}, \frac{L}{W}, \frac{B}{W}, \frac{R}{W}, \frac{S}{W}, \frac{t}{W}, \alpha, N\right) \tag{4}$$

Experimental models. The simulation of the flow around the PCLW was carried out in a channel with a width, length, and height of 0.49 m to 1.115 m, 3.2 m, and 0.5 m, respectively. In Fig. 1, the PCLW models and their geometric features are shown.

The geometric features and the range of experimental parameters of the PLCW are presented in Table 1.

Extreme gradient boosting (XGB). XGB[31–33] is a robust supervised learning solution to regression, classification, and ranking problems in a fast and accurate way. XGB is a more generalized form of gradient-boosting decision trees. It utilizes parallel processing, resolves missing values efficiently, prevents overfitting, and performs well on datasets of different sizes.

For a given dataset with n examples and m features $D = \{f(x_i, y_i)\}$ ($|D| = n$, $x_i \in R^m$, $y_i \in R$), XGB consists of an ensemble of K classification and regression trees (CARTs). The final prediction is formulated as follows[31]:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F \tag{5}$$

$\hat{y}_i$ is the final predictive value, F is the list of CARTs, and $f_k(x_i)$ is the function of input in the k-th decision tree. In the XGB, the objective function consists of two components: regularization and training error, which are defined as follows[31]:

$$X_{obj} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{6}$$

where $\sum_{i=1}^{n} l(y_i, \hat{y}_i)$ calculates the difference between the predicted value and the observed value of the loss function. $\sum_{k=1} \Omega(f_k)$ calculates the regularization component, which is:

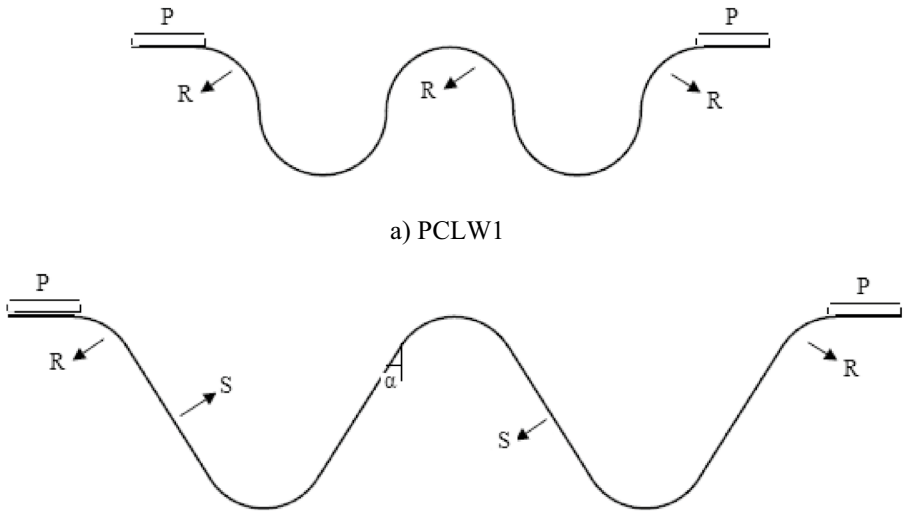


Figure 1. A big picture of PCLW1 and PLCW2.

Q (l/s)	H/W	α(°)	L(cm)	P(cm)	N	W(cm)	S(cm)	D(cm)
60–100	0.2–0.9	10	90–225	15.75	2–3	10–20	5–40	2.5–10

Table 1. The geometric features of PLCW used in simulations.

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \|w\|^2 \tag{7}$$

where γ is the leaf penalty coefficient, T is the total number of a leaf node, λ guarantees that the scores of a leaf node are not too large, and w is the scores of a leaf node. XGB employs the gradient boosting strategy, appends one new tree at each iteration, and modifies the preceding test results by fitting the residuals of the previous prediction:

$$y_i^{(K)} = \sum_{i=1}^{K} f_k(x_i) = \hat{y}_i^{K-1} + f_K(x_i) \tag{8}$$

Integrating Eq. (1) and (2), the objective function for the t-th tree can be written as[31]:

$$L^{(K)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(K-1)} + f_K(X_i)) + \Omega(f_k) \tag{9}$$

Taking the Taylor expansion of the loss function up to the second order, Eq. (9) can be approximated as follows:

$$L^{(K)} = \sum_{i=1}^{n} [l(y_i, \hat{y}_i^{(K-1)} + f_K(X_i)) + \frac{1}{2} h_i f_K^2(X_i)] + \Omega(f_k) \tag{10}$$

$$\begin{aligned} X_{obj} &= \sum_{i=1}^{n} \left[g_i f_K(x_i) + \frac{1}{2} h_i f_K^2(x_i) \right] + \Omega(f_k) \\ &= \sum_{i=1}^{n} \left[g_i w_q(x_i) + \frac{1}{2} h_i w_q^2(x_i) \right] + \Omega(f_k) + \lambda T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \\ &= \sum_{j=1}^{n} \left[\left(\sum\nolimits_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum\nolimits_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \lambda T \end{aligned} \tag{11}$$

where $g_i = \partial \hat{y}^{K-1} l(y_i, \hat{y}^{K-1})$ and $h_i = \partial^2 \hat{y}^{K-1} l(y_i, \hat{y}^{K-1})$ are the first and second-order gradient statistics of the loss function.

The optimal weight w_j of leaf j, and the objective function of a tree can be written as follows:

$$w_i = -\frac{G_i}{H_i + \lambda} \tag{12}$$

where $G_i = \sum_{i \in I_j} g_i$ and $H_i = \sum_{i \in I_j} h_i + \lambda$.

$$w_j = -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j}{H_j + \lambda} + \gamma T \tag{13}$$

the weak fitting model will be intensified as follows:

$$f_k(x_i) = f_{k-1}(x_i) + \sum_{j=1}^{T} w_j . \eta \tag{14}$$

where η is the learning rate. XGB appends new trees at each iteration by continuously dividing features. Appending a new tree to the model is learning a new function $f_k(X, \theta_k)$ to fit the residual of previous prediction. Once K trees are learned, the strong fitting model $F(x_i)$ used to predict:

$$F(x_i) = f_0(x_i) + \sum_{k=1}^{K} \sum_{j=1}^{T} w_j . \eta \tag{15}$$

where, $F(x_i)$ is the strong-fitting model.

Figure 2 shows the working principle of XGB.

Since the hyper-parameters of XGB are often set empirically, optimal tuning of parameters is essential for designing robust XGB. In this paper, we used the LSHADE algorithm to tune the XGB parameters including the number of decision trees (K), learning rate (η), maximum depth (md), minimum child weight (mcw), gamma value (γ), sub-sample (ss). Table 2 lists the XGB parameters and their range used in the implementation.

LSHADE. Success-history-based parameter adaptation for differential evolution (SHADE)[34] is an adaptive evolutionary optimization strategy. LSHADE[35] enhances SHADE with a linear population size reduction technique, which gradually reduces the size of the population using a linear function. LSHADE starts its optimiza-

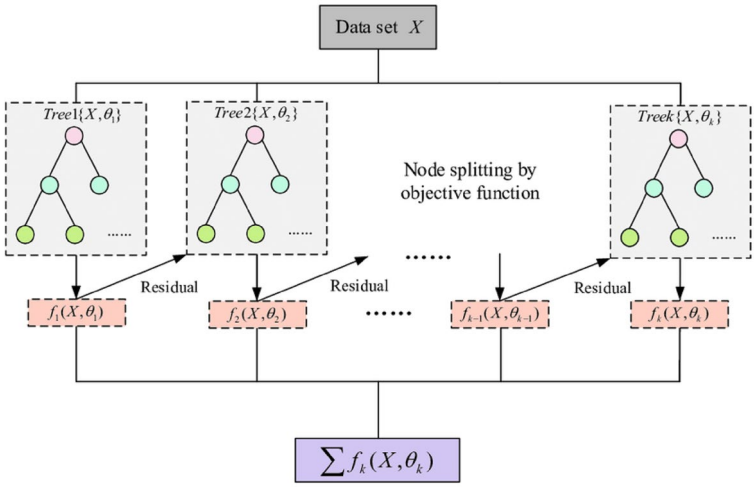


Figure 2. A big picture of the XGB method.

Parameter	Range
Number of decision trees (K)	$[1,\infty]$
Learning rate (η)	[0, 1]
Gamma (γ)	$[0,\infty]$
Sub sample (ss)	$[1,\infty]$
Maximum depth (md)	$[0,\infty]$
Minimum child weight (mcw)	$[0,\infty]$

Table 2. The parameters of the XGB algorithm that need to be tuned.

tion process with a randomly generated population of real parameter vectors. The algorithm repeats a process of trail vector generation and selection until some termination conditions are satisfied.

LSHADE-XGB (LXGB). The incentive mechanism of LXGB is to improve the classification performance of XGB by integrating the LSHADE optimization algorithm with XGB. Figure 3 shows the working principle of the LXGB algorithm.

Assessment metrics. RMSE, RRMSE and NSE metrics were used to evaluate the performance of LXGB approach (Eqs. 16–18).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i)^2} \tag{16}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(Y_i - X_i)^2}{\sum_{i=1}^{n}(Y_i - \overline{X})^2} \tag{17}$$

$$RRMSE = \frac{RMSE}{\sum_{i=1}^{n} Y_i} \tag{18}$$

RMSE: Root mean square error; NSE: Nash–Sutcliffe model efficiency coefficient; RRMSE: Relative root mean square error. where X_i is the predicted values, Y_i is the observed values, and $\overline{X}$ is the average of X.

Results and discussion

The C_d of PCLW1 and PCLW2 weirs was estimated using the hybrid LXGB approach. At first, all available data were normalized to remove or correct outliers[36].

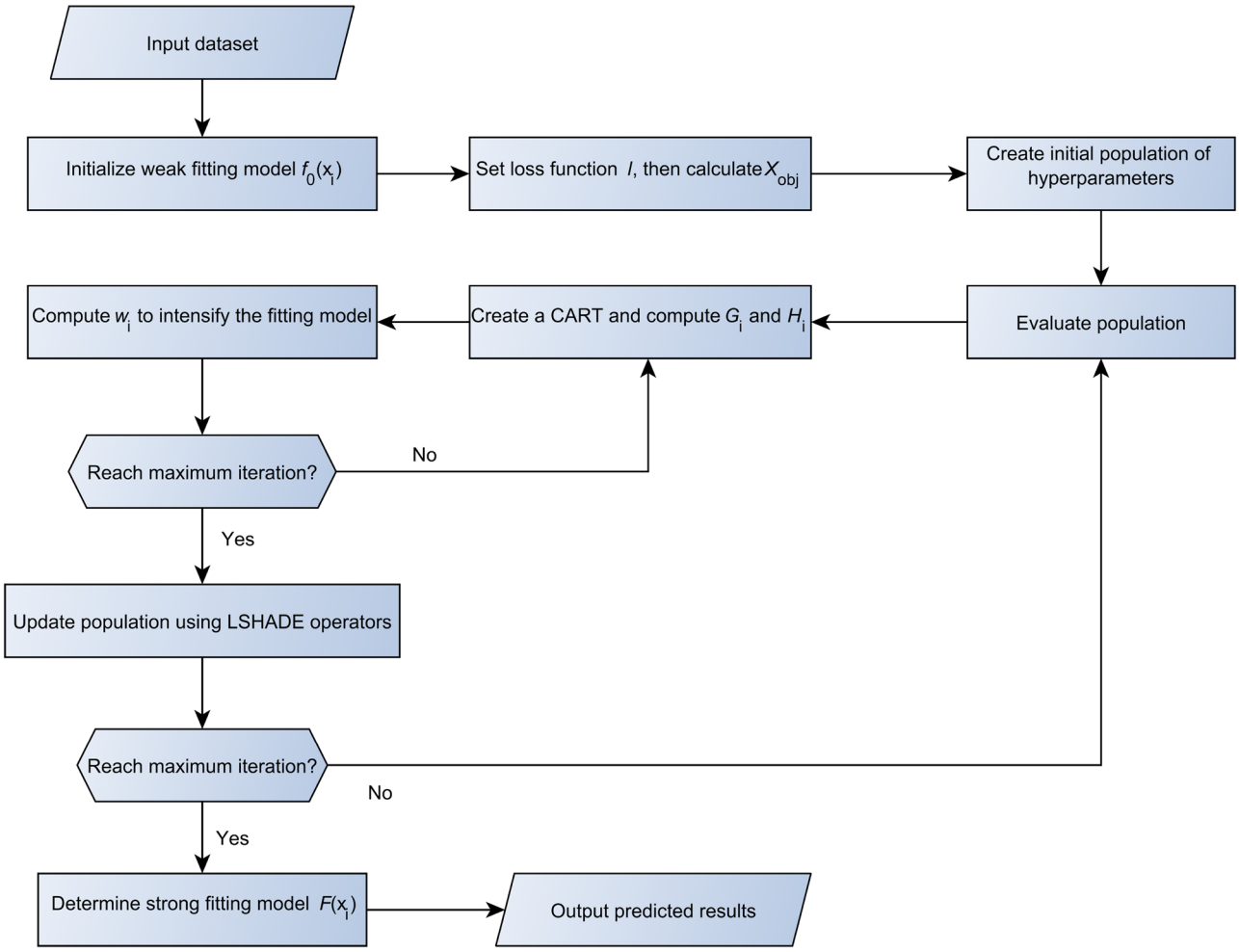


Figure 3. Principle of the LXGB algorithm.

$$X_n = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{19}$$

where X_{min} is the minimum data, X represents the raw data, X_{max} is the maximum data, and X_n is the normalized data.

The ratio of the weir length to the weir height (L/W), the ratio of the channel width to the weir height (B/W), the ratio of the weir thickness to weir height (t/W), the number of cycles (N), the radius to the weir height (R/W), the ratio of the straight section between the weir curves length to the weir height (S/W), the ratio of the, the ratio of the hydraulic head to the weir height (H/W), were considered as input parameters of the LXGB approach. 132 datasets, including geometric and hydraulic parameters, were selected. The data were randomly divided into two parts: 80% (106 data) for training the model and 20% (26 data) for testing it.

Seven models with different variables were examined to introduce the most influential input parameters in estimating the C_d of PCLW1 and PCLW2 weirs. Tables 3 and 4 and Figs. 4 and 5 present various input variables.

In Tables 5 and 6, the evaluation criteria for different input variables to estimate the C_d are presented. A part of the modeling process by the LXGB approach is presented in Fig. 6.

The results show the accuracy of the presented LXGB approach in estimating the C_d of PCLW1 and PCLW2 models of PCLW. Mahmoud et al.[37] concluded that the ANFIS-PSO and MLP-FA (multi-layer perceptron and firefly optimization algorithm) methods are the most accurate in estimating the C_d of triangular labyrinth weirs, respectively. In a similar study, Majediasl and Fuladipanah[38] concluded that the SVM model produces the most exact results in predicting the C_d of labyrinth weir with $RMSE = 0.0118$. Shafiei et al.[21] reported that the ANFIS-FFA model is quite accurate in estimating the C_d of the labyrinth weir. Karami et al.[10] showed that the ELM method with $RMSE = 0.006$ has acceptable efficiency in estimating the C_d of the labyrinth weir. In a similar study, the effectiveness of the least-squares support vector machine-bat algorithm (LSSVM-BA) method was used to investigate the discharge of a curved labyrinth weir[39]. The results of the studies showed that the SVM-based model gave accurate results in estimating the C_d of the arched labyrinth weir with values of RMSE = 0.013 and $R^2 = 0.970$[40]. Multi-layer perceptron neural network (MLPNN) managed to estimate the discharge over the triangular arced labyrinth weirs of RMSE = 0.00385 and $R^2 = 0.999$[41].

Model	Input variables
K_1	*H/W*, *L/W*, *R/W*, *B/W*, *t/w*, *N*
K_2	*H/W*, *L/W*, *R/W*, *N*
K_3	*L/W*, *R/W*, *B/W*, *t/W*, *N*
K_4	*R/W*, *B/W*, *t/W*, *N*
K_5	*B/W*, *t/W*, *N*
K_6	*H/W*, *R/W*, *L/W*
K_7	*H/W*, *N*

Table 3. Combinations of the variables of weir in the PCLW1 plan.

Model	Input variables
K_1	*H/W*, *L/W*, *R/W*, *B/W*, *S/W*, *t/w*, *N*
K_2	*H/W*, *L/W*, *R/W*, *S/W*, *N*
K_3	*L/W*, *R/W*, *B/W*, *t/W*, *S/W*, *N*
K_4	*R/W*, *B/W*, *t/W*, *N*
K_5	*B/W*, *t/W*, *N*
K_6	*H/W*, *R/W*, *L/W*
K_7	*H/W*, *N*

Table 4. Combinations of the variables of weir in the PCLW2 plan.

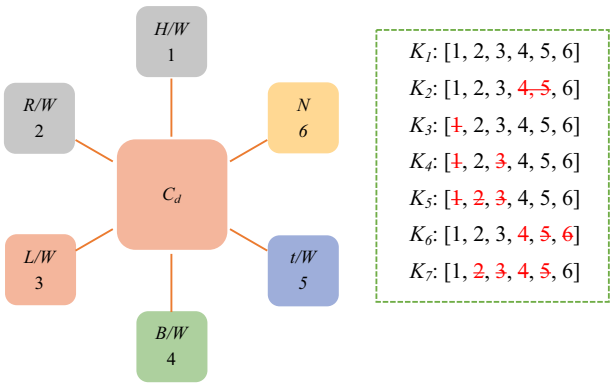


Figure 4. Combination of input variables (PCLW1 model).

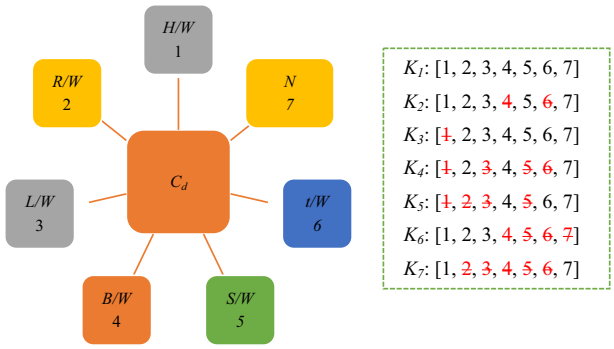


Figure 5. Combination of input variables (PCLW2 model).

Model	Train			Test		
	RMSE	RRMSE	NSE	RMSE	RRMSE	NSE
K_1	0.025	0.026	0.963	0.031	0.032	0.939
K_2	0.013	0.015	0.973	0.017	0.019	0.952
K_3	0.044	0.044	0.948	0.047	0.048	0.925
K_4	0.082	0.082	0.830	0.085	0.086	0.815
K_5	0.092	0.094	0.801	0.010	0.106	0.770
K_6	0.008	0.009	0.991	0.013	0.015	0.972
K_7	0.036	0.037	0.952	0.044	0.045	0.928

Table 5. Performance of LXGB on the PCLW1 plan with different combination models.

Model	Train			Test		
	RMSE	RRMSE	NSE	RMSE	RRMSE	NSE
K_1	0.028	0.029	0.963	0.033	0.034	0.935
K_2	0.016	0.018	0.973	0.019	0.020	0.948
K_3	0.047	0.049	0.948	0.049	0.051	0.922
K_4	0.083	0.085	0.830	0.087	0.089	0.810
K_5	0.093	0.095	0.801	0.010	0.112	0.766
K_6	0.009	0.012	0.991	0.015	0.012	0.970
K_7	0.038	0.040	0.952	0.047	0.047	0.925

Table 6. Performance of LXGB on the PCLW2 plan with different combination models.

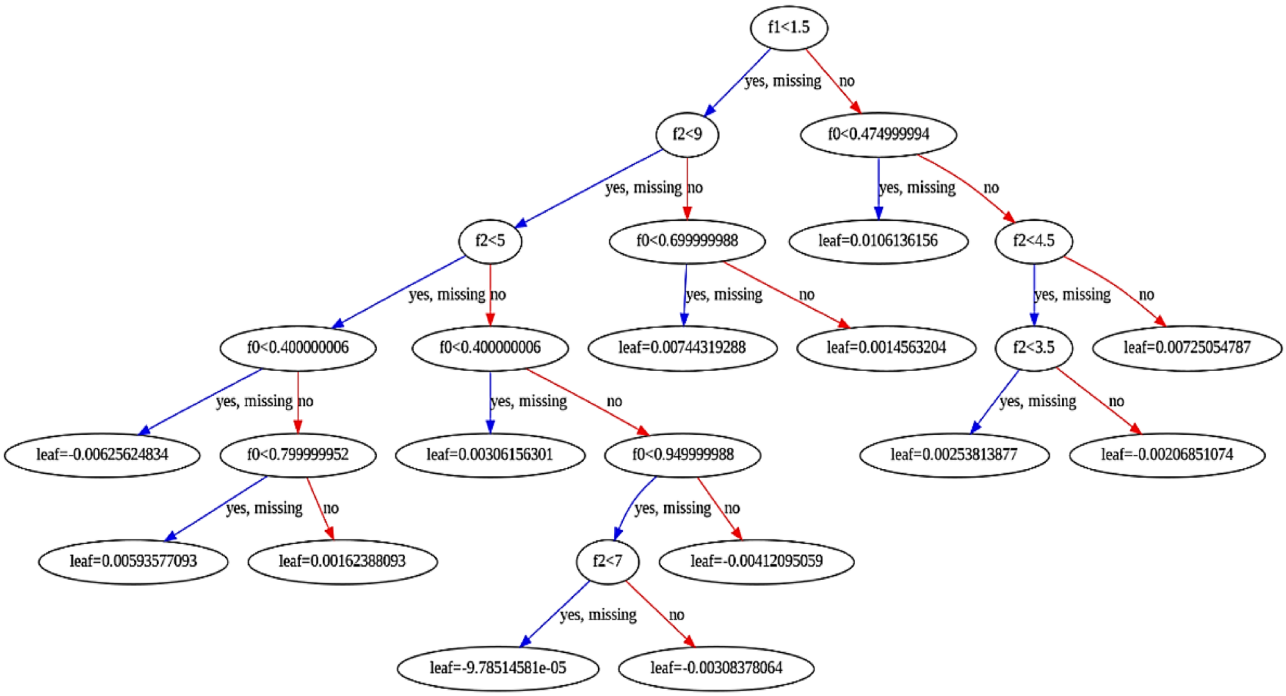


Figure 6. Structure of the tree generated by LXGB algorithm.

The results of the estimated and observed C_d of the PCLW1 and PCLW2 models of pseudo-cosine labyrinth weirs were compared in Figs. 7 and 8. According to the results, the K_6 model with the input variables of (R/W), (L/W), and (H/W), had the optimal values of statistical indicators. The C_d of PCLW1 and PCLW2 weirs increases with the increase of the weir height. In a similar study, it was concluded that with the increase in the weir height, the C_d of the triangular duckbill labyrinth weir increases, which is in agreement with the results of the present study[7]. The increase in the effective length of the labyrinths at a specified width, due to the radius increases of

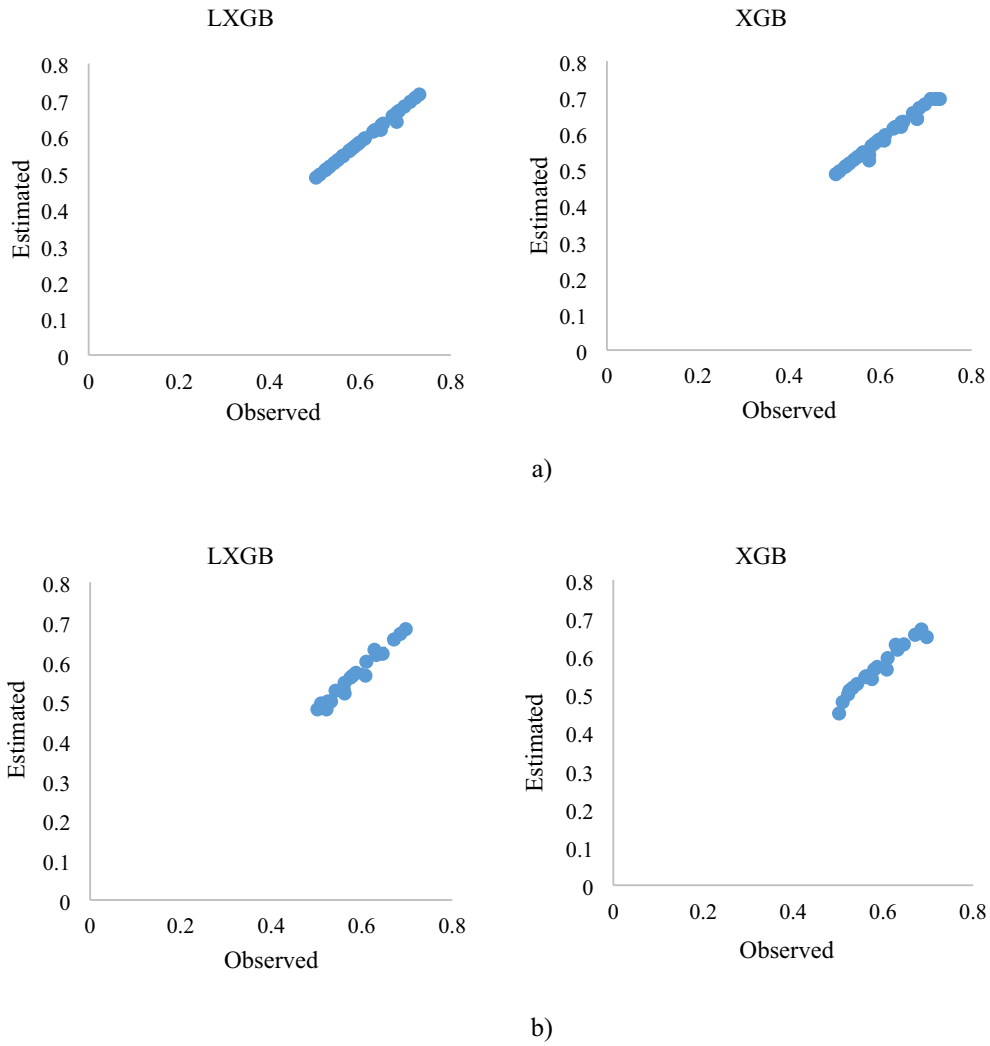


Figure 7. C_d changes of the PCLW1 (**a**) training phase, (**b**) test phase (model K_6).

PCLW1 and PCLW2 weirs causes an increase in the Cd. The studies showed that increasing the radius causes a reduction in eddy flows, turbulence, and a sudden increase in water height during the weir[39,40,42]. The results of the investigations showed that with the increase of R/W, the C_d increases in the arched labyrinth weir, which is consistent with the results of the present study[41]. Also, the K_2 model (H/W, L/W, R/W, N) is in the second rank, which shows that length, weir height, radius, and the number of cycles have a more significant impact on C_d of PCLW1 and PCLW2 weirs. By increasing the number of labyrinth weir cycles, discharge and Cd increase, which is consistent with the results of the present study[40,43]. Figure 9 shows the importance of the influential input parameters in estimating the C_d of PCLW.

Emami et al.[44] predicted the C_d of a curved plan labyrinth weirs using the WOA-ANFIS method, and the input parameters H/W and θ (weir arc angle) were introduced as the most effective parameters in estimating the C_d. Majediasal and Fuladipanah[38], investigated the support vector machine (SVM) method for C_d of sharp-crested triangular labyrinth weirs and concluded that the input combination, including geometric parameters (θ, h/w, L/B), has the best results. Mohammadi et al.[45] reported that the parameters H_t/P, W/P (the ratio of the weir width to the height), R/W, W/LC (the ratio of the weir width to the effective length) as input variables have the most accuracy and efficiency in estimating the C_d of U-shaped labyrinth weirs. Haghiabi et al.[46] indicated the C_d of triangular labyrinth weirs using the ANFIS system and concluded that the ANFIS has a proper implementation in C_d estimation. Studies showed that the H/W parameter is the most influential parameter on the C_d of a labyrinth and arced labyrinth weirs[47].

Table 7 compares the performances of the XGB and LXGB on the test dataset. The results show the superiority of the LXGB compared with the XGB algorithm in terms of performance measures. This issue proves that combining the LSHADE with XGB improves the estimation performance.

In Table 8, the values of the evaluation criteria for estimating the C_d of labyrinth weirs with different plans have been compared with the results of other studies. The results for LXGB are generated with the PCLW1 plan. The comparisons show the appropriate accuracy of the LXGB approach in estimating the C_d of labyrinth weirs with $R^2 = 0.97$ and $RMSE = 0.014$.

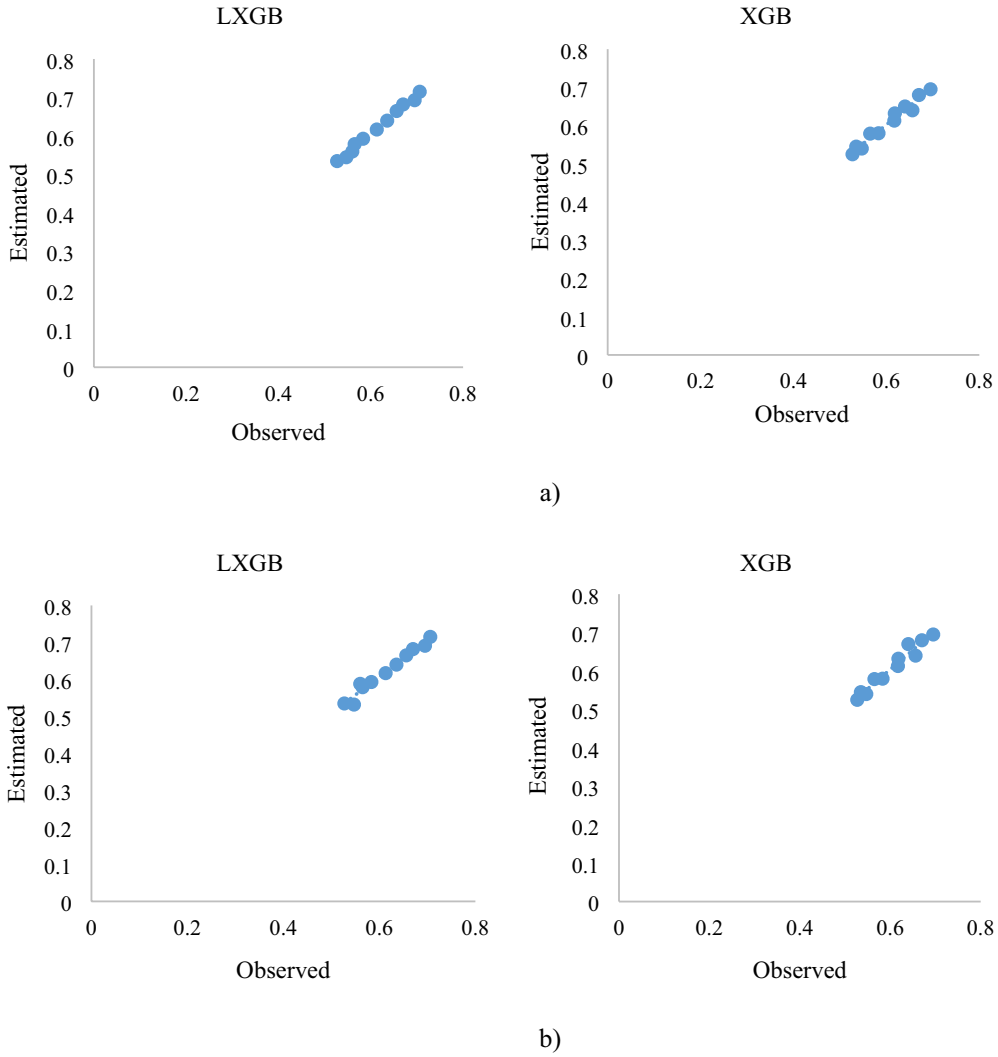


Figure 8. C_d changes of the PCLW2 (**a**) training phase, (**b**) test phase (model K_6).

Conclusion

This study introduces a novel design for labyrinth weirs called pseudo-cosine labyrinth weirs (PCLW). The LXGB was used to estimate the C_d of the PCLW weir. Seven models with different combinations of appropriate input parameters were introduced. A proper model was defined by analyzing the estimation results. The superior model estimates C_d by considering input parameters H/W, R/W, and L/W. LXGB was achieved in estimating the C_d of PCLW overflows by obtaining values of $R^2 = 0.971$, $RMSE = 0.014$, and $NSE = 0.97$. The results demonstrated that the proposed LXGB algorithm generated more significant results than previous studies in estimating the C_d of labyrinth weirs. Such a cost-effective prediction model may have significant practical application, as it can be an economical alternative to the expensive laboratory solution, which is costly and time-consuming. The proposed model is useful to correct the design of water transfer systems.

LSHADE Coding

Input:

Maximum number of iterations G, Population size N

Output:

Optimal solution,

// **Control parameters assignments based on historical memory**

Initialize algorithm parameters; G=1, $N_G = N_{init}$, archive A={}, |A|= N_{init}

Generate initial population P_G={X_1G, X_2G, …, X_NG};

Set M_CR=0.5, M_F=0.5;

while (*the stopping conditions are not satisfied*){

SCR = ∅, SF = ∅;

//Reproduction of trial vectors by using current-to-pbest/1/bin

for i = 1 to N do

ri = Select from [1,H] randomly;

$$CR_i = \begin{cases} 0 & if \ M_{CR,r_i} = \perp \\ randn_i(M_{CR,r_i}, 0.1) & else \end{cases}$$

$$F_{i,G} = randc_i\left(M_{F,r_i},\ 0.1\right);$$

Generate trial vector $u_{i,G}$ based on the current-to-pbest/1/bin;

//Survival for next generation

for i = 1 to *N* do

if $f(u_{i,G}) \le f(x_{i,G})$ then

$$x_{i,G} = u_{i,G};$$

else

$$x_{i,G+1} = x_{i,G}$$

if $f(u_{i,G}) < f(x_{i,G})$ then

$$A \leftarrow x_{i,G};$$

$$S_{CR} \leftarrow CR_{i,G},\ S_F \leftarrow F_{i,G};$$

If necessary, remove randomly selected solutions from the archive so that the archive size is |A|.

// Historical-memory update

Update memories M_{CR} and M_F;

// Linear population size reduction

Calculate N_{G+1};

if ($N_G < N_{G+1}$) then

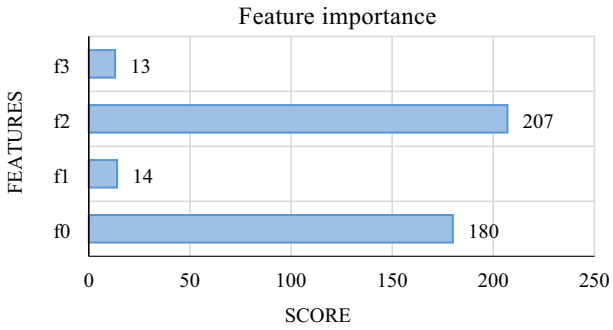
Sort individuals in *P* based on their fitness values and remove lowest ($N_G - N_{G+1}$) members;

Resize archive size |A| according to new |P|;

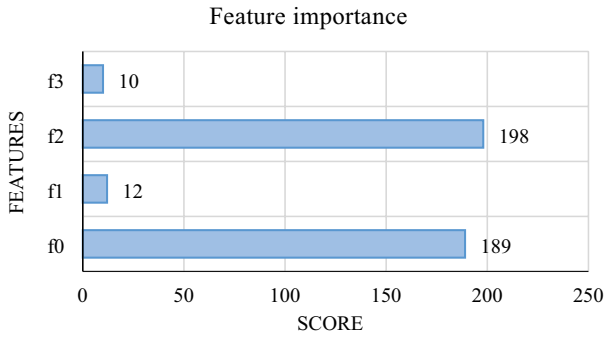
G++;

Select the best solution;

}



a) Feature importance on the PCLW1 plan



b) Feature importance on the PCLW2 plan

Figure 9. Importance of features on the $C_d$estimation on the plan (**a**) PCLW1 and (**b**) PCLW2; f_0: H/W, f_1: R/W, f_2: L/W, f_3: N.

Method	Plan	R^2	RMSE	NSE
XGB	PCLW1	0.953	0.025	0.940
LXGB	PCLW2	0.970	0.013	0.972
XGB	PCLW1	0.954	0.027	0.941
LXGB	PCLW2	0.970	0.015	0.970

Table 7. Performance evaluation of the XGB and LXGB algorithms on test dataset.

Method	R^2	RMSE	NSE
SAELM	0.950	–	0.960
ANFIS-FFA	0.940	0.039	0.901
GEP	0.967	0.017	0.780
ANN	0.920	0.045	–
LXGB	0.970	0.014	0.971

Table 8. Performance evaluation of the LXGB approach and similar methods.

Data availability

The datasets generated and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

Received: 16 May 2023; Accepted: 22 July 2023
Published online: 29 July 2023

References

1. El Bedawy, R. Water resources management: Alarming crisis for Egypt. *J. Mgmt. Sustain.* **4**, 108 (2014).

2. Zomorodian, M. *et al.* The state-of-the-art system dynamics application in integrated water resources modeling. *J. Environ. Manage.* **227**, 294–304 (2018).
3. Verma, S., Verma, M. K., Prasad, A. D., Mehta, D. J., & Islam, M. N. Modeling of uncertainty in the estimation of hydrograph components in conjunction with the SUFI-2 optimization algorithm by using multiple objective functions. *Model. Earth Syst. Environ.* 1–19 (2023).
4. Singh, D., & Kumar, M. Hydraulic design and analysis of piano key weirs: A review. *Arab. J. Sci. Eng.*. 1–15 (2021).
5. Achour, B., Amara, L. & Mehta, D. New theoretical considerations on the gradually varied flow in a triangular channel. *LARHYSS J.* **50**, 7–29 (2022).
6. Miller, D. S. (ed.) *Discharge characteristics: IAHR hydraulic structures design manuals* 8th edn. (Routledge, 2017).
7. Emami, S., Arvanaghi, H. & Parsa, J. Numerical investigation of geometric parameters effect of the labyrinth weir on the discharge coefficient. *J. Rehabil. Civil Eng.* **6**(1), 1–9 (2018).
8. Emami, S., Parsa, J., Emami, H. & Abbaspour, A. An ISaDE algorithm combined with support vector regression for estimating discharge coefficient of W-planform weirs. *Water Supply.* **21**(7), 3459–3476 (2021).
9. Roushangar, K., Alami, M. T., Majedi Asl, M. & Shiri, J. Modeling discharge coefficient of normal and inverted orientation labyrinth weirs using machine learning techniques. *ISH J. Hydraul. Eng.* **23**(3), 331–340 (2017).
10. Karami, H., Karimi, S., Bonakdari, H. & Shamshirband, S. Predicting discharge coefficient of triangular labyrinth weir using extreme learning machine, artificial neural network and genetic programming. *Neural Comput. Appl.* **29**, 983–989 (2018).
11. Bilhan, O., Emiroglu, M. E. & Kisi, O. Use of artificial neural networks for prediction of discharge coefficient of triangular labyrinth side weir in curved channels. *Adv. Eng. Softw.* **42**(4), 208–214 (2011).
12. Parsaie, A., & Haghiabi, A. H. Prediction of side weir discharge coefficient by genetic programming technique. *Jordan J. Civil Eng.* **11(1)**, (2017).
13. Emiroglu, M. E., Bilhan, O. & Kisi, O. Neural networks for estimation of discharge capacity of triangular labyrinth side-weir located on a straight channel. *Expert Syst. Appl.* **38**(1), 867–874 (2011).
14. Honar, T., Tarazkar, M. H. & Tarazkar, M. R. Estimating of side weir discharge coefficient by using Neuro-Fuzzy (ANFIS). *J. Water Soil Conserv.* **17**(2), 169–176 (2011).
15. Aghdarimoghaddam, A. & Nodoshan, J. Geometry optimization of triangle labyrinth spillway using anfis models and genetic algorithms. *J. Model. Eng.* **5**(19), 57–68 (2009).
16. Azamathulla, H. M. D. & Wu, F. Support vector machine approach for longitudinal dispersion coefficients in natural streams. *Appl. Soft Comput.* **11**(2), 2902–2905 (2011).
17. Azamathulla, H. M., Haghiabi, A. H. & Parsaie, A. Prediction of side weir discharge coefficient by support vector machine technique. *Water Sci. Technol. Water Supply.* **16**(4), 1002–1016 (2016).
18. Bilhan, O., Aydin, M. C., Emiroglu, M. E. & Miller, C. J. Experimental and CFD analysis of circular labyrinth weirs. *J. Irrig. Drain. Eng.* **144**(6), 04018007 (2018).
19. Safarrazavizadeh, M., Esmaeilivaraki, M. & Biabani, R. Experimental study on flow over sinusoidal and semicircular labyrinth weirs. *ISH J. Hydraul. Eng.* **27**(1), 304–313 (2021).
20. Bonakdari, H., Ebtehaj, I., Gharabaghi, B., Sharifi, A., & Mosavi, A. Prediction of discharge capacity of labyrinth weir with gene expression programming. In *Intelligent Systems and Applications: Proceedings of the 2020 Intelligent Systems Conference (IntelliSys)*. **1**, 202–217 (2021).
21. Shafiei, S., Najarchi, M. & Shabanlou, S. A novel approach using CFD and neuro-fuzzy-firefly algorithm in predicting labyrinth weir discharge coefficient. *J. Braz. Soc. Mech. Sci. Eng.* **42**, 1–19 (2020).
22. Norouzi, P., Rajabi, A. & Shabanlou, S. Estimation of labyrinth weir discharge coefficient using self-adaptive extreme learning machine. *J. Water Soil Sci.* **32**(1), 39–52 (2022).
23. Wang, F., Zheng, S., Ren, Y., Liu, W. & Wu, C. Application of hybrid neural network in discharge coefficient prediction of triangular labyrinth weir. *Flow Meas. Instrum.* **83**, 102108 (2022).
24. Chen, W. *et al.* Accurate discharge coefficient prediction of streamlined weirs by coupling linear regression and deep convolutional gated recurrent unit. *Eng. Appl. Comput. Fluid Mech.* **16**(1), 965–976 (2022).
25. Ahmad, F., Hussain, A. & Ansari, M. A. Development of ANN model for the prediction of discharge coefficient of an arced labyrinth side weir. *Model. Earth Syst. Environ.* **9**, 1–8 (2022).
26. Emami, H., Emami, S. & Parsa, J. A Walnut optimization algorithm applied to discharge coefficient prediction on labyrinth weirs. *Soft. Comput.* **26**(22), 12197–12215 (2022).
27. Safari, S. *et al.* Evaluation of ANN, GEP, and regression models to estimate the discharge coefficient for the rectangular broad-crested weir. *Polish J. Environ. Stud.* **31**(5), 4817 (2022).
28. Mohammadi, M. & Yasi, M. Investigation of labyrinth weir with arc plan. *J. Agric. Sci. Technol.* **11**(41), 1–12 (2007).
29. Henderson, F. M. *Open channel flow* (Macmillan Publishing, 1996).
30. Novak, P., Guinot, V., Jeffrey, A. & Reeve, D.E. Hydraulic modelling- an introduction. *Spon Press, an Imprint of Taylor & Francis, London and New York*. 599 (2010).
31. Guo, R. *et al.* Degradation state recognition of piston pump based on ICEEMDAN and XGBoost. *Appl. Sci.* **10**(18), 6593 (2020).
32. Thongsuwan, S., Jaiyen, S., Padcharoen, A. & Agarwal, P. ConvXGB: A new deep learning model for classification problems based on CNN and XGBoost. *Nucl. Eng. Technol.* **53**(2), 522–531 (2021).
33. Gu, Y., Zhang, D. & Bao, Z. A new data-driven predictor, PSO-XGBoost, used for permeability of tight sandstone reservoirs: A case study of member of chang 4+ 5, western Jiyuan Oilfield, Ordos Basin. *J. Petrol. Sci. Eng.* **199**, 108350 (2021).
34. Tanabe, R., & Fukunaga, A. Success-history based parameter adaptation for differential evolution. *IEEE congress on evolutionary computation*. 71–78 (2013).
35. Tanabe, R., & Fukunaga, A. S. Improving the search performance of SHADE using linear population size reduction. *IEEE congress on evolutionary computation (CEC)*. 1658–1665 2014.
36. Larose, D. T. & Larose, C. D. *Discovering knowledge in data: an introduction to data mining* (Wiley, 2014).
37. Mahmoud, A., Yuan, X., Kheimi, M. & Yuan, Y. Interpolation accuracy of hybrid soft computing techniques in estimating discharge capacity of triangular labyrinth weir. *IEEE Access* **9**, 6769–6785 (2021).
38. Majediasl, M. & Fuladipanah, M. Application of the evolutionary methods in determining the discharge coefficient of triangular labyrinth weirs. *JWSS-Isfahan Univ. Technol.* **22**(4), 279–290 (2019).
39. Hu, Z. *et al.* Using soft computing and machine learning algorithms to predict the discharge coefficient of curved labyrinth overflows. *Eng. Appl. Comput. Fluid Mech.* **15**(1), 1002–1015 (2021).
40. Roushangar, K., Alami, M. T., Shiri, J. & Asl, M. M. Determining discharge coefficient of labyrinth and arced labyrinth weirs using support vector machine. *Hydrol. Res.* **49**(3), 924–938 (2018).
41. Zounemat-Kermani, M., Kermani, S. G., Kiyaninejad, M. & Kisi, O. Evaluating the application of data-driven intelligent methods to estimate discharge over triangular arced labyrinth weir. *Flow Meas. Instrum.* **68**, 101573 (2019).
42. Achour, B., Amara, L. & Mehta, D. Control of the hydraulic jump by a thin-crested sill in a rectangular channel new experimental consideration. *LARHYSS J.* **50**, 31–48 (2022).
43. Zadghorban, M., Masoudian, M., Esmaeilivaraki, M. & Gharagezlou, M. Investigation of cyliderical weir roughness on hydraulic characteristics. *Iran. J. Watershed Manag. Sci. Eng.* **12**(40), 69–80 (2018).

44. Emami, S., Parsa, J. & Emami, H. Estimation of discharge coefficient of curved plan-form labyrinth weirs using a Hybrid WOA-ANFIS method. *Iran. J. Irrigat. Drainage* **14**(5), 1664–1676 (2020).
45. Mohammadi, F., Hassanzadeh, Y. & Roushangar, K. Determining the discharge coefficient of one-cycle sharp-crested u-shape weirs using kernel-based SVM approach. *Iran. J. Irrigat. Drainage* **14**(5), 1722–1736 (2020).
46. Haghiabi, A. H., Parsaie, A. & Ememgholizadeh, S. Prediction of discharge coefficient of triangular labyrinth weirs using adaptive neuro fuzzy inference system. *Alex. Eng. J.* **57**(3), 1773–1782 (2018).
47. Heydari, M., Dosti, M., & Safari, H. Optimizing the flow coefficient of trapezoidal zigzag overflows using the intelligent algorithm of gradual cooling. *10th International Seminar on River Engineering, Shahid Chamran University, Ahvaz, Iran* (2014).

Author contributions

S.E., H.E., and J.P. designed the study, analyzed and discussed the data and results. S.E. and H.E. performed the experiments and simulations. S.E. and H.E. prepared the hybrid proposed model/materials/equipment's/irrigation system. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.E.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.