Journal of
Hydroinformatics

Journal of Hydroinformatics Vol 26 No 1, 175 doi: 10.2166/hydro.2023.268

Discharge modeling and characteristic analysis of semi-circular side weir based on the soft computing method

Shanshan Li[a], Guiying Shen[a,*], Abbas Parsaie[b], Guodong Li[a] and Dingye Cao[a]
[a] State Key Laboratory of Eco-hydraulics in Northwest Arid Region of China, Xi'an University of Technology, Xi'an 710048, China
[b] Faculty of Water Sciences Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran
*Corresponding author. E-mail: 2200421250@stu.xaut.edu.cn

ABSTRACT

In this study, a support vector machine (SVM) and three optimization algorithms are used to develop a discharge coefficient (C_d) prediction model for the semi-circular side weir (SCSW). After that, we derived the input and output parameters of the model by dimensionless analysis as the ratio of the flow depth at the weir crest point upstream to the diameter (h_1/D), the ratio of main channel width to diameter (B/D), the ratio of side weir height to diameter (P/D), upstream of side weir Froude number (F_r), and C_d. The sensitivity coefficients for dimensionless parameters to C_d were calculated based on Sobol's method. The research shows that SVM and Genetic Algorithm (GA-SVM) have high prediction accuracy and generalization ability; the average error and maximum error were 0.08 and 2.47%, respectively, which were about 95.72 and 60.86% lower compared with the traditional empirical model. The first-order sensitivity coefficients S_1 and global sensitivity coefficients S_i of h_1/D, B/D, P/D, and F_r were 0.35, 0.07, 0.13, and 0.02; 0.63, 0.25, 0.30, and 0.32, respectively. h_1/D has a significant effect on C_d. In particular, when $h_1/D < 0.24$ and $0.48 < F_r < 0.58$, $0.67 < F_r < 0.72$, the discharge capacity of the SCSW is relatively large.

Key words: dimensionless parameters, discharge characteristics, intelligent model, semi-circular side weir, Sobol's method

HIGHLIGHTS

- We developed an effective and high-accuracy model for predicting the C_d of SCSW.
- The importance of dimensionless parameters on C_d was quantified by Sobol's method.
- It explored the flow characteristics of semi-circular side weir.

1. INTRODUCTION

As one of the most common diversion structures, side weirs are used for flow control, drainage networks, irrigation, and wastewater channels (Zahiri *et al.* 2013). In recent years, with the change in extreme weather and a significant increase in storm floods, side weirs have been used as common equipment in sewer networks and irrigation systems to divert excess water flow from channels to other channels (Uyumaz *et al.* 2014). Semi-circular labyrinth side weirs are widely used due to their long overflow front length, stable overflow structure, and facilitation of sediment removal. Also, semi-circular side weir (SCSW) flow as a spatially variable flow has more parameters affecting the discharge coefficient (C_d). Therefore, it is important to accurately evaluate the influence and variation law of different factors on the C_d for the design and operation of this structure.

At present, most scholars mainly use traditional empirical methods to check the discharge capacity of SCSWs. Haghshenas & Vatankhah (2021) proposed discharge calculation equations for SCSW, in which the mean and maximum errors of the best model were 1.87 and 6.31%, respectively. Mamand & Raheem (2018) used SPSS software to fit the empirical equation of the SCSW, and the coefficients of determination (R^2) in the form of multivariate linear regression and multivariate power regression were 0.8498 and 0.8584, respectively. Khalili & Honar (2017) gave the calculation equation of the C_d of the SCSW by using the model experiments and dimensional analysis. The research shows that the C_d of the SCSW was higher than that of the rectangular side weir. However, the discharge is affected by the plane position of the weir sill, the

---

shape of the weir, the upstream and downstream flow conditions, and different flow resistances generated, resulting in different expressions of the C_d, which are not convenient for users. Also, the discharge coefficients were determined according to empirical equations, which were limited by certain datasets, effective parameter interactions, high uncertainty, numerous assumptions, and other defects (Tao *et al.* 2022), resulting in insufficient mining of physical properties among parameters and limited calculation accuracy.

In recent years, many scholars have attempted to use soft computing techniques for solving the problems of large calculations and inconvenient use of empirical equations (Haghbin & Sharafati 2022; Shen *et al.* 2022; Gharehbaghi *et al.* 2023; Parsaie *et al.* 2023; Seyedian *et al.* 2023; Yarahmadi *et al.* 2023). Jamei *et al.* (2021) developed three linear models for predicting the C_d of the triangular side orifices. The research shows that the intelligent model can accurately evaluate the discharge capacity of the side orifices under free-flow conditions. Tao *et al.* (2022) used three machine learning models to estimate the C_d prediction models of the gate under free-flow and submerged-flow conditions. The results show that the model has higher accuracy for the free-flow condition. Ismael *et al.* (2021) used neural network technology for predicting the C_d of inclined cylindrical weirs with different diameters; the root mean square error (RMSE) of the radial basis function network model was reduced by 9 and 41% compared with the cascade-forward neural network and the back-propagation neural network (BPNN) in the testing stage, respectively. However, with the wide application of intelligent models in weir flow, it has been gradually discovered that this technology has problems such as overfitting and easily falling into local optimum. Therefore, researchers began to try to optimize the hyperparameters of the model through optimization algorithms to derive the best model parameters to improve the forecast accuracy and stability of the model. For example, Haghbin *et al.* (2022) developed a hybrid data-driven approach to evaluate the C_d of step spillways, and the optimized model improved the performance index to 86.13%. Pradeep & Samui (2022) used a neural network technology hybrid optimization algorithm to predict rock strain, and the results showed that the optimized model was better than other single models in the training and testing phases. Chen *et al.* (2022) aimed to predict the discharge coefficient of streamlined weirs, and the results showed that the hybrid deep data-driven algorithms provide more accurate results than the classical ones. Simsek *et al.* (2023) used the artificial neural network (ANN) to predict the discharge coefficient of trapezoidal broad-crested weir; the study results showed that the Froude number significantly increases the performance of the models in estimating C_d values, and the ANN method was more successful in determining C_d than other methods. Balouchi & Rakhshandehroo (2018) used the soft computing models to evaluate the discharge coefficient for combined weir-gate, and multilayer perceptron was considered superior; it had better statistical indices of RMSE, mean absolute error (MAE), and R^2 (0.027, 0.022, and 0.984, respectively).

However, the prediction model needs to meet the requirements of high accuracy and stability due to the large discharge and complex physical parameters of the SCSW. According to the current literature, research shows that a high-precision SCSW C_d prediction model has not been developed yet. Therefore, it is important to develop an accurate and stable prediction model for the C_d of SCSW in this study. In addition, there is also great interest in the interaction characteristics between model inputs and outputs. Zhang *et al.* (2013) used Sobol's method to analyze the sensitivity of potential hydrological processes under different hydrological models and climatic conditions. Nossent *et al.* (2011) successfully applied the Sobol sensitivity method to the prioritization of input parameters of complex environmental models. However, most scholars pay more attention to the stability and accuracy of the weir flow prediction model, and the interactions and variation relationships between input parameters and discharge coefficients have not been explored in depth. Hence, this paper not only establishes the discharge coefficient prediction model for the SCSW but also provides a new method for the accurate calculation of the discharge of the structure. More importantly, based on predecessors, the influence of dimensionless parameters on the discharge coefficient is quantified, and this study fills the research gap in this area.

In summary, this study aims to systematically evaluate the effects of the hydraulic parameters of SCSWs on the C_d. First, the particle swarm optimization (PSO) algorithm, genetic algorithm (GA), and sparrow search algorithm (SSA) are used to optimize the hyperparameters c and γ of the support vector machine (SVM) and establish three different models for predicting the C_d of SCSWs. Then, the accuracy and generalization ability of the intelligent and traditional empirical models are compared using different performance indexes. On this basis, Sobol's method is used to explore the interaction and change process between hydraulic parameters and C_d and analyze the change law of hydraulic parameters and C_d. The sensitivity of different hydraulic parameters to C_d is quantified to provide an essential reference basis for the design and promotion of SCSWs.

2. DATA AND MODELS

2.1. Experimental data

In this study, the dataset was obtained from Haghshenas & Vatankhah (2021). The experimental device consisted mainly of a pumping station with a recirculation system that provided a horizontal rectangular channel 12 m long, 0.25 m wide, and 0.5 m deep. The weir upstream discharge Q_1 and side weir discharge Q_w were determined by triangular and rectangular weirs, respectively, and through the accuracy of $\pm 0.5\%$ electromagnetic flowmeter calibration. The SCSW was installed on the main rectangular channel wall 6 m away from the inlet, and the downstream and upstream water depths of the weir were measured at the centerline of the main channel using a point gauge with an accuracy of 0.1 mm. The SCSWs were made of 10-mm-thick plexiglass sheets with a crest thickness of 1 mm, and the plan layout is shown in Figure 1. Three different weir heights ($P = 5$, 10, and 15 cm; the weir crest height, P, varied from 5 to 15 cm for each value of weir diameter) and three different weir diameters ($D = 25$, 30, and 40 cm) were measured in laboratory measurements. Q_1 varied from 14.7 to 42.1 L/s, Q_w varied from 3.0 to 25.8 L/s, and the diverted discharge ratio was $0.14 \leq Q_w/Q_1 \leq 0.73$. A total of 155 runs were carried out under free-flow conditions, and the data characteristics are shown in Table 1.

2.2. Dimensional analysis

As can be seen from Figure 1, the variables that may affect the SCSW discharge include the following: h_1 ($h_1 = y_1 - p$) is the depth of flow relative to the crest point of the upstream weir, D is the side weir diameter, P is the side weir crest height, V_1 is the mean velocity at upstream of the side weir, y_1 is the flow depth at the upstream end of the side weir, B is main channel width, ρ is water density, μ is water viscosity, and g is gravitational acceleration. The discharge of SCSW can be expressed as Equation (1).

$$Q_w = f_1(h_1, B, D, P, V_1, y_1, \rho, \mu, g) \tag{1}$$

According to the Buckingham-π theorem, the above parameters were dimensionally analyzed (Haghshenas & Vatankhah 2021; Saffar *et al.* 2021), and D, g, ρ were used as three independent variables; the dimensionless parameters that affect C_d can be expressed as Equation (2).

$$C_d = f_2\left(\frac{h1}{D}, \frac{B}{D}, \frac{P}{D}, F_r = \frac{V_1}{\sqrt{gy_1}}, R_e = \frac{\rho D V_1}{\mu}\right) \tag{2}$$

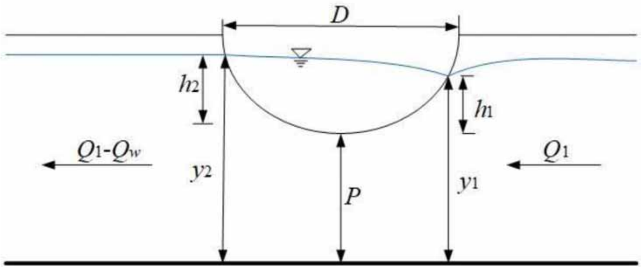

Figure 1 | Plane structure of SCSW.

Table 1 | Statistics of data characteristics

Statistical parameters	***B/D***	***P/D***	**h_1/D**	**F_r**	**C_d**
Maximum	1	0.6	0.469	0.815	0.780
Minimum	0.625	0.125	0.156	0.174	0.565
Mean	0.799	0.299	0.304	0.433	0.663
Middle quartile	0.833	0.250	0.305	0.420	0.652
SD	0.155	0.136	0.083	0.153	0.056

where F_r is the Froude number upstream of the side weir and R_e is the Reynolds number. For side weirs, the flow of water is usually turbulent, and the influence of dynamic viscosity (R_e) on the hydraulic characteristics of flow is negligible (Norouzi *et al.* 2020); the dimensionless parameters affecting the C_d can be expressed as Equation (3).

$$C_d = f_3\left(\frac{h1}{D}, \frac{B}{D}, \frac{P}{D}, F_r\right) \tag{3}$$

2.3. Support vector machine

In this study, the dataset is small, the sample uncertainty is high, and the sample parameters are highly nonlinear. Therefore, a suitable large-scale, fast, and robust model is selected. Meanwhile, SVM is a powerful supervised learning technique that can provide reliable and robust predictions (Najafzadeh & Oliveto 2020). Considering that the PSO and GA belong to the traditional swarm intelligence algorithm, and the SSA belongs to the new swarm intelligence algorithm by using the same dataset to compare the hyperparameter changes between the three algorithms, the stability and reliability of the model can be better determined.

SVM is a classification technique proposed by Vapnik based on the statistical learning theory and structural risk minimization (SRM) (Cortes & Vapnik 1995) and is now widely used for high-accuracy prediction due to its advantages in solving nonlinear problems (Ahmad *et al.* 2014; Parsaie *et al.* 2019; Najafzadeh & Niazmardi 2021; Parsaie *et al.* 2021). Its purpose is to generate a decision boundary between two classes, which is called a hyperplane, and the separating hyperplane is determined by the orthogonal vector w and the bias b. Its direction is as far away from the nearest data points in each class, and these nearest points are called support vectors (Huang *et al.* 2018); its model structure is shown in Figure 2. The solution to the nonlinear problem can be achieved by mapping the data to a higher dimensional feature space with the help of kernel functions (Najafzadeh *et al.* 2016). There are two very important parameters C and γ in the SVM model, and the parameter C represents the penalty. The value of C affects the prediction accuracy and the value of γ affects the partitioning of the feature space; the parameter γ has a greater impact on the results than the penalty factor C. Therefore, to obtain suitable C and γ, three optimization algorithms are used in this study to optimize the values of C and γ globally to obtain the best performance prediction model for the C_d of the SCSW. The optimization process is shown in Figure 3.

2.4. SVM and PSO

PSO is a classic population search algorithm, and its calculation equation is as follows (Huang & Dun 2008; Ardjani *et al.* 2010; Cuong-Le *et al.* 2022):

$$v_{i+1} = \omega v_i + c_1 r_1 (p_{\text{best}} - x_i) + c_2 r_2 (g_{\text{best}} - x_i) \tag{4}$$

$$x_{i+1} = x_i + v_{i+1} \tag{5}$$

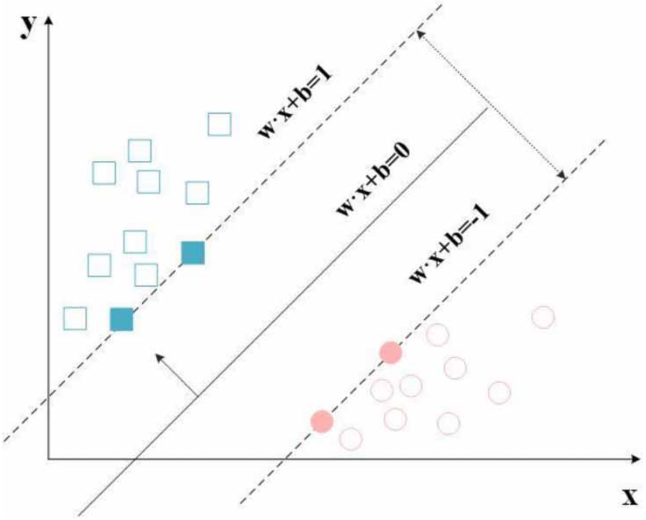

Figure 2 | SVM plan structure diagram.

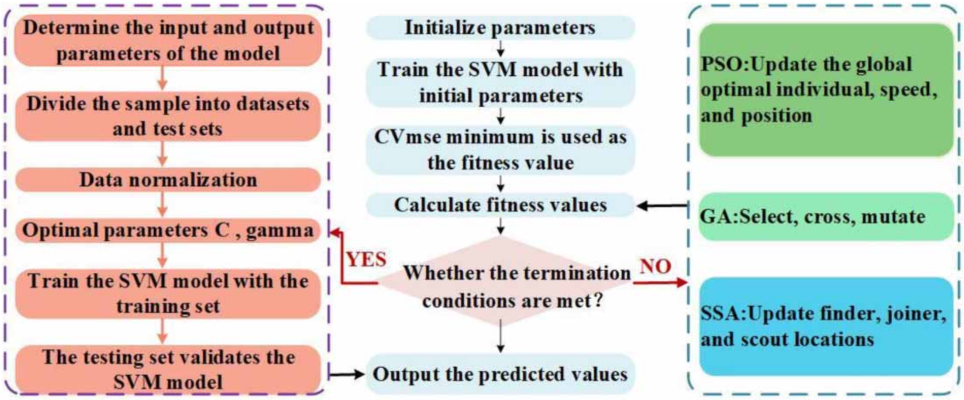

Figure 3 | Process diagram of optimization algorithm.

where x_i is the position of each particle, v_i is the velocity of each particle, p_{best} is the particle optimal value, g_{best} is the global optimal value, r_1 and r_2 are random numbers between 0 and 1, and c_1 and c_2 are acceleration factors.

The particle swarm regards the two parameters of C and γ of the SVM as two particle swarms and first sets the parameters of population size and iteration number for population and velocity initialization, inputs the randomly generated C and γ into the SVM model for training; the mean square error of model cross-validation (CV_{mse}) is used as the model fitness function, the minimum fitness of the particle represents the optimal particle position at this time, and the optimization algorithm ends when the iteration number meets the set value.

2.5. SVM and GA

The GA is an adaptive optimization method with a global search function that uses random search to efficiently guide the parameter space to encode each individual. The key technology of the algorithm consists of five elements: encoding of parameters, initialization of the population, calculation of the fitness function, layout of genetic operations, and control of the parameter arrangement (Li & Kong 2014). Therefore, through continuous evolution from generation to generation, an optimally adapted individual can eventually be obtained. It has the advantages of global optimality, implicit parallelism, high stability, and wide availability (Li & Kong 2014; Guan *et al.* 2021).

The basic steps of the GA:

(1) *Encoding*: The GA represents the solution data in the solution space as genotypic string structure data in the genetic space before searching, and the different combinations of these string structure data constitute the different points.
(2) *Initial population generation*: N initial string structure data are randomly generated, each string structure data is called an individual, N individuals form a population, and the GA uses these N string structure data as initial points to start evolution.
(3) *Adaptability evaluation*: Adaptability indicates the strengths and weaknesses of individuals or solutions. The fitness function is defined in different ways for different problems.

Finally, the optimal solution is obtained by three basic operations: selection, crossover, and variation.

2.6. SVM and SSA

The SSA is a new intelligent optimization algorithm that simulates the foraging and anti-predation behavior of sparrows (Xue & Shen 2020; Yan *et al.* 2022). At present, it has been widely used in related fields. Throughout the foraging process, there are three behaviors: discoverer, joiner, and alerter. Among them, the identities of the discoverer and joiner are changed dynamically. The location update of the discoverer is shown in Equation (6).

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t} \exp\left(-\dfrac{i}{\alpha C_{\max}}\right), & R_2 < \text{ST} \\ X_{i,j}^{t} + \text{QL}, & R_2 \geq \text{ST} \end{cases} \tag{6}$$

where $X_{i,j}^{t}$ is the position of the ith individual in the region dimension after the tth iteration in the sparrow population; a is a uniform random number, $a \in (0,1]$; $C_{\max}$ is the maximum number of iterations; R_2 is a random number with a warning value of [0,1]; ST is the security threshold with an interval of [0.5, 1]; Q is a random number and obeys the standard normal distribution; and L is a $1 \times d$ dimensional matrix. The location update of the joiner is shown in Equation (7).

$$X_{i,j}^{t+1} = \begin{cases} Q\exp\left(-\dfrac{X_w - X_{i,j}^{t}}{i^2}\right) & i > n/2 \\ X_p^{t+1} + |X_{i,j}^{t} - X_p^{t+1}|A^{+}L, & i \leq n/2 \end{cases} \tag{7}$$

where X_p^{t+1} is the best position of the discoverer in the $t+1$ iteration; X_w is the worst position in the current sparrow population; $A^{+} = A^{T}(AA^{T})^{-1}$, A is $1 \times \mathrm{d}$ matrix with element 1 or -1. n is the population size. When $i > n/2$, the ith joiner with low fitness is not fed and needs to be foraged; conversely, when $i \leq n/2$, the joiner will forage near the optimal position. The location update of the alerter is shown in Equation (8).

$$X_{i,j}^{t+1} = \begin{cases} X_{\mathrm{b}}^{t} + \beta|X_{i,j}^{t} - X_{\mathrm{b}}^{t}|, & f_i > f_g \\ X_{i,j}^{t} + K\left(\dfrac{|X_{i,j}^{t} - X_w^{t}|}{(f_i - f_w) + \varepsilon}\right), & f_i = f_g \end{cases} \tag{8}$$

where X_{b}^{t} is the global optimum position; β is a random number that conforms to the standard normal distribution. K is a uniform random number, $K \in [-1,1]$; f_i, f_w, and f_g are the current position fitness value, the worst position fitness value, and the optimal position fitness value, respectively; ε is the minimum constant to the denominator is not 0. $f_i > f_g$ indicates that sparrows are at the edge of the population and are vulnerable to attack; $f_i = f_g$ indicates that sparrows are in the middle of the population, warning of danger, and adjust the search strategy in time to avoid attacks.

2.7. Sobol's sensitivity analysis method

The Sobol method (Sobol 1990), as a global sensitivity analysis method based on variance decomposition, obtained the importance of the input parameters on the output results by calculating the first-order sensitivity and the global sensitivity of the input parameters (Lu *et al.* 2018). The objective function $f(x)$ of the model is decomposed as the sum of 2^p increasing terms:

$$f(x_1, x_2, \ldots, x_n) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{i=1}^{n}\sum_{j=i+1}^{n} f_{ij}(x_i, x_j) + \cdots + f_{i,\ldots,n}(x_{i,\ldots,n}) \tag{9}$$

where f_0 represents the constant in the objective function, and each integral variable in the formula is 0, then the expression is

$$\int_0^1 f_{i1is}(x_{i1}, x_{i2}, \ldots, x_{is})dx_{ik} = 0 \tag{10}$$

where $1 \leq i_1 < \ldots < i_s \leq n, \ 1 \leq k \leq n$ (11)

Total variance: $V(Y) = \int f^2(x)dx - f_0^2$ (12)

Partial variance: $V_{i1,\, i2,\, \ldots,\, is} = \int f^2 i_1, i_2, \ \ldots, i_s dxi_1 dxi_2 \ldots xi_s$ (13)

First-order sensitivity coefficient: $\mathrm{S_i} = \dfrac{\mathrm{V_i}}{\mathrm{V(Y)}}$ (14)

Global sensitivity coefficient: $S_{Ti} = 1 - \dfrac{V_{\sim i}}{V(Y)}$ (15)

where $V(Y)$ represents the sum of the parameters on the output results of the model objective function $f(\mathrm{x})$; $V_{i1,\ i2,\ \ldots,\ is}$

represents the influence of the interaction of the parameter combination on the model output results. V_i represents the influence of the ith parameter on the output result of the model objective function $f(x)$; and $V_{\sim i}$ represents the sum of the variance caused by all parameters except the ith parameter.

2.8. Evaluation index

This study used several statistical methods to evaluate model performance. The parameters are RMSE, correlation coefficient (R), mean absolute percentage error (MAPE), standard deviation (SD), scatter index (SI), developed discrepancy ratio (DDR), bias coefficient (Bias). Methods are defined in the following equations.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(O_i - P_i)^2}{N}} \tag{16}$$

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{Oi - Pi}{Oi}\right| \times 100 \tag{17}$$

$$R = \frac{\sum_{i=1}^{N}(O_i - \overline{O}_i)(P_i - \overline{P}_i)}{\sqrt{\sum_{i=1}^{N}(O_i - \overline{O}_i)^2 \sum_{i=1}^{N}(P_i - \overline{P}_i)^2}} \tag{18}$$

$$\mathrm{SD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Oi - \overline{Oi})^2} \tag{19}$$

$$\mathrm{SI} = \frac{\mathrm{RMSE}}{Oa} \tag{20}$$

$$\mathrm{DDR} = \frac{Pi}{Oi} - 1 \tag{21}$$

$$\mathrm{Bias} = \frac{1}{n}\sum_{i=1}^{n}(Pi - Oi)^2 \tag{22}$$

where O_i and P_i represent the experimental and predicted values of C_d, respectively, and O_a and P_a represent the mean of the experimental and predicted values, respectively.

3. RESULTS AND DISCUSSION

3.1. Model comparison

In this study, 109 experimental datasets were selected as the training set and the remaining 46 sets were used as the testing set. h_1/D, B/D, P/D, and F_r were used as model inputs and C_d as model outputs. The global optimization of the hyperparameters C and γ of SVM was performed by three optimization algorithms, PSO, GA, and SSA; the specific parameter settings of each model are shown in Table 2, and the performance indexes of all models were finally obtained as shown in Tables 3 and 4. When the SVM model is used to calculate the C_d of the SCSW, the RMSE, MAPE, SD, and R were 0.047, 0.076, 0.073, and 0.897 in the training phase, respectively. The RMSE, MAPE, SD, and R were 0.045, 0.072, 0.062, and 0.926 in the testing phase, respectively. The PSO-SVM, GA-SVM, and SSA-SVM are significantly superior in each evaluation index in the training and testing phases than SVM, indicating that all three optimization algorithms can effectively improve the performance of SVM through global optimization search.

Figure 4 shows the scatter plot for the experimental and predicted values for all models. Larger values of R indicate the better fitting ability of the models, and the closer the predicted and experimental values are to the trend line (1:1). As can be seen from Figure 4, compared with SVM, the R of PSO-SVM, GA-SVM, and SSA-SVM increased by about 6.65, 9.11, and 7.23% in the training phase, respectively. Also, the R increased by about 2.83, 4.04, and 2.42% in the testing phase, respectively. It can be seen that among the three optimization models, GA-SVM has better generalization ability and

Table 2 | Parameter settings of all models

Model	Parameter	Value	*C*	γ
PSO-SVM	Particle swarm size	20	0.1	6.72
	Number of iterations	30		
	Inertia factor	0.9		
	Acceleration constants	2		
	Speed range	[−1,1]		
GA-SVM	Population size	20	4.05	4.40
	Number of iterations	30		
	Crossover probability	0.5		
	Mutation probability	0.1		
SSA-SVM	Number of sparrows	20	0.1	4.33
	Number of iterations	30		
	warning value ST	0.6		
	Proportion of discoverers	0.7		
	Proportion of detectors	0.2		

Table 3 | All model performance indexes in the training stage

Model	RMSE	MAPE (%)	SD	*R*	SI	Bias
SVM	0.047	0.076	0.073	0.897	0.071	0.0130
PSO-SVM	0.021	0.053	0.043	0.961	0.031	0.0028
GA-SVM	0.014	0.037	0.041	0.987	0.022	0.0008
SSA-SVM	0.019	0.048	0.044	0.967	0.024	0.0031

Table 4 | All model performance indexes in the testing stage

Model	RMSE	MAPE (%)	SD	*R*	SI	Bias
SVM	0.045	0.072	0.062	0.926	0.069	0.0120
PSO-SVM	0.017	0.016	0.046	0.953	0.026	0.0010
GA-SVM	0.009	0.008	0.043	0.965	0.014	0.0004
SSA-SVM	0.016	0.016	0.047	0.949	0.031	0.0008

prediction stability. Figure 5 shows the Taylor plots for all the models; the longitudinal distance from the origin represents the SD, the purple radial lines indicate the correlation coefficient (R), and the green circular arcs show the RMSE. As the circle section expanded, this parameter value increased. Moreover, the SD, R, and RMSE of the training and testing phases were specified by a single point, and the model closest to the reference point was considered the best model. It can be seen that SVM has the worst prediction effect and GA-SVM has the best prediction result, where PSO-SVM and SSA-SVM have almost the same effect. Therefore, GA-SVM can be used as the optimal intelligent prediction model for the C_d of SCSW.

3.2. Comparison with empirical equations

Figure 6 shows the DDR values for all models, which can be used to evaluate the distribution of errors in detail. It can be seen that the overall error of GA-SVM is small, indicating that the model has high prediction accuracy. Meanwhile, Haghshenas & Vatankhah (2021) used the dimensional analysis technique to fit the SCSW discharge calculation model, and the average error and maximum error of the best model were 1.87 and 6.31%, respectively. The mean and maximum errors of GA-SVM were 0.08 and 2.47%, respectively, which were about 95.72 and 60.86% lower compared to the traditional empirical model, indicating that the intelligent model has higher accuracy in predicting the C_d of SCSWs. Figure 7 shows the error

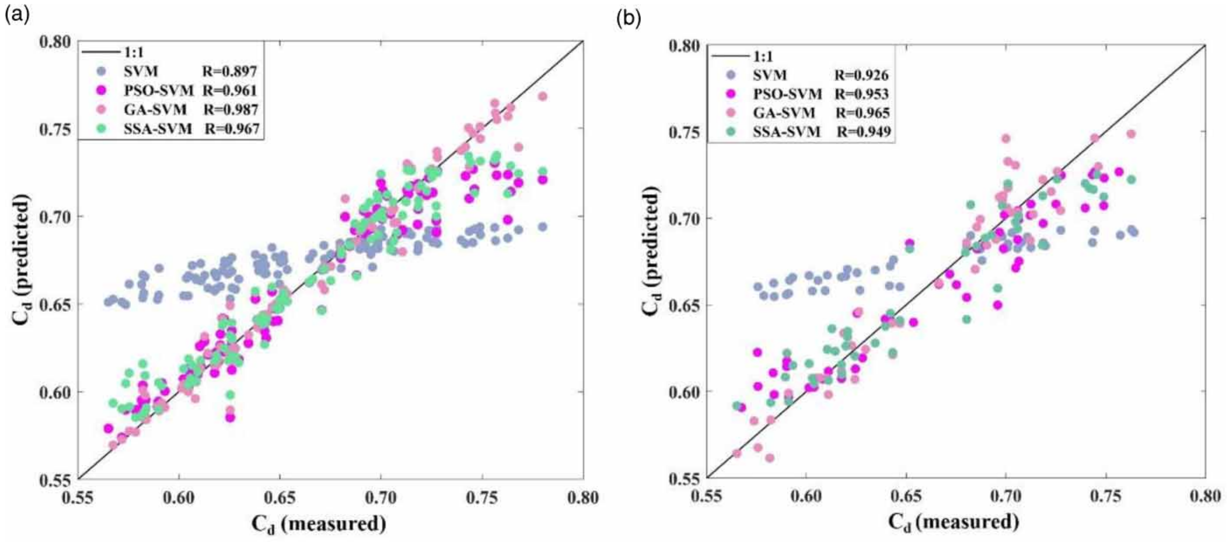

Figure 4 | Scatter plot of experimental and predicted values of C_d. (a) Training stage and (b) testing stage.

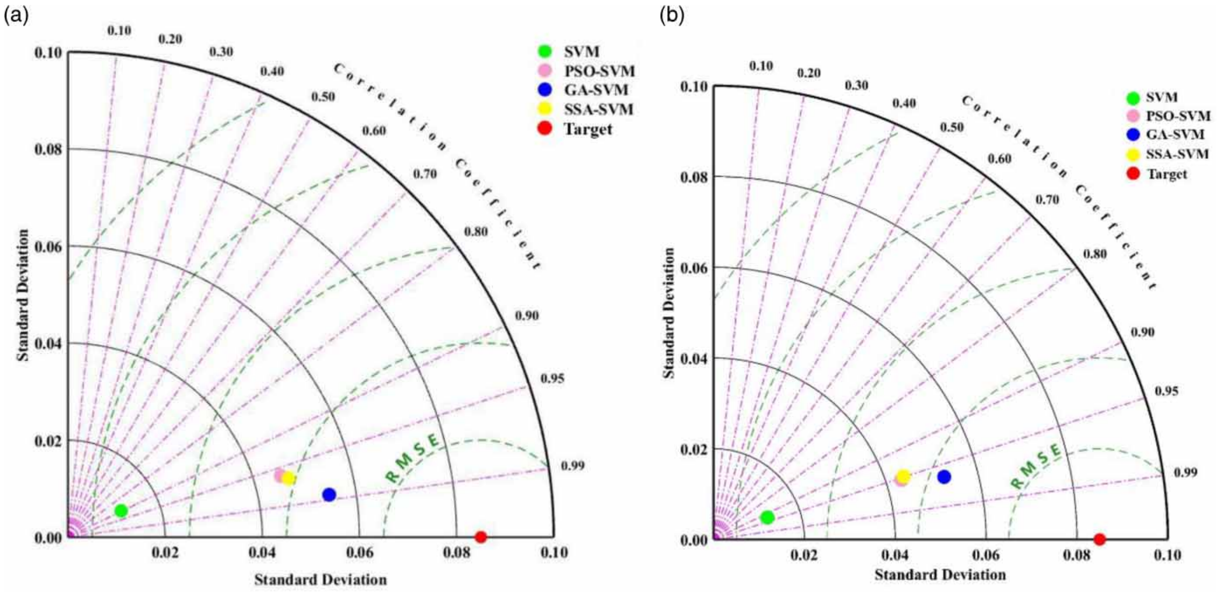

Figure 5 | The Taylor diagram for all models. (a) Training stage and (b) testing stage.

density plot of GA-SVM in the testing phase; 91.31% of the prediction errors were below 2%, and the errors were mainly concentrated at $C_d = 0.7$, indicating that GA-SVM has high accuracy and stability in predicting the C_d.

3.3. Quantitative analysis of parameters

From the above analysis, it can be seen that GA-SVM can be used as a prediction model for the C_d of SCSW. Therefore, the model input parameters were further quantified and analyzed using Sobol's method. As can be seen from Figure 8, the sensitivity coefficients of the dimensionless parameters and the first-order sensitivity coefficients S_1 of h_1/D, B/D, P/D, and F_r were 0.35, 0.07, 0.13, and 0.02, respectively, indicating that when only the influence of a single parameter on C_d was

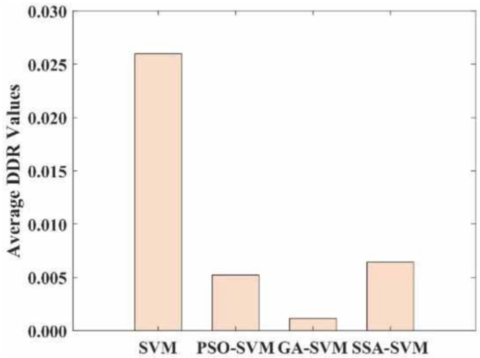

Figure 6 | DDR values for all models.

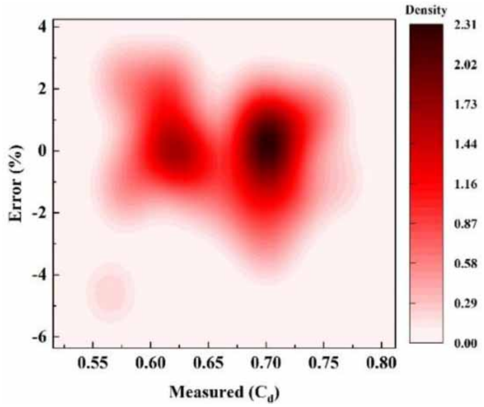

Figure 7 | Error density plot of GA-SVM in the testing phase.

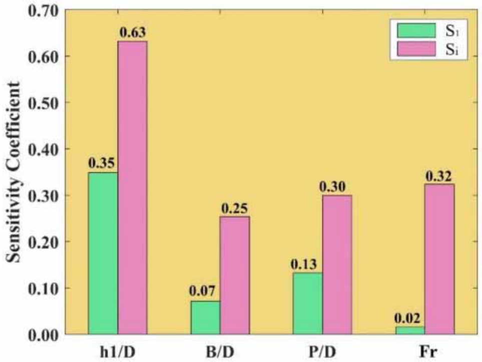

Figure 8 | Sensitivity coefficients of dimensionless parameters.

considered, the influence of h_1/D on C_d was the largest, followed by P/D, and F_r has the least effect on C_d. When a parameter interacts with other parameters, the global sensitivity coefficients S_i for h_1/D, B/D, P/D, and F_r were 0.63, 0.25, 0.30, and 0.32, respectively. Also, h_1/D has the greatest effect on C_d, indicating that h_1/D is an important parameter affecting C_d. However, the effect of F_r on C_d after interacting with other parameters was only inferior to h_1/D, indicating that the ratio of flow velocity to flow depth plays an important role in the assessment of C_d under the influence of geometric parameters. Therefore, h_1/D and F_r should be considered important parameters when assessing the discharge capacity of SCSW.

3.4. Parameter sensitivity analysis

Figure 9 shows the variation relationship between the dimensionless parameters h_1/D, B/D, P/D, and F_r and the predicted C_d for the SCSW. As can be seen from Figure 9(a), C_d decreases with the increase of h_1/D. As D increases, the trend of C_d decreases more obviously; where the variation trend of $D = 0.25$ m and $D = 0.30$ m is very close. As can be seen from Figure 9(b), C_d increases with the increase of F_r, and the trend increases more slowly when $B/D = 0.625$. The trend increases more obviously when $B/D = 0.833$. For the same F_r, C_d shows a decreasing trend as B/D increases.

3.5. Analysis of discharge characteristics

Figure 10 shows the variation of the most influential input parameters against the predicted C_d. It can be seen that when $h_1/D < 0.24$, $0.48 < F_r < 0.58$, and $0.67 < F_r < 0.72$, the C_d of the SCSW is larger and has a higher discharge capacity of the SCSW at this time. Also, when $F_r < 0.50$ and $0.40 < h_1/D < 0.47$, the C_d of SCSW is smaller, and the discharge capacity of SCSW is relatively small at this time.

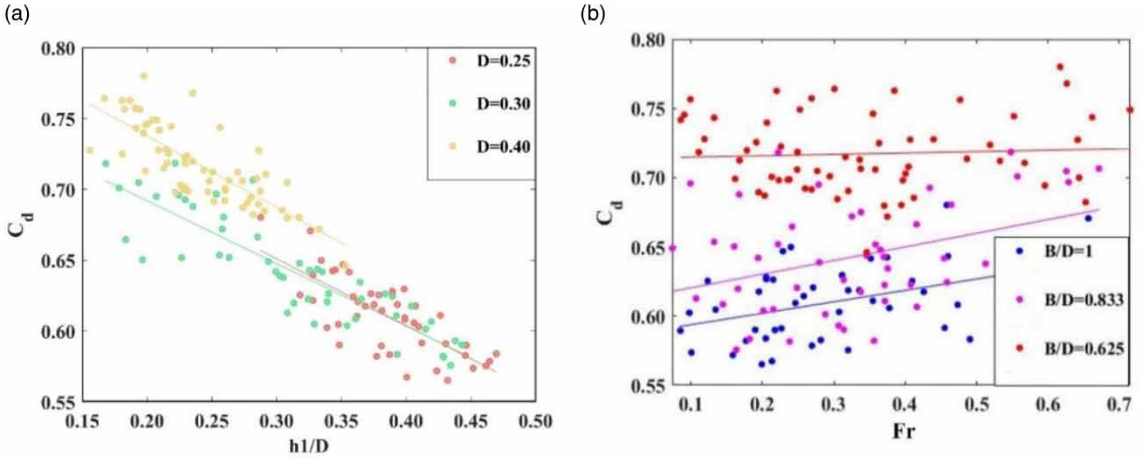

Figure 9 | Variation of the dimensionless parameters versus predicted discharge coefficient. (a) The variation between h_1/D and predicted C_d. (b) The variation between F_r and predicted C_d.

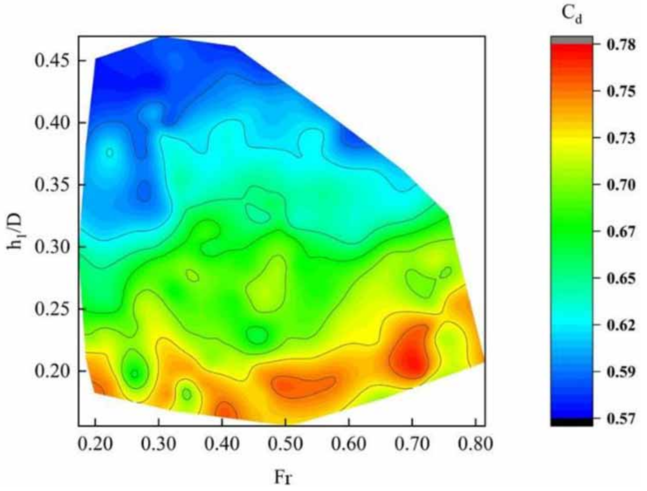

Figure 10 | Variation of the most influential inputs versus predicted C_d.

4. CONCLUSION

In order to achieve accurate water measurement and reasonable distribution of water resources in small channels, a semi-circular labyrinth side weir is used as an efficient and greater discharge capacity control structure. In this study, PSO-SVM, GA-SVM, and SSA-SVM optimization models were developed based on SVM. Then, Sobol's method was introduced to calculate the sensitivity coefficients of different dimensionless parameters h_1/D, B/D, P/D, and F_r to C_d. This paper evaluates the effect of different factors on the discharge capacity of SCSW. The parameter variation range of various discharge capacities is proposed, and the variation law between different parameters and C_d is analyzed. The following conclusions were drawn.

1. In the current study, GA-SVM can be used as an efficient and high-accuracy prediction model for the C_d of SCSW. In the testing phase, $R = 0.987$, MAPE = 0.037%, RMSE = 0.014, SD = 0.041, SI = 0.022, and Bias = 0.008, and 91.31% of the prediction errors were below 2%; the model has high generalization ability, stability, and prediction accuracy, and this model effectively solves the problems of large computational complexity and difficult coefficient correction in traditional empirical models.
2. The quantitative analysis showed that the S_1 and S_i of h_1/D, B/D, P/D, and F_r were 0.35, 0.07, 0.13, and 0.02; and 0.63, 0.25, 0.30, and 0.32, respectively; h_1/D was the most important parameter affecting C_d, the effect of F_r on C_d after interacting with other parameters was only inferior to h_1/D, and C_d decreased as h_1/D increased. As D increased, C_d decreased the greater the trend. As the diameter of the side weir increases, the lateral flow will increase significantly in the subcritical flow regime.
3. When $h_1/D < 0.24$, $0.48 < F_r < 0.58$, and $0.67 < F_r < 0.72$, the C_d of SCSW is greater. Meanwhile, when $F_r < 0.50$ and $0.40 < h_1/D < 0.47$, the C_d of the SCSW is relatively small. This can provide an important reference basis for the application of SCSW in practical engineering.

In addition, in this study, the width of the main channel is constant. Therefore, it is necessary to further explore the influence of the width change of the main channel on the discharge coefficient of the SCSW.

FUNDING

This study was partly supported by the National Natural Science Foundation-sponsored project (grant 52079107), the Natural Science Basic Research Project of Shaanxi Province (grant 2023-JC-QN-0395), and the General Special Scientific Research Project of Shaanxi Province (grant 22JK0470).

DATA AVAILABILITY STATEMENT

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

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Ahmad, A. S., Hassan, M. Y., Abdullah, M. P., Rahman, H. A., Hussin, F., Abdullah, H. & Saidur, R. 2014 A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renewable and Sustainable Energy Reviews* **33**, 102–109.

Ardjani, F., Sadouni, K. & Benyettou, M. 2010 Optimization of SVM multiclass by particle swarm (PSO-SVM). In: *2010 2nd International Workshop on Database Technology and Applications*. IEEE, pp. 1–4.

Balouchi, B. & Rakhshandehroo, G. 2018 Using physical and soft computing models to evaluate discharge coefficient for combined weir-gate structures under free flow conditions. *Iranian Journal of Science and Technology, Transactions of Civil Engineering* **42**, 427–438.

Chen, W., Sharifrazi, D., Liang, G., Band, S. S., Chau, K. W. & Mosavi, A. 2022 Accurate discharge coefficient prediction of streamlined weirs by coupling linear regression and deep convolutional gated recurrent unit. *Engineering Applications of Computational Fluid Mechanics* **16** (1), 965–976.

Cortes, C. & Vapnik, V. 1995 Support-vector networks. *Machine Learning* **20** (3), 273–297.

Cuong-Le, T., Nghia-Nguyen, T., Khatir, S., Trong-Nguyen, P., Mirjalili, S. & Nguyen, K. D. 2022 An efficient approach for damage identification based on improved machine learning using PSO-SVM. *Engineering with Computers* **38**, 3069–3084.

Gharehbaghi, A., Ghasemlounia, R., Afaridegan, E., Haghiabi, A. H., Mandala, V., Azamathulla, H. M. & Parsaie, A. 2023 A comparison of artificial intelligence approaches in predicting discharge coefficient of streamlined weirs. *Journal of Hydroinformatics* **25** (4), 1513–1530.

Guan, S., Wang, X., Hua, L. & Li, L. 2021 Quantitative ultrasonic testing for near-surface defects of large ring forgings using feature extraction and GA-SVM. *Applied Acoustics* **173**, 107714.

Haghbin, M. & Sharafati, A. 2022 A review of studies on estimating the discharge coefficient of flow control structures based on the soft computing models. *Flow Measurement and Instrumentation* **83**, 102119.

Haghbin, M., Sharafati, A., Aghamajidi, R., Asadollah, S. B. H. S., Noghani, M. H. M. & Jalón, M. L. 2022 Determination of discharge coefficient of stepped morning glory spillway using a hybrid data-driven method. *Flow Measurement and Instrumentation* **85**, 102161.

Haghshenas, V. & Vatankhah, A. R. 2021 Discharge equation of semi-circular side weirs: An experimental study. *Flow Measurement and Instrumentation* **81**, 102041.

Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y. & Xu, W. 2008 A distributed PSO–SVM hybrid system with feature selection and parameter optimization. *Applied Soft Computing* **8** (4), 1381–1391.

Huang, S., Cai, N. & Pacheco, P. P. *et al.* 2018 Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics & Proteomics* **15** (1), 41–51.

Ismael, A. A., Suleiman, S. J., Al-Nima, R. R. O. & Al-Ansari, N. 2021 Predicting the discharge coefficient of oblique cylindrical weir using neural network techniques. *Arabian Journal of Geosciences* **14** (16), 1–8.

Jamei, M., Ahmadianfar, I., Chu, X. & Yaseen, Z. M. 2021 Estimation of triangular side orifice discharge coefficient under a free flow condition using data-driven models. *Flow Measurement and Instrumentation* **77**, 101878.

Khalili, M. & Honar, T. 2017 Discharge coefficient of semi-circular labyrinth side weir in subcritical flow. *Water SA* **43** (3), 433–441.

Li, X. Z. & Kong, J. M. 2014 Application of GA-SVM method with parameter optimization for landslide development prediction. *Natural Hazards and Earth System Sciences* **14** (3), 525–533.

Lu, R., Wang, D., Wang, M. & Rempala, G. A. 2018 Estimation of Sobol's sensitivity indices under generalized linear models. *Communications in Statistics – Theory and Methods* **47** (21), 5163–5195.

Mamand, B. S. & Raheem, A. M. 2018 Discharge coefficients for different types of side weirs. *Zanco Journal of Pure and Applied Sciences* **30** (1), 24–31.

Najafzadeh, M. & Niazmardi, S. 2021 A novel multiple-kernel support vector regression algorithm for estimation of water quality parameters. *Natural Resources Research* **30** (5), 3761–3775.

Najafzadeh, M. & Oliveto, G. 2020 Riprap incipient motion for overtopping flows with machine learning models. *Journal of Hydroinformatics* **22** (4), 749–767.

Najafzadeh, M., Etemad-Shahidi, A. & Lim, S. Y. 2016 Scour prediction in long contractions using ANFIS and SVM. *Ocean Engineering* **111**, 128–135.

Norouzi, R., Arvanaghi, H., Salmasi, F., Farsadizadeh, D. & Ghorbani, M. A. 2020 A new approach for oblique weir discharge coefficient prediction based on hybrid inclusive multiple model. *Flow Measurement and Instrumentation* **76**, 101810.

Nossent, J., Elsen, P. & Bauwens, W. 2011 Sobol'sensitivity analysis of a complex environmental model. *Environmental Modelling & Software* **26** (12), 1515–1525.

Parsaie, A., Haghiabi, A. H., Emamgholizadeh, S. & Azamathulla, H. M. 2019 Prediction of discharge coefficient of combined weir-gate using ANN, ANFIS and SVM. *International Journal of Hydrology Science and Technology* **9** (4), 412–430.

Parsaie, A., Haghiabi, A. H., Latif, S. D. & Tripathi, R. P. 2021 Predictive modelling of piezometric head and seepage discharge in earth dam using soft computational models. *Environmental Science and Pollution Research* **28** (43), 60842–60856.

Parsaie, A., Dehdar-Behbahani, S., Chadee, A. A. & Haghiabi, A. H. 2023 Estimating the energy dissipation of flow passing over triangular and trapezoidal plan weirs using the GMDH model. *Water Practice and Technology* **18** (5), 1115–1124.

Pradeep, T. & Samui, P. 2022 Prediction of rock strain using hybrid approach of ANN and optimization algorithms. *Geotechnical and Geological Engineering* **40** (9), 4617–4643.

Saffar, S., Babarsad, M. S., Shooshtari, M. M. & Riazi, R. 2021 Prediction of the discharge of side weir in the converge channels using artificial neural networks. *Flow Measurement and Instrumentation* **78**, 101889.

Seyedian, S. M., Haghiabi, A. H. & Parsaie, A. 2023 Reliable prediction of the discharge coefficient of triangular labyrinth weir based on soft computing techniques. *Flow Measurement and Instrumentation* **92**, 102403.

Shen, G., Li, S., Parsaie, A., Li, G., Cao, D. & Pandey, P. 2022 Prediction and parameter quantitative analysis of side orifice discharge coefficient based on machine learning. *Water Supply* **22** (12), 8880–8892.

Simsek, O., Gumus, V. & Ozluk, A. 2023 Prediction of discharge coefficient of the trapezoidal broad-crested weir flow using soft computing techniques. *Neural Computing and Applications* **35**, 17485–17499.

Sobol, I. M. 1990 On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe Modelirovanie* **2** (1), 112–118.

Tao, H., Jamei, M., Ahmadianfar, I., Khedher, K. M., Farooque, A. A. & Yaseen, Z. M. 2022 Discharge coefficient prediction of canal radial gate using neurocomputing models: An investigation of free and submerged flow scenarios. *Engineering Applications of Computational Fluid Mechanics* **16** (1), 1–19.

Uyumaz, A., Danandeh Mehr, A., Kahya, E. & Erdem, H. 2014 Rectangular side weirs discharge coefficient estimation in circular channels using linear genetic programming approach. *Journal of Hydroinformatics* **16** (6), 1318–1330.

Xue, J. & Shen, B. 2020 A novel swarm intelligence optimization approach: Sparrow search algorithm. *Systems Science & Control Engineering* **8** (1), 22–34.

Yan, H., Zhang, J., Zhou, N., Shi, P. & Dong, X. 2022 Coal permeability alteration prediction during CO_2 geological sequestration in coal seams: A novel hybrid artificial intelligence approach. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources* **8** (3), 1–11.

Yarahmadi, M. B., Parsaie, A., Shafai-Bejestan, M., Heydari, M. & Badzanchin, M. 2023 Estimation of Manning roughness coefficient in alluvial rivers with bed forms using soft computing models. *Water Resources Management* **37**, 3563–3584.

Zahiri, A., Azamathulla, H. M. & Bagheri, S. 2013 Discharge coefficient for compound sharp crested side weirs in subcritical flow conditions. *Journal of Hydrology* **480**, 162–166.

Zhang, C., Chu, J. & Fu, G. 2013 Sobol′'s sensitivity analysis for a distributed hydrological model of Yichun River Basin, China. *Journal of Hydrology* **480**, 58–68.

First received 4 July 2023; accepted in revised form 10 November 2023. Available online 23 November 2023