

RESEARCH ARTICLE | APRIL 02 2025

Modeling the discharge coefficient of labyrinth sluice gates using hybrid support vector regression and metaheuristic algorithms

Aliasghar Azma ; Alistair G. L. Borthwick ; Reza Ahmadian ; Yakun Liu  ; Di Zhang 



Physics of Fluids 37, 045117 (2025)

<https://doi.org/10.1063/5.0260738>



Articles You May Be Interested In

Application of smart sluice gates for water distribution in support of sustainable smart agriculture

AIP Conf. Proc. (August 2023)

Steady flow past a sluice gate

Phys. Fluids (December 1986)

Hydraulic jump efficiency on sluice gate with different operating systems and dissipaters

AIP Conf. Proc. (March 2023)



Physics of Fluids
Special Topics
Open for Submissions

[Learn More](#)

Modeling the discharge coefficient of labyrinth sluice gates using hybrid support vector regression and metaheuristic algorithms

Cite as: Phys. Fluids **37**, 045117 (2025); doi:10.1063/5.0260738

Submitted: 26 January 2025 · Accepted: 12 March 2025 ·

Published Online: 2 April 2025



View Online



Export Citation



CrossMark

Aliasghar Azma,¹  Alistair G. L. Borthwick,^{2,3}  Reza Ahmadian,⁴  Yakun Liu,^{1,a)}  and Di Zhang¹ 

AFFILIATIONS

¹School of Hydraulic Engineering, Faculty of Infrastructure Engineering, Dalian University of Technology, Dalian 116024, China

²Institute for Infrastructure and Environment, School of Engineering, The University of Edinburgh, Edinburgh, UK

³School of Engineering, Mathematics and Computing, University of Plymouth, Plymouth, UK

⁴Hydro-environmental Research Centre, School of Engineering, Cardiff University, Cardiff, UK

^{a)} Author to whom correspondence should be addressed: liuyakun@dlut.edu.cn

ABSTRACT

Gates and weirs are frequently used hydraulic structures employed for controlling water flow rates in irrigation and drainage networks. Therefore, accurately estimating the discharge coefficient (C_d) is important for precise flow measurement. The present study used intelligent predictive models for modeling C_d in labyrinth sluice gates. For this purpose, key dimensionless parameters and reliable experimental datasets were used. The support vector regression (SVR) model was hybridized with particle swarm optimization (PSO) and genetic algorithms (GA). The statistical metrics and graphical plots evaluated the performance of the generated models. Three commonly used statistical indicators, namely root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2), were used for quantitatively evaluating the performance of the proposed models. The SVR-PSO model achieved the lowest values of RMSE (0.0287) and MAE (0.0209) and the highest value of R^2 (0.9732), indicating that it was more accurate than SVR-GA (RMSE = 0.0324, MAE = 0.0257, R^2 = 0.9685) and SVR (RMSE = 0.0575, MAE = 0.0468, R^2 = 0.8958) on the testing data. The findings revealed that the hybrid SVR methods were more accurate than the standalone SVR model. In addition, regarding the value of the objective function criterion (OBF), the SVR-PSO (OBF = 0.0245) and SVR-GA (OBF = 0.0273) had lower OBF values and provided more precise estimates of the C_d compared to existing nonlinear regression-based formulas and existing data-driven approaches. Finally, sensitivity and SHapley Additive exPlanations (SHAP) analyses determined the relative importance of each input variable for the prediction of C_d .

© 2025 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International (CC BY-NC-ND) license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>). <https://doi.org/10.1063/5.0260738>

I. INTRODUCTION

Gates and weirs are important hydraulic structures in regulating discharge flow in open channels and controlling water levels in supply channels and waterways.^{1,2} These structures are used to control water flow and measurements of discharge rates. Therefore, accurate estimation of the discharge coefficient (C_d) is important for the efficient design of gates and operation. This coefficient quantifies the relationship between the measured and theoretical flow, and the values of C_d illustrate the efficiency of the gates and weirs. There are various types of gates, such as standard, side, skew, and labyrinth sluice gate configurations.³ The selection of an appropriate gate type depends on specific operational needs. Therefore, hydraulic engineers have extensively examined various types of gates and weirs in different geometric

configurations and hydraulic conditions.⁴ Labyrinth sluice gates were introduced as appropriate hydraulic structures for irrigation systems and flood control infrastructure.⁵

Generally, three modeling approaches, including physical modeling, numerical simulation, and data-driven techniques, are proposed in the literature for the prediction of C_d in gates and weirs. Each of the modeling methods has strengths and limitations for the estimation of C_d . Physical modeling used scaled hydraulic models that experimentally measured C_d under controlled laboratory conditions, capturing real flow conditions.⁶ The experimental data were used for validation of other methods; however, physical modeling requires the cost of building a model and conducting experiments, is time-consuming, and is under the influence of scale effects.⁷ Numerical modeling simulates

the hydraulic behavior of flow and detailed insights of flows, extracting the various parameters of flow in complex geometries of hydraulic structures.⁸ Numerical simulations are cost-effective compared to physical models; however, they need substantial computational resources, expertise in computational fluid dynamics modeling, and experimental validation to mitigate uncertainties. Data-driven techniques can find nonlinear relationships from large datasets without any prior knowledge about physical processing. However, they need high-quality data and hyperparameter tuning to achieve reliable accuracy. However, their interpretability poses a persistent challenge.

Traditionally, the estimation of C_d is conducted using experimental works using physical hydraulic models. However, experimental work is time-consuming and costly. Recently, implementations of data-driven models compared to experimental work for modeling hydraulic parameters have led to savings in cost and time.⁹ These methods have been used to model C_d for various types of gates and weirs.¹⁰ Azamathulla *et al.* proposed the gene expression programming (GEP) model to estimate C_d in side sluice gates.¹¹ The GEP model was developed by Azimi *et al.* for C_d estimation in side weirs.¹² They investigated effective parameters to predict C_d . Ghorbani *et al.* simulated C_d for vertical sluice gates¹³ while considering the effect of various sill shapes. For C_d prediction, they used gradient-boosting techniques and experiments. Roushangar *et al.* presented a hybrid kernel extreme learning machine and gray wolf optimization (KELM-GWO) model for C_d prediction in radial gates.¹⁴ For the prediction of C_d in rotary gates, Marashi *et al.* coupled artificial neural networks (ANNs) and SVMs with genetic algorithms (GA) and simulated annealing (SA). Compared to empirical methods, Marashi *et al.* showed that their data-driven models produced accurate predictions.¹⁵ For simulating C_d in weir gates, Parsaie *et al.* assessed the adaptive neuro-fuzzy inference system (ANFIS) and ANN.^{16,17} They concluded that the ANFIS performed rather better than the ANN. Using genetic programming

(GP), Salmasi and Abraham generated predictive models for calculating C_d in inclined sluice gates.¹⁸ Their GP results demonstrated more accurate than conventional regression methods.

Other noteworthy contributions include those of Sahib *et al.*,¹⁹ who applied ANN to predict C_d for a combined trapezoidal weir and rectangular gate, achieving high accuracy through iterative parameter optimization, and Nouri *et al.*,²⁰ who evaluated various data-driven approaches, including model trees, SVM, and ANN, for compound rectangular broad-crested weirs, identifying support vector regression (SVR) as the most reliable approach. As mentioned earlier, various data-driven methods have been proposed for the prediction of C_d . However, some studies highlighted the accuracy and capability of SVR-based models for accurate predictions of C_d in weirs and gates, summarized in Table I.

As shown in Table I, the effectiveness of SVR for estimating C_d is highlighted. Furthermore, Karami *et al.* and Zaji *et al.* demonstrated that integrating metaheuristic algorithms with SVR improved performance compared to standalone SVR.^{21,25} Therefore, it is essential to evaluate the efficiency of optimization algorithms for predicting C_d . SVR has been a popular data-driven model for C_d prediction in weirs and gates with satisfactory results. However, the tuning parameters of SVR are necessary to avoid being trapped in a local minimum. Therefore, evolutionary algorithms are effective methods for determining internal SVR parameters. The capability of popular and widely used evolutionary algorithms, such as GA and particle swarm optimization (PSO), for determining the values of the parameters in an SVR model has been successfully demonstrated in recent studies in various civil and water engineering fields.^{27–30}

Numerous studies have shown that more accurate outcomes arise from combining the SVR model with GA and PSO. For the prediction of suspended sediment load in a watershed basin, Rahgoshay *et al.* examined SVR-GA, MARS,³⁰ and model tree approaches. The SVR-GA

TABLE I. The application of SVR-based models for the estimation of C_d .

Authors	Methods	Type of weirs/gates	Key findings
Karami <i>et al.</i> ²¹	SVR, SVR-FA, PCA, RSM	Triangular labyrinth weir	The statistical metrics showed that the SVR-Firefly model possesses the greatest capability for simulation compared to the other models.
Roushangar <i>et al.</i> ²²	SVR	Labyrinth and arced labyrinth weirs	The results obtained demonstrated that SVM-based models can identify discharge coefficients.
Azimi <i>et al.</i> ²³	SVR	Side weir	Based on the simulation results, the superior model is reasonably accurate.
Zaji and Bonakdari ²⁶	SVR	Modified labyrinth side weir	The SVR model outperformed nonlinear regression models in predicting the C_d .
Zaji <i>et al.</i> ²⁵	SVR, SVR-FA	Modified labyrinth side weir	The SVR-FF model demonstrated enhanced prediction accuracy compared to the standalone SVR model.
Zaji <i>et al.</i> ²⁴	SVR	Modified oblique side weirs	The findings revealed that SVR utilizing a radial basis kernel function surpassed SVR, which employs polynomial kernel functions.

model turned out to have smaller errors than the others. For the prediction of the transverse mixing coefficient in streams, Nezaratian *et al.* found that GA might enhance the accuracy of the SVR model.³¹ The SVR-PSO method for calculating the sediment load during floods was developed by Kazemi *et al.* Their findings indicated that this approach effectively provided dependable predictions of flood features in basins.³² Yong *et al.*³³ integrated SVR models with meta-heuristic algorithms, including the whale optimization algorithm, differential evolution, and PSO, for daily reference evapotranspiration estimation in Malaysia. Accordingly, the results highlighted the superior hybrid SVR compared to a single SVR and found the SVR-PSO model was more accurate than others. Mozaffari *et al.*³⁴ predicted groundwater levels in aquifers using SVR-GA, and the results outperformed Bayesian and standalone SVR models. The above-mentioned studies revealed the potential and efficacy of data-driven models in the prediction of C_d for weirs and gates. Nevertheless, there is a noticeable gap in research about the application of integration of SVR base models with metaheuristic algorithms for the prediction of C_d in labyrinth sluice gates. In addition, to the best of the authors' knowledge the hybridization of SVR with evolutionary algorithms has not yet been used for the accurate estimation of C_d in labyrinth sluice gates. Therefore, based on the successful hybridization of GA and PSO with the SVR method to improve power prediction accuracy, the SVR-PSO and SVR-GA models were adopted to estimate C_d for labyrinth sluice gates. The integration of GA and PSO with SVR (i.e., SVR-GA and SVR-PSO) for modeling various engineering problems was illustrated successfully in many studies. These metaheuristic algorithms (i.e., GA and PSO) are selected for their ability to handle complex hydraulic engineering problems and for automatic optimization of internal parameters of the SVR model for precise prediction.

In summary, the previous literature confirmed the successful estimation of the discharge coefficient (C_d) of various gates and weirs using the SVR model. However, only a few studies have combined the SVR model with evolutionary algorithms to predict the C_d . In addition, according to the literature review, the application of the combined SVR model with two widely used optimization algorithms (GA and PSO) in various civil and water engineering problems has successfully improved the accuracy of predictions by identifying the optimal parameters of the SVR model. Hence, the present work investigated the capacity of the hybridization of the SVR model with two popular and efficient metaheuristic algorithms, including GA and PSO, to acquire the internal parameters of the SVR model for the estimation of C_d in labyrinth sluice gates. The performance of the produced models was evaluated using different statistical metrics and various graphs. The possibility of hybrid SVR models for precise estimation of C_d in labyrinth sluice gates was investigated in this work.

II. MATERIAL AND METHODS

The experimental dataset and suggested methods for modeling of C_d in labyrinth sluice gates are described in the following subsections.

A. Experimental dataset

Hashem *et al.*⁵ have supplied experimental data on the discharge coefficient of labyrinth sluice gates. Their laboratory experiments were conducted in a rectangular hydraulic channel of 10 m in length, 0.30 m in width, and 0.50 m in height. Each 0.004 m thick steel plate was used

as a gate; the labyrinth gates were set 5.50 m downstream from the canal inlet.

An ultrasonic depth sensor of accuracy ± 0.10 mm was used to measure the flow depth. The study involved 187 experimental tests involving labyrinth sluice gates, each with a triangular cross section. The parameter study involved three apex angles ($\theta = 45^\circ$, 60° , and 90°) and four gate opening heights ($G = 0.02$, 0.03 , 0.04 , and 0.05 m). In the tests, the discharge varied from 0.0057 to 0.0261 m³/s, and the upstream water depth varied from 0.072 to 0.045 m. The tests were conducted under free-flow conditions, enabling a thorough analysis of the labyrinth gate behavior. The gates were tested for exactly the same parameters in one-cycle ($N = 1$) and two-cycle ($N = 2$) configurations; leaf-plan paths of the gates were observed for both one-cycle and two-cycle configurations. Hashem *et al.* used the data to develop a robust empirical equation for estimating C_d . Notably, Hashem *et al.* reported that the efficiency of the labyrinth gates diminished as the apex angle increased.⁵ The C_d increases as the H/G ratio increases. However, the number of cycles (N) was found to have minimal influence on C_d . Figure 1 shows a definition sketch of the labyrinth gate and related key variables.

B. Dimensional analysis

The dimensional analysis provided by Hashem *et al.* indicated that the following relationship can be used to determine C_d for the labyrinth sluice gate:^{5,35}

$$C_d = f\left(\frac{H}{G}, \theta, N, Re, We\right), \quad (1)$$

where H is the upstream water depth, G is the gate opening, θ defines the apex angle ($\theta = \frac{L}{l}$), N is the number of cycles in the labyrinth gate. Re and We are denoted as Reynolds and Weber numbers, respectively. Furthermore, L represents the overall length of the labyrinth gate, while l denotes the length of the labyrinth gate's projection onto a vertical surface. It is worth mentioning that based on the laboratory conditions of Hashem *et al.*,⁵ the values of Re and We exceed 2000 and exceed 50, respectively, and can be negligible on C_d .³⁶ Therefore, the primary nondimensional parameters influencing C_d may be rearranged and simplified to give

$$C_d = f\left(\frac{H}{G}, \theta, N\right). \quad (2)$$

The statistical parameters of influential variables in Eq. (2) are provided in Table II.

Figure 2 presents the correlation matrix for key non-dimensional variables influencing the discharge coefficient (C_d) in labyrinth sluice gates. This figure visually represents the strength and direction of the relationships between variables affecting C_d .

As seen in Fig. 2, parameter H/G has a strong direct correlation with C_d . This suggests that as H/G increases, the discharge coefficient tends to increase. The parameter θ shows a negative correlation with C_d . This inverse relationship suggests that larger apex angles lead to reduced discharge. The parameter N appears to have a minimal effect on C_d , as reflected by the weak correlation.

C. Data-driven models and metaheuristic algorithms

Data-driven models employ advanced computational techniques that discover hidden patterns and relationships within datasets.

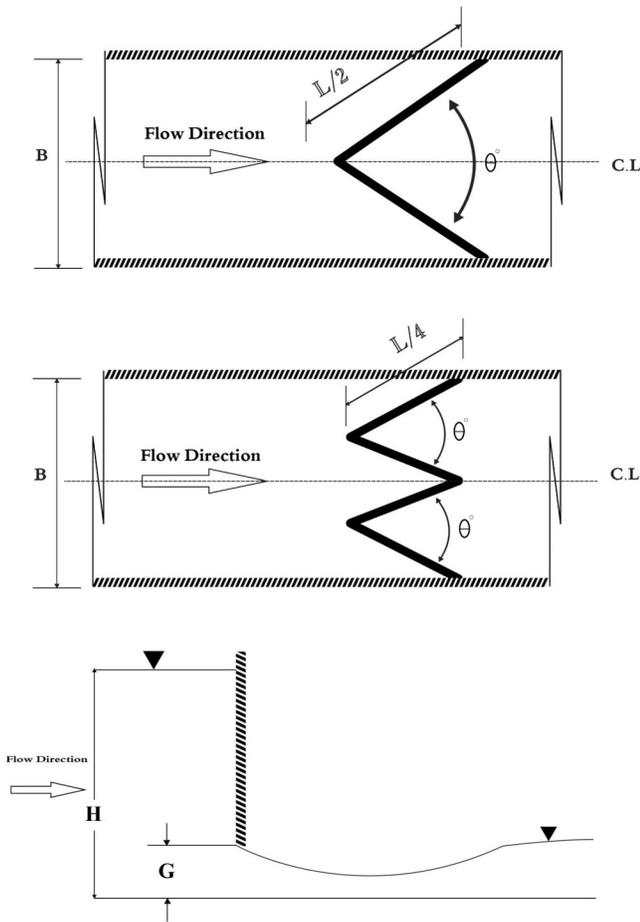


FIG. 1. The schematic view of the labyrinth sluice gate (adopted from Ref. 35).

Data-driven models are widely used and effective for finding the relationship between input and output parameters and extracting knowledge from datasets.³⁷ Metaheuristic algorithms are optimization methods that search the problem space for the optimal solution. They are particularly useful for solving complex problems where traditional optimization methods may fail.

TABLE II. The values of the main statistical parameters of Eq. (2).

Statistical index	C_d	$\frac{H}{G}$	θ	N
Minimum	0.548	2.075	0.785	1.000
Maximum	1.344	22.500	1.570	2.000
Average	0.803	7.061	1.082	1.481
Standard deviation	0.177	4.612	0.316	0.501
Kurtosis	0.333	1.491	-1.148	-2.016
Skewness	0.944	1.408	0.663	0.076
Median	0.767	5.500	1.047	1.000

Such algorithms provide optimal solutions by iteratively exploring and exploiting the search space to optimize an objective function.³⁸ Metaheuristics often are inspired by natural processes, such as biological evolution (genetic algorithm) and swarm behavior modeling (particle swarm optimization). The hybridization of data-driven models with metaheuristic algorithms can enhance the predictive accuracy of developed predictive models.³⁹ Metaheuristics schemes can optimize the internal parameters of data-driven models and improve their overall performance.⁴⁰

The present study used SVR as a data-driven model and PSO and GA metaheuristic algorithms for modeling C_d . The efficiency of these algorithms is demonstrated for solving complex problems, particularly in hydraulic engineering applications. In addition, the SVR model presented a good performance in earlier research that supported its application for discharge coefficient prediction in labyrinth sluice gates. SVR was chosen in this investigation because of its efficacy in modeling nonlinear interactions, proficiency in generalizing from limited datasets, and solid theoretical basis in machine learning.^{41,42} However, the accuracy results of SVR depend on the appropriate tuning of its hyperparameters.⁴³ Usually, the SVR parameters are explored using a conventional approach, such as the grid search method. However, this method is computationally costly and may fail to find the best values of the hyperparameters of the SVR model. Employing intelligent tuning techniques to determine the internal parameters of SVR can be implemented by metaheuristic algorithms such as PSO and GA. The determination of SVR parameters by metaheuristic algorithms frequently leads to improved accuracy and lower computing costs.⁴³

1. Support vector regression (SVR)

SVR is a supervised learning methodology designed for regression modeling. It is a modified version of the SVM algorithm that is mostly used for classification tasks.⁴⁴ SVR seeks to find a function, $f(x)$, that describes the relationship between the input

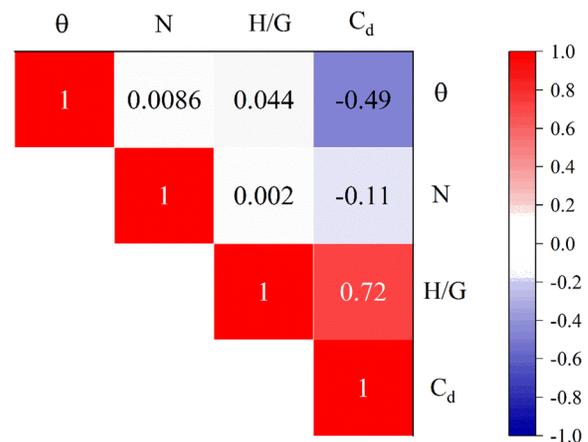


FIG. 2. Correlation matrix for key non-dimensional variables for the labyrinth sluice gate.

features, x , and the desired values, y , in order to reduce estimation errors. The principal expression in this methodology can be expressed as follows:⁴⁴

$$f(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b, \quad (3)$$

where $f(\mathbf{x})$ represents the function relating the output and input parameters, \mathbf{w}^T denotes the transpose of the weighting vector applied to the input data, φ signifies a nonlinear transformation function that maps the input variable x into an m -dimensional feature vector, and term b refers to the bias factor. The optimization problem for SVR can be expressed as⁴⁵

$$\min_{\mathbf{w}, b, \xi_i, \xi_i^*} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*), \quad (4)$$

subject to the constraints

$$\begin{cases} y_i - (\mathbf{w}^T \mathbf{x}_i + b) \leq \epsilon + \xi_i, \\ (\mathbf{w}^T \mathbf{x}_i + b) - y_i \leq \epsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0, \forall i = 1, 2, \dots, n. \end{cases}$$

where y_i is the target value for the i -th data sample, ϵ is the width of the ϵ -insensitive tube, ξ_i, ξ_i^* are slack variables for the i -th data sample, and C is the regularization parameter that controls the trade-off between model complexity and prediction error. Employing Lagrange multipliers, α_i, α_i^* and the Karush–Kuhn–Tucker (KKT) condition, Eq. (3) may be written as⁴⁵

$$f(x) = \sum_{i=1}^m (\alpha_i - \alpha_i^*) k(\mathbf{x}_i, \mathbf{x}_j) + b, \quad (5)$$

where k represents the kernel function.

Various kernel functions can be used to analyze and interpret data effectively in SVR modeling. Linear, sigmoid, polynomial, and radial basis functions (RBF) are among the routinely used kernel functions. The choice of appropriate kernel function is a crucial issue for the performance of the SVR model to detect complex data patterns. This work adopted RBF expressed as follows:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right), \quad (6)$$

where σ denotes the kernel parameter.

This work chooses the RBF kernel for SVR because of its strong capacity to describe the complex and nonlinear interactions in hydraulic engineering applications like discharge coefficient prediction. Moreover, the RBF has advantages such as being less sensitive to outliers and noise in experimental data, better model generalization, and a good history demonstrated in modeling hydraulic engineering, especially C_d prediction.^{46–48}

The performance of the SVR method depends on three primary variables, including $C, \sigma,$ and ϵ . A metaheuristic optimization algorithm can be used to assign the best values for these parameters. For the optimization process, the root mean squared error (RMSE) fitness function is used.

The RMSE fitness function is defined as follows:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (C_{di}^{obs} - C_{di}^{pre})^2}, \quad (7)$$

where C_{di}^{obs} and C_{di}^{pre} refer to measured and estimated values of C_d . The m parameter represents all measured data samples.

2. Genetic algorithm (GA)

GA is a metaheuristic optimization algorithm that is inspired by the process of natural selection and the genetic evolution of creatures.⁴⁹ The main procedures of the GA algorithm are as follows:⁵⁰ (i) Initialization: A set of potential solutions (chromosomes) is randomly generated within the problem search area. (ii) Selection: The fittest individuals are selected according to a fitness function value, which evaluates the quality of each candidate. (iii) Crossover (recombination): Selected individuals undergo crossover to produce offspring, combining features from parent solutions to generate diversity. (iv) Mutation: A mutation induces random changes in offspring, allowing the algorithm to explore diverse areas of the search space. (vi) Termination: The process repeats over generations until convergence criteria, like a maximum iteration limit or desired fitness level, are met.

3. Particle swarm optimization (PSO)

PSO was originally inspired by the social patterns exhibited by birds and fish.⁵¹ It relies on a group of particles passing the search area to find optimal solutions based on their individual experiences and interactions with others. The main steps of PSO are as follows:^{52,53} (i) Initialization: A collection of particles is dispersed within the search area at random. Every particle signifies a possible solution. (ii) Velocity and Position Update: Every particle modifies its position and velocity according to its individual best position (p_{best}) and the global best position (g_{best}) found by the swarm. (iii) Iteration: The movement of particles is influenced by cognitive (individual) and social (group) components, enabling exploration and exploitation of the search space. (iv) Termination: The algorithm stops when particles converge to a solution or a specified condition is met.

The principal equations of the PSO algorithm for updating the velocity and position of each particle are given in Eqs. (2) and (3), respectively,

$$\begin{aligned} v_i^{t+1} &= wv_i^t + c_1 r_1 (pbest_i - x_i^t) + c_2 r_2 (gbest - x_i^t), \\ x_i^{t+1} &= x_i^t + v_i^{t+1}, \end{aligned}$$

where v_i^{t+1} is the updated velocity of particle i at iteration $t + 1$, w is the inertia weight, c_1 and c_2 are the acceleration coefficients; r_1 and r_2 are random values that have uniform distribution between 0 and 1, x_i^t and x_i^{t+1} are the current position of particle i at iteration t and the new position by adding the updated velocity to the current position.

III. ASSESSMENT OF MODEL ACCURACY

The suggested data-driven models were evaluated using three common statistical indices. In this study, the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) were selected as the main statistical indicators to evaluate the performance of the data-driven models. These three metrics are used widely in machine learning and hydraulic modeling.⁵⁴ These statistical indices are defined as follows:

Coefficient of determination (R^2)

$$R^2 = \left(\frac{\sum_{i=1}^m \left[(C_{di}^{obs} - \overline{C_d^{obs}}) (C_{di}^{pre} - \overline{C_d^{pre}}) \right]}{\sqrt{\sum_{i=1}^m (C_{di}^{obs} - \overline{C_d^{obs}})^2} \cdot \sqrt{\sum_{i=1}^m (C_{di}^{pre} - \overline{C_d^{pre}})^2}} \right)^2. \quad (8)$$

Root mean square error (RMSE)

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (C_{di}^{obs} - C_{di}^{pre})^2}. \quad (9)$$

Mean absolute error (MAE)

$$MAE = \frac{1}{m} \sum_{i=1}^m |C_{di}^{obs} - C_{di}^{pre}|. \quad (10)$$

For an ideal model, RMSE and MAE values are zero, whereas R^2 is unity.

IV. MODELING C_d USING DATA-DRIVEN APPROACHES

The discharge coefficient (C_d) was simulated using the non-dimensional relationship given by Eq. (2). The dataset was partitioned into training and testing portions, assigning 70% for model creation and reserving 30% for validation. In the present study, the two meta-heuristic algorithms, PSO and GA, were employed to tune the three internal variables of SVR algorithms (i.e., C , σ , and ϵ) to predict the C_d of the labyrinth sluice gate. The RMSE fitness function was adopted for the optimization process.

A. The main steps of the development of the SVR-PSO model

The SVR-PSO approach is developed to determine SVR method parameters using the GA algorithm. The SVR-PSO model development procedure is as follows:

- **Initialization:** The PSO algorithm commences by initializing a random population of particles. These particles represent potential solutions, each associated with specific SVR parameters (e.g., kernel function parameter σ , regularization term C , and epsilon ϵ). The initial values are randomly assigned within predefined bounds.
- **Initial SVR model training:** The model uses the training dataset to predict target values. Based on the predicted outputs, the model performance is evaluated using a fitness function (i.e., RMSE) to determine the accuracy of SVR-PSO for each particle parameter set.
- **Velocity and position update:** Regarding the stopping criterion, the velocity and position of each particle are updated iteratively. A velocity update equation is used to determine the next movement direction of the particle by considering its personal best position, the global best position, and its present velocity. A position update then adjusts the particle's position in the parameter search space based on the updated velocity.
- **Solution refinement:** The best solution is refined during each iteration. This involves updating the global best solution found by the swarm and the personal best solution for each particle.

- **Termination:** The optimization process iterates until the termination criterion is achieved. The termination criterion may consider predefined maximum iterations or reaching a low error value.

Integrating SVR with PSO can adjust the SVR parameters and result in higher accuracy for the prediction of C_d . The flow chart of the modeling procedure is illustrated in Fig. 3.

B. The main steps of the development of the SVR-GA model

The hybrid SVR-GA model is developed to determine the SVR parameters using the GA algorithm. The development of the SVR-GA model involves the following main steps:

- **Initialization:** The GA algorithm starts by randomly creating an initial population of decision variables of the SVR model, including kernel function parameter (σ), regularization term (C), and epsilon (ϵ).
- **Initial SVR model training:** The SVR model is trained using an initial parameter, and the corresponding fitness function value (i.e., RMSE) is calculated for each individual in the population.
- **Crossover operation:** The crossover operator is applied to the existing population (selected parent) to create offspring by combining the genetic material (SVR parameters). The resulting offspring are used for training the SVR model and computing their fitness function values.

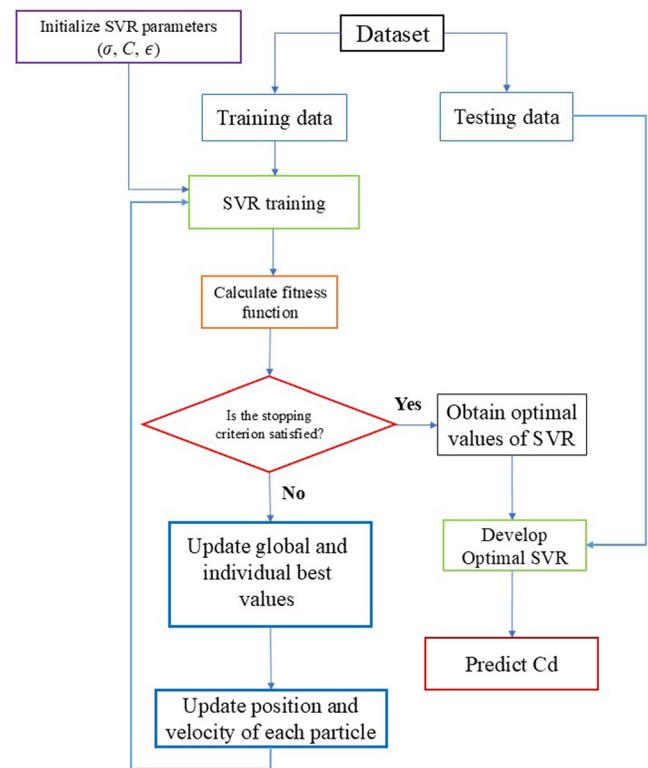


FIG. 3. Flowchart of SVR-PSO for estimation of value of C_d .

- **Mutation operation:** The mutation operator introduces variability in the population by making random variations to the offspring. The mutated population is used for training the SVR model, and fitness function values for each mutant are calculated.
- **Population sorting and repository update:** The total population, consisting of parents, their offspring, and mutants, is sorted according to its fitness function values. The set of the best solutions is updated to ensure the maintenance of the best parameter sets.
- **Termination:** The optimization process continues until the termination criterion is satisfied. This criterion may reach the predefined maximum number of iterations or meet a specified error limit.

This SVR-GA approach can efficiently explore the parameter space and enhance model performance. The flow chart of the SVR-GA model is illustrated in Fig. 4.

V. RESULTS AND DISCUSSION

In this study, an optimized SVR model was proposed for the prediction of C_d . The optimal values obtained for the SVR-PSO and SVR-GA models, including C , σ , and ϵ obtained during the optimization process are tabulated in Table III.

Table IV presents the values of the three statistical performance indicators, R^2 , RMSE, and MAE, used to evaluate the SVR-based models for the prediction of C_d .

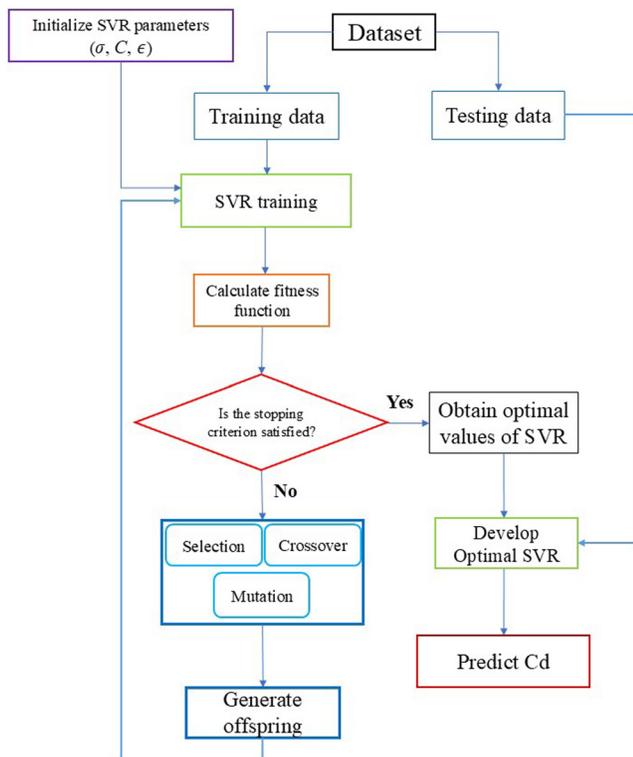


FIG. 4. Flowchart of SVR-GA for estimation of the value of C_d .

TABLE III. The optimal values of C , σ , and ϵ for the hybridized SVR models.

Approach	C	ϵ	σ
SVR-GA	56.0973	0.0182	2.6242
SVR-PSO	63.6851	0.0100	2.3548

As observed in Table IV, the SVR model achieves an R^2 value of 0.9184 for the training dataset, indicating a good fit between the predicted and actual values. However, the RMSE of 0.0538 and MAE of 0.0436 suggest that the SVR model has a relatively higher prediction error compared to the hybrid models. For the testing dataset, the value of R^2 for the SVR model decreases slightly to 0.8958, and the RMSE and MAE values increase to 0.0575 and 0.0468, respectively. These metrics for the SVR model reflect less accurate predictions in the testing phase compared to other hybrid models. The SVR-GA hybrid model exhibits significant improvement over the standalone SVR. In the training dataset, the SVR-GA model achieves an R^2 value of 0.9736, demonstrating a much closer fit to the actual values. The corresponding RMSE and MAE values are 0.0285 and 0.0224, respectively, indicating lower prediction errors. In addition, in testing datasets, the SVR-GA model maintains its accuracy with an R^2 value of 0.9685, an RMSE value of 0.0324, and an MAE value of 0.0257. These metrics confirm the more accurate SVR-GA model compared to the standalone SVR model. Based on the statistical metrics values, it confirmed that the SVR-PSO model has better accuracy compared to both the SVR-GA and the standalone SVR.

In the training dataset, the SVR-PSO approach reached the highest R^2 value of 0.9778 and the lowest RMSE (0.0270) and MAE (0.0200). For the testing dataset, the SVR-PSO model again demonstrates the best performance with an R^2 value of 0.9732. In addition, the RMSE and MAE values are 0.0287 and 0.0209, respectively. These metrics confirm the outstanding predictive accuracy of the SVR-PSO model in both the training and testing datasets. Compared to the SVR model, the SVR-PSO approach achieved a 49.8% reduction in RMSE for the training dataset and a 50.6% reduction for the testing dataset. Similarly, the SVR-GA model achieved significant RMSE reductions of 47.0% (training) and 43.7% (testing) compared to a single SVR model. Both the SVR-GA and SVR-PSO models exhibit strong generalization, as indicated by the minimal change in accuracy between the training and testing stages. The values of statistical metrics reveal that the GA and PSO optimization algorithms are significantly effective in improving the precision of the SVR model by optimizing its parameters. The

TABLE IV. Values of R^2 , RMSE, and MAE statistical indices for the SVR-based models in determining C_d for a labyrinth weir.

Model	R^2	RMSE	MAE
SVR (training data)	0.9184	0.0538	0.0436
SVR (testing data)	0.8958	0.0575	0.0468
SVR-GA (training data)	0.9736	0.0285	0.0224
SVR-GA (testing data)	0.9685	0.0324	0.0257
SVR-PSO (training data)	0.9778	0.0270	0.0200
SVR-PSO (testing data)	0.9732	0.0287	0.0209

16 April 2025 09:41:55

SVR-PSO model has the best performance for the prediction of C_d . Among the evaluation of SVR-based models, SVR-PSO is the most accurate model for predicting C_d , as it achieves the highest R^2 value and the lowest RMSE and MAE values for both the training and testing stages, compared to SVR and SVR-GA models.

For further analysis of developed hybrid SVR models, several graphical techniques are used to examine the performance of proposed models for the prediction of C_d . Therefore, three common graphical evaluation plots, including scatterplots, Taylor, and violin plots, were used. A scatterplot is a type of data visualization used to display relationships or correlations between measured and predicted values obtained from SVR-based models for the prediction of C_d . The scatterplot visually compares the results of each model to the measured values. The accurate model has the results near the 45° line (dashed point line), indicating that the results of predictions align with the measured values. Figure 5 displays scatterplots of SVR-based models for training and testing datasets.

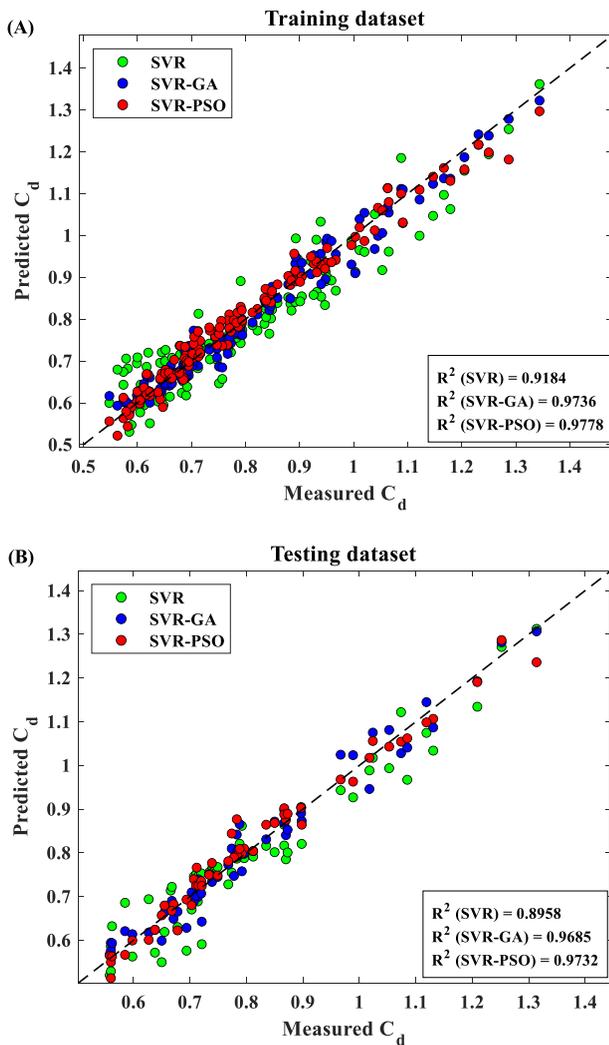


FIG. 5. Scatter plots of SVR-based models in (a) training and (b) testing stages.

The scatterplot for the training and testing datasets shows that the predicted values of SVR-PSO are closer to the measured C_d values compared to the SVR and SVR-GA models. Therefore, it is suggested that the SVR-PSO is the most accurate of the three models. On the other hand, the SVR-GA model also performs well but slightly less accurately than the SVR-PSO model. Finally, the standalone SVR model has the most scattered predictions, indicating that it is the least accurate of the three models.

Taylor and violin plots are advanced statistical graphs that have been widely used in evaluating machine learning models in hydraulic and environmental modeling. The Taylor plot is a statistical diagram used to graphically summarize the performance of the SVR models by comparing their results to observed data. It simultaneously displays three key metrics, including the correlation coefficient, the RMSE, and the standard deviation.^{55,56} In this plot, the observed data serve as a reference, and the results of other models are represented as points whose position refers to their correlation and standard deviation relative to the observed values. A model that is closer to the reference point (the observed values) is more accurate. This allows for the easy evaluation of multiple models simultaneously, making it easier to determine the best model.

Figure 6 shows the Taylor plots obtained for the model evaluation of the discharge coefficient for the labyrinth gate.

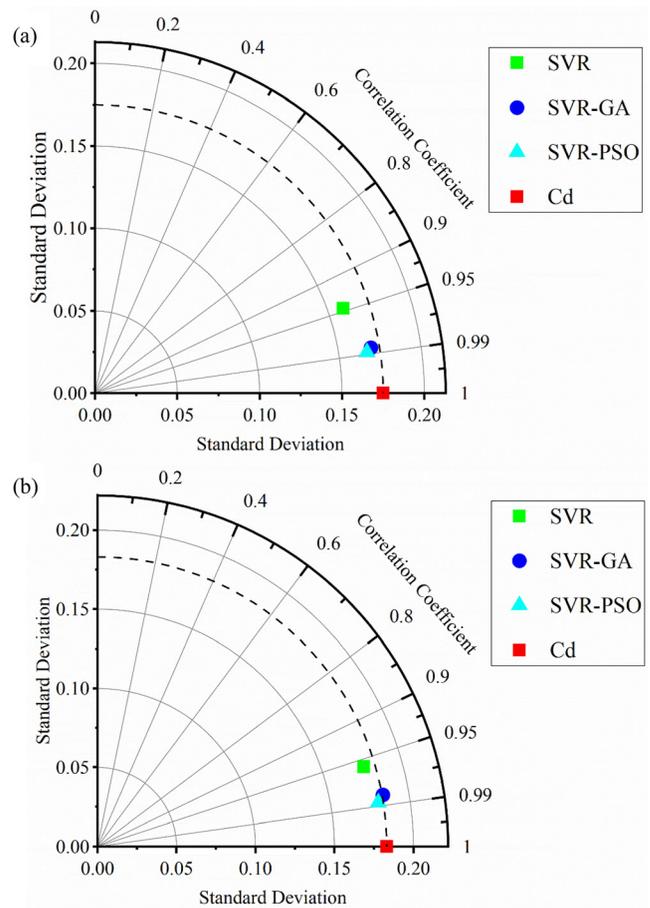


FIG. 6. Taylor plots of SVR-based models in (a) training and (b) testing stages.

16 April 2025 09:41:55

As seen in the Taylor plots, the SVR-PSO results lie closer to the observed C_d , confirming the better performance of the SVR-PSO approach than the other models.

The Taylor plot is used to evaluate the performance of three SVR-based models in predicting C_d for training and testing stages [Figs. 6(a) and 6(b)]. SVR-PSO lies closest to the observed point, followed by SVR-GA, then SVR, of training data with minimal error. SVR-PSO again positions closest to the observed point, followed by SVR-GA, with SVR being the farthest. Figure 6 illustrates that SVR-PSO outperforms SVR-GA and SVR in both training and testing, with points closest to the observed C_d , reflecting the highest accuracy for the prediction of C_d .

A violin plot is a data visualization tool that detailed representation of the distribution of data.⁵⁷ Violin plots allow a more comprehensive visualization of data distribution. By visualizing multiple models simultaneously, the violin plot provides a clear comparison of prediction accuracy and consistency. Figure 7 presents violin plots for

evaluating the performance of different predictive models used for estimating the C_d in labyrinth sluice gates. The violin plot helps in comparing the distribution of predicted C_d values for different models by assessing model consistency by examining the spread and shape of each violin. By simultaneously comparing the predicted data with the observed data, the model performance is assessed, helping to identify which model best aligns with the target distribution. The violin plot is a data visualization tool that indicates the distribution and spread of a dataset. Figure 7 shows the violin plots obtained for the training and testing datasets using the three SVR approaches. Differences in violin shapes between observed and predicted data reveal that the model with similar shapes has a good fit.

As observed in violin plots, the shape of SVR-PSO during the training and testing stages is closer to the C_d values.

It is confirmed that the SVR-PSO model had better results than the SVR and SVR-GA models in both training and testing phases, as indicated by its violin plot closely aligning with the observed data. The symmetrical shape of the SVR-PSO violin further reflects minimal error and high precision. Overall, graphical representations demonstrated that the SVR-PSO model is the most dependable and accurate model for the prediction of values of C_d .

VI. SHAPLEY ADDITIVE EXPLANATIONS (SHAP) ANALYSIS

SHAP method explains the contribution of every feature to the predictions of a machine learning model based on cooperative game theory.⁵⁸ Therefore, this approach leads to enhanced model interpretability of the black-box models, such as the SVR technique. This analysis is an organized assessment of the effect of every feature on the predicted performance of the model. The SHAP value for a feature x_i in the prediction of $f(x)$ is determined as follows:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)],$$

where Φ_i is the SHAP value for input variable i , S denotes the subset of all input variables excluding feature i , and $f(S)$ represents the output model when only the features in S are considered. N is the total number of input variables. In the present study, SHAP analysis was used to identify the rank of each feature based on its importance in predicting C_d . Figure 8 shows the SHAP analysis results related to the SVR-PSO model.

The input variables $x_3=H/G$, $x_1=\theta$, and $x_2=N$. It is clear that H/G and N variables have the highest and lowest effects on C_d , respectively.

VII. SENSITIVITY ANALYSIS

Sensitivity analysis is an essential technique for evaluating the influence of input parameters on model predictions. In this study, a leave-one-variable-out sensitivity analysis was performed using SVR-PSO models to assess the effect of each input variable on the discharge coefficient. This method involves systematically removing one variable at a time and analyzing the impact on model accuracy, which is a widely used approach in machine learning and hydraulic modeling. The sensitivity analysis using SVR-PSO models was conducted to determine the effect of each input factor on C_d . Therefore, the SVR-PSO models were developed by removing each input variable from Eq. (2). The results of the sensitivity analysis indicated the specific effect of excluding each input variable on the output parameter.⁵⁹ The results

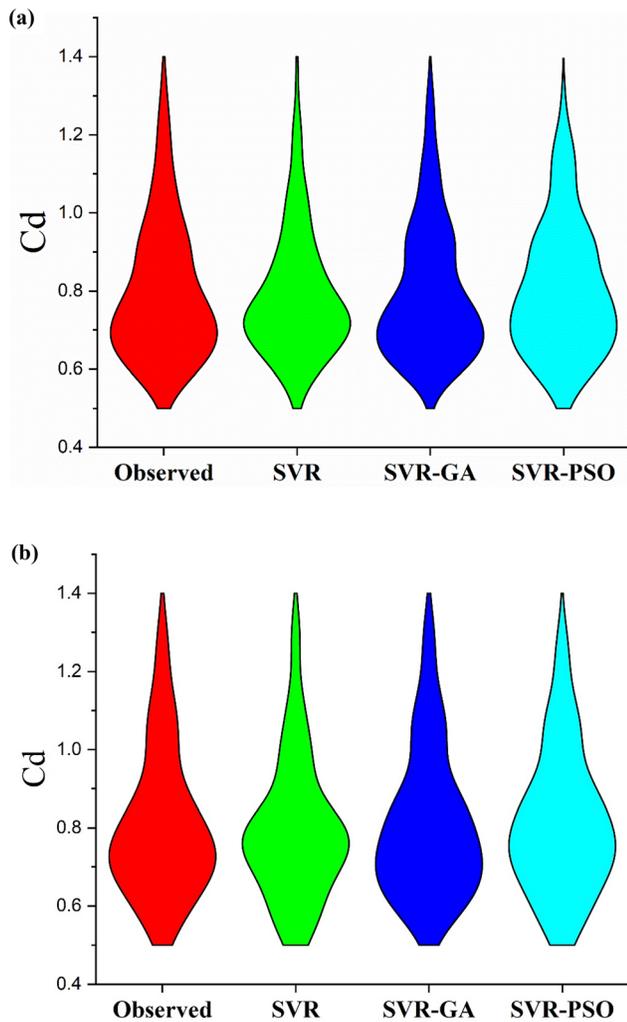


FIG. 7. Violin plots of SVR-based models in (a) training and (b) testing stages.

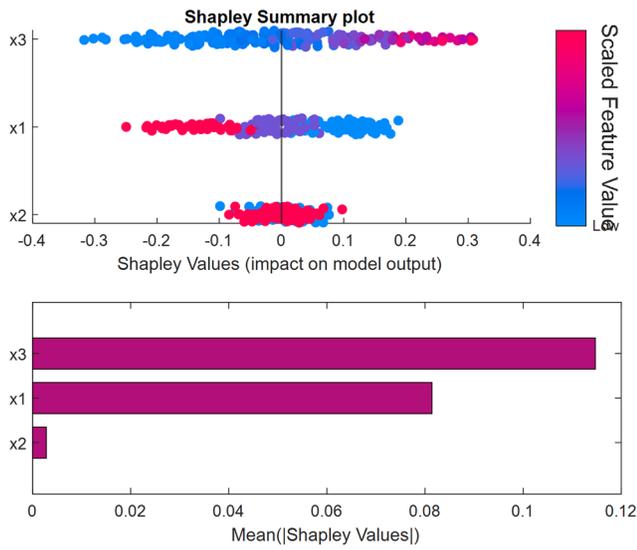


FIG. 8. The results of SHAP analysis for the estimation of C_d .

TABLE V. Sensitivity analysis of the SVR-PSO model for the estimation of C_d .

Model	Remove variable	R^2	RMSE	MAE
$C_d = f(\theta, N)$	$\frac{H}{G}$	0.3460	0.1570	0.1192
$C_d = f\left(\frac{H}{G}, N\right)$	θ	0.6956	0.1161	0.0878
$C_d = f\left(\frac{H}{G}, \theta\right)$	N	0.9372	0.0425	0.0383

of the sensitivity analysis are tabulated in Table V for the testing dataset.

As shown in Table V, it is clear that the H/G variable has the most significant effect on C_d , with an $R^2 = 0.3460$, $RMSE = 0.1570$, and an $MAE = 0.1192$. Furthermore, the N variable has the least influence on C_d ($R^2 = 0.9372$, $RMSE = 0.0425$, and $MAE = 0.0383$). It is important to note that this finding is consistent with the findings of Hashem *et al.*⁵

TABLE VI. The regression-based equation for prediction of C_d .

Approach	Equation
LR ³⁵	$C_d = 0.9776 - 0.2977 \times \theta - 0.0368 \times N + 0.02841 \times \left(\frac{H}{G}\right)$
SPR ³⁵	$C_d = 0.8361 - 0.543 \times \theta + 0.10224 \times \frac{H}{G} - 0.0155 \times N + 0.2033 \times \theta^2 - 0.001669 \times \left(\frac{H}{G}\right)^2 - 0.03384 \times \left(\theta \times \frac{H}{G}\right) - 0.00271 \times \left(N \times \frac{H}{G}\right)$
NLR ⁵	$C_d = 0.511 \times (H/G)^{0.268} \times \theta^{-0.566} \times N^{-0.066}$

TABLE VII. Values of R^2 , RMSE, and MAE statistical indices for the SVR-based models in determining C_d .

Model	R^2	RMSE	MAE
LR ³⁵	0.8012	0.0787	0.0622
SPR ³⁵	0.9509	0.0396	0.0317
NLR ⁵	0.9060	0.0548	0.0436
SVR-GA (present study)	0.9721	0.0296	0.0226
SVR-PSO (present study)	0.9742	0.0285	0.0202

VIII. COMPARISON RESULTS WITH EXISTING PREVIOUS STUDIES FOR ESTIMATION OF C_d

The results of the present study are compared with regression-based models and data-driven methods for the prediction of C_d that existed in the literature. It is worth mentioning that the previous study conducted by Hashem *et al.* used the same datasets as used in the present study.^{5,35} This comparison with existing models makes it possible to evaluate the accuracy of proposed SVR-based models compared to the existing methods for the estimation of C_d in labyrinth sluice gates.

A. Comparison results with existing regression-based models for estimation of C_d

The recently proposed regression-based equations for the prediction of C_d in labyrinth sluice gates are listed in Table VI.

The values of statistical metrics for the regression-based equations for the estimation of values of C_d are listed in Table VII for all datasets.

As seen in Table VII, both hybrid SVR models have the highest accuracy compared to other regression-based models.

B. Comparison results with existing data-driven models for estimation of C_d

The results of the data-driven models suggested for C_d prediction are listed in Table VIII. The previous study by Hashem *et al.*⁵ provided the results of K^* , decision tree (DT), and M5P algorithms for training and testing datasets. For comparison of the previously suggested data-driven results with the outcomes of SVR-GA and SVR-PSO provided in this study, the objective function (OBF) criterion is used. The OBF criterion is defined as follows:⁶⁰

TABLE VIII. The R2, RMSR, and MAE values of existing data-driven models for the prediction of C_d .

Model	Dataset	R ²	RMSE	MAE	Train/test ratio
K* ³⁵	Training	0.9752	0.0391	0.0291	50/50
K* ³⁵	Testing	0.9409	0.0579	0.0436	
M5P ³⁵	Training	0.9502	0.0390	0.0307	50/50
M5P ³⁵	Testing	0.9469	0.0422	0.0348	
DT ³⁵	Training	0.9357	0.0427	0.0325	50/50
DT ³⁵	Testing	0.5676	0.1234	0.0709	
SVR-GA (Present study)	Training	0.9736	0.0285	0.0224	70/30
SVR-GA (Present study)	Testing	0.9685	0.0324	0.0257	
SVR-PSO (Present study)	Training	0.9778	0.0270	0.0200	70/30
SVR-PSO (Present study)	Testing	0.9732	0.0287	0.0209	

$$\begin{aligned}
 OBF = & \left(\frac{n_{\text{training data}}}{n_{\text{all data}}} \times \frac{RMSE_{\text{training data}} + MAE_{\text{training data}}}{R^2_{\text{training data}} + 1} \right) \\
 & + \left(\frac{n_{\text{testing data}}}{n_{\text{all data}}} \times \frac{RMSE_{\text{testing data}} + MAE_{\text{testing data}}}{R^2_{\text{testing data}} + 1} \right) \quad (11)
 \end{aligned}$$

where $n_{\text{training data}}$, $n_{\text{testing data}}$, and $n_{\text{all data}}$ refer to the number of training samples, testing, and total datasets. The best optimal approach reaches the lowest value of the OBF criterion. As observed in Eq. (11), the OBF criterion simultaneously considered the R², RMSE, and MAE values for both the training and testing datasets. Table VIII provides the values of the main parameters of statistical metrics (i.e., R², RMSE, and MAE) for the calculation of the values of the OBF criterion for data-driven models.

Regarding the OBF criterion and the statistical metrics tabulated in Table VIII, the values of OBF for the proposed data-driven models are calculated and displayed in Fig. 9.

As observed in Fig. 9, the values of the OBF criterion for the SVR-PSO and SVR-GA models have minimum values compared to K*, DT, and M5P methods. In addition, the SVR-PSO model has the lowest value of OBF (=0.0245) compared to other models, confirming the best performance of SVR-PSO for the estimation of C_d .

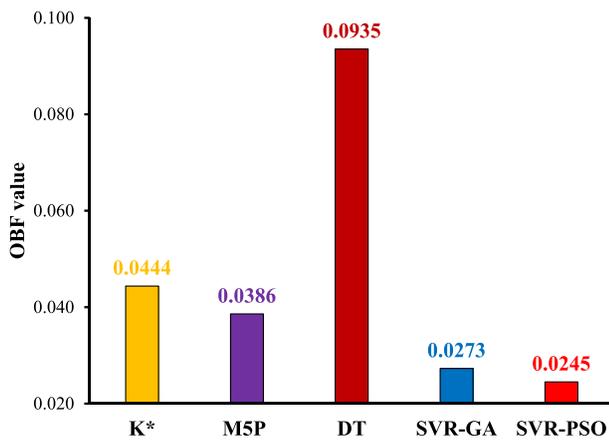


FIG. 9. The values of OBF values for the proposed data-driven models for the prediction of C_d .

IX. SUMMARY AND CONCLUSIONS

The present research examined the capability of hybridizing the SVR model with two widely used metaheuristic algorithms, including PSO and GA algorithms, to predict the discharge coefficient (C_d) of the labyrinth sluice gate. For this purpose, the SVR models were developed using reliable datasets and dimensionless parameters for the estimation of C_d in the labyrinth sluice gate. Graphical plots and statistical evaluations assessed the SVR, SVR-GA, and SVR-PSO model accuracy. The highest correlation coefficient ($R^2 = 0.9742$) and the lowest RMSE (0.0285) and MAE (0.0202) are achieved by SVR-PSO, which also has the greatest overall performance. This suggests that PSO efficiently optimizes SVR parameters, resulting in more precise C_d predictions. In comparison to standalone SVR, SVR-GA also enhances efficacy; however, it fails to attain the same level of accuracy as SVR-PSO. The highest RMSE and lowest correlation coefficient are observed in the standalone SVR model, which implies that parameter tuning through GA or PSO improves prediction accuracy considerably. According to these tests, the SVR-PSO model is the most accurate at estimating C_d . Furthermore, the SVR models' outcomes were compared to previously published regression-based equations and data-driven models. The research results indicated that the accuracy of the standalone SVR model was significantly enhanced by the integration of SVR with metaheuristic optimization algorithms, including GA and PSO. Moreover, sensitivity and SHAP analysis revealed that the H/G parameter was the main factor of C_d estimation, which matched the findings of earlier investigations. The present work showed the possibilities of integrating SVR with optimization techniques for hydraulic parameter modeling in hydraulic engineering. For future research, the application of white box data-driven models, such as multivariate adaptive regression splines and group method handling for predicting C_d , can be investigated. In addition, the hybridization of deep learning and metaheuristic optimization can be investigated to assess the predictive capability for C_d prediction. The proposed methodology in the present research could be extended to other hydraulic structures for the determination of hydraulic parameters. The primary limitation of this study is the reliance on laboratory data; therefore, it is recommended that the proposed method be assessed using prototype data.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant Nos. 52179060 and 52479060).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Aliasghar Azma: Investigation (equal). **Alistair G. L. Borthwick:** Investigation (equal). **Reza Ahmadian:** Investigation (equal). **Yakun Liu:** Funding acquisition (equal); Resources (equal). **Di Zhang:** Investigation (equal).

DATA AVAILABILITY

The data that support the findings of this study are available within the article.

REFERENCES

- ¹X. Wang, X. Yan, H. Duan, X. Liu, and E. Huang, "Experimental study on the influence of river flow confluences on the open channel stage-discharge relationship," *Hydrol. Sci. J.* **64**(16), 2025–2039 (2019).
- ²A. Azma, M. Tavakol Sadrabadi, Y. Liu, M. Azma, D. Zhang, Z. Cao, and Z. Li, "Boosting ensembles for estimation of discharge coefficient and through flow discharge in broad-crested gabion weirs," *Appl. Water Sci.* **13**(2), 45 (2023).
- ³B. M. Crookston and B. P. Tullis, "Hydraulic design and analysis of labyrinth weirs. I: Discharge relationships," *J. Irrig. Drain. Eng.* **139**(5), 363–370 (2013).
- ⁴A. Parsaie and A. H. Haghiabi, "Improving modelling of discharge coefficient of triangular labyrinth lateral weirs using SVM, GMDH and MARS techniques," *Irrig. Drain.* **66**(4), 636–654 (2017).
- ⁵T. Hashem, A. Y. Mohammed, and T. J. Alfatlawi, "Hydraulic characteristics of labyrinth sluice gate," *Flow Meas. Instrum.* **96**, 102556 (2024).
- ⁶M. Bijankhan, V. Ferro, and S. Kouchakzadeh, "New stage-discharge relationships for radial gates," *J. Irrig. Drain. Eng.* **139**(5), 378–387 (2013).
- ⁷P. Novak, A. I. B. Moffat, C. Nalluri, and R. Narayanan, *Hydraulic Structures* (CRC Press, 2010).
- ⁸B. M. Savage and M. C. Johnson, "Flow over ogee spillways: Physical and numerical model case study," *J. Hydraul. Eng.* **127**(8), 640–649 (2001).
- ⁹G. Yang, R. Xu, Y. Tian, S. Guo, J. Wu, and X. Chu, "Data-driven methods for flow and transport in porous media: A review," *Int. J. Heat Mass Transfer* **235**, 126149 (2024).
- ¹⁰M. Haghbin and A. Sharafati, "A review of studies on estimating the discharge coefficient of flow control structures based on the soft computing models," *Flow Meas. Instrum.* **83**, 102119 (2022).
- ¹¹H. M. Azamathulla, Z. Ahmad, and A. Ab. Ghani, "Computation of discharge through side sluice gate using gene-expression programming," *Irrig. Drain.* **62**(1), 115–119 (2013).
- ¹²H. Azimi, H. Bonakdari, and I. Ebtehaj, "A highly efficient gene expression programming model for predicting the discharge coefficient in a side weir along a trapezoidal canal," *Irrig. Drain.* **66**(4), 655–666 (2017).
- ¹³M. A. Ghorbani, F. Salmasi, M. K. Saggi, A. S. Bhatia, E. Kahya, and R. Norouzi, "Deep learning under H₂O framework: A novel approach for quantitative analysis of discharge coefficient in sluice gates," *J. Hydroinf.* **22**(6), 1603–1619 (2020).
- ¹⁴K. Roushangar, A. Alirezazadeh Sadaghiani, and S. Shahnazi, "Novel application of robust GWO-KELM model in predicting discharge coefficient of radial gates: A field data-based analysis," *J. Hydroinf.* **25**(2), 275–299 (2023).
- ¹⁵A. Marashi, S. Kouchakzadeh, and H. A. Yonesi, "Rotary gate discharge determination for inclusive data from free to submerged flow conditions using ENN, ENN-GA, and SVM-SA," *J. Hydroinf.* **25**(4), 1312–1328 (2023).
- ¹⁶A. Parsaie, A. H. Haghiabi, S. Emamgholizadeh, and H. M. Azamathulla, "Prediction of discharge coefficient of combined weir-gate using ANN, ANFIS and SVM," *Int. J. Hydrol. Sci. Technol.* **9**(4), 412–430 (2019).
- ¹⁷A. Parsaie, A. H. Haghiabi, M. Saneie, and H. Torabi, "Predication of discharge coefficient of cylindrical weir-gate using adaptive neuro fuzzy inference systems (ANFIS)," *Front. Struct. Civ. Eng.* **11**, 111–122 (2017).
- ¹⁸F. Salmasi and J. Abraham, "Expert system for determining discharge coefficients for inclined slide gates using genetic programming," *J. Irrig. Drain. Eng.* **146**(12), 06020013 (2020).
- ¹⁹J. H. Sahib, L. K. Al-Waeli, and A. H. Jaber Al Rammahi, "Utilization of ANN technique to estimate the discharge coefficient for trapezoidal weir-gate," *Open Eng.* **12**(1), 142–150 (2022).
- ²⁰M. Nouri, P. Sihag, O. Kisi, M. Hemmati, S. Shahid, and R. M. Adnan, "Prediction of the discharge coefficient in compound broad-crested-weir gate by supervised data mining techniques," *Sustainability* **15**(1), 433 (2022).
- ²¹H. Karami, S. Karimi, M. Rahmanimanes, and S. Farzin, "Predicting discharge coefficient of triangular labyrinth weir using support vector regression, support vector regression-firefly, response surface methodology and principal component analysis," *Flow Meas. Instrum.* **55**, 75–81 (2017).
- ²²K. Roushangar, M. T. Alami, J. Shiri, and M. M. Asl, "Determining discharge coefficient of labyrinth and arced labyrinth weirs using support vector machine," *Hydrol. Res.* **49**(3), 924–938 (2018).
- ²³H. Azimi, H. Bonakdari, and I. Ebtehaj, "Design of radial basis function-based support vector regression in predicting the discharge coefficient of a side weir in a trapezoidal channel," *Appl. Water Sci.* **9**, 1–12 (2019).
- ²⁴A. H. Zaji, H. Bonakdari, and S. Shamsirband, "Support vector regression for modified oblique side weirs discharge coefficient prediction," *Flow Meas. Instrum.* **51**, 1–7 (2016).
- ²⁵A. H. Zaji, H. Bonakdari, S. R. Khodashenas, and S. Shamsirband, "Firefly optimization algorithm effect on support vector regression prediction improvement of a modified labyrinth side weir's discharge coefficient," *Appl. Math. Comput.* **274**, 14–19 (2016).
- ²⁶A. H. Zaji and H. Bonakdari, "Optimum support vector regression for discharge coefficient of modified side weirs prediction," *INAE Lett.* **2**, 25–33 (2017).
- ²⁷S. Doroudi, A. Sharafati, and S. H. Mohajeri, "Estimation of daily suspended sediment load using a novel hybrid support vector regression model incorporated with observer-teacher-learner-based optimization method," *Complexity* **2021**(1), 5540284.
- ²⁸S. Doroudi, A. Sharafati, and S. H. Mohajeri, "Satellite-based estimation of daily suspended sediment load using hybrid intelligent models," *Hydrol. Sci. J.* **68**(2), 307–324 (2023).
- ²⁹M. M. Hameed, M. A. Abed, N. Al-Ansari, and M. K. Alomar, "Predicting compressive strength of concrete containing industrial waste materials: Novel and hybrid machine learning model," *Adv. Civil Eng.* **2022**(1), 5586737.
- ³⁰M. Rahgoshay, S. Feiznia, M. Arian, and S. A. A. Hashemi, "Modeling daily suspended sediment load using improved support vector machine model and genetic algorithm," *Environ. Sci. Pollut. Res.* **25**, 35693–35706 (2018).
- ³¹H. Nezaratian, J. Zahiri, M. F. Peykani, A. Haghiabi, and A. Parsaie, "A genetic algorithm-based support vector machine to estimate the transverse mixing coefficient in streams," *Water Qual. Res. J.* **56**(3), 127–142 (2021).
- ³²M. S. Kazemi, M. E. Banihabib, and J. Soltani, "A hybrid SVR-PSO model to predict concentration of sediment in typical and debris floods," *Earth Sci. Inf.* **14**(1), 365–376 (2021).
- ³³S. L. S. Yong, J. L. Ng, Y. F. Huang, C. K. Ang, N. Ahmad Kamal, M. Mirzaei, and A. Najah Ahmed, "Enhanced daily reference evapotranspiration estimation using optimized hybrid support vector regression models," *Water Resour. Manage.* **38**, 4213–4229 (2024).
- ³⁴S. Mozaffari, S. Javadi, H. K. Moghaddam, and T. O. Randhir, "Forecasting groundwater levels using a hybrid of support vector regression and particle swarm optimization," *Water Resour. Manage.* **36**(6), 1955–1972 (2022).
- ³⁵T. Hashem, I. K. Harith, N. H. Alrubaye, A. Y. Mohammed, and M. L. Hussien, "Unlocking precision in hydraulic engineering: Machine learning insights into labyrinth sluice gate discharge coefficients," *J. Hydroinf.* **26**(11), 2883–2901 (2024).
- ³⁶O. Machiels, M. Piroton, A. Pierre, B. Dewals, and S. Erpicum, "Experimental parametric study and design of Piano Key Weirs," *J. Hydraul. Res.* **52**(3), 326–335 (2014).
- ³⁷T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning* (Springer, 2009).
- ³⁸F. Glover and G. Kochenberger, *Handbook of Metaheuristics* (Springer, 2003).
- ³⁹A. Azma, Y. Liu, M. Eftekhari, and D. Zhang, "Comparison of hybrid deep learning models for estimation of the time-dependent scour depth downstream of river training structures," *Phys. Fluids* **36**(10), 101911 (2024).

- ⁴⁰M. Alrashidi and S. Rahman, "Short-term photovoltaic power production forecasting based on novel hybrid data-driven models," *J. Big Data* **10**(1), 26 (2023).
- ⁴¹V. N. Vapnik, *The Nature of Statistical Learning Theory* (Springer, New York, 2000).
- ⁴²A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Stat. Comput.* **14**(3), 199–222 (2004).
- ⁴³Z. Li, Z. Ge, Q. Deng, Z. Zhou, C. Zhu, L. Liu, and Z. Yao, "Predicting the depth of rock cutting by abrasive water jet using support vector machine optimized with whale optimization algorithm," *Phys. Fluids* **36**(12), 126633 (2024).
- ⁴⁴H. Tabari, C. Martinez, A. Ezani, and P. Hosseinzadeh Talaei, "Applicability of support vector machines and adaptive neurofuzzy inference system for modeling potato crop evapotranspiration," *Irrig. Sci.* **31**, 575–588 (2013).
- ⁴⁵J. Shawe-Taylor and S. Sun, "A review of optimization methodologies in support vector machines," *Neurocomputing* **74**(17), 3609–3618 (2011).
- ⁴⁶B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (MIT Press, 2002).
- ⁴⁷C. W. Hsu, C. C. Chang, and C. J. Lin, *A Practical Guide to Support Vector Classification* (Department of Computer Science, National Taiwan University, 2003).
- ⁴⁸Y. B. Dibike, D. P. Solomatine, M. B. Abbott, and M. Bruen, "Model induction with support vector machines: Introduction and applications," *J. Comput. Civil Eng.* **15**(3), 208–216 (2001).
- ⁴⁹K. S. Mohamed, "Bio-Inspired machine learning algorithm: Genetic algorithm," in *Machine Learning for Model Order Reduction* (Springer, Cham, 2018).
- ⁵⁰M. A. Albadr, S. Tiun, M. Ayob, and F. Al-Dhief, "Genetic algorithm based on natural selection theory for optimization problems," *Symmetry* **12**(11), 1758 (2020).
- ⁵¹J. Kennedy, "The particle swarm: Social adaptation of knowledge," in *Proceedings of 1997 IEEE International Conference on Evolutionary Computation (ICEC'97)* (IEEE, 1997), pp. 303–308.
- ⁵²Y. Zhang, S. Wang, and G. Ji, "A comprehensive survey on particle swarm optimization algorithm and its applications," *Math. Probl. Eng.* **2015**(1), 931256.
- ⁵³M. Samadi, H. Sarkardeh, E. Jabbari, and S. M. E. Azimi, "Optimization of concrete slab floor thickness of stilling basins according to pressure fluctuations," *Iran. J. Sci. Technol. Trans. Civil Eng.* **49**, 1655–1672 (2025).
- ⁵⁴G. Guo, S. Li, Y. Liu, Z. Cao, and Y. Deng, "Prediction of cavity length using an interpretable ensemble learning approach," *Int. J. Environ. Res. Public Health* **20**(1), 702 (2022).
- ⁵⁵K. E. Taylor, "Summarizing multiple aspects of model performance in a single diagram," *J. Geophys. Res.* **106**(D7), 7183–7192, <https://doi.org/10.1029/2000JD900719> (2001).
- ⁵⁶N. Murtaza, D. Khan, A. Rezzoug, Z. U. Khan, B. Benzougagh, and K. M. Khedher, "Scour depth prediction around bridge abutments: A comprehensive review of artificial intelligence and hybrid models," *Phys. Fluids* **37**(2), 021306 (2025).
- ⁵⁷J. L. Hintze and R. D. Nelson, "Violin plots: A box plot-density trace synergism," *Am. Stat.* **52**(2), 181–184 (1998).
- ⁵⁸S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," [arXiv:1705.07874](https://arxiv.org/abs/1705.07874).
- ⁵⁹M. Samadi, M. H. Afshar, E. Jabbari, and H. Sarkardeh, "Prediction of current-induced scour depth around pile groups using MARS, CART, and ANN approaches," *Mar. Georesour. Geotechnol.* **39**(5), 577–588 (2021).
- ⁶⁰A. R. Ghanizadeh, F. Safi Jahanshahi, and S. S. Naseralavi, "Intelligent modeling of unconfined compressive strength of cement stabilised iron ore tailings: A case study of Golgohar mine," *Eur. J. Environ. Civil Eng.* **28**(8), 1759–1787 (2024).