Journal of Hydroinformatics

Journal of Hydroinformatics Vol 25 No 4, 1488 doi: 10.2166/hydro.2023.050

An interpretable machine learning model for predicting cavity water depth and cavity length based on XGBoost–SHAP

Tiexiang Mo, Shanshan Li* and Guodong Li

State Key Laboratory of Eco-hydraulics in Northwest Arid Region of China, Xi'an University of Technology, Xi'an, Shaanxi 710048, China
*Corresponding author. E-mail: shanshanli@xaut.edu.cn

SL, 0000-0002-4730-3709

ABSTRACT

In contrast to the traditional black box machine learning model, the white box model can achieve higher prediction accuracy and accurately evaluate and explain the prediction results. Cavity water depth and cavity length of aeration facilities are predicted in this research based on Extreme Gradient Boosting (XGBoost) and a Bayesian optimization technique. The Shapley Additive Explanation (SHAP) method is then utilized to explain the prediction results. This study demonstrates how SHAP may order all features and feature interaction terms in accordance with the significance of the input features. The XGBoost–SHAP white box model can reasonably explain the prediction results of XGBoost both globally and locally and can achieve prediction accuracy comparable to the black box model. The cavity water depth and cavity length white box model developed in this study have a promising future application in the shape optimization of aeration facilities and the improvement of model experiments.

Key words: Bayesian optimization, cavity length, cavity water depth, SHAP, white box model, XGBoost

HIGHLIGHTS

- SHAP can accurately evaluate the prediction results of XGBoost.
- SHAP considers the role of both single features and interactive features.
- Bayesian optimization can significantly improve XGBoost performance.
- Local interpretation can visualize the impact of all features.
- The cavity water depth is more complex than the cavity length.

1. INTRODUCTION

In high-head flow buildings, the phenomenon of cavitation erosion is often caused by high flow velocity, high pressure, and large flow rate, which will cause erosion and damage to the flow surface of the building (Glazov 1984; Wu & Chao 2011). Currently, engineers use several measures to reduce cavitation damage, including designing overcurrent buildings with appropriate shapes, ensuring the flatness of overcurrent surfaces, and implementing an aeration corrosion reduction scheme. Among them, the experimental study shows that the aeration corrosion reduction scheme has the characteristics of good operability and significant corrosion reduction effect and has the best comprehensive performance in the above scheme (Pfister & Hager 2010; Bai *et al.* 2016). Cavity water depth and cavity length are key indicators for measuring the corrosion reduction effect of aeration facilities. A longer cavity length and lower backwater level indicate more efficient water aeration, resulting in an improved corrosion reduction effect (Glazov 1984; Wu & Ruan 2008). Many factors affect the cavity water depth and the cavity length, including the shape of the aeration facility, the flow conditions, etc., so it is difficult to directly observe and analyze the main influencing factors (Brujan & Matsumoto 2012; Tsuru *et al.* 2017).

At present, machine learning technology has a large number of applications in the field of water conservancy projects, such as the prediction of weir flow rate coefficient, scour depth, and water quality assessment (Parsaie *et al.* 2017; Azamathulla *et al.* 2019). Relevant research results show that machine learning algorithms can achieve good prediction effects in water conservancy project applications. Machine learning models are divided into the black box model and the white box

model. Black box models include Random Forest (RF), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), etc. Most machine learning search results are based on black box patterns (Azimi *et al.* 2016; Pan *et al.* 2022). Some black box models can achieve excellent fitting and prediction results in classification and regression problems. For example, some scholars introduced genetic algorithm (GA) and Bayesian optimization to adjust the hyperparameters of XGBoost and found that the R^2 (coefficient of determination) of the model prediction results was as high as 0.941 (Gu *et al.* 2022; Kim *et al.* 2022). However, the internal mechanism of black box models is too complex, it is difficult for researchers to explain the prediction results and the prediction results are not convincing (Mi *et al.* 2020). White box models are also called interpretable machine learning models, which overcome the defect of poor interpretation performance of black box models. White box models are divided into two categories, intrinsic interpretable machine learning model and ex-post interpretable machine learning model. Intrinsic interpretable machine learning models include the Generalized Addictive Model (GAM), Explainable Boosting Machine (EBM), etc., where GAM ignores the interaction terms of input features and the prediction accuracy of the model is too low (Agarwal *et al.* 2021). Some intrinsic interpretable machine learning models can achieve excellent prediction results and make reasonable explanations for the results, such as GAMI-Net (Yang *et al.* 2021). In addition, there is an interpretable machine learning model based on physical meaning. Research demonstrated that the use of high-level concepts aid in evolving equations that are easier to interpret by domain specialists (Babovic 2009). To simulate groundwater levels, scholars have proposed a method of embedding physical constraints into machine learning models. Research shows that the physically constrained hybrid model exhibits better adaptability and generalization abilities when compared to pure deep learning models (Cai *et al.* 2021, 2022). In the field of rainfall–runoff simulation, there are also physically meaningful machine learning models such as Model Induction Knowledge Augmented-System Hydrologique Asiatique (MIKA-SHA) and Machine Learning Rainfall–Runoff Model Induction (ML-RR-MI). They can help hydrologists better understand catchment dynamics (Chadalawada *et al.* 2020; Herath *et al.* 2021). Ex-post interpretable machine learning models interpret the prediction results of black box models through complex ex-post analysis methods, such as Partial Dependence Plot (PDP), Accumulated Local Effect (ALE), and Shapley Additive Explanation (SHAP) (Maxwell *et al.* 2021; Mangalathu *et al.* 2022). SHAP is an ex-post interpretation method that draws on the idea of game theory. By calculating the marginal contribution value of all input features and feature interaction items in the model, namely the Shapley Value, it is used to measure the influence of features and interaction items, so as to realize the interpretation of the black box model (Meddage *et al.* 2022).

In this study, the cavity water depth and cavity length prediction model based on XGBoost–SHAP was established by collecting experimental data. The model uses Bayesian optimization to search the hyperparameters of the model, uses SHAP to explain the global and local interpretation of the model prediction results, and analyzes the rationality of the interpretation results according to the experimental conclusions.

2. METHODS

2.1. Extreme gradient boosting

In 2016, Chen Tianqi proposed XGBoost, a tree ensemble model for classification and regression based on Gradient Boosting Decision Tree (GBDT) (Chen & Guestrin 2016). The fundamental concept of XGBoost is identical to that of GBDT; however, significant improvements have been added, such as the use of second-order derivatives to improve the accuracy of the loss function, regular terms to prevent overfitting, and block storage to facilitate parallel processing. A loss function and a regularization term make up the objective function of XGBoost. Given the training set data $T = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$, loss function $l(y_i, \widehat{y_i})$, and regularization $\Omega(f_k)$, the overall objective function can be written as

$$obj(\phi) = \sum_{i}^{n} l(y_i, \widehat{y_i}) + \sum \Omega(f_k) \tag{1}$$

where n denotes the total number of samples, y_i denotes the actual value of ith samples, $\widehat{y_i}$ denotes the predicted value of ith samples, f_k denotes the model for kth tree, and $\sum \Omega(f_k)$ denotes the complexity of kth trees.

The objective function can be transformed into the following form:

$$obj(\phi) = \sum_{i=1}^{n} l(y_i, \widehat{y_i}^{t-1} + f_t(x_i)) + \Omega(f_t) + Const \tag{2}$$

where *Const* is a constant term. The final objective function can be obtained by second-order Taylor expansion and removing the constant term, regularizing expansion and removing the constant term, and combining the coefficient of the first term and the coefficient of the second term (Pan *et al.* 2022). Perform second-order Taylor expansion on the objective function:

$$obj(\phi) = \sum_{i=1}^{n} \left[l(y_i, \hat{y}_i^{t-1}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) + Const \tag{3}$$

where $g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$, $h_i = \partial^2_{\hat{y}^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$. Next, the objective function is further simplified:

$$obj(\phi) = \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) = \sum_{i=1}^{T} \left[\left(\sum_{i \in I_j} g_j \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \tag{4}$$

Let $G_j = \sum_{i \in I_j} g_j$, $H_j = \sum_{i \in I_j} h_j$, the objective function is:

$$obj(\phi) = \sum_{j=1}^{T} \left[G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] + \gamma T \tag{5}$$

The objective function for each leaf node j is:

$$f(w_j) = G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \tag{6}$$

Equation (6) is a quadratic function about w_j. The tree's structure is better and the minimal *obj* is the objective function's ideal solution when the target value *obj* is smaller. Since each leaf node's objective function in the XGBoost objective function is independent of the others, when one leaf node gets its maximum value, so does the objective function as a whole. Equation (7) can be changed to produce the ideal version of the objective function by substituting $w_j = -G_j/(H_j + \lambda)$:

$$obj(\phi) = -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma T \tag{7}$$

2.2. Bayesian optimization

Hyperparameters have a significant impact on the accuracy and generalizability of the model during the training process. One of the key responsibilities of machine learning model modeling has evolved into finding the best hyperparameters (Yang *et al.* 2022). The fundamental principle of Bayesian optimization is to estimate the posterior distribution of the objective function using the data and the Bayesian theorem, and then choose the subsequent sample hyperparameter combination based on the distribution. The data of the preceding sample points are fully utilized, and the optimization process finds the parameter that optimizes the overall improvement of the result by learning the shape of the objective function (Sui & Yu 2020). The following is the Bayesian theorem:

$$P(f|D_i) = \frac{P(D_i|f)P(f)}{P(D_i)} \tag{8}$$

where f denotes the unknown objective function, $D_i = (x_1 y_1), (x_2 y_2), \ldots, (x_n y_n)$ denotes the observed set, x_i denotes the observed vector, and $y_i = f(x_i) + \varepsilon_i$ denotes the observed value error. $P(D_i|f)$ denotes the likelihood distribution of y, $P(f)$ is the prior probability distribution of f. $P(D_i)$ denotes the marginal likelihood distribution of marginal f, and $P(f|D_i)$ denotes the posterior probability distribution of f. The most commonly used kernel function for Gaussian processes is the radial basis function RBF, which measures the distance between any two points by generating a covariance matrix. The basic form is as follows:

$$l(x_i, x_j) = \alpha^2 \exp\left(\frac{\|x_i - x_j\|}{2\beta^2} \right) \tag{9}$$

where α and β are hyperparameters of the Gaussian kernel, x_i and x_j are different values of a hyperparameter. Prior Function and Acquisition Function are the two main processes in Bayesian optimization. The Prior Function process primarily employs Gaussian process regression, as illustrated in Figure 1.

2.3. Shapley Additive Explanation

SHAP calculates the contribution of each feature and interaction term in the model to determine their influence and provides an explanation for the black box model (Meddage *et al.* 2022). Shapley Value is another name for the marginal contributions of different features and interactions (Mangalathu *et al.* 2022). Shapley Value can evaluate the impact of various features on the model outcomes and, in turn, provide an explanation for the black box model's prediction outcomes. The ex-post interpretation model g is written as follows for a single sample x:

$$g(x) = \phi_0 + \sum_{i=1}^{M} \phi_i \tag{10}$$

where M is the number of features in the black box model f ; ϕ_0 is the average of the predicted values of all samples of f, also known as base value ; ϕ_i is the Shapley value of ith feature that needs to be calculated, and it is also the core of the whole SHAP method. At the same time, the post-interpretation model g needs to satisfy the following three properties:

1. Local accuracy, where the predicted value of model g for a single sample is equal to the predicted value of the black box model for a single sample, $g(x) = f(x)$.
2. Missingness. Missing sample features have no bearing on the ex-post interpretation model g in the case of a single sample with missing values, $\phi_i = 0$.
3. Consistency. The ϕ_i of the feature changes with the contribution of the feature in the new model for a single sample when the complexity of the model changes, for as from RF to XGBoost. In 1938, Lyold Shapely demonstrated that ϕ_i satisfies the definition $g(x) = \phi_0 + \sum_{i=1}^{M} \phi_i$ and that the three aforementioned conditions each have a sole solution.

The prediction outcome of the black box model $f(x)$ can be used to replace the $g(x)$ in Equation (10) in accordance with the local accuracy of the ex-post interpretation model g, that is:

$$f(x) = \phi_0 + \sum_{i=1}^{M} \phi_i \tag{11}$$

The model prediction result can be considered as the sum of all the features, as shown in Equation (11), and ϕ_i is understood as the influence of each feature on $f(x)$. The following is the ϕ_i calculation equation:

$$\phi_i = \sum_{S \subseteq \{M \setminus x_i\}} \frac{|S|!(|M| - |S| - 1)!}{|M|!} \{f(x_{S \cup \{i\}}) - f(x_S)\} \tag{12}$$

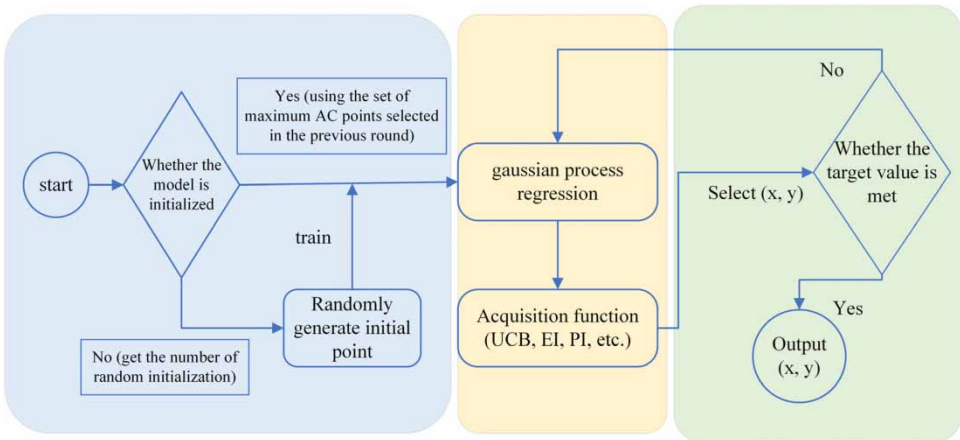

Figure 1 | Gaussian optimization process.

where x_i joins the model and does not join the model under various feature combinations, the simulation results will change, as shown by the expected value of Equation (12). Among them, M denotes the feature set; s denotes the feature subset of $\{M\backslash x_i\}$; the value of S has many cases, corresponding to various feature combinations; and $f(x_{S\cup\{i\}})$ and $f(x_S)$ denote the output results of the model under various feature combinations when x_i is added to the model and when it is not added to the model. $(|S|!(|M| - |S| - 1)!)/(|M|!)$ denotes the probability corresponding to various feature combinations, '| |' denotes the number of elements in the set, and '!' denotes factorial. Numerous estimated Shapley Value calculation techniques have been created because the actual calculation of the Shapley Value takes a lot of time. Different approximation techniques, such as TreeSHAP for tree models, DeepSHAP for neural networks, and model-independent Kernel SHAP, are developed in accordance with the peculiarities of the black box model to be interpreted (Mi *et al.* 2020). Because of its quick calculation times and accurate interpretation of black box model prediction results, TreeSHAP is the most popular of them all (Meddage *et al.* 2022). Based on TreeSHAP, this study aims to propose an interpretable machine learning model.

3. EXPERIMENTAL DATA AND EVALUATION INDEX

3.1. Experiments and data

The experimental data of Shen (Chun-ying 2010) and Zhang (Li-heng 2006) were gathered in this study to ensure that the predictions made by machine learning models were applicable to all situations. In two sets of testing, the aeration facilities take the shape of a bucket and aeration tank combination, as shown in Figure 2. Where i is the bottom slope of the flume, Δ is the height of the ridge, Q is the flow rate, I is the slope of the flip bucket, V is the flow velocity on the ridge, h is the water depth on the ridge, Fr is the Froude number, L is the cavity length value, θ is the impact angle of the water tongue, d is the water depth value of the cavity. Among them, i, Δ, I, and Q are preset values for different test conditions, and the remaining variables are observed values. A total of 270 sets of experimental data were collected from the two groups of experiments. The data range is shown in Table 1.

3.2. Prediction effect analysis of different models

Results from previous studies demonstrate that the XGBoost Set Learning method performs better than SVM and RF in the regression prediction of complex nonlinear relational issues (Gu *et al.* 2022). In this study, a black box model based on

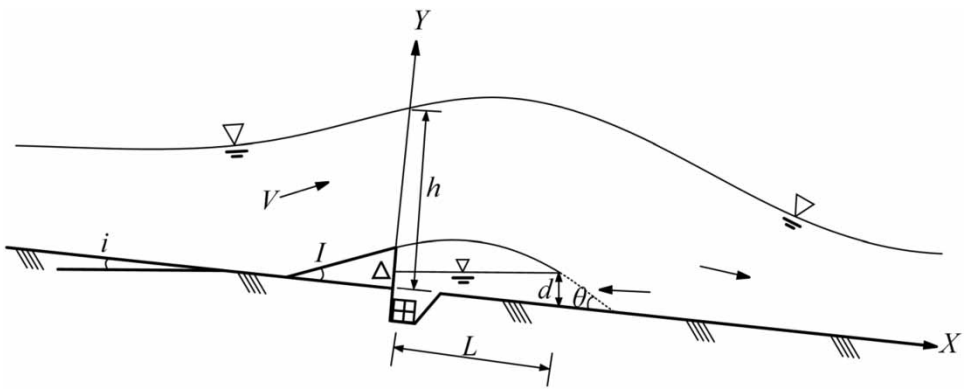

Figure 2 | Aeration facility diagram.

Table 1 | Experimental data

i	Δ (cm)	*Q* (L/s)	*I*	*V* (m/s)	*h* (cm)	Fr	*L* (cm)	θ (°)	*d* (cm)
0.077	1.0–3.0	1.7–5.2	0.1–0.2	1.00–1.75	1.50–3.40	2.46–3.27	4.9–20.2	6.0–20.0	0.40–1.85
0.087	1.0–4.0	1.7–142.7	0.1–0.2	1.07–5.60	1.35–8.50	2.55–6.13	5.0–71.7	6.0–20.0	0.00–1.70
0.096	1.0–3.0	1.7–5.2	0.1–0.2	1.07–1.75	1.25–3.25	2.62–3.90	5.2–26.5	5.5–19.0	0.40–1.30
0.105	1.5–4.0	14.4–141.7	0.1–0.2	1.70–5.90	2.80–8.00	3.27–6.67	13.7–73.0	7.8–14.1	0.00–2.60
0.122	1.5–4.0	13.4–138.0	0.1–0.2	1.80–5.90	2.50–7.80	3.61–7.20	14.4–75.0	7.5–13.6	0.00–2.50

XGBoost was established to predict the cavity water depth and the cavity length. The grid hyperparameters of XGBoost are searched using the Bayesian optimization technique, and the four main hyperparameters of n_estimators, max_depth, colsample_bytree and min_child_weight are optimized, respectively.

With regard to model evaluation, in this study, the determination coefficient (R^2), root mean square error (RMSE), and mean absolute error (MAE) as measurement indicators. The methodology for calculating the indicators is as follows:

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - y)^2} \tag{13}$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{14}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |(y_i - \hat{y}_i)^2| \tag{15}$$

where n is the total number of samples, y_i is the real observed value, y is the average of the real observed value, and $\hat{y}_i$ is the predicted value. SSE is the sum of the remaining squares, the error between the estimated value and the real value, reflecting the degree of fit of the model. SST is the sum of the squares of the total deviation, namely the error between the mean and the real value, reflecting the degree of deviation from the mathematical expectation. *MSE* is the mean square error.

Figure 3 shows the R^2 score diagram of the training set of the model, the test set with Bayesian optimization, and the test set without Bayesian optimization under different test set ratios. Among them, Bayesian optimization can improve the cavity water depth model's performance by 1%. Due to the high R^2 score of the cavity length model training set, there is still a 0.4% performance improvement even though Bayesian optimization has a relatively low-performance improvement on the model. In conclusion, because Bayesian optimization can run an iterative memory search simultaneously on multiple hyperparameters, it is faster and more efficient than classical grid search and random search in model hyperparameter optimization.

Tables 2 and 3 record the R^2, RMSE, and MAE values of the cavity water depth and the cavity length prediction models in five test series. Among these, the validation set records the test values of the five-fold cross validation index combined with Bayesian optimization, and the test set records the final predictive effect of the model. As can be observed, the test set of the two models has the greatest R^2 score when the proportion of the training set is 70%, the R^2 of the cavity water depth is 0.919, and the R^2 of the cavity length is as high as 0.987. The RMSE and MAE of the two models are relatively small when the

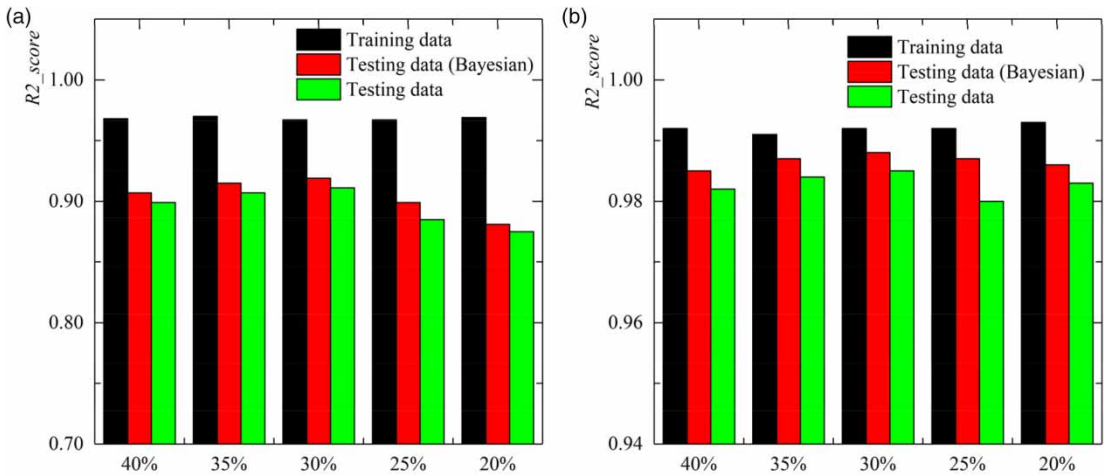

Figure 3 | R^2 score plots of different test set proportions: (a) prediction of cavity water depth and (b) prediction of cavity length.

Table 2 | Index values of cavity water depth at different test set ratios

The proportion of the test set (%)	Training dataset			Validation dataset			Testing dataset		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
40	0.968	0.128	0.096	0.948	0.195	0.141	0.921	0.197	0.145
35	0.970	0.122	0.093	0.934	0.186	0.140	0.915	0.212	0.157
30	0.967	0.127	0.093	0.915	0.236	0.177	0.919	0.208	0.157
25	0.967	0.131	0.095	0.900	0.223	0.164	0.899	0.222	0.167
20	0.969	0.125	0.095	0.880	0.218	0.154	0.881	0.245	0.170

Table 3 | Index values of cavity length at different test set ratios

The proportion of the test set (%)	Training dataset			Validation dataset			Testing dataset		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
40	0.992	0.090	0.071	0.987	0.114	0.086	0.985	0.124	0.091
35	0.991	0.093	0.074	0.980	0.119	0.094	0.987	0.116	0.088
30	0.992	0.090	0.071	0.994	0.079	0.064	0.987	0.116	0.086
25	0.992	0.087	0.069	0.997	0.077	0.061	0.988	0.108	0.083
20	0.993	0.086	0.067	0.992	0.091	0.068	0.986	0.091	0.080

training set ratio is 70%, showing that the model error is relatively minor. When combined with R^2, it can be demonstrated that relatively good prediction precision may be attained when the two models choose 70% of the training set. The model's optimized hyperparameter is displayed in Table 4.

4. RESULTS AND DISCUSSION

On the basis of XGBoost and Bayesian optimization, a black box prediction model for the cavity water depth and the cavity length was created in the final section, and the ideal super parameter combination and ideal training set ratio of the model were examined. In this section, the model's prediction outcomes are explained both globally and locally in terms of SHAP, and the validity of the interpretation results is examined by contrasting them with the findings of the tests.

4.1. Global explanation

Global interpretation refers to the interpretation of the input features and feature interactions of the model as a whole, that is, the shapely value of all features and feature interactions is calculated. Firstly, the absolute Shapley values of all samples of features are calculated, and then the absolute Shapley values of all samples are averaged, which can be used as Shapley values to measure the importance of features, as shown in the following equation.

$$I_j = \frac{1}{n}\sum_{i=1}^{n} |\phi_j^{(i)}| \tag{16}$$

where n is the total number of samples with different features and $|\phi_j^{(i)}|$ is the absolute Shapley value of a single sample. Figure 4 is the ranking diagram of the feature importance of the cavity water depth and the cavity length. Figure 4(a)

Table 4 | Hyperparameters of the model

Predicted labels	max_depth	n_estimators	colsample_bytree	min_child_weight
Cavity water depth	3.00	20.00	0.90	5.37
Cavity length	6.57	20.00	0.50	3.04

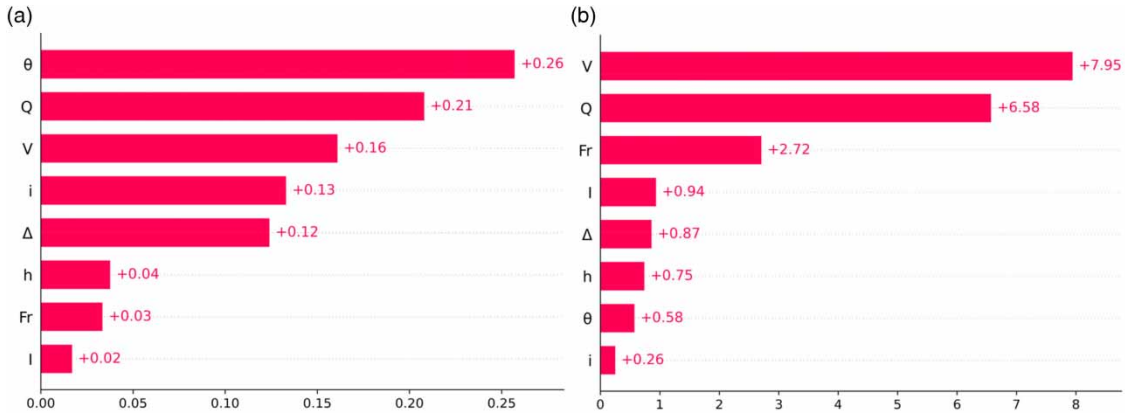

Figure 4 | Mean (|SHAP value|) (average impact on model output magnitude): (a) mean (|SHAP value|) of cavity water depth and (b) mean (|SHAP value|) of cavity length.

shows that the main variables influencing the prediction of the cavity water depth are the impact angle of the water tongue θ, the flow rate Q, the flow velocity V, the flume's bottom slope i, and the height of the ridge Δ which is consistent with the physical model's experimental finding (Chun-ying 2010). In comparison to variables like Q and V, θ is relatively flexible and, in theory, has the biggest impact on cavity backwater (Li-heng 2006). Fr and I, on the other hand, have less of an impact on the cavity water depth. V, Q, and Fr are the primary determining parameters for the cavity length. Theoretically, it is also possible to assess that the length of the cavity and the ejection distance of the water tongue are directly related to the flow velocity on the ridge.

The SHAP summary map for estimating labels is shown in Figure 5. The summary plot combines feature effects and feature importance. Figure 5 provides the Shapley value for each sample and the matching feature value. According to the correlation between the Shapley value of the sample and its feature value, the impact of features on prediction outcomes can be assessed. Using flow velocity as an example, the larger V value corresponds to a smaller Shapley value, indicating that the more the cavity water depth is negatively impacted, the smaller the corresponding cavity water depth value is. The positive influence on the cavity length is stronger as V increases, showing that the cavity length grows as flow velocity increases. Similar effects on the cavity length are produced by the flow rate Q, Froude number Fr, and flip bucket slope I, which are in close agreement with the experimental findings.

Interaction effect refers to the effect between the feature interaction terms in addition to the main features. The Shapley interaction index in game theory is defined as:

$$\phi_{i,j} = \sum_{S \subseteq \{i,j\}} \frac{|S|!(M - |S| - 2)!}{2(M-1)!} \delta_{ij}(S) \tag{17}$$

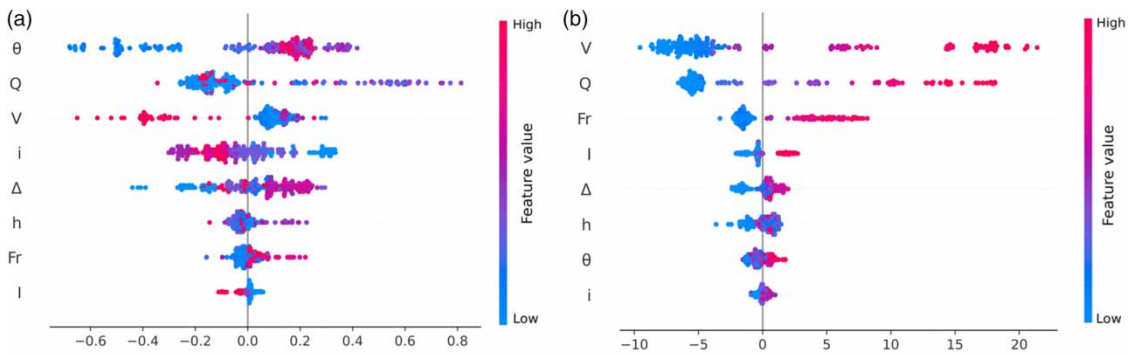

Figure 5 | SHAP summary plot: (a) SHAP value of cavity water depth and (b) SHAP value of cavity length.

when $i \neq j$ and:

$$\delta_{ij}(S) = \widehat{f}_x(S \cup \{i, j\}) - \widehat{f}_x(S \cup \{i\}) - \widehat{f}_x(S \cup \{j\}) + \widehat{f}_x(S) \tag{18}$$

Equation (18) produces a pure interaction effect by deducting the feature's main effect. A matrix of dimension $M \times M$ is created when SHAP interaction values of all the all of the features are calculated, where M is the total number of features.

The interaction between all of the features of the two prediction models is depicted in Figure 6. The primary diagonal displays how the main features affected the outcome of the prediction. The color denotes the size of the sample feature value, while the horizontal axis denotes the size of the SHAP value. Figure 6 also shows the density of the sample distribution, which can be seen as the abbreviated version of the SHAP summary plot in Figure 5. The remaining data display the relationship between several features and the SHAP value's size. Figure 7 illustrates how Q and V and θ and V interact in the two prediction models using an interaction diagram created by SHAP between the two features. Figures 7(a) and 7(b) show

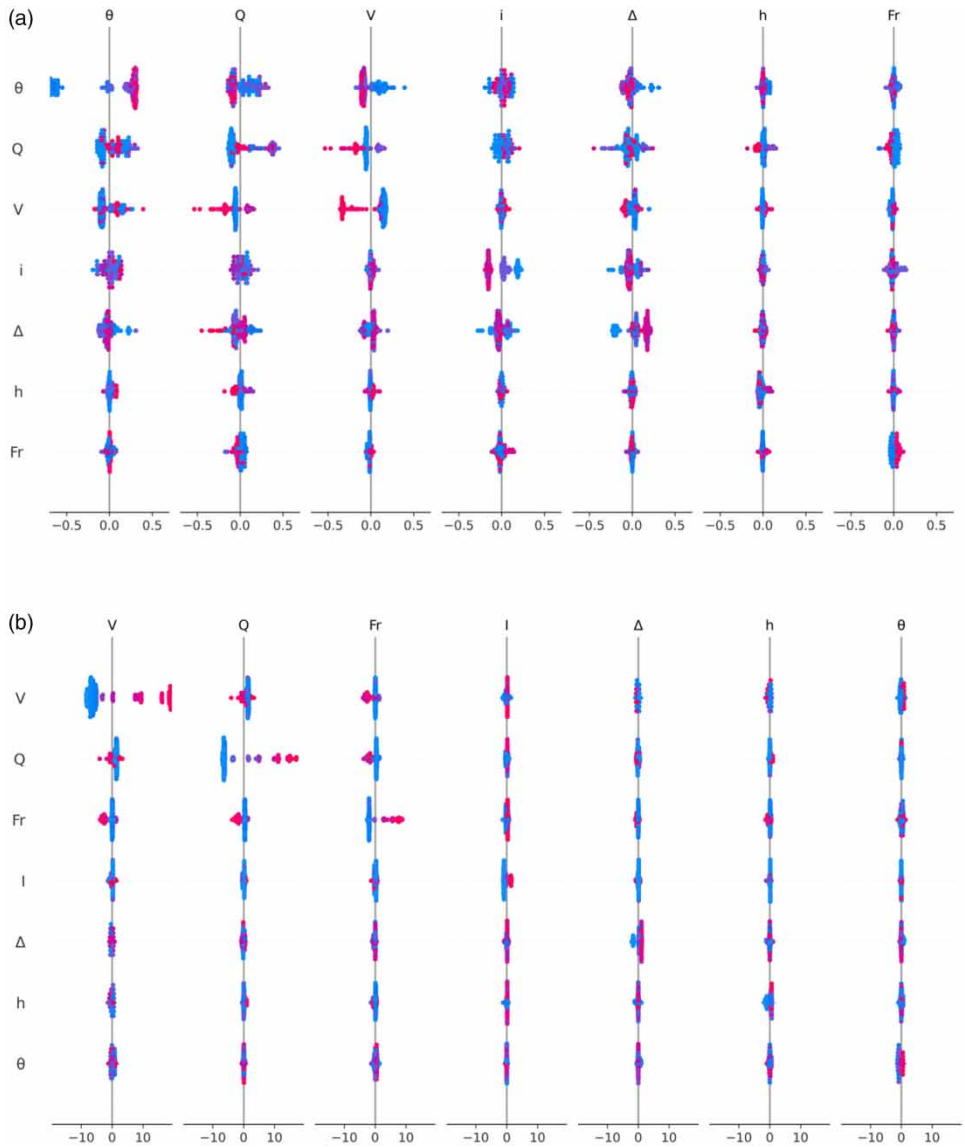

Figure 6 | Interaction diagram of feature values: (a) interaction values for predicting cavity water depth and (b) interaction values for predicting cavity length.

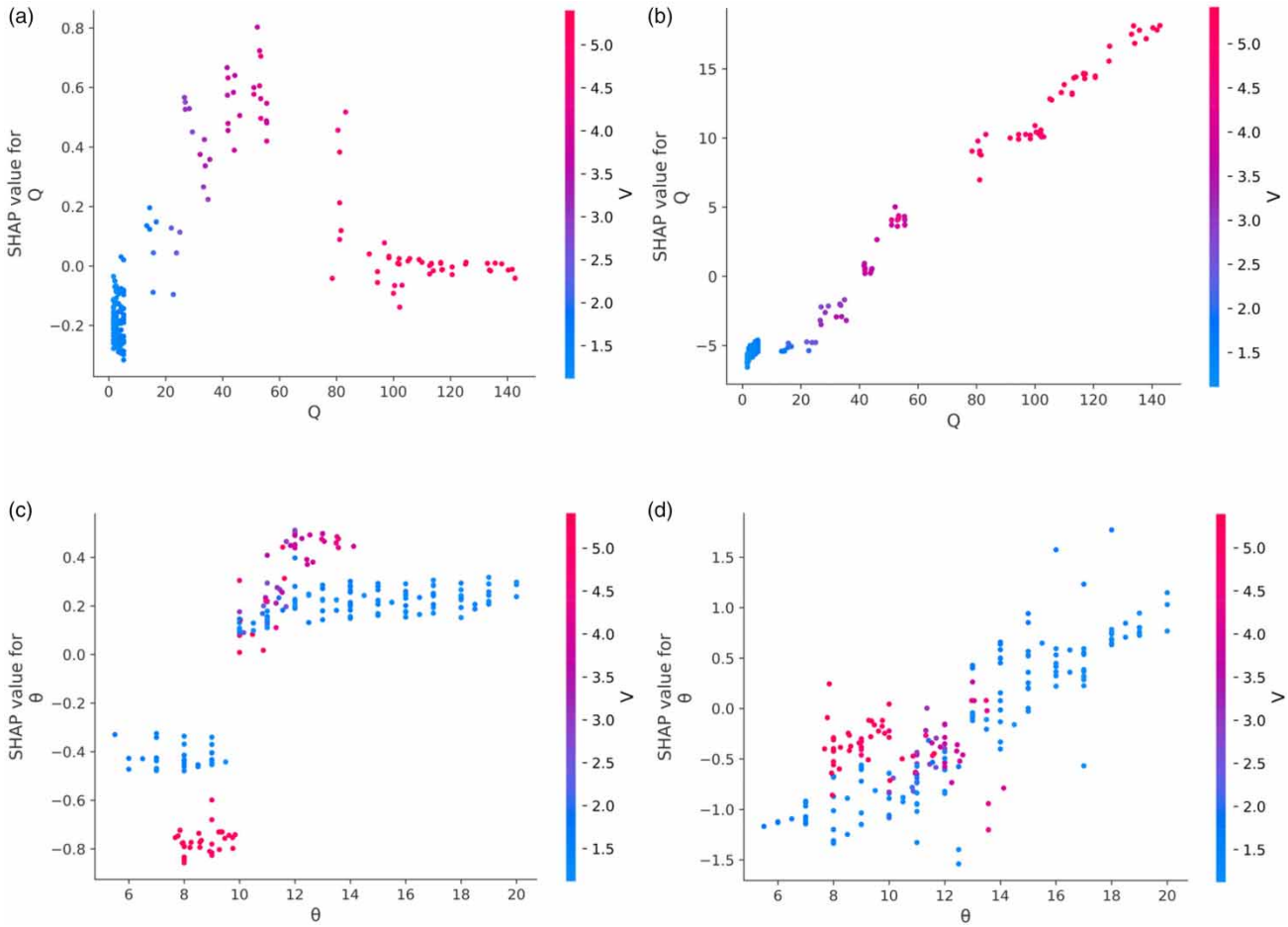

Figure 7 | Plot of SHAP feature dependence: (a) SHAP interaction values for Q and V of cavity; (b) SHAP interaction values for Q and V of water depth cavity length; (c) SHAP interaction values for θ and V of cavity water depth; and (d) SHAP interaction values for θ and V of cavity length.

that the cavity length is positively associated with Q and that as Q grows, the corresponding V also increases. When Q is too little or too large, the SHAP value for the cavity water depth is low or even negative, indicating that these values can hinder the creation of the cavity water accumulation. In this experiment, the SHAP value corresponding to several features has an evident mutation when Q approaches 60–80 cm^3 /s. According to certain research, there is a critical value for a particular flow condition, aerator size, and bottom slope that must be met in order to maintain the stability of the cavity; if the flow condition changes by less than the critical value, the cavity will vanish (Bai *et al.* 2016). It demonstrates that the critical condition range of Q in this experiment is between 60 and 80 cm^3/s. The feature interaction diagram of θ and V show that when θ is greater than 11°, the smaller or larger V causes serious cavity ponding, indicating that the impact angle of the water tongue being less than 11° is one of the critical conditions for reducing the harm of cavity ponding in this test facility. Figure 7(d) demonstrates that when θ is between 6° and 10°, V greater than 4.0 m/s can completely eliminate the effect of cavity backwater and will not interfere with the formation of cavity length, providing the aeration facilities with the ideal hydraulic conditions for efficient operation.

4.2. Local explanation

SHAP has the ability to interpret a single input sample in addition to the overall interpretation. The interpretation diagrams of the cavity water depth and cavity length corresponding to the same sample are illustrated in Figures 8(a)–8(d). The sample's cavity water depth, using subgraph a as an illustration, is 1.399 cm, while the average cavity water depth for all samples is 1.1 cm. The values of all features are displayed on the left numerical axis. The global interpretation study shows that more

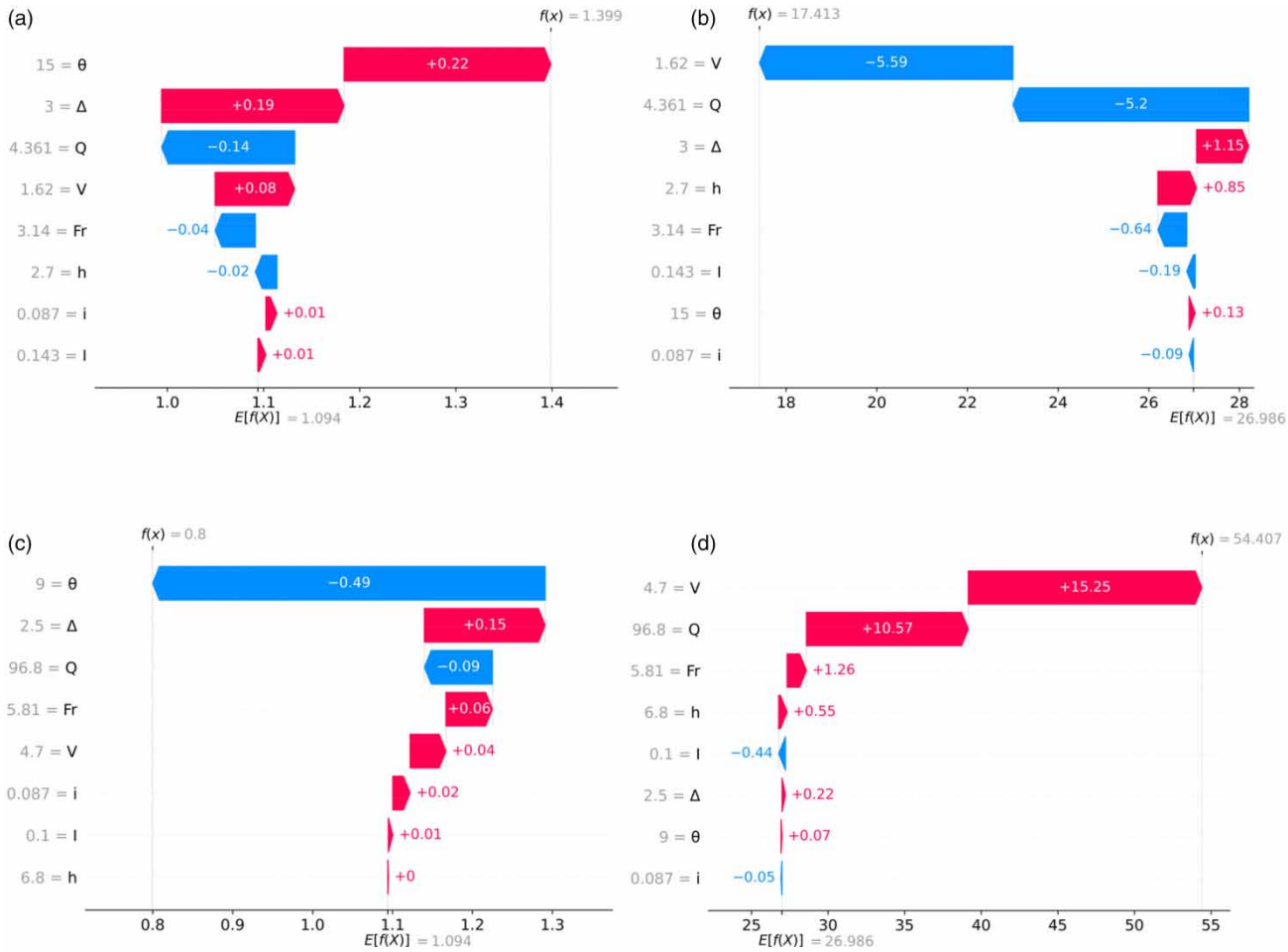

Figure 8 | Local interpretation diagram: (a) sample local interpretation map when cavity water depth is 1.399 cm; (b) sample local interpretation map when cavity length is 17.413 cm; (c) sample local interpretation map when cavity water depth is 0.8 cm; and (d) sample local interpretation map when cavity length is 54.407 cm.

significant cavity backwater will be produced when θ is greater than 11°, which will influence the aeration effect of the aeration facility. Researchers can better comprehend the extent to which the related features have an affect by using the local interpretation diagram. According to Figure 8(a), the prediction results are positively impacted with 0.22 when θ is 15°, demonstrating that the greater θ is the primary contributor to the development of cavity water accumulation. Second, the ridge height of 3 cm also has a positive effect of 0.19 on the cavity water depth. The greatest obstacle to the development of cavity backwater, as shown in Figure 8(d), is the smaller θ. Local interpretations of the cavity length are shown in Figures 8(b) and 8(d), respectively. As can be observed, smaller V and Q are the major factors hindering the creation of cavity length, whereas greater V and Q are the major factors promoting cavity length production, which is consistent with the conclusion of the global interpretation.

5. CONCLUSION

This study proposes an XGBoost–SHAP model for predicting the cavity depth and cavity length of aeration facilities. Unlike intrinsic interpretable machine learning models such as physics-informed machine learning models, XGBoost–SHAP belongs to the category of ex-post interpretability models. However, the results demonstrate that it can still effectively explain the nonlinear relationship between the influencing factors of the cavity water depth and the cavity length. The XGBoost–SHAP model proposed in this study provides a novel approach for exploring interpretable machine learning models in research. The main conclusions are as follows:

a. A Bayesian optimization algorithm is used to adjust the four hyperparameters of XGBoost, and R^2, RMSE, MAE are used as indicators to measure the performance of the model. The results show that the model's performance was considerably improved after the application of Bayesian optimization. The R^2 score of the cavity water depth model has increased by about 1%, and the R^2 score of the cavity length model has increased by about 0.4%.
b. Global interpretation results show that the main factors affecting the cavity water depth and the cavity length are not identical. Among them, the impact angle of the water tongue θ is the most important factor affecting the cavity water depth, the flow velocity V is the most important factor affecting the cavity length. Interpretation results of the interaction feature terms show that the aeration facility can obtain a larger cavity length and eliminate the effect of the cavity water depth when θ is in the range of 6°–10° and V is greater than 4.0 m/s. Global interpretation results are basically consistent with the results of the aeration experiment.
c. Local interpretation can sort the features according to the weight of different features of samples, and show the influence value of all features on the prediction results. It shows that the SHAP–XGBoost model can predict the corresponding cavity water depth and cavity length for any given hydraulic conditions and aeration facility size. The SHAP–XGBoost model established in this study studies the nonlinear relationship between the hydraulic conditions and the size of the aeration facility. It can also optimize the aeration experiment scheme and reduce the time and space cost required for physical experiments.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Ganggui Guo for helping collect data for this work. This work was supported by the National Natural Science Foundation of China [grant number 52079107]. It is also supported by the Natural Science Foundation of Shaanxi Province [grant number 2023-JC-QN-0395] and the Natural Science Foundation of Shaanxi Provincial Department of Education [grant number 22JK0470].

ETHICAL APPROVAL

This article does not contain any studies with human participants or animals performed by any of the authors.

INFORMED CONSENT

Informed consent was obtained from all individual participants included in the study.

AUTHOR CONTRIBUTIONS

T.M. performed the methodology, conceptualized the study, did formal analysis, did investigation, did data curation, wrote the original draft; S.L. did project administration, supervised the study, collected resources, wrote, reviewed, and edited the article; G.L. acquired funds, did project administration, supervised the study, collected resources, wrote, reviewed, and edited the article.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Agarwal, R., Melnick, L., Frosst, N. & Zhang, X. 2021 Neural additive models interpretable machine learning with neural nets. *Advances in Neural Information Processing Systems* **34**, 4699–4711.

Azamathulla, H. M., Shafai Bajestan, M., Azhdary Moghaddam, M. & Rashki Ghaleh Nou, M. 2019 Estimation of scour depth around submerged weirs using self-adaptive extreme learning machine. *Journal of Hydroinformatics* **21** (6), 1082–1101. https://doi.org/10.2166/hydro.2019.070.

Azimi, H., Bonakdari, H., Ebtehaj, I. & Michelson, D. G. 2016 A combined adaptive neuro-fuzzy inference system–firefly algorithm model for predicting the roller length of a hydraulic jump on a rough channel bed. *Neural Computing and Applications* **29** (6), 249–258. https://doi.org/10.1007/s00521-016-2560-9.

Babovic, V. 2009 Introducing knowledge into learning based on genetic programming. *Journal of Hydroinformatics* **11** (3–4), 181–193. https://doi.org/10.2166/hydro.2009.041.

Bai, R., Zhang, F., Liu, S. & Wang, W. 2016 Air concentration and bubble characteristics downstream of a chute aerator. *International Journal of Multiphase Flow* **100** (87), 156–166.

Brujan, E.-A. & Matsumoto, Y. 2012 Collapse of micrometer-sized cavitation bubbles near a rigid boundary. *Microfluidics and Nanofluidics* **13** (6), 957–966.

Cai, H., Shi, H., Liu, S. & Babovic, V. 2021 Impacts of regional characteristics on improving the accuracy of groundwater level prediction using machine learning: the case of central eastern continental United States. *Journal of Hydrology: Regional Studies* **37**, 100930. https://doi.org/10.1016/j.ejrh.2021.100930.

Cai, H., Liu, S., Shi, H., Zhou, Z., Jiang, S. & Babovic, V. 2022 Toward improved lumped groundwater level predictions at catchment scale: mutual integration of water balance mechanism and deep learning method. *Journal of Hydrology* **613**, 128495. https://doi.org/10.1016/j.jhydrol.2022.128495.

Chadalawada, J., Herath, H. M. V. V. & Babovic, V. 2020 Hydrologically informed machine learning for rainfall-Runoff modeling: a genetic programming-based toolkit for automatic model induction. *Water Resources Research* **56** (4). https://doi.org/10.1029/2019wr026933

Chen, T. & Guestrin, C. 2016 XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Chun-ying, S. 2010 *Study on Hydraulic Characteristics of Aerator Under Small Slope and low Froude Number*. *Master Thesis*, Kunming University of Science and Technology, Kunming, China. (in Chinese).

Glazov, A. 1984 Calculation of the air-capturing ability of a flow behind an aerator ledge. *Hydrotechnical Construction* **18** (11), 554–558.

Gu, Z., Cao, M., Wang, C., Yu, N. & Qing, H. 2022 Research on mining maximum subsidence prediction based on genetic algorithm combined with XGBoost model. *Sustainability* **14** (16), 10421. https://doi.org/10.3390/su141610421.

Herath, H. M. V. V., Chadalawada, J. & Babovic, V. 2021 Hydrologically informed machine learning for rainfall–runoff modelling: towards distributed modelling. *Hydrology and Earth System Sciences* **25** (8), 4373–4401. https://doi.org/10.5194/hess-25-4373-2021.

Kim, D., Kwon, K., Pham, K., Oh, J.-Y. & Choi, H. 2022 Surface settlement prediction for urban tunneling using machine learning algorithms with Bayesian optimization. *Automation in Construction* **140**, 104331. https://doi.org/10.1016/j.autcon.2022.104331.

Li-heng, Z. 2006 *Experimental Investigation on the Backwater in Cavity Pocket of Flow Aeration Types to Prevent Cavitation Ersion*. *Master Thesis*, Sichuan University, Chengdu, China. (in Chinese).

Mangalathu, S., Karthikeyan, K., Feng, D.-C. & Jeon, J.-S. 2022 Machine-learning interpretability techniques for seismic performance assessment of infrastructure systems. *Engineering Structures* **250**. https://doi.org/10.1016/j.engstruct.2021.112883

Maxwell, A. E., Sharma, M. & Donaldson, K. A. 2021 Explainable boosting machines for slope failure spatial predictive modeling. *Remote Sensing* **13** (24), 4991. https://doi.org/10.3390/rs13244991.

Meddage, D. P. P., Ekanayake, I. U., Weerasuriya, A. U., Lewangamage, C. S., Tse, K. T., Miyanawala, T. P. & Ramanayaka, C. D. E. 2022 Explainable machine learning (XML) to predict external wind pressure of a low-rise building in urban-like settings. *Journal of Wind Engineering and Industrial Aerodynamics* **226**, 105027. https://doi.org/10.1016/j.jweia.2022.105027.

Mi, J.-X., Li, A.-D. & Zhou, L.-F. 2020 Review study of interpretation methods for future interpretable machine learning. *IEEE Access* **8**, 191969–191985. https://doi.org/10.1109/access.2020.3032756.

Pan, S., Zheng, Z., Guo, Z. & Luo, H. 2022 An optimized XGBoost method for predicting reservoir porosity using petrophysical logs. *Journal of Petroleum Science and Engineering* **208**, 109520. https://doi.org/10.1016/j.petrol.2021.109520.

Parsaie, A., Haghiabi, A. H., Saneie, M. & Torabi, H. 2017 Prediction of energy dissipation of flow over stepped spillways using data-driven models. *Iranian Journal of Science and Technology, Transactions of Civil Engineering* **42** (1), 39–53. https://doi.org/10.1007/s40996-017-0060-5.

Pfister, M. & Hager, W. H. 2010 Chute aerators. I: air transport characteristics. *Journal of Hydraulic Engineering* **136** (6), 352–359.

Sui, G. & Yu, Y. 2020 Bayesian contextual bandits for hyper parameter optimization. *IEEE Access* **8**, 42971–42979. https://doi.org/10.1109/access.2020.2977129.

Tsuru, W., Konishi, T., Watanabe, S. & Tsuda, S.-i. 2017 Observation of inception of sheet cavitation from free nuclei. *Journal of Thermal Science* **26** (3), 223–228.

Wu, J.-h. & Chao, L. 2011 Effects of entrained air manner on cavitation damage. *Journal of Hydrodynamics, Ser. B* **23** (3), 333–338.

Wu, J. & Ruan, S. 2008 Cavity length below chute aerators. *Science in China Series E: Technological Sciences* **51** (2), 170–178.

Yang, Z., Zhang, A. & Sudjianto, A. 2021 GAMI-Net: an explainable neural network based on generalized additive models with structured interactions. *Pattern Recognition* **120**, 108192. https://doi.org/10.1016/j.patcog.2021.108192.

Yang, E., Yang, Q., Li, J., Zhang, H., Di, H. & Qiu, Y. 2022 Establishment of icing prediction model of asphalt pavement based on support vector regression algorithm and Bayesian optimization. *Construction and Building Materials* **351**, 128955. https://doi.org/10.1016/j.conbuildmat.2022.128955.

First received 4 March 2023; accepted in revised form 1 June 2023. Available online 12 June 2023