APPLICATION OF SOFT COMPUTING

Stacking ensemble-based hybrid algorithms for discharge computation in sharp-crested labyrinth weirs

Khabat Khosravi1,2 • Mir Jafar Sadegh Safari3 • Zohreh Sheikh Khozani4 • Brian Crookston5 • Ali Golkarian1

Accepted: 23 March 2022 / Published online: 15 April 2022 © The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Labyrinth weirs are utilized to transport a greater discharge during floods in contrast to conventional weirs due to their increased weir crest length. Nevertheless, due to the increased geometric complexity of labyrinth weirs, determination of accurate discharge coefficients and accordingly, head-discharge ratings are quite essential issues in practical application. Hence, as a first step the present study proposes the following eight standalone algorithms: decision table (DT), Kstar, least median square (LMS), M5 prime (M5P), M5 rule (M5R), pace regression (PR), random forest (RF) and sequential minimal optimization (SMO). Then, applying the stacking (ST) algorithm, these standalone models were hybridized to predict the discharge coefficient ($C_d$) for sharp-crested labyrinth weirs. Potential/effective variables were constructed in the form of several independent dimensionless parameters (i.e., $h/W$, $L/B$, $L/h$, Froude number (Fr), $B/W$ and $L/W$) to predict $C_d$ as an output. The accuracy of the developed models was examined in terms of different statistical visually based and quantitative-based error measurement criteria. The results illustrate that $h/W$ and $B/W$ parameters have the highest and lowest effect on the $C_d$ prediction, respectively. According to NSE, all developed algorithms provided accurate performances, while ST-Kstar had the highest prediction power.

Keywords Discharge coefficient • Hybridization • Labyrinth weir • Stacking algorithm • Machine learning

1 Introduction

1.1 Predict the discharge coefficient necessity

Weirs are among the simplest form of spillway that are widely used in water resources engineering structures, including dam projects. When possible, weirs are installed perpendicular to the flow, aligned with the channel axis and used to control and measure the water level as well as discharge. Due to the effects of seasonality and climate change, droughts and floods are becoming more severe; indeed many flood control structures require discharge capacity upgrades (FEMA 2013). Hence, accurate information about the flow discharge is necessary for accurate planning of watershed management, irrigation and water usage, flood modeling, etc.

Investigations in hydraulic laboratories have been heavily utilized in the past as direct measurement of flow discharge in the field is a difficult and time-consuming task. So far, many experimental studies have been applied
using different weir types with a variety of shapes to measure flow discharge to investigate their efficiencies. Labyrinth weirs were first introduced by Gentilini (1940) and later developed by Taylor (1968) and Hay and Taylor (1970). This nonlinear weir has an advantage over straight overflow weirs; although capacity of this type of weir varies with head, but overall, their discharge capacity can easily exceed twice comparing to the linear weirs with the same width (Tullis et al. 1995). Thus, labyrinth weirs are a common type of spillway that is used for dam reservoirs and is even more efficient than ogee spillway (Suprapto 2013). Taylor (1968) conducted the first study of applying nonlinear labyrinth weirs with various forms including triangular, rectangular, and trapezoidal in a laboratory condition and declared that trapezoidal shape is preferred due to its balance of hydraulics and constructability. Houston (1982) used a monograph approach, which was proposed previously by Hay and Taylor (1970) to design labyrinth weirs and declared that this method leads to about 25% error if the project-specific geometry and conditions deviate from the underlying data. The study conducted by Tullis et al. (1995) showed that discharge capacity is strongly dependent on the total head, effective length of the weir crest and the corresponding coefficient of discharge. Emiroglu and Kisi (2013) reported that the coefficient of discharge has a strong relationship with main channel hydraulic flow and geometric weir characteristics. Due to the constraints of a physical model study (cost, facilities, etc.) and to facilitate usage by industry, different empirical equations were developed based on the regression analysis to predict discharge \( Q \). One of the most well-known and widely used equations is \( Q = (2/3)C_d\sqrt{2ghLh^{1.5}} \), where \( C_d \), \( g \), \( L \) and \( h \) are coefficient of discharge, gravitational acceleration, crest length of the weir and piezometric head over the crest, respectively. According to this equation, \( L \) and \( h \) are the readily available parameters and \( g \) is a constant value, thus the only challenging parameter that has a significant effect on the result and its calculation is the experimentally determined \( C_d \).

Weirs have been regularly studied by researchers for decades, including published values of \( C_d \). For straight weirs, classical studies include (Rehbock 1929) who reported that \( h \) and its ratio to the weir height \( W \) (\( h/W \)) strongly effects \( C_d \) values (\( C_d = 0.611 + 0.08(h/w) \)). Kindsvater and Carter (1959) proposed several equations as a function of \( h/W \) and weir width over the channel width (\( b/B \)) (\( C_d = 0.602 + 0.075(h/w) \)). Also Kandaswamy and Rouse (1957) conducted experiments, for different ranges of \( h/W \) as \( h/W \leq 5 \), \( 5 < h/W < 15 \) and \( h/W \geq 15 \). Swamee (1988) proposed an equation in the form of \( C_d = 0.611 + 0.075(h/W) \). According to the relevant literature, so far, different methods, different effective parameters and also different equations for a specific condition were proposed, which makes it a challenging task to select a more appropriate method for hydrologic analyses. Furthermore, modern risk analyses consider experimental uncertainties which is a challenge of these proposed models. In addition to physical modeling, \( C_d \) can also be predicted through computational fluid dynamic (CFD) models (Babaali et al. 2015; Su et al. 2015), but this approach needs high-accuracy calibration data with detailed information of the dataset (i.e., boundary conditions for energy, momentum and continuity law, nappe behavior for local pressure over weir’s crest, tailwater submergence, weirs geometry, crest shape and so on) since model development and calibration is a difficult task. Also, due to the complexity of the process (three-dimensional flow over the weir), it is very difficult to have an exact prediction using an analytical approach (Crookston and Tullis 2013).

### 1.2 Literature review and related works

Therefore, to expand hydraulic estimations beyond physical and CFD models, the soft computing (SC) approach has gained more attention in solving and predicting complicated and nonlinear hydrological and hydraulic phenomena. Several advantages of SC approaches are nonlinearity structures, the ability to handle big datasets, considering data with different scales, prediction of phenomena with complicated processes and a robust predictor that can allow some incomplete or missing data. Prediction capability of SC approaches strongly depends on the size of the dataset and especially data quality. Up to now, different SC models have been applied to predict \( C_d \) for weirs. Artificial neural network (ANN) is the most widely used algorithm in the field of water resources engineering; however, it shows low convergence speed and prediction power in the testing phase, especially, when range of test data is out of training data. Thus, ANN combined with fuzzy logic and adaptive neuro-fuzzy inference systems (ANFIS) was developed. Azamathulla et al. (2016) compared the prediction performance of ANN, support vector machine (SVM) and ANFIS for discharge coefficient of side weirs and stated that the SVM has a better performance than ANN and ANFIS algorithms. Parsaie et al. (2019a) did a similar study for the combined weir-gate and reported similar results. Salazar and Crookston (2019) specifically considered \( C_d \) for arced trapezoidal labyrinth spillways using neural networks (NN) and random forests (RF) algorithms. Karami et al. (2018) examined simulation power of ANN, genetic programming (GP) and extreme learning machine (ELM) for \( C_d \) for of triangular labyrinth weirs and reported that ELM outperforms other algorithms followed by ANN and GP. Norouzi et al. (2019) stated that ANN has a higher
prediction power than SVM. Bonakdari et al. (2020) predicted $C_d$ using gene expression programming (GEP) and showed the superiority of GEP over the nonlinear regression method (NLR). The group method of data handling (GMDH) approach was used by Ebtehaj et al. (2015) to predict $C_d$ and the results were compared with ANN and nonlinear regression equations. Their results showed the superiority of GMDH over ANN and NLR. Parsaie et al. (2019b) compared the prediction performance of GMDH, GP and multivariate adaptive regression splines (MARS) to the mathematical modeling of discharge coefficient of nonlinear weirs with triangular plan. They revealed that the MARS model has a higher computation power over GMDH and GP. ANFIS, SVM, GMDH and other similar algorithms have a weakness which leads to higher error when they are applied in a standalone framework. These algorithms must be optimized through metaheuristic algorithm to find the optimum operator values, especially weights in the membership function. ELM also needs a large dataset to have a high prediction capability (unlike in this study). The SVM is a robust algorithm but it is susceptible to hyper-parameter selection (Ahmad et al. 2018). Thus, Zaji et al. (2016) developed fire optimization-based support vector regression (SVR-FF) for $C_d$ prediction and reported that firefly metaheuristic algorithm enhanced the SVR model about 10%. Ebtehaj et al. (2018) utilized genetic algorithm (GA) for the optimum selection of ANFIS membership functions and the evolutionary design of a GMDH model structure to achieve more accurate prediction of coefficient of discharge. The results revealed that the ANFIS-GA has a higher prediction power than GMDH-GA.

Recently, a new type of SC algorithms called data mining models (i.e., tree-based, rule-based, lazy-based models and so on) is developed to overcome the weakness of the aforementioned more-traditional algorithms. Higher prediction capability of data mining algorithms over their traditional counterparts were reported in literature. Specifically, Akhbari et al. (2017) stated that M5 tree algorithm has a better performance than ANN algorithms. Hussain and Khan (2020) declared random forest (RF) model outperforms ANN and SVM for streamflow forecasting. Khosravi et al. (2019a) reported that optimized ANFIS hybrid algorithm with metaheuristic algorithms is an improvement over standalone decision trees algorithms (M5Prime (M5P), random tree (RT), RF and reduced error pruning tree (REPT)) and thus the hybridized data mining algorithm may outperform optimized traditional algorithms.

Although not specific to weirs, hybrid approach has been applied to other complex water-related problems. For example, Khosravi et al. (2018) applied standalone [i.e., REPT, M5P and instance-based learning (IBK)] and hybrid models [i.e., bagging-M5P, random committee-REPT (RC-REPT) and random subspace-REPT (RS-REPT)] as well as Salih et al. (2020) developed M5P, attribute selected classifier (ASC), M5Rule (M5R), and KStar for predicting suspended sediment load. Khosravi et al (2019b) used IBK and locally weighted learning (LWL) to predicted fluoride concentrations in groundwater. Khosravi et al. (2020) hybridized decision tree algorithm using bagging algorithm for bed load sediment transport rate prediction and reported that the bagging algorithm enhanced performance of standalone algorithms. Bui et al. (2020) applied hybridized algorithms of cross-validation parameter selection (CVPS) and randomizable filtered classification (RFC) with decision tree algorithms for water quality index prediction. Thus, there is evidence that such an approach could be applied with success to labyrinth weir hydraulics.

2 Research gap and main objective

The field of machine learning is advancing rapidly; algorithms of many types have been used with success to predict behaviors and patterns in fields outside of geoscience with limited application to this important field. For example, least median square (LMS), pace regression (PR), sequential minimal optimization (SMO), Kstar, decision table (DT) and M5 Rule (M5Rare a kind of data mining algorithm that are rarely applied in geoscience. LMS has been successfully used for outliers detection in housing appraisal (Morano and Tajani, 2014). DT has been applied to develop a precautionary approach to problems in behavior, life history and recruitment variability (MacCall 1999). Stacking algorithm (ST) is a kind of ensemble-based data mining algorithm which rarely applied in hydrology and water resources but applied to the other field like predicting the cost of highway construction projects (Meharie et al. 2021). Algorithms applied in geoscience but not to flood infrastructure include PR, which was applied to predict suspended sediment load. SMO was applied to landslide susceptibility mapping (Pham et al. 2019) and KStar applied for scour depth prediction (Khosravi et al. 2021). M5R used to predict suspended sediment load (Salih et al. 2020). Thus, the literature shows that some of these models successfully developed and applied in a different field of study may provide helpful insights to labyrinth weir hydraulics and this study. Also included herein are the M5 Prime (M5P) and random forest (RF) algorithms, frequently applied with success in geoscience (Peng et al. 2020).

Available conventional approaches for discharge coefficient computation were developed by applying a classic regression approach. They are mostly over-fitted models established on limited amount of data. To this end, the
main objective of the present study is to identify a robust, reliable and accurate method for coefficient of discharge prediction for the complex labyrinth weir. To accomplish this, the prediction power of eight novel standalone algorithms and eight hybrid algorithms was investigated. Of the standalone algorithms this study included: LMS, PR, sequential minimal optimization SMO, Kstar, DT, M5R, M5P and RF. The eight new hybrid algorithms paired the ST with those standalone algorithms (i.e., ST-LMS, ST-Pace, ST-SMO, ST-Kstar, ST-DT, ST-M5R, ST-M5P and ST-RF) for coefficient of discharge prediction at sharp-crested labyrinth weirs. To the best of the authors’ knowledge, the majority of these algorithms have not been explored in geosciences. Our study was able to accurately predict discharge coefficients and, using the widely used ST-RF) for coefficient of discharge prediction at sharp-
crested labyrinth weirs. The Kumar et al. (2011) experiments were carried out in a flume with 12 m length, 0.28 m width and 0.41 m depth. A triangular labyrinth weir made of a thin steel plate with six different vertex angles \( (\theta = 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ \) and \( 180^\circ \) ) was located 11 m downstream of the channel entrance (Fig. 1). Flow supplied to the flume was measured using a volumetric sump (located at the flume exit). A point-gauge with accuracy of \( \pm 0.1 \) mm was used to measure the head of water over the crest of the weirs \( (h) \). More information about flume setup and experimental method can be found in Kumar et al. (2011).

The Kumar et al. (2011) dataset was randomly separated into two subgroups in a ratio of 70:30, with 70% (86 sets) of data were used for model development while 30% (37 sets) reserved as a testing dataset for validation of developed models. This approach is considered by the authors to be the most common method in modeling while there is not a universal guideline for preparation of training and testing dataset. Descriptive statistics of the training and testing datasets for input parameters are tabulated in Table 1.

3.2 Dataset collection

Determination of the best input parameters have a significant effect on the result. To enhance a model’s prediction power, based on the correlation coefficient between inputs and output parameters, seven input combinations were constructed to find the optimal or most effective scenario. As the first step, input parameters with the highest correlation coefficient \( (r) \) were used as a single input. The hypothesis was to identify the parameter with the highest ability to accurately predict \( C_d \). Next, the parameter with the second highest \( r \) value was added to the first input and to this end, input No. 2 was constructed. This method continued until the last of the seven parameters corresponding to the lowest \( r \) was added (Table 2).

To find the most effective input combination, the model’s default operator was applied. Efficiency of all constructed input combinations was examined in terms of the root mean square error \( (RMSE) \); the lower the \( RMSE \) the more effective the input parameter combination.

3.4 Model’s parameter optimization

In addition to data quality, the length of the dataset, the model’s prediction power and the effectiveness of input parameters and optimized value for each operator have
significant effects on the modeling prediction accuracy. Optimum values for each model’s operator vary from study to study and data to data; hence, there is not an optimum value for all cases. For this study on labyrinth weirs, a trial-and-error approach was applied to determine the optimum model operators via the Waikato Environment for Knowledge Analysis (WEKA 3.9) software. Default values were first applied to each developed model and their performance was checked through RMSE. Next, higher and lower values were applied, and their performances were checked again until from the range of values the results the optimum values were identified with lowest RMSE.

3.5 Model theory background

3.5.1 Decision table (DT)

A decision table (DT) model is described as a table expressing the comprehensive collection of mutually exclusive restricted expressions in a predetermined difficulty area (Wets 1998). The table determines the requirements that are deemed related to the decision-making procedure under study. The universe of discussion is defined for each situation as $D_i$ where $i$ is the collection of all feasible values that the condition can achieve. The elements of condition are important parts of the DT model. The other section of the DT model contains the action that demonstrates the decision output. The DT has three main features: stability, exclusivity and entirely; therefore, because of these characteristics the trend of every parameter can be estimated.

The DT is composed of four sections as follows (Van-thienen and Wets 1995):

### Table 1

| Parameters | Training | Testing |
|------------|----------|---------|
| $\theta$ (degree) | 180.00 30.00 103.45 50.69 180.00 | 180.00 30.00 100.00 50.20 180.00 |
| $h/w$ | 0.72 0.09 0.38 0.16 0.72 | 0.67 0.12 0.36 0.17 0.67 |
| $L/B$ | 3.86 1.00 1.75 1.00 3.86 | 3.86 1.00 1.78 1.01 3.86 |
| $L/h$ | 135.25 3.89 19.61 24.22 135.25 | 98.36 4.18 20.44 22.00 98.36 |
| $Fr$ | 3.21 0.61 1.18 0.71 3.21 | 3.26 0.62 1.22 0.73 3.26 |
| $B/W$ | 3.04 2.59 2.76 0.14 3.04 | 3.04 2.59 2.76 0.14 3.04 |
| $L/W$ | 11.76 2.69 4.95 3.16 11.76 | 11.76 2.69 5.06 3.20 11.76 |
| $Cd$ | 0.91 0.54 0.72 0.07 0.91 | 0.86 0.57 0.72 0.07 0.86 |
| $Q$ (m$^3$/s) | 0.01 0.00 0.01 0.00 0.01 | 0.01 0.00 0.01 0.00 0.23 |

### Table 2

| No | Input variables | Output |
|----|-----------------|--------|
| 1  | $h/w$           | $Cd$   |
| 2  | $h/w, L/h$      | $Cd$   |
| 3  | $h/w, L/h, Fr$  | $Cd$   |
| 4  | $h/w, L/h, Fr, L/W$ | $Cd$ |
| 5  | $h/w, L/h, Fr, L/W, \theta$ | $Cd$ |
| 6  | $h/w, L/h, Fr, L/W, \theta, B/W$ | $Cd$ |
| 7  | $h/w, L/h, Fr, L/W, \theta, B/W, L/B$ | $Cd$ |
1. The factors of condition are the parameters that are relevant to the decision-making procedure. They show the objects requiring information to make the right decision.
2. Condition states are conceptual statements that decide the related value sets for a specific condition.
3. The action of items identifies the outcomes of the decision-making process.
4. The action values are potential values which can be taken by a given action.

3.5.2 Kstar

The Kstar algorithm recommended by Cleary and Trigg (1995) is a basic classifier approach relating to the lazy trainer’s group. These lazy learners delay the generalization process and the model construction phase until obtaining a request (Witten et al. 2016). The Kstar model employs entropy-based estimation dependent on the possibility of transforming one case into another by selecting randomly between all available transformations. As a distance evaluation, entropy contains many advantages [e.g., the uniformity of this approach in real or symbolic properties also when we face missing values of these properties (Madhusudana et al. 2016)]. To illustrate, suppose \( I \) as a set of instances and \( T \) as the transformation of each instance \( (t \in T) \) maps to another instance \( t : I \rightarrow I \). In order to finish mapping instances to each other, \( \sigma \) is defined (which is belong to \( T \)). Now, \( P \) contains all prefix code from \( T^r \), \( r \) which is completed by \( \sigma \). The \( T^r \) members describe a transformation as:

\[
\tilde{t}(a) = t_n(t_{n-1}(\ldots t_1(a)\ldots)\text{where} \tilde{t} = t_1\ldots t_n)
\]

The probability function of \( T^r \) is denoted as \( P \):

\[
0 \leq \frac{p(t_u)}{p(\tilde{t})} \leq 1
\]

\[
\sum_u p(t_u) = p(\tilde{t})
\]

\[
\sum_{t \in P} p(\tilde{t}) = 1
\]

In addition, the probability of the total path from \( a \) instance to \( b \) instance is defined as \( P^* \):

\[
P^*(b|a) = \sum_{t \in P; \tilde{t}(a) = b} p(\tilde{t}) = 1
\]

3.5.3 Least median square (LMS)

The least median square (LMS) model was proposed by Rousseeuw (1984) and is based on the design from Hampel (1971). This model is well known for being significantly skewed by only a single outlier. Often a significant decrease in the residual sum of squares is achieved by making compromises in data matching to achieve a great benefit in fitting outliers. In such cases, only a specific outlying variable could result in surface regression that does not pass through the majority of the data. To avoid this limitation, the LMS method is applicable. The LMS technique includes decreasing the square residual. Since the outlying observations do not change the median squared residual, the response outliers do not influence the fitted surface and are easily detected in the respective residual plots. The LMS estimator, \( \hat{\theta}_{LMS} \), is determined as:

\[
\hat{\theta}_{LMS} = \arg\min_{\theta} \{|e_i(\theta)|\}_{h,n}
\]

in which \( \theta \) is a \( p \times 1 \) variable of unknown parameter, \( h : n \) shows the \( h \)-th order statistic of \( n \) values. To reach the full breakdown level when the outputs are in general place, the range of \( h \) is selected as: \( h = \left[ \frac{n}{2} \right] + \left\lfloor \frac{n+1}{2} \right\rfloor \), which \( \lfloor \cdot \rfloor \) is the maximum integer function (Agull6 1997).

3.5.4 M5Prime (M5P)

Quinlan (1992) proposed a method that builds linear, piecewise-based tree models as M5Prime (M5P). The decision tree built by this model has multivariate linear models; therefore, the M5P is known as a flexible method (Zhan et al. 2011). The first step of M5P modeling is building the tree model via calculating the standard deviation of the output values at training cases. In this step, the M5P selects one of the trials, which maximizes the anticipated reduction of errors. To compute the standard deviation reduction (SDR) Eq. (9) is used as:

\[
SDR = sd(T) - \sum_{T_i} \frac{|T_i|}{|T|} \times sd(T_i)
\]

where \( T \) denotes the set of samples, \( T_i \) the \( i \)-th subsect of samples that consequences from the node splitting that are then matched with the chosen parameter, and \( sd(T) \) and \( sd(T_i) \) are the standard deviation of \( T \) and \( T_i \), respectively.

The next step is pruning the tree, which makes an approximation of the estimated error that the test dataset would encounter at each node (Wang and Witten 1997). The final step is to apply a smoothing procedure repair for the sharp interruption that happened between adjacent linear models at the pruned tree leaves. This step is especially applicable for modeling some phenomena when a small number of training samples is available, also when the algorithm estimates very different values for the target parameter in the model (Quinlan 1992).
3.5.5 M5Rule (M5R)

Numeric and normal data can be estimated by using a rule-based learning algorithm nominated as M5Rule (M5R). This model was suggested by Frank and Witten (1998) and produces the rules from the M5 tree-based algorithm on the partial and regression tree (PART) model. The rule algorithm acts by performing the construction of the model tree and attempts to select the best rule at each step (Ayaz et al. 2015). A separate-conquer approach is the principal idea of this technique. The fundamental difference between M5P and M5R is the variation in the tree structure in such a way that constructs the tree as partially and full, respectively. Building of the M5R model for this study contains the following five steps:

1. A M5 tree trainer is used with the training dataset.
2. Pruning the tree (same as M5)
3. Selecting the best leaf as rule
4. The step 3 continues until all cases are considered. Each data can be included by various rules at almost the same time.
5. The M5R builds full trees as mentioned previously.

3.5.6 Pace regression (PR)

Pace regression (PR) model presented by Wang (2000) is one of the novel regression models that fixes many disadvantages of this group of algorithms, especially the problem of subset selection. The PR model assesses the impact of each parameter and uses clustering analysis to boost the statistical basis to predict their relation to overall regression, and therefore enhances the regression of classical ordinary squares. Over regular circumstances, the PR is demonstrably optimal if the number of coefficients tends to be infinite. Wang and Witten (2002) demonstrated that the PR model has the best performance in comparison to the other regression models for reducing a model’s dimension when high-dimensional data is available. In other words, the PR model can solve the problem of dimensionality designation. Theoretically, when the number of free variables is infinitely high, it’s optimal to decrease the expected estimation loss (Wang 2000).

3.5.7 Random forest (RF)

RF model was introduced by Breiman (2001) as an ensemble of decision-making trees, combining the predictions of several individual trees using base rules. The general principles of group training techniques assume that their accuracy is higher than other training algorithms. Traditional decision trees tend to over-fit training data, while RF overcomes this deficiency through introducing random elements in a growing process by partial selection of parameters in the tree framework. Established trees are merged to provide a robust estimator concerning the utilized dataset. The combination of several prediction models is more accurate than one model, groups reduce class weaknesses, and they also increase the strength of individual and unique sets of classes (Kotsiantis 2011). The RF technique describes a set of conditions or constraints that are organized in a hierarchical manner and grow successively from the root node to the terminal nodes or leaves. The decision-making process in each internal node of the root node is repeated according to the tree rule until the condition of the previous stop is obtained. Each of the final nodes or leaves is connected to a simple regression model that is used only in the node. When the tree process is completed, pruning is used to improve the generalization capacity of the trees by reducing the complexity of the model structure. To avoid matching different RTSs, the RF algorithm reduces the diversity of trees by creating different subsets of training data known as bagging. Bagging is a technique used to create training data by random sampling the original training dataset with replacements. Liaw and Wiener (2002) showed that the performance of RF model can be represent as:

(a) Obtaining raw data from the bootstrap samples during training
(b) For each sample the unpruned regression tree is developed with some adjustments as the sample of the estimator is randomly picked at each node and the strongest split between all variables is selected.
(c) Collecting trees estimates to predict new results, using median regression values.

3.5.8 Sequential minimal optimization (SMO)

Sequential minimal optimization (SMO) is a simple methodology that can fix the quadratic programming (QP) problem of support vector machine (SVM) quickly without any additional storage of the matrix and without using the stages of numerical QP optimization. SMO breaks down the QP problem into QP sub-problems, employing Osuna’s method to guarantee convergence (Platt 1999). Just two variables are chosen for optimization at each stage and the others are fixed temporarily. This method contains the following steps (Candel et al. 2010):

- Analytical optimization stage for both Lagrange multipliers to determine new values.
- Selection method to pick those two Lagrange multipliers such that the optimal convergence is expedited in each stage.
• Stop criteria to evaluate effectively when an ideal solution has been obtained.
• An update system that efficiently updates the values involved in the selection and optimization of two Lagrange multipliers each time the vector of Lagrange multipliers is changed.
• A method using these subsequent elements to obtain an optimized response to the SVM problem.

The \( z_i \) of SVM learning demonstrated a dual quadratic form as minimizing:

\[
\min L_D(x) = \sum_{i=1}^{n} z_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} z_i z_j y_i y_j \langle x_i, x_j \rangle
\]

subject to \( s.t. \quad 0 \leq z_i \leq C \) and \( \sum_{i=1}^{n} z_i y_i = 0, \forall i \)

where \( C \) is a parameter that governs the tradeoff between maximizing the margin and reducing the error in the learning stage and \( \langle x_0, y_n \rangle \) are the classifier labels.

### 3.5.9 Stacking (ST)

The principle of ensemble method is to combine the estimations of different basic predictors to enhance the strength and generalizability of an estimator. Stacked (ST) generalization is a general term that describes any strategy to serve information from one series of generalizers to another before the final estimate is produced (Wolpert 1992). Based on this method, the meta model (called level 1) is learned on out of fold forecasts of the first model (called level 0) producing the final estimations of the overall ensemble model. Therefore, a stacked generalization method can produce more precise estimations in comparison to the best model of level 0. The process of data division and learning can be defined in four stages as shown in Fig. 5 and in the following steps:

1. **First stage** Split the dataset into training and holdout sections.
2. **Second stage** Learn each base model of level 0 on the training dataset and testing them on the holdout dataset.
3. **Third stage** Apply holdout dataset estimations as input variables and goal values as outputs to learn the level 1 trainer (meta-model).
4. **Fourth stage** Repeat steps one through three sequentially until all the data have been used to make unfolded forecasts.

This ensemble model can be considered as a way of collectively predicting the errors of all main generalizers when operating on a specific training dataset, and then modifying residual predictions by applying the Level 1 method. However, it can be seen as a super multilayer perceptron that employs the base method as hidden later neural units and level 1 method as an output layer to improve precision and robustness.

### 3.6 Model evaluation and comparison

Efficiencies of each developed algorithm must be evaluated because without the model’s prediction power validation stage, modeling results would be unreliable and would not be scientifically sound (Chung and Fabbri 2003). Also, as training datasets are used for model building processes, the results of this section only show how well the developed algorithms fit the corresponding dataset. Thus, testing datasets were applied for the model validation stage. Two of the most common approaches for model evaluation and comparison are visually and quantitatively based methods. Visually based methods are comprised of line graphs, scatter plots and box plots. These approaches benefit from fast, interesting and desirable comparison and can quickly provide more information about accuracy prediction of maximum, minimum, median, first and third quartiles which cannot be driven using quantitative metrics. But these metrics suffer from lack of information about models performance classification and their ranking. Therefore, different quantitative approaches including RMSE, mean absolute error (MAE), the Nash–Sutcliffe efficiency (NSE) and percentage of bias (PBIAS) were computed and applied as follows:

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} [C_{d,\text{Obs}} - C_{d,\text{Pre}}]}, \quad 0 \leq \text{RMSE} < \infty
\]

\[
\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |C_{d,\text{Obs}} - C_{d,\text{Pre}}|, \quad 0 \leq \text{MAE} < \infty
\]

\[
\text{NSE} = 1 - \frac{\sum_{i=1}^{N} (C_{d,\text{Obs}} - C_{d,\text{Pre}})^2}{\sum_{i=1}^{N} (C_{d,\text{Obs}} - \bar{C}_{d,\text{Obs}})^2}, \quad -\infty < \text{NSE} < 1
\]

\[
\text{PBIAS} = \frac{\sum_{i=1}^{N} [C_{d,\text{Obs}} - C_{d,\text{Pre}}]}{\sum_{i=1}^{N} C_{d,\text{Obs}}} * 100, \quad -\infty \leq \text{PBIAS} \leq \infty
\]

where \( C_{d,\text{Obs}} \) and \( C_{d,\text{Pre}} \), are measured and predicted coefficient of discharge values and \( \bar{C}_{d,\text{Obs}} \) is the mean of measured coefficients of discharge.

Also the reliable nonparametric Friedman test was applied to reveal whether there is significant difference among models. The null hypothesis \( (H_0) \) assumes no difference between the performances of the models. If the

\[
\begin{align*}
C_{d,\text{Obs}} & \quad \text{observed discharge} \\
C_{d,\text{Pre}} & \quad \text{predicted discharge} \\
\bar{C}_{d,\text{Obs}} & \quad \text{mean of measured discharge}
\end{align*}
\]
p-value is less than the significance level ($\alpha = 0.05$), and chi square value ($\chi^2$) is higher than standard value (3.841), the null hypothesis is rejected.

4 Results and analysis

4.1 Relative parameter importance

Each input parameter has a different relative effectiveness on the results. Seven dimensionless input parameters were considered herein for the modeling process based on the aforementioned literature review and theory of the discharge over weirs. Effectiveness of these parameters is evaluated using the Pearson correlation coefficient ($r$). As shown in Fig. 2, the results demonstrated that $h/W$ has the highest impact on the modeling of $C_d$ ($r = 0.713$) followed by $L/h$ ($r = 0.537$), $Fr$ ($r = 0.318$), $L/B$ ($r = 0.122$), $L/W$ ($r = 0.119$), $\theta$ ($r = 0.112$) and $B/W$ ($r = 0.019$).

4.2 Best input combination

To identify the best-input combination for $C_d$ computations, seven scenarios were examined on eight standalone models of LMS, PR, SMO, Kstar, DT, MSR, MSP and RF, and their hybrid counterparts based on ST algorithm. The examined scenarios are given in Table 3, listed from one to seven combinations where the incorporated parameters in order of the presentation are $h/W$, $L/h$, $Fr$, $L/W$, $\theta$, $B/W$ and $L/B$. The models’ performances in terms of accurate prediction are given in Table 4 as a heatmap. It is important to involve most significant parameters, as irrelevant parameters lead to a complex structure that may lower prediction accuracy.

The results obtained in Table 3 indicate that an increase in the number of incorporated parameters into the model improves the model’s performance significantly. For instance, for the best standalone model of Kstar, it gives the $RMSE$ of 0.043 for input No.1 (single input parameter), while it reaches the $RMSE$ of 0.0085 and 0.0084 with six and seven input parameters, respectively. It shows 80% promotion in Kstar accuracy in $C_d$ computation. Such an improvement is even more tangible in the hybrid models. The results indicate the ST-Kstar model as the most robust model (shown in Table 3), with $RMSE$ of 0.042 and 0.0046 for one and six input variables, respectively. It indicates almost 90% improvement in its computation accuracy, due to incorporation of more input parameters into the model’s structure.

4.3 Comparison of models

The $C_d$ values from experimental data and the values computed by the standalone models of LMS, PR, SMO, Kstar, DT, MSR, MSP and RF, and hybrid models of ST-LMS, ST-PR, ST-SMO, ST-Kstar, ST-DT, ST-MSR, ST-MSP and ST-RF are compared in terms of line graphs and scatter plots in Fig. 3. As shown in Fig. 3, although a few standalone models provide accurate performances, they generate large scatter as their results are not well fitted to the best-fit line. For instance, DT and SMO models have significant over- and underestimations, while Kstar and RF give more accurate computations. As shown in Fig. 3, hybridization of the standalone models with ST ensemble algorithm improves model performances for the majority of cases. Hybrid models of ST-Kstar and ST-RF are superior predictors where most of the data remained on the best-fit line, thus showing their high performance in $C_d$ computation. Also shown for the case ST-PR, ST-SMO and ST-DT models, the hybridization technique cannot significantly improve their performances as large scatter are still shown, thus demonstrating less accurate performances of these models.

Comparison of the developed models in this study for $C_d$ computation is shown via Fig. 4. The use of violin plots is helpful to understand the distribution of the data in the studied models results. It uses density curves where their widths are attributed to the frequency of data in a specific region. To this end, a model which has most similar violin plot shape to the measured counterpart generates more accurate computations.

As shown in Fig. 4, the ST-Kstar is in excellent agreement with the measured violin plot; the ST-RF models have approximately similar violin plot shapes to the measured one, although the ST-RF model has a wider distribution in the upper quartile. It shows that ST-RF is not as accurate for the higher $C_d$ values. Generally, it can be concluded that hybrid models outperform their corresponding standalone models. The red dash line noted in Fig. 4 shows the median of the data. For the hydride models, in most of the cases, the median line is located at...
the middle of the violin plot, while for some of the standalone models such as M5P, M5R, ST-M5P and ST-M5R the median lines are placed in a lower level, indicating their under-estimation of results. In terms of the maximum value, M5R, Kstar, ST-M5R and ST-Kstar were accurate predictors while only Kstar and ST-Kstar models had the ability to predict minimum $C_d$ value in a high accuracy.

Furthermore, comparison of the developed models for $C_d$ computation was conducted in terms of four statistical performance indices of RMSE, MAE, NSE and PBIAS, which report modeling performance quantitatively, and are shown in Fig. 5. Among the standalone models, Kstar and RF provided much better results than the remaining models considered herein. Consistently their hybrid versions as ST-Kstar and ST-RF are found superior to their alternatives. In terms of NSE, all developed algorithms due to NSE score higher than 0.75 have a very good performance (Ayele et al. 2017), but ST-Kstar outperforms all other models with RMSE, MAE, NSE and PBIAS of 0.011, 0.008, 0.976 and 0.027, respectively. Indeed, the hybridization algorithm significantly improved the performance of some models yet for some cases hybridization has no considerable enhancement in the model’s performance. For instance, a considerable improvement is seen in Fig. 5 for DT and LMS models where their RMSE with 0.023 and 0.020 values are decreased to 0.016 and 0.014 in ST-DT and ST-LMS model, respectively. It shows 30% and 25% promotion in ST-DT and ST-LMS models accuracies in contrast to DT and LMS standalone models. This scenario is vice versa for M5P and SMO models as hybridization.

According to PBIAS result, Kstar, ST-Kstar and M5R algorithms under-estimated $C_d$ values (negative values) while the remainder of the algorithms were over-estimators.

To show whether the differences among models performance are significant, the Friedman test was applied (Table 4). The results revealed that although models’ performances are different they are not statistically different, as Sig is higher the 0.05 and thus the null hypothesis is rejected.

Table 3 Heatmap for determination of the best input combination based on RMSE

| Inputs No. | 1   | 2   | 3   | 4   | 5   | 6   | 7   |
|-----------|-----|-----|-----|-----|-----|-----|-----|
| LMS       | 0.052 | 0.08 | 0.043 | 0.049 | 0.046 | 0.03 | 0.044 |
| PR        | 0.049 | 0.049 | 0.039 | 0.024 | 0.021 | 0.019 | 0.019 |
| SMO       | 0.052 | 0.052 | 0.041 | 0.03 | 0.031 | 0.025 | 0.024 |
| Kstar     | 0.043 | 0.022 | 0.015 | 0.01 | 0.0088 | 0.0085 | 0.0084 |
| DT        | 0.042 | 0.042 | 0.027 | 0.037 | 0.01 | 0.01 | 0.01 |
| M5R       | 0.043 | 0.032 | 0.027 | 0.035 | 0.018 | 0.016 | 0.016 |
| M5P       | 0.043 | 0.033 | 0.027 | 0.025 | 0.024 | 0.021 | 0.021 |
| RF        | 0.022 | 0.021 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| ST-LMS    | 0.05 | 0.075 | 0.041 | 0.043 | 0.04 | 0.026 | 0.031 |
| ST-Pace   | 0.048 | 0.048 | 0.039 | 0.023 | 0.022 | 0.019 | 0.019 |
| ST-SMO    | 0.051 | 0.051 | 0.04 | 0.03 | 0.03 | 0.025 | 0.024 |
| ST-Kstar  | 0.042 | 0.022 | 0.006 | 0.005 | 0.0047 | 0.0046 | 0.0046 |
| ST-DT     | 0.041 | 0.038 | 0.025 | 0.026 | 0.01 | 0.01 | 0.01 |
| ST-M5R    | 0.042 | 0.031 | 0.025 | 0.02 | 0.017 | 0.015 | 0.015 |
| ST-M5P    | 0.042 | 0.033 | 0.025 | 0.025 | 0.024 | 0.019 | 0.019 |
| ST-RF     | 0.022 | 0.021 | 0.015 | 0.0094 | 0.009 | 0.008 | 0.0074 |

** Red light shows the worst input while the green colors illustrate the best inputs
However, additional insights can be gained by evaluating the models’ performances in terms of accurate discharge computation based on the results obtained for $C_d$ as shown in Fig. 6. To this end, the predicted $C_d$ values are incorporated into the $Q = (2/3)C_d \sqrt{2ghL}1.5$ to compute $Q$ passing over the weir. From a first glance on Fig. 6, it can be understood that all machine learning models are successful for adjusting the $C_d$ parameter for discharge computation. Although small scatter is seen for some models such as DT, LMS, M5R, PR and SMO, most of the models provide promising results where data are well-placed on the best fit line. Similar to results reported in the prior section, ST-Kstar and ST-RF models stay ahead in competition with other models in providing accurate results. To sum up, all developed algorithms predicted $C_d$ values accurately with corresponding coefficient of determination higher than 0.99 for whole of the cases.

### 5 Discussion

Due to topographical characteristics, site-specific constraints, project economics, and performance goals, it is often an essential issue to address discharge capacity of hydraulic structures. For many new and rehabilitation projects, enlargement of the weir crest length is a viable option to increase its discharge capacity while balancing these issues. For the case of conventional weirs, for greater discharges during the floods, water levels must also be greater and may cause unacceptable levels of upstream flooding and damage. In the case of flood infrastructure such as embankment dams and levees, this increased upstream elevation may result in overtopping, embankment erosion, breaching, and significant downstream flooding and corresponding consequences. To this end, implementation of labyrinth weirs can be considered to overcome the afflux problem in conventional weirs. An important problem for application of the labyrinth weir comes from the determination of its discharge coefficient. Recommended approaches for computation of discharge coefficient in labyrinth weirs were established from selected hydraulic and geometric variables. However, the existing benchmarks are generated on experimental data through developing a best fit relationship applying classical regression methods. Following the same methodology in published literature and through considering a variety of dimensionless parameters, robust machine learning algorithms are utilized in this study to develop rigorous models for discharge coefficient computation in sharp-crested labyrinth weirs.

This study applied eight standalone models of LMS, PR, SMO, Kstar, DT, M5R, M5P and RF, and their eight hybridized version constructed using the ST algorithm to develop ST-LMS, ST-PR, ST-SMO, ST-Kstar, ST-DT, ST-M5R, ST-M5P and ST-RF models. The results indicate that all models developed in this study have acceptable/very good performance for discharge coefficient computation. For the worst case in the models of DT and SMO, they provide errors less than 2%, while for the best models of ST-Kstar and ST-RF, 0.8% and 0.9% errors are found, respectively, for discharge coefficient computation showing the robustness of the models developed in this study. It can be asserted as a significant promotion in discharge coefficient computation in sharp-crested labyrinth weirs. This is completely due to different computation capability, flexibility and complexity of each algorithm, which is due to different structures of each model. Also, higher performance of hybridized algorithms may be due to increasing in flexibility and nonlinearity of each model. It has to be emphasized that, according to Kumar et al. (2011), ±5% error in discharge computation in sharp-crested labyrinth weirs is acceptable. The range of the error in discharge computation in this study for different models are found from 1.6% to 3%, where for the best models of ST-Kstar and ST-RF are 1.6% 1.8%, respectively (Fig. 7). This result also shows superiority of the present study over the approach proposed by Kumar et al. (2011).

Although this is the first attempt that applied new data mining algorithms to predict $C_d$ values it is difficult to make a direct comparison to traditional machine learning algorithms from other researchers with the same data of the present study. For example, Bonakdari et al. (2020) used GEP and NLR algorithm and reported RMSE of 0.021 and 0.040, respectively. Our study shows 47.6% and 72.5% higher performance, respectively, using ST-Kstar algorithm. Also, Akhbari et al. (2017) stated that M5 tree model with a $R^2 = 0.831$ has high prediction accuracy, which is in disagreement with the result of this study that showed 15% higher prediction capability than M5 tree model. These results show the successful application of the new machine learning algorithms proposed in the present study for discharge coefficient computation in sharp-crested labyrinth weirs. These discrepancies between prior studies may be linked to the details of training and implementation, which as shown herein are critical steps that can heavily influence results.

Our finding in determining relative importance of each input parameters on the result is in accordance with Roushangar et al. (2018) who stated that $h/W$ is the most influential parameter on $C_d$ prediction. Akhbari et al. (2017) stated that $h/B$ and $Fr$ are two most effective input parameters.
Fig. 3  Line-graph and scatter plot of measured Vs. prediction $C_d$ in testing phase
parameters. Also, Azimi et al. (2017) declared that $Fr$ parameter, among single input parameters, has the highest effectiveness, which leads to lowest error. Bonakdari et al. (2020) stated that $\theta$ is the less effective parameter in $Cd$ prediction. To sum up, parameter importance results vary from study to study and its importance depends on the conditions which controlled the experiments.

In terms of identifying the best input combination, except prediction accuracy, the number of input parameters incorporated in the molding process is important, as sometimes, measuring many input parameters is time-consuming. Hence, a model which leads to a slightly lower accuracy with less input parameters is preferable to a model with a slightly higher accuracy with greater number of inputs. For example, ST-Kstar with input No. 5 with
RMSE of 0.0047, is preferable than input No.6 and 7 with RMSE of 0.0046.

It has been known that credibility of a hydraulic model is significantly attributed to the range of data used for the model development. On the other hand, there are a limited number of experimental studies on sharp-crested labyrinth weirs in the literature. Consequently, conducting experimental studies in large channels, adopting wide ranges of crest length, crest height and vertex angle needed, particularly at field scale, for further advancement of these models. Incorporation of the large number of parameters in the model structure arises a difficulty to use the model as a practical tool. To this end, future studies may consider the use of fewer parameters for simplifying the developed models and generating explicit models.
6 Conclusions

Weirs as flow measurement structures are used for many purposes such as flood control, irrigation planning and controlling the flow discharge. Weirs are also widely implemented in water management and hydro-system projects. Discharge capacity would be estimated using coefficient of discharge, but accurate determination of this parameter can be a challenging task. The present study used different soft computing algorithms to predict coefficient of discharge using various readily available parameters as model inputs. The main findings of the present study can be summarized as follows:

Fig. 3 continued

Fig. 4 Violin plot used for model performance
Fig. 5  Model validation: a RMSE, b MAE, c NSE and d PBIAS
1. All developed algorithms have a very good performance, while, ST-Kstar algorithm outperforms its alternatives.

2. Hybrid ST-Kstar model has improved prediction performance of standalone Kstar about 0.82% and provides almost 8.3% higher performance compared to the SMO, with lowest prediction power.
3. \(h/W\) has the highest impact on the modeling of \(C_d\) \((r = 0.713)\) followed by \(L/h\) \((r = 0.537)\), \(Fr\) \((r = 0.318)\), \(L/B\) \((r = 0.122)\), \(L/W\) \((r = 0.119)\), \(\theta\) \((r = 0.112)\) and \(B/W\) \((r = 0.019)\). Result shows 80% promotion in Kstar model accuracy in \(C_d\) computation when effective input combination was applied compared to the input No.1. This reaches up to 90% for ST-Kstar algorithm.

4. Relative importance of input parameters differs from study to study. While \(L/h\), \(Fr\), \(L/B\) and \(L/W\) are the most important parameters in predicting the coefficient of discharge.

5. Best input combination is found as a model in which all input parameters involved except of \(B/W\) which its incorporation to the model, decreased modeling process performance.

6. Utilizing predicted \(C_d\) value by soft computing techniques, the computed discharges provide \(R^2\) values higher than 0.99, near to the unity.

7. The novel approaches proposed in the present study outperform the traditional and nonlinear regression models.

8. Kstar, ST-Kstar and MSR underestimated \(C_d\) values while rest of algorithms overestimated.

---

**Fig. 6** continued

**Fig. 7** Average percentage error for discharge prediction

---

\[
y = 1.0211x - 0.0001 \\
R^2 = 0.9984
\]

\[
y = -4.8387x^2 + 1.0671x - 0.0002 \\
R^2 = 0.9959
\]

\[
y = 1.0239x - 0.0001 \\
R^2 = 0.9989
\]

\[
y = 1.0146x - 7E-05 \\
R^2 = 0.9974
\]

---
Current findings show that both new standalone and hybrid algorithms are cost-effective tools not only for coefficient of discharge prediction. Relying on the promising results of this study, it is expected that the applied algorithms in this study can be implemented in a variety of hydrology and hydraulic problems.

Author contributions KK: was involved in conceptualization, methodology, software, writing original draft. MJSS: helped in writing original draft, review and editing. ZSK: contributed to the conceptualization, writing original draft, data curation, analysis. BC: was involved in methodology, writing original draft, review and editing. AG: contributed in the conceptualization, review and editing.

Funding There is not any funding for this paper.

Data availability Available from the corresponding author upon reasonable request.

Declarations

Conflict of interest There is not any conflict of interest among authors.

References

Agulló J (1997) Exact algorithms for computing the least median of squares estimate in multiple linear regression. In: JSTOR. pp 133–146
Ahmad MW, Moursheed M, Rezgui Y (2018) Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression. Energy 164:465–474. https://doi.org/10.1016/j.energy.2018.08.207
Akhrabi A, Hossein Zaji A, Azimi H, Vafaeifard M (2017) Predicting the discharge coefficient of triangular plan form weirs using radian basis function and M5’ methods
Ayaz Y, Kocamaz AF, Karakoç MB (2015) Modeling of compressive strength and UPV of high-volume mineral-admixed concrete using rule-based M5 rule and tree model MSP classifiers. Constr Build Mater 94:235–240. https://doi.org/10.1016/j.conbuildmat.2015.06.029
Ayele GT, Teshale EZ, Yu B et al (2017) Streamflow and sediment yield prediction for watershed prioritization in the upper Blue Nile river basin. Ethiopia Water (switzerland) 9:782. https://doi.org/10.3390/w9100782
Azamathulla HM, Ghani AA, Zakaria NA (2009) ANFIS-based approach to predicting scour location of spillway. Proc Int Civ Engr Water Manag 162:399–407. https://doi.org/10.1680/wama.2009.162.6.399
Azamathulla HM, Hagiabi AH, Parsaie A (2016) Prediction of side weir discharge coefficient by support vector machine technique. Water Sci Technol Water Supply 16:1002–1016. https://doi.org/10.2166/ws.2016.014
Azimi H, Bonakdari H, Etehaji I (2017) Sensitivity analysis of the factors affecting the discharge capacity of side weirs in trapezoidal channels using extreme learning machines. Flow Meas Instum 54:216–223. https://doi.org/10.1016/j.flowmesinst.2017.02.005
Babaali H, Shamsai A, Vosoughifar H (2015) Computational modeling of the hydraulic jump in the stilling basin with convergence walls using CFD codes. Arab J Sci Eng 40:381–395. https://doi.org/10.1007/s13369-014-1466-z
Bonakdari H, Etehaji I, Gharabaghi B, Sharifi A (2020) Prediction of discharge capacity of labyrinth weir with gene expression programming prediction of discharge capacity of labyrinth weir with gene expression programming. arxiv.org. https://doi.org/10.20944/preprints202001.0313.v1
Breiman L (2001) Random forests. Mach Learn 45:5–32
Bui DT, Khosravi K, Tiefenbacher J et al (2020) Improving prediction of water quality indices using novel hybrid machine-learning algorithms. Sci Total Environ. https://doi.org/10.1016/j.scitotenv.2020.137612
Candel D, Naneufel R, Concha C, Allende H (2010) A sequential minimal optimization algorithm for the all-distances-support vector machine. In: lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). pp 484–491
Chung CJF, Fabbri AG (2003) Validation of spatial prediction models for landslide hazard mapping. Nat Hazards 30:451–472. https://doi.org/10.1023/B:NHAZ.000007172.62651.2b
Cleary JG, Trigg LE (1995) K*: an instance-based learner using an entropic distance measure. Mach Learn Proc 1995:108–114
Crookston BM, Tullis BP (2013) Hydraulic design and analysis of labyrinth weirs. I: discharge relationships. J Irrig Drain Eng 139:363–370. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000558
Etehaji I, Bonakdari H, Gharabaghi B (2018) Development of more accurate discharge coefficient prediction equations for rectangular side weirs using adaptive neuro-fuzzy inference system and generalized group method of data handling. Meas J Int Meas Confed 116:473–482. https://doi.org/10.1016/j.measurement.2017.11.023
Etehaji I, Bonakdari H, Zaji AH et al (2015) Gene expression programming to predict the discharge coefficient in rectangular side weirs. Appl Soft Comput J 35:618–628. https://doi.org/10.1016/j.asoc.2015.07.003
Emiroglu ME, Kisi O (2013) Prediction of discharge coefficient for trapezoidal labyrinth side weir using a neuro-fuzzy approach. Water Resour Manag 27:1473–1488. https://doi.org/10.1007/s11269-012-0249-0
FEMA (2013) Selecting and accommodating inflow design floods for dams. Washington D.C.
Frank E, Witten IH (1998) Generating accurate rule sets without global optimization. Proc Fifteenth Int Conf Mach Learn 144–151. https://doi.org/1-55860-556-8
Gentilini B (1940) Stramazzi con cresla a planta obliqua e a zig-zag (Weirs with planta oblique crest and zig-zag). Italian
Hampel FR (1971) A general qualitative definition of robustness. Ann Math Stat 42:1887–1896. https://doi.org/10.1214/aoms/1177693054
Hay N, Taylor G (1970) Performance and design of labyrinth weirs. J Irrig Drain Eng 96:233–2357
Hussain D, Khan AA (2020) Machine learning techniques for landslide hazard assessment. Nat Hazards Earth Syst Sci 20:1002–1016. https://doi.org/10.5194/nhess-20-1002-2020
Kandaswamy P, Rouse H (1957) Characteristics of flow over terminal convergence walls using CFD codes. Arab J Sci Eng 40:381–395. https://doi.org/10.1007/s13369-014-1466-z
Khoshavani KS, Rouse H (2009) Computational methods for the design of labyrinth dams. Washington D.C.
Karami H, Karimi S, Bonakdari H, Shamshirband S (2018) Predicting discharge coefficient of triangular labyrinth weir using extreme learning machine, artificial neural network and genetic programming. Neural Comput Appl 29:983–989
Khosravi K, Barzegar R, Miraki S, et al (2019a) Stochastic Modeling of Groundwater Fluoride Contamination: Introducing Lazy
