参考著作No.6

Journal of Civil Structural Health Monitoring
July 2018, Volume 8, Issue 3, pp 431–442 | Cite as

Estimation of scour depth at bridges with complex pier foundations using support vector regression integrated with feature selection

Authors Authors and affiliations

Nhat-Duc Hoang, Kuo-Wei Liao, Xuan-Linh Tran

Journal of Civil Structural Health Monitoring
https://doi.org/10.1007/s13349-018-0287-2

参考著作No.6

ORIGINAL PAPER

Estimation of scour depth at bridges with complex pier foundations using support vector regression integrated with feature selection

Nhat-Duc Hoang[1] · Kuo-Wei Liao[2] · Xuan-Linh Tran[1]

Received: 28 December 2017 / Accepted: 23 May 2018

Abstract

This study aims at establishing machine learning models based on the support vector regression (SVR) for estimating local scour around complex piers under steady clear-water condition. A data set consisting of scour depth measurement cases has been collected to construct the prediction models. The data set includes eight influencing factors that consider aspects of pier geometry, flow property, and river bed material. Moreover, to enhance the performance of the SVR model, filter and wrapper feature selection strategies are used. The research finding is that all feature selection approaches can help to improve the prediction accuracy compared with the SVR model that uses all available features. Notably, the feature selection method based on the variable neighborhood search (VNS) algorithm achieves the best performance (MAPE = 21.65%, R^2 = 0.85). Accordingly, the prediction model produced by SVR and VNS can be useful for assisting decision makers in the task of structural health monitoring as well as the design phase of bridges.

Keywords Scour depth prediction · Bridge Scour · Complex pier foundations · Support vector regression · Feature selection · Variable neighborhood search

1 Introduction

Bridge scour is generally defined as the removal of sediment (e.g., sand and gravel) from around bridge abutments or piers [1]. Scour which is caused by swiftly moving water can scoop out scour holes; this leads to the deterioration of the integrity of a bridge structure [2, 3]. About 60% of bridge failures in the United States are related to scour [4]. More importantly, scour failures have the tendency to happen quickly without any prior warning and it is very difficult to monitor them during flood events [4, 5].

River bed scouring can be basically categorized into the three types: general scour, contraction scour, and local scour [6]. In recent history, it has been observed that the majority of bridge failures has been caused by local scour of the streambed [7–9]; therefore, this particular type of scour is considered to be the most important part for bridge safety analysis. As pointed out by [9], besides bridge safety, failures caused by scour often lead to considerable costs, including direct expenditures for repairing damaged bridges and indirect costs due to the impact on transportation (e.g., maintaining traffic flow without the bridge and of the cost for the time lost utilizing alternate routes) and on the economy of local communities. Accordingly, structural health monitoring of bridges regarding to scour is a crucial problem and has attracted an increasing attention of many scholars and hydraulic engineers [10–13].

In addition, most of the pier scour research works has focused on dealing with the scour with uniform piers [12, 14]; the impacts of the pile caps and pile groups on the scour depth had not been taken into account. In real-world

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13349-018-0287-2) contains supplementary material, which is available to authorized users.

✉ Xuan-Linh Tran
tranxuanlinh@dtu.edu.vn

Nhat-Duc Hoang
hoangnhatduc@dtu.edu.vn

Kuo-Wei Liao
kliao@ntu.edu.tw

1 Faculty of Civil Engineering, Institute of Research and Development, Duy Tan University, R. 809 – No. 03 Quang Trung, Da Nang 550000, Vietnam

2 Department of Bioenvironmental Systems Engineering, National Taiwan University, No.1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan

Published online: 02 June 2018

circumstance, due to geotechnical and economical reasons, most of the bridges do not feature the property of uniform piers; their cross-sectional dimension varies over the length of the pier. The complex piers are composed of several components, i.e., column, pile cap, and pile group. This imposes a significant challenge for predicting the bridge scour.

Needless to say, models that can accurately estimate the scour depth at bridge piers are highly desirable. The reason is that under-predicting the scour depths can lead to catastrophic consequences, because the bridge foundation design is not sufficient for ensuring the required support to the structure [9]. On the contrary, overestimating the scour depth results in uneconomical designs of the bridge foundations.

Both theoretical and experimental studies have pointed out that bridge scour is affected by various influencing factors (i.e., pier geometry, property of water flow, and characteristics of the river bed) [15]. Moreover, the underlying functions that relate to the influencing factors and the scour depth are highly non-linear and difficult to express them explicitly [16]. These facts lead to the sophisticated procedures for estimating the scour depth using empirical formulae [15, 17–19] and the inaccuracy in scour depth prediction of those approaches [9, 20].

Within this context, various scholars have relied on machine learning to establish scour depth prediction models from experimental data sets. Muzzammil [21] carried out a comparative study on the performance of artificial neural network (ANN), adaptive network-based fuzzy inference system, and multiple linear regression used for estimating scour depth at uniform bridge abutments. Scour depth prediction models based on ANN and fuzzy ANN have been extensively utilized [16, 22–26]. Najafzadeh and Azamathulla [27] proposed a quadratic polynomial of group method of data handling network for estimating scour depth around bridge piers. Genetic Programming and Gene-Expression Programming for modeling of local scours have been proposed in [28–30].

As can be seen from the literature, studies on scour depth prediction at bridges with complex piers are still very limited. Our current research aims at contributing to the body of knowledge by proposing a scour depth estimation model for bridge structures with complex piers that employs the Support Vector Regression (SVR). SVR is widely known as a powerful and reliable tool for non-linear modeling [31, 32]; however, this machine learning method has been rarely applied in the problem of interest.

In addition, feature or influencing factor selection on experimental/real-world data sets has also been left unexplored in scour depth modeling. Feature selection is indeed very crucial; as summarized by [33], the procedure of feature selection can help to enhance the prediction performance of the machine learning models, construct more cost-effective models, and build a more comprehensible model with fewer variables involved. Therefore, this study equips the SVR with the feature selection algorithms. Both the widely known filter and wrapper strategies of feature selection are employed. Notably, a new wrapper method based on the stochastic search of variable neighborhood search (VNS) is proposed in this study. The research finding is that the VNS-based wrapper method is useful in identifying a highly relevant subset of scour depth influencing factors and enhancing the SVR prediction accuracy.

The rest of this paper is organized in the following manner. The second section describes notable factors that influence the scour depth at bridges with complex pier foundations. Research material and method is stated in the third section, followed by the proposed framework description. Experimental result and comparison are reported in the next section. Several conclusions of this study are provided in the final part.

2 Influencing factors and the collected data set of scour depth at bridges with complex pier foundations

Notably, for predicting scour depth at non-uniform piers, many important impact factors should be considered. In this study, those factors are categorized into three groups: the pier geometry, flow property, and material characteristic at the riverbed.

2.1 Pier geometry

In this category of factors, the pier width perpendicular to the flow direction (b_c), pile-cap width (b_{pc}), and soil covering height (level of the top surface of the pile cap below the surrounding bed level (Y) are generally recognized as influential variables [18]. As a common perception, a larger pier width (b_c) leads to an increase of the scour depth (d_s). [34] simulated scouring of non-uniform piers and found that when the non-uniform ratio of b_c/b_{pc} is greater, the pier scour depth is smaller.

Melville and Raudkivi [35] divided non-uniform piers into three separated sections based on the soil covering height (Y): (Zone 1) the section in which the pile cap is below the bottom of the scour hole ($Y/b_c > 2.4$), (Zone 2) the section in which the pile-cap top is within the scour hole ($2.4 \geq Y/b_c \geq 0$), and (Zone 3) the section in which the pile-cap top is above the bed level ($Y/b_c < 2.4$). The authors experimentally found that compared to the scour results of a uniform pier, Zone 1 does not show influence on the scour, Zone 2 reduces the scour, and Zone 3 increases the scour depth. Moreover, to characterize the pile-cap feature, the longitudinal extension of pile-cap face out from pier face (L_u) can be employed [18].

2.2 Flow property

Inevitably, this group of variables has to consider the mean velocity of the approach flow (V). Depending on the magnitude of velocity (V), the scour can be further divided into two types, namely, clear-water and live-bed scour [35]. In the first type, the scour depth increases as a function of the flow velocity without sediment movement. In the second type, the flow velocity surpasses the critical mean velocity for particle motion (V_c); this leads to the transports of sediment across the bed surface and complicates the scour status.

As shown in the previous works of Melville [15, 19], when the flow velocity (V) exceeds the threshold velocity (V_c), the scour depth first decreases and then increases to another peak. As a consequence, the average scour depth of the live-bed scour is smaller than that of the clear-water scour depth. Therefore, for safety purpose, scour depth in the clear-water condition is often considered in bridge safety evaluation. Furthermore, the flow depth (y), which is the distance from the water surface to the river bed, should be employed to estimate the scour depth [36]. It is generally known that when the ratio of y/b_c is greater, the impact of the water flow on the scour depth is greater and vice versa. [36] showed that if the ratio of $y/b_c > 3$ or 4 (i.e., deep water), the impact of the change of flow depth on the scour depth can be ignored.

2.3 Characteristics of river bed material

The third group of variables, including the median grain size (d_{50}), river bed material geometric standard deviation (σ_g), and the critical velocity of sediment movement (V_c), characterizes the feature of the river bed material [27, 36]. In general, the greater the size of the river better material, the smaller the local scour depth and vice versa. The reason is that large size of the river bed material provides better resistance against the scour effect. Besides the size of the river bed material, the roughness and the size distribution of the material may also impose influence on the local scour depth; these two factors can be inferred via the critical velocity of sediment movement (V_c) and the geometric standard deviation (σ_g).

2.4 The collected data set

This study has collected 170 data samples of clear-water scour from existing experimental works and four more data samples conducted in the Hydrotech Research Institute of the National Taiwan University [37]. For more details regarding the data measurement and collection, the readers are guided to the previous works [18, 35, 38, 39].

In total, 174 data instances can be used for model construction and verification. The eight influencing factors, including the flow depth y, the pier width perpendicular to the flow direction b_c, the pile-cap width b_{pc}, the longitudinal extension of pile-cap face out from pier face L_u, the soil covering height Y, the ratio of the mean velocity to the critical velocity of sediment movement V/V_c, the median grain size d_{50}, and the river bed material geometric standard deviation σ_g, are employed to estimate the scour depth d_s of complex pier foundations.

It is noted that the bed material used in the experiment is sand. Moreover, the properties of the river bed, including the erodibility, are characterized by the median grain size (d_{50}), river bed material geometric standard deviation (σ_g), and the critical velocity of sediment movement (V_c). The statistical descriptions of all variables are provided in Table 1. Scatter plots of the variables are illustrated in Fig. 1. The collected data set used in this study is provided in the Appendix.

3 Research material and method

3.1 Support vector regression (SVR)

Support vector regression (SVR), introduced by Vapnik [40], is a competent tool for non-linear function approximation. The learning process of SVR guarantees to convert to a single a global optimum. Moreover, this algorithm often demonstrates outstanding prediction performance due to its implementation of the structural risk minimization principle which considers both the training error and the generalization of the model [41, 42, 43]. SVR is notably characterized by the usage of kernels for dealing with nonlinear data, absence of local minima in model learning, and

Table 1 Data description

Variable	Notation	Unit	Min	Average	Std.	Max
IF_1	y	m	0.13	0.22	0.12	0.60
IF_2	b_c	m	0.01	0.04	0.03	0.15
IF_3	b_{pc}	m	0.05	0.10	0.06	0.37
IF_4	L_u	m	0.01	0.13	0.41	2.78
IF_5	Y	m	− 0.67	− 0.02	0.12	0.21
IF_6	V/V_c	–	0.53	0.89	0.13	1.18
IF_7	d_{50}	mm	0.06	0.62	0.22	1.00
IF_8	σ_g	–	1.00	1.12	0.12	1.30
Y	d_s	m	0.01	0.09	0.05	0.34

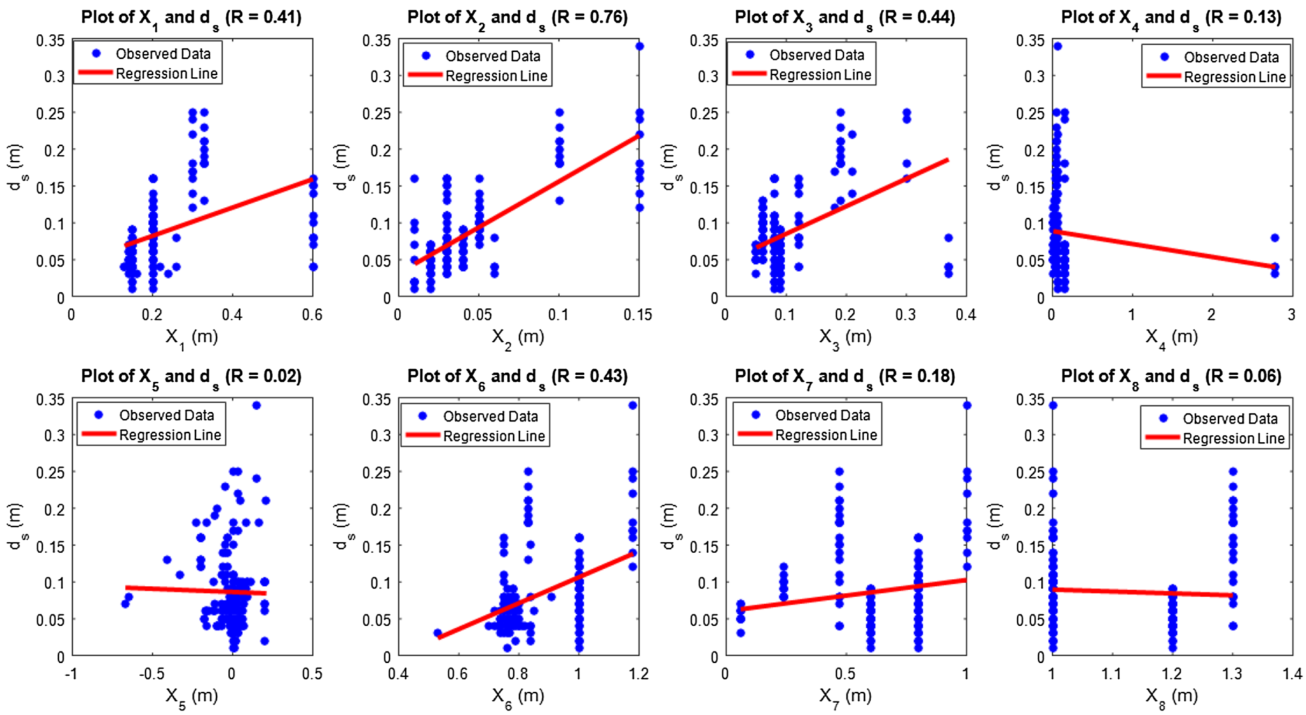


Fig. 1 Scatter plots of scour depth influencing factors

sparseness of the solution [42]. A non-linear function is learned by the algorithm via a kernel function K which maps the data into high-dimensional kernel-induced feature space (see Fig. 2).

In SVR, the goal is to learn a function $f(x)$ that best describes the mapping between a set of L training samples $(x_1, y_1),\ldots, (x_L, y_L)$. Similar to the conventional linear regression, SVR algorithm attempts to locate a hyperplane $f(x)$ with the smallest structural risk in the high-dimensional feature space [31]:

$$f(x) = w^T \phi(x) + b \tag{1}$$

where $\phi(x)$ is a non-linear mapping from the original input space to the feature space. w and b are model parameters and estimated via the following minimization problem:

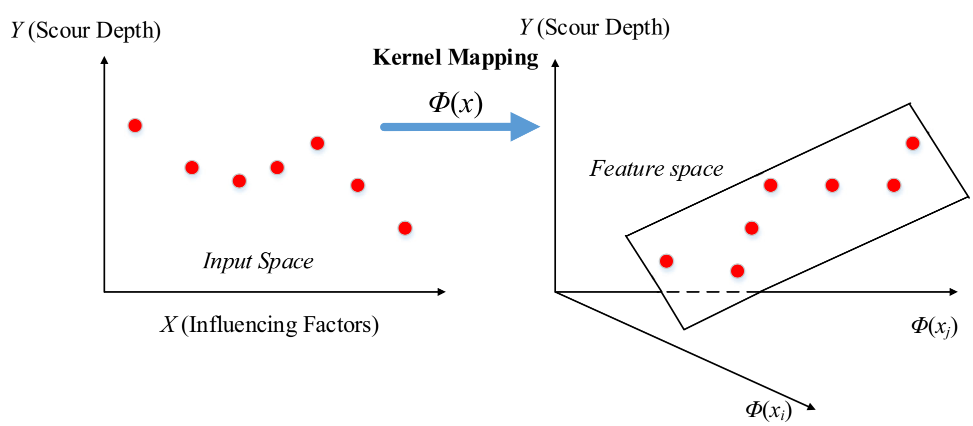
$$\text{Min. } \frac{1}{2}||w||^2 + C\sum_{i=1}^{L}\left(\xi_i + \xi_i^*\right) \tag{2}$$

subjected to

$$\begin{cases} y_i - (\langle w, \phi(x_i)\rangle + b) \le \varepsilon + \xi_i \\ (\langle w, \phi(x_i)\rangle + b) - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases}$$

where C is the the penalty factor that determines the trade of between model accuracy and model complexity; ξ_i and ξ_i^* denote the slack variables. i is the index of training data sample. ε is the error toleration threshold meaning that the data sample x_i is not penalized as long as its error does not exceed this threshold [44].

Fig. 2 Illustration of the SVR learning process



After solving the above optimization problem with the Lagrangian and Karush–Kuhn–Tucker conditions for optimality, the final SVR model can be expressed as

$$f(x)_i = \sum_{i=1}^{L} \left(\alpha_i - \alpha_i^*\right) K(x_i, x) + b \tag{3}$$

where α_i and α_i^* are nonzero Lagrangian multipliers. $K(x_i,x)$ denotes a kernel function. The commonly used kernel function is the radial basis function (RBF) [37, 45, 46] is shown as follows:

$$K(x_i, x_j) = \exp\left(-\gamma ||x_i - x_j||^2\right) \tag{4}$$

where γ is a free parameter of the RBF.

3.2 Feature selection

In general, feature selection (or input variable selection) methods can be divided into two types: filter and wrapper algorithms [47]. Filter algorithms evaluate the relevance of each feature before the utilization of any machine learning algorithm; the relevance of each feature is quantified using statistical properties of features. On the other hand, wrapper methods determine the relevance of features according to the accuracy of the prediction model that employs such features on the training data.

ReliefF [48] is a popular filter method for feature analysis. Based on probability and information theories, ReliefF is able to detect conditional dependencies between features and provide a unified view on the relevance of features in function approximation tasks [49]. This method assigns a weight value for each input feature that indicates its importance; the higher the weight, the more relevant the input feature. Due to its efficiency, this algorithm is selected to be used in this study.

In the case of the wrapper strategy, these methods are implemented by first determining the machine learning algorithm (SVR in this study), the performance criterion (e.g., root mean square error), and the search strategy. Based on the defined search strategy, an iterative process is performed; the machine learning algorithm is trained at each iteration with the available data set and the performances of the model according to different subsets of features are calculated. After the searching process, a desirable subset of feature is determined according to the pre-defined performance criterion.

Sequential Forward Selection (SFS), Sequential Backward Selection (SBS), and Metaheuristic-Based Feature Selection (MFS) are the three commonly used wrapper algorithms due to their ease of implementation [50, 51]. SFS starts from the empty set of feature and sequentially add a certain feature that can reduce the prediction error of the machine learning model. Meanwhile, SBS operates in the opposite direction of SFS. SBS begins with the full set of features and sequentially remove a feature that helps to improve the model prediction accuracy.

The main drawback of SFS is that this algorithm cannot cast out features that become redundant after the advent of other features. On the other hand, the main disadvantage of SBS is its inability to consider the relevance of a feature after it has been rejected. MFS formulates the feature selection as an optimization problem in which the decision variables are the selected subset of features. MFS tends to identify a better subset of features due to its ability to avoid the aforementioned drawbacks of SFS and SBS. Nevertheless, MFS is often criticized for their computational burden due to the entwinement of the learning and feature selection tasks [52, 53]. Therefore, this study employs the Variable Neighborhood Search (VNS) as a single solution-based optimizer instead of other population-based optimizers (e.g., genetic algorithm and particle swarm optimization) for performing feature selection with the SVR model. The aim of this algorithm selection is to seek for quality solutions of feature subsets with reasonable computational cost.

3.3 Variable neighborhood search (VNS)

Variable Neighborhood Search (VNS), proposed by Mladenović, Hansen [54], is a single solution-based metaheuristic method for solving a set of combinatorial optimization and global optimization problems. VNS initially searches within small neighborhoods until a local optimum is encountered, at which point the search process switches to a larger neighborhood, which might help to escape from the local optimum [55].

In VNS, k neighborhood relation N_1, N_2,…,N_k is employed, which are ordered according to increasing size. The algorithm begins with the neighborhood N_1 and performs neighborhood descend-based movements until it reaches a local minimum. If no further improvement is found using a neighborhood N_i, VNS continues the search in an enlarged neighborhood N_{i+1}. If an improvement is achieved, VNS returns to the neighborhood N_1. The VNS algorithm is demonstrated in Fig. 3.

4 The proposed SVR-based model for estimating scour depth at bridges with complex pier foundations

This section of the study describes the structure of the SVR-based scour depth prediction at bridges with complex pier foundations. An overview of the prediction model is shown in Fig. 4. The proposed model can be divided into three consecutive steps: feature selection, model training,

---

Variable Neighborhood Search:
Define objective function (f)
Define solution boundary (B)
Define a neighborhood set $N_1, N_2, \ldots, N_k$
Define the maximum number of iteration (*Iter*)
Randomly initial a solution $x \in B$
$g = 1$ // Iteration counter
While $g <$ Iter
  Choose the most improving neighbor x^* of x in N_i
  If f(x*) < f(x)
    $x = x^*$
    $i = 1$
  Else $i = i + 1$
  End if
End while
Return x

Fig. 3 VNS algorithm

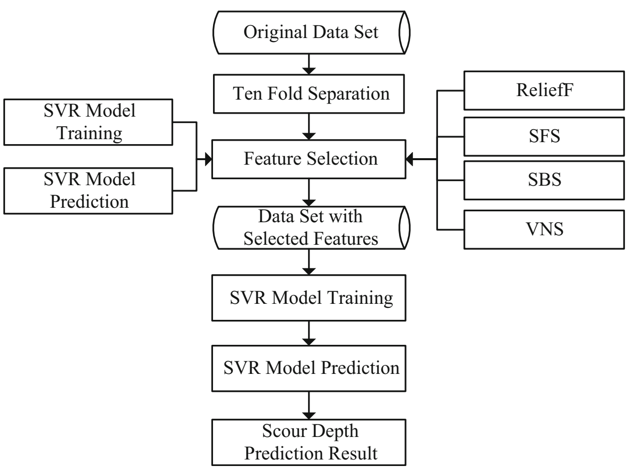


Fig. 4 Proposed SVR-based scour depth prediction model

and model prediction. It is noted that the model is programmed and operates in MATLAB environment.

It is noted that the original data set has been normalized using the Z-score method. Accordingly, in the first step, the relevance of scour depth influencing factors is evaluated via one among four approaches: ReliefF, SFS, SBS, and VNS. It is noted that to reliably assess the relevance of input factors and avoid the randomness due to data selection, the feature selection process is carried out on the basis of a tenfold cross-validation process. Herein, the original data set is separated into ten mutually exclusive folds in which each fold in turn is utilized as testing data and the other nine folds are utilized for model construction.

Hence, the following cost function is employed to evaluate the quality of a feature subset:

$$f = \frac{\sum_{i=1}^{10} \mathrm{RMSE}_i^{\mathrm{Testing}}}{10} \tag{5}$$

where $\mathrm{RMSE}_i^{\mathrm{Testing}}$= the model error measured in terms of Root Mean Square Error (RMSE) for the ith testing data fold.

It is worth noticing that the implementation of ReliefF requires the setting of the number of nearest neighbor (K_n) [48]; in this study, K_n is set to be 5% of the total number of the data samples. Furthermore, for the case of VNS-based feature selection, a general form of a solution candidate is $X = [x_1, \ldots x_i, \ldots, x_D]$, where x_i is an integer of the range [1, D]; D = 8 is the number of available influencing factors in the current data set. The valid subset of features is formed by selecting unique values of X. For instance, a solution candidate $X = [1, 2, \ldots, 8]$ indicates that all influencing factors are selected; a solution candidate X = [1, 2, 3, 5, 3, 6, 7, 8] indicates means that the factor 4 is not included. In addition, the neighborhood set N of the VNS algorithm is selected to be [0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 4, 8].

When the most appropriate subset of scour depth influencing factors has been identified by a method of feature selection, the data set with selected features is established by excluding rejected features from the original data set. Based on the newly formed data set, the training and predicting phases of SVR are consecutively carried out. It is noted that the training phase of the SVR mode requires the specification of the penalty parameter C and the RBF parameter γ; in this study, these two parameters are selected via a tenfold cross-validation-based grid search procedure with the parameter set of [0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50, 100]. In addition, the error toleration threshold (ε) is fixed to be 0.1.

5 Experimental result and comparison

This section reports the proposed SVR-based scour depth prediction model accompanied with the four feature selection algorithms (ReliefF, SFS, SBS, and VNS). It is noted that the Root Mean Squared Error (RMSE), the Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2) are used to quantify the prediction accuracy of the prediction model. RMSE indicates the deviation between the output values actually observed and the output values computed from a trained model. MAPE is the deviation between actual and prediction values divided by the actual value, and is presented in terms of a percentage. In addition, R^2 shows the proportion of the variability in the output variable explained by the model.

The feature selection outcome produced by ReliefF is shown in Fig. 5. As can be seen from the result, ReliefF rejects IF_7 (d_{50}) and IF_8 (σ_g) from the final feature subset.

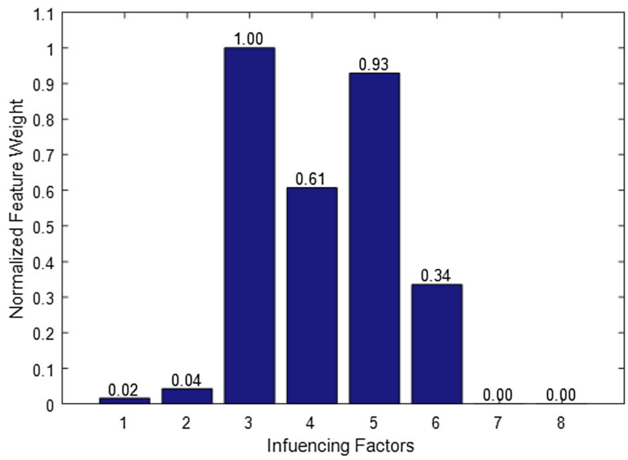


Fig. 5 Feature evaluation using ReliefF method

IF_3 (b_{pc}), IF_4 (L_u), IF_5 (Y), and IF_6 (V/V_c) receive comparatively high weighting values; meanwhile, the weights of $IF_1(y)$ and IF_2 (b_c) are relatively low. Based on such outcome, the SVM coupled with ReliefF method is run using two scenarios: one with a feature subset of (1, 2, 3, 4, 5, 6) and one a feature subset of (3, 4, 5, 6). Tenfold cross-validation processes point out that the first scenario (RMSE = 0.021; MAPE = 24.559%; R^2 = 0.822) is better than the second scenario (RMSE = 0.021; MAPE = 27.095%; R^2 = 0.801). Therefore, the first scenario of feature subset is selected for the SVR model supported by ReliefF; this model is denoted as SVR–ReliefF.

Based on experiments, it is interesting to find that the two methods of SFS and SBS yield the same result of feature selection (see Tables 2, 3). These two methods reach a consensus that IF_2, IF_4, IF_5, IF_7, and IF_8 are useful for scour depth prediction; IF_1, IF_3, and IF_6 are recognized as not useful features. These results of the two wrapper methods are quiet contradictory to the feature subset found by the filter approach of ReliefF. The reason is that ReliefF suggested that IF_7 and IF_8 are redundant. Notably, the model performance using SFS/SBS, denoted as SVR–SFS/SBS (RMSE = 0.021; MAPE = 22.743%; R^2 = 0.828), demonstrates certain improvement compared with SVR–ReliefF (RMSE = 0.021; MAPE = 24.559%; R^2 = 0.822).

Table 2 Feature selection using SFS method

Iteration	Influencing factors								RMSE
	1	2	3	4	5	6	7	8	
1		x							0.034
2		x			x				0.030
3		x			x			x	0.024
4		x		x	x			x	0.022
5		x		x	x		x	x	0.021

Note 'x' = a selected feature

Table 3 Feature selection using SBS method

Iteration	Influencing factors								RMSE
	1	2	3	4	5	6	7	8	
1	x	x	x	x	x	x	x	x	0.0245
2	x	x	x	x	x		x	x	0.0229
3	x	x		x	x		x	x	0.0226
4		x		x	x		x	x	0.0214

Note 'x' = a selected feature

The feature selection outcome performed by VNS optimization algorithm is reported in Table 4. Since VNS is a stochastic search, the VNS-based feature selection is performed in 20 runs to reliably evaluate the feature importance. The maximum number of iteration is set to be 300. A typical run of the VNS-based feature selection is shown in Fig. 6. Moreover, the average selection score (ASC) of a feature (see Fig. 7) can be computed as the average time that the feature is selected by VNS.

As can be seen from the result, IF_1, IF_2, IF_4, IF_5, IF_7, and IF_8 receive high ASC; notably, the ASC values of IF_2, IF_4, and IF_5 are 1 meaning that those features are absolutely relevant for scour depth prediction. It is suspected that IF_6 and IF_3 are redundant, since their ASC values are relatively small (0.05 and 0.15). Therefore, two scenarios of feature subsets are employed for the SVR model supported by VNS (denoted that SVR–VNS): subset 1 which includes the features of (1, 2, 3, 4, 5, 7, 8) with IF_6 being excluded and subset 2 which includes the features of (1, 2, 4, 5, 7, 8) with IF_6 and IF_3 being excluded. Experiments point out that SVR–VNS with the subset 1 (RMSE = 0.018; MAPE = 21.653%; R^2 = 0.852) is better than that with the subset 2 (RMSE = 0.020; MAPE = 23.933%; R^2 = 0.845). Therefore, SVR–VNS with the subset 1 is selected to be used for result comparison with other models.

It is noted that from the preliminary inspection shown in Fig. 1, the factors of X_4 (the longitudinal extension of pile-cap face out from pier face) and X_5 (the soil covering height) have low linear correlations with the scour depth. The correlation coefficients for X_4 and X_5 are 0.13 and 0.02, respectively. However, analysis results using the feature selection methods of SBS and SFS show that these two factors are useful for scour depth prediction (as illustrated in Table 2, 3). In addition, the VNS optimization process has revealed that X_4 and X_5 are highly relevant for scour depth estimation; their average selection scores are both 1 (see Fig. 7). These facts imply that the relation between these two factors with the scour depth is non-

Table 4 Feature selection using VNS algorithm

IF	Run																				ASC
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1		x		x	x	x	x		x	x		x	x	x			x		x	x	0.65
2	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	1.00
3											x	x						x			0.15
4	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	1.00
5	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	1.00
6																x					0.05
7				x	x	x		x	x				x		x	x		x		x	0.50
8	x	x	x	x	x		x			x	x	x	x	x	x		x		x	x	0.75

Note ‘x’ = a selected feature

ASC average selection score, *IF* influencing factors

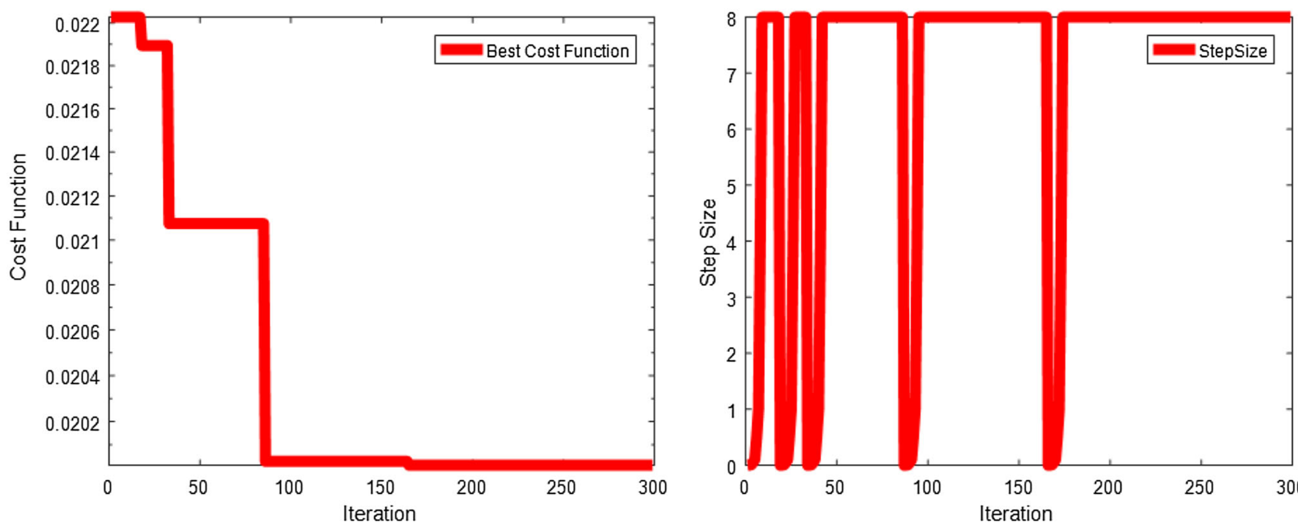


Fig. 6 Typical run of VNS-based feature selection

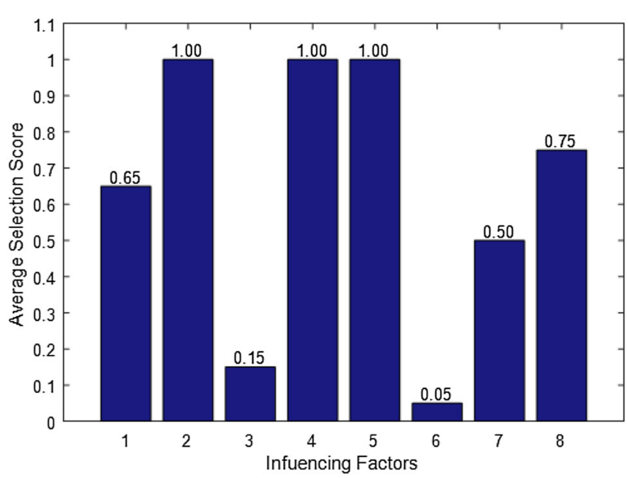


Fig. 7 Feature evaluation using VNS algorithm

linear. Therefore, the application of SVR in analyzing the functional mapping between the influencing factors and the output of scour depth is very suitable. It is because SVR can effectively model non-linear relations between input and output variables ([56]; [57]).

Table 5 summarizes the experimental result comparison that reports the performance of the four model used in this study. SVR denotes the model that employs all available eight features for scour depth prediction. As can be observed, all models equipped with feature selection methods (SVR–ReliefF, SVR–SFS/SBS, and SVR–VNS) show improvement compared with SVR. It can be seen that wrapper approaches (SVR–SFS/SBS and SVR–VNS) are better than the filter approach of ReliefF. Among them, SVR–VNS achieves the most desirable outcome in terms of all performance measurements (RMSE, MAPE, and R^2) (see Fig. 8). Illustrations of the scour depth prediction result yielded by the most desirable SVR–VNS model are shown in Fig. 9.

Table 5 Prediction result summary

Phase	Performance	SVR		SVR–ReliefF		SVR–SFS/SBS		SVR–VNS	
		Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
Training	RMSE	0.009	0.000	0.009	0.000	0.013	0.001	0.012	0.001
	MAPE (%)	14.203	1.223	14.420	0.861	16.896	1.033	16.655	0.725
	R^2	0.971	0.003	0.971	0.003	0.941	0.007	0.948	0.004
Testing	RMSE	0.021	0.009	0.021	0.009	0.021	0.008	0.018	0.010
	MAPE (%)	25.475	13.140	24.559	9.634	22.744	8.197	21.653	6.978
	R^2	0.814	0.085	0.822	0.087	0.828	0.119	0.852	0.152

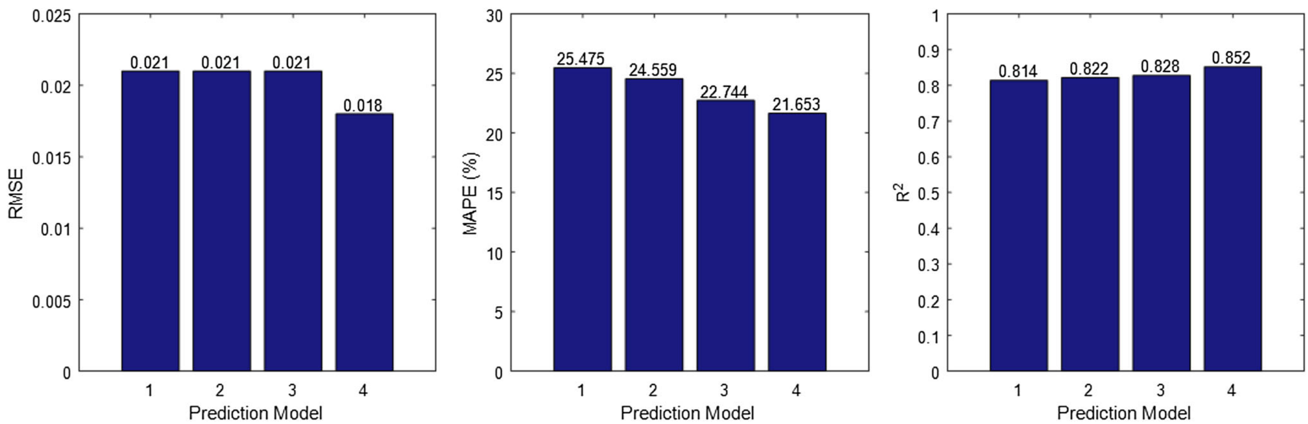


Fig. 8 Model comparison in terms of RMSE, MAPE, and R^2. *Note* Prediction models 1, 2, 3, and 4 denote SVR, SVR–ReliefF, SVR–SFS/SBS, and SVR–VNS, respectively

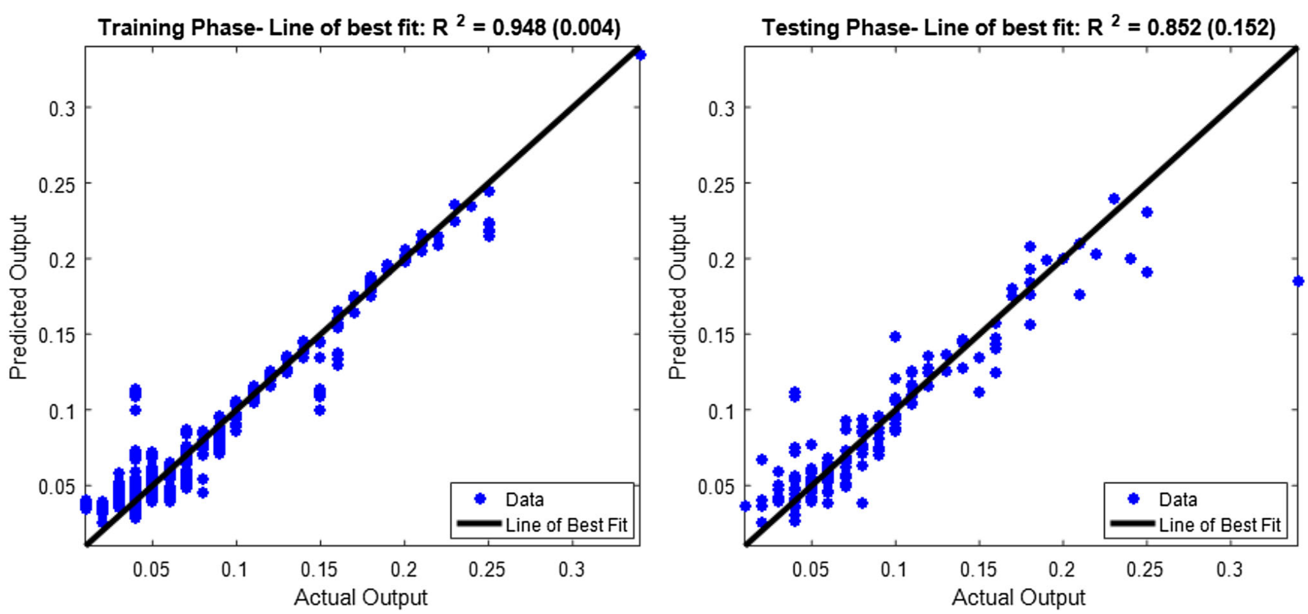


Fig. 9 Prediction results of SVR–VNS

In addition, the performance of the proposed SVR–VNS is compared with the conventional formula-based approaches including the HEC-18 Richardson and Davis [6], Melville and Coleman [19], and Ataie-Ashtiani et al. 18] approaches. The result comparison is reported in Table 6. It is clearly shown that the newly proposed method

Table 6 Result comparison between SVR–VNS and formula-based approaches

Prediction methods	Performance measurement		
	MAPE	RMSE (%)	R^2
SVR–VNS	21.65	0.02	0.85
HEC-18	57.50	0.43	0.76
Melville and Coleman [19]	102.76	0.71	0.59
Ataie-Ashtiani et al. [18]	53.72	0.35	0.77

(MAPE = 21.65%, RMSE = 0.02, R^2 = 0.85) is superior to other formula-based models in all performance measurement metrics.

6 Conclusion

Prediction models for scour depth prediction at complex pier foundations based on SVR and feature selection methods are investigated in this study. The SVR, the machine learning algorithm based on the statistical learning theory, is employed or constructing mapping functions used for the prediction of scour depth. To construct and verify these machine learning based approaches, a data set containing 174 records of scour depth measurement has been collected for this study.

Moreover, both filter and wrapper feature selection strategies have been integrated into the proposed SVR prediction model. The filter method is based on the ReliefF algorithm; meanwhile, the wrapper feature selection strategy employs the SFS, SBS, and VNS algorithms. Experimental results have shown that the SVR models integrated with feature selection methods are better than the SVR that uses all available input factors. Notably, the SVR–VNS demonstrates the best outcome; this method achieves an outstanding results with MAPE = 21.65% and R^2 = 0.85, which are considered to very desirable, since the problem of scour depth estimation at complex pier foundations has been proved to be very challenging.

Accordingly, the prediction model produced by SVR and VNS can be helpful to support decision makers in design phase of bridge which is susceptible to scouring. Based on the estimated results obtained from the proposed SVR–VNS, if the predicted scour depth surpasses the pre-specified critical value, scour prevention measures must be performed for the monitored bridge. The future improvements of the current study may include: (1) collecting more data samples to enhance the applicability of current prediction model; (2) investigating the performance of the SVR models used with other types of kernel function (e.g., hybrid kernel functions) in scour depth estimation; and (3) utilizing other advance machine learning approaches to establish more accurate modeling tools.

References

1. Mueller DS, Wagner CR (2005) Field observations and evaluations of streambed scour at bridges. Office of Engineering Research and Development Federal Highway Administration, McLean
2. Kallias AN, Imam B (2016) Probabilistic assessment of local scour in bridge piers under changing environmental conditions. Struct Infrastruct Eng 12:1228–1241. https://doi.org/10.1080/15732479.2015.1102295
3. Warren LP (2011) Scour at Bridges: stream stability and scour assessment at bridges in Massachusetts US Geological Survey
4. Deng L, Cai CS (2010) Bridge scour: prediction, modeling, monitoring, and countermeasures. Pract Period Struct Des Constr 15:125–134. https://doi.org/10.1061/(ASCE)SC.1943-5576.0000041
5. Landers MN (1992) Bridge Scour Sata Management. Published in Hydraulic Engineering: saving a threatened resource—in search of solutions. In: Proceedings of the Hydraulic Engineering sessions at Water Forum'92 Baltimore, Maryland, August 2–6, 1992 Published by American Society of Civil Engineers
6. Richardson EV, Davis SR (2001) Evaluating scour at bridges (HEC-18) Technical Rep No NHI 01-001. FHWA, Washington, DC
7. Hong J-H, Chiew Y-M, Lu J-Y, Lai J-S, Lin Y-B (2012) Houfeng bridge failure in Taiwan. J Hydraul Eng 138:186–198. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000430
8. Raikar RV, Dey S (2005) Scour of gravel beds at bridge piers and abutments. Proc Inst Civ Eng Water Manag 158:157–162. https://doi.org/10.1680/wama.2005.158.4.157
9. Toth E (2015) Asymmetric error functions for reducing the underestimation of local scour around bridge piers: application to neural networks models. J Hydraul Eng 141:04015011. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000981
10. Liao K-W, Hoang N-D, Gitomarsono J (2018) A probabilistic safety evaluation framework for multi-hazard assessment in a bridge using SO-MARS learning model. KSCE J Civ Eng 22:903–915. https://doi.org/10.1007/s12205-018-1291-0
11. Prendergast LJ, Hester D, Gavin K (2016) Determining the Presence of scour around bridge foundations using vehicle-induced vibrations. J Bridge Eng 21:04016065. https://doi.org/10.1061/(ASCE)BE.1943-5592.0000931
12. Wang C, Yu X, Liang F (2017) A review of bridge scour: mechanism, estimation, monitoring and countermeasures. Nat Hazards 87:1881–1906. https://doi.org/10.1007/s11069-017-2842-2
13. Zarafshan A, Iranmanesh A, Ansari F (2012) Vibration-based method and sensor for monitoring of bridge scour. J Bridge Eng 17:829–838. https://doi.org/10.1061/(ASCE)BE.1943-5592.0000362
14. Park C-W, Park HI, Cho Y-K (2017) Evaluation of the applicability of pier local scour formulae using laboratory and field data. Mar Georesour Geotechnol 35:1–7. https://doi.org/10.1080/1064119X.2014.954658
15. Melville B (2008) The physics of local scour at bridge piers. In: Proceedings of the Fourth International Conference on Scour and Erosion, Tokyo, Japan

16. Azimi H, Bonakdari H, Ebtehaj I, Ashraf Talesh SH, Michelson DG, Jamali A (2017) Evolutionary Pareto optimization of an ANFIS network for modeling scour at pile groups in clear water condition. Fuzzy Sets Syst 319:50–69. https://doi.org/10.1016/j.fss.2016.10.010
17. Arneson LA, Zevenbergen LW, Lagasse PF, Clopper PE (2012) Evaluating scour at bridges Publication No FHWA HIF 12-003. Federal Highway Administration, Washington, DC
18. Ataie-Ashtiani B, Baratian-Ghorghi Z, Beheshti AA (2010) Experimental investigation of clear-water local scour of compound piers. J Hydraul Eng 136:343–351. https://doi.org/10.1061/(ASCE)0733-9429(2010)136:6(343)
19. Melville BW, Coleman SE (2000) Bridge scour. Water Resources Publications, Littleton, Colo
20. Etemad-Shahidi A, Rohani MS (2014) Prediction of scour at abutments using piecewise regression. Proc Inst Civ Eng Water Manag 167:79–87. https://doi.org/10.1680/wama.11.00100
21. Muzzammil M (2010) ANFIS approach to the scour depth prediction at a bridge abutment. J Hydroinform 12:474–485. https://doi.org/10.2166/hydro.2010.004
22. Cheng M-Y, Cao M-T (2014) Hybrid intelligent inference model for enhancing prediction accuracy of scour depth around bridge piers. Struct Infrastruct. https://doi.org/10.1080/15732479.2014.939089
23. Choi S-U, Choi B, Lee S (2017) Prediction of local scour around bridge piers using the ANFIS method. Neural Comput Appl 28:335–344. https://doi.org/10.1007/s00521-015-2062-1
24. Firat M, Gungor M (2009) Generalized regression neural networks and feed forward neural networks for prediction of scour depth around bridge piers. Adv Eng Softw 40:731–737. https://doi.org/10.1016/j.advengsoft.2008.12.001
25. Hosseini R, Amini A (2015) Scour depth estimation methods around pile groups KSCE J. Civ Eng 19:2144–2156. https://doi.org/10.1007/s12205-015-0594-7
26. Zounemat-Kermani M, Beheshti A-A, Ataie-Ashtiani B, Sabbagh-Yazdi S-R (2009) Estimation of current-induced scour depth around pile groups using neural network and adaptive neuro-fuzzy inference system. Appl Soft Comput 9:746–755. https://doi.org/10.1016/j.asoc.2008.09.006
27. Najafzadeh M, Barani G-A, Hessami-Kermani M-R (2014) Group method of data handling to predict scour at downstream of a ski-jump bucket spillway. Earth Sci Inf 7:231–248. https://doi.org/10.1007/s12145-013-0140-4
28. Guven A, Gunal M (2008) Genetic programming approach for prediction of local scour downstream of hydraulic structures. J Irrig Drain Eng 134:241–249. https://doi.org/10.1061/(ASCE)0733-9437(2008)134:2(241)
29. Azamathulla HM, Ghani AA, Zakaria NA, Guven A (2010) Genetic programming to predict bridge pier scour. J Hydraul Eng 136:165–169. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000133
30. Azamathulla HM (2012) Gene expression programming for prediction of scour depth downstream of sills. J Hydrol 460:156–159. https://doi.org/10.1016/j.jhydrol.2012.06.034
31. Kang F, Li J (2016) Artificial bee colony algorithm optimized support vector regression for system reliability analysis of slopes. J Comput Civ Eng 30:04015040. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000514
32. Prayogo D, Susanto YTT (2018) Optimizing the prediction accuracy of friction capacity of driven piles in cohesive soil using a novel self-tuning least squares support vector machine. Adv Civ Eng 2018:9. https://doi.org/10.1155/2018/6490169
33. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn Res 3:1157–1182
34. Imamoto H, Ohtoshi K (1987) Local Scour around a non-uniform circular pier. In: Proceedings of IAHR Congress, Lausanne, Switzerland, pp 304–309
35. Melville BW, Raudkivi AJ (1996) Effects of foundation geometry on bridge pier scour. J Hydraul Eng 122:203–209. https://doi.org/10.1061/(ASCE)0733-9429(1996)122:4(203)
36. Raudkivi AJ, Ettema R (1983) Clear water scour at cylindrical piers. J Hydraul Eng 109:338–350. https://doi.org/10.1061/(ASCE)0733-9429(1983)109:3(338)
37. Liao K-W, Muto Y, Lin J-Y (2017) Scour depth evaluation of a bridge with a complex pier foundation. KSCE J Civ Eng. https://doi.org/10.1007/s12205-017-1769-1
38. Coleman SE (2005) Clearwater local scour at complex piers. J Hydraul Eng 131:330–334. https://doi.org/10.1061/(ASCE)0733-9429(2005)131:4(330)
39. Sheppard DM, Renna R (2005) Florida bridge scour manual Florida DOT. Tallahassee, Florida
40. Vapnik VN (1998) Statistical learning theory. Wiley, New York
41. Kazem A, Sharifi E, Hussain FK, Saberi M, Hussain OK (2013) Support vector regression with chaos-based firefly algorithm for stock market price forecasting. Appl Soft Comput 13:947–958. https://doi.org/10.1016/j.asoc.2012.09.024
42. Salcedo-Sanz S, Rojo-Álvarez JL, Martínez-Ramón M, Camps-Valls G (2014) Support vector machines in engineering: an overview Wiley interdisciplinary reviews. Data Min Knowl Disc 4:234–267. https://doi.org/10.1002/widm.1125
43. Vu DT, Hoang N-D (2016) Punching Shear capacity estimation of FRP-reinforced concrete slabs using a hybrid machine learning approach. Struct Infrastruct Eng 12:1153–1161
44. Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. Stat Comput 14:199–222. https://doi.org/10.1023/b:stco.0000035301.49549.88
45. Cheng M-Y, Hoang N-D (2017) Estimating construction duration of diaphragm wall using firefly-tuned least squares support vector machine. Neural Comput. https://doi.org/10.1007/s00521-017-2840-z
46. Tien Bui D, Tuan TA, Hoang N-D, Thanh NQ, Nguyen DB, Van Liem N, Pradhan B (2016) Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. Landslides. https://doi.org/10.1007/s10346-016-0711-9
47. Unler A, Murat A (2010) A discrete particle swarm optimization method for feature selection in binary classification problems. Eur J Oper Res 206:528–539. https://doi.org/10.1016/j.ejor.2010.02.032
48. Robnik-Šikonja M, Kononenko I (2003) Theoretical and empirical analysis of ReliefF and RReliefF. Mach Learn 53:23–69. https://doi.org/10.1023/a:1025667309714
49. Robnik-Šikonja M, Kononenko I (1997) An adaptation of relief for attribute estimation in regression machine learning. In: Proceedings of the Fourteenth International Conference (ICML'97), Morgan Kaufmann, pp 296–304
50. Hoang N-D, Tien Bui D, Liao K-W (2016) Groutability estimation of grouting processes with cement grouts using differential flower pollination optimized support vector machine. Appl Soft Comput 45:173–186. https://doi.org/10.1016/j.asoc.2016.04.031
51. Liu H, Motoda H (2007) Computational methods of feature selection. CRC Press, Taylor & Francis Group, Boca Raton
52. Kohavi R, John GH (1997) Wrappers for feature subset selection. Artif Intell 97:273–324. https://doi.org/10.1016/S0004-3702(97)00043-X
53. Reunanen J (2003) Overfitting in making comparisons between variable selection methods. J Mach Learn Res 3:1371–1382

54. Mladenović N, Hansen P (1997) Variable neighborhood search. Comput Oper Res 24:1097–1100. https://doi.org/10.1016/S0305-0548(97)00031-2
55. Hoos HH, Stützle T (2004) Stochastic local search foundations and applications. Morgan Kaufmann/Elsevier, Amsterdam
56. Cheng M-Y, Prayogo D, Wu Y-W (2018) Prediction of permanent deformation in asphalt pavements using a novel symbiotic organisms search–least squares support vector regression. Neural Comput Appl. https://doi.org/10.1007/s00521-018-3426-0
57. Wang J, Zhong D, Wu B, Shi M (2018) Evaluation of compaction quality based on SVR with CFA: case study on compaction quality of earth-rock dam. J Comput Civil Eng 32:05018001. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000742