A Hybrid Intelligence Approach to Enhance the Prediction Accuracy of Local Scour Depth at Complex Bridge Piers

Dieu Tien Bui 1,2, Ataollah Shirzadi 3, Ata Amini 4, Himan Shahabi 5,6, Nadhir Al-Ansari 7,*, Shahriar Hamidi 8, Sushant K. Singh 9, Binh Thai Pham 10,*, Baharin Bin Ahmad 11 and Pezhman Taherei Ghazvinei 12

1 Geographic Information Science Research Group, Ton Duc Thang University, Ho Chi Minh City 700000, Vietnam; buitiendieu@tdtu.edu.vn
2 Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City 700000, Vietnam
3 Department of Rangeland and Watershed Management, Faculty of Natural Resources, University of Kurdistan, Sanandaj 66177-15175, Iran; a.shirzadi@uok.ac.ir
4 Kurdistan Agricultural and Natural Resources Research and Education Center, AREEO, Sanandaj 66177-15175, Iran; ata_amini@yahoo.com or a.amini@areeo.ac.ir
5 Department of Geomorphology, Faculty of Natural Resources, University of Kurdistan, Sanandaj 66177-15175, Iran; h.shahabi@uok.ac.ir
6 Board Member of Department of Zrebar Lake Environmental Research, Kurdistan Studies Institute, University of Kurdistan, Sanandaj 66177-15175, Iran
7 Department of Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, 971 87 Lulea, Sweden
8 Department of Water Science and Engineering, Faculty of Agriculture, University of Kurdistan, Sanandaj 66177-15175, Iran; shahriari133@gmail.com
9 Lead AI-ML Scientist, Department of Health, Insurance & Life Sciences, Data & Analytics, Virtusa Corporation, Irvington, NJ 07111, USA; sushantorama@gmail.com
10 Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam
11 Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia (UTM), Johor Bahru 81310, Malaysia; baharina@utm.my
12 Department of Civil Engineering, Technical and Engineering College, Ale Taha University, Tehran 1488836164, Iran; taherei@gmail.com
* Correspondence: nadhir.alansari@ltu.se (N.A.-A.); phamthaibinh2@duytan.edu.vn (B.T.P.)

Received: 2 October 2019; Accepted: 23 December 2019; Published: 3 February 2020

Abstract: Local scour depth at complex piers (LSCP) cause expensive costs when constructing bridges. In this study, a hybrid artificial intelligence approach of random subspace (RS) meta classifier, based on the reduced error pruning tree (REPTree) base classifier, namely RS-REPTree, was proposed to predict the LSCP. A total of 122 laboratory datasets were used and portioned into training (70%: 85 cases) and validation (30%: 37 cases) datasets for modeling and validation processes, respectively. The statistical metrics such as mean absolute error (MAE), root mean squared error (RMSE), correlation coefficient (R), and Taylor diagram were used to check the goodness-of-fit and performance of the proposed model. The capability of this model was assessed and compared with four state-of-the-art soft-computing benchmark algorithms, including artificial neural network (ANN), support vector machine (SVM), M5P, and REPTree, along with two empirical models, including the Florida Department of Transportation (FDOT) and Hydraulic Engineering Circular No. 18 (HEC-18). The findings showed that machine learning algorithms had the highest goodness-of-fit and prediction accuracy (0.885 < R < 0.945) in comparison to the other models. The results of sensitivity analysis by the proposed model indicated that pile cap location (Y) was a more sensitive factor for LSCP among other factors. The result also depicted that the RS-REPTree ensemble model (R = 0.945) could well enhance the prediction power of the REPTree base classifier (R = 0.885). Therefore, the proposed model can be useful as a promising technique to predict the LSCP.
Keywords: scour depth; complex piers; pile cap; machine learning algorithms; ensemble models

1. Introduction

Local scour is responsible for most bridge failures around the world every year. In a streambed, the flow interferes with bridge piers and leads to the creation of multiple vortices, which remove sediment in the vicinity of the piers, and a scour hole is formed [1]. When the scour hole deepens sufficiently, it causes bridge failure. The failures significantly increase the costs of temporary maintenance and also ecological impacts on downstream ecosystems, such as spawning beds [2]. Because of the complicated process of scour around bridge piers, the local scour depth at complex pier (LSCP) is a complicated phenomenon and hence its accurate predictions are a critical issue for the design of bridge foundations. In other words, overestimation of LSCP may lead to extra construction costs and even bridge failure around their foundations [3].

An accurate prediction of LSCP is a hot topic in river engineering because overestimated and underestimated predictions lead to an increase in the dimensions of the bridges, resulting in an increase of the construction costs and bridge failure, respectively [4]. Therefore, a reliable prediction of LSCP for a safe, economic and technically sound structure is of paramount importance. In a river, when a high volume of water flows, scouring of particles around the base of the bridges occurs, and then a scour hole appears around bridge piers. If the LSCP is not predicted correctly, the bottom level of the local scour hole will exceed the original level of the pier foundation. As a result, as time passes and the volume of water flowing increases, local scour depth develops and the bridge’s base loses strength, and eventually it will be destroyed [2]. Almost 53% of all bridge failures are attributed to flood and scour [5].

Over the past decades, the mechanisms and prediction of scour hole occurrence at the simple pier and a group of piles have been widely investigated. Due to economic and technical issues, the piers with complex geometry have developed to become the most common foundation type of bridge piers in alluvial streambeds [6]. The term “complex pier” (CPs) is used in contrary to the simple pier. By definition, the complex pier is a term that defines a special kind of non-uniform pier that is comprised of a column, a pile cap, and a pile group [7]. At piers with complex geometry, due to scouring during a flood, the pile cap position with respect to the initial stream bed level changes. As a result, the influence of pile cap may be changed from a protective to intensifying role at the scour process when it is entirely buried and exposed to the flow, respectively [7]. Such roles increase the complicity of scour mechanisms and prediction at CPs [8].

To estimate local scour depth at complex pier ($y_s$), a few empirical methods have been proposed including the FHWA design methodology, Hydraulic Engineering Circular No. 18. (HEC-18) [9], the Florida Department of Transportation (FDOT) bridge mechanisms scour manual [2,10,11]. In addition, a procedure was proposed by Amini and Mohammad [7] which, based on field data, gives reasonable estimates of the scour depth at CPs [12]. For calculations of scour depth, the HEC-18 and FDOT methods apply a superposition procedure to combine the effect of each element of CPs. However, the methods presented by Lee and Hong [1], Amini et al. [6], and Arneson et al. [9] provided relations for an equivalent width ($b_e$) for that around a CP to be used in simple pier equations where $b_e$ is the diameter of a circular simple pier that produces scour depth equal to the CP, for the same sediment and flow conditions. Apart from HEC-18 and FDOT methods, Mueller and Wagner [13] used field data to examine the efficacy of 20 bridge pier scour depth estimation methods and found that these methods predict the scour depth inaccurately with a large number of overestimations.

In recent years, ensemble machine learning models have become popular among environmental researchers not only for classification issues to generate susceptibility maps [14–36] but also for regression problems to simulate and predict an environmental variable such as wastewater hydraulics [37], saturated hydraulic conductivity [38], shear strength of soft soil [39], soil moisture [40],
and soil temperature [41]. The advantages of artificial intelligence (AI) have encouraged numerous researchers to use methods and techniques based on AI to estimate the depth of scour [42–45]. Based on artificial neural networks (ANNs), some local scour depth estimation methods have been proposed. In case of bridge scouring, Cheng and Cao [46] for predicting local scour depth at simple bridge piers, proposed an intelligent fuzzy radial basis function neural network inference model (IFRIM). Their model was a hybrid of the fuzzy logic, the artificial bee colony algorithm and radial basis function neural network. Najafzadeh et al. [47] presented a group method of data handling, using the back propagation algorithm and quadratic polynomial. They found that the AI-based model provides accurate predictions of scour at simple piers. In the case of local scour prediction at pile group, Zounemat-Kermani et al. [48] and Hosseini et al. [49] reported the accuracy of the ANNs and the neuro-fuzziness system in comparison with empirical methods.

However, contrary to the simple piers and pile groups, due to the complication of the scour mechanism, and the variation of influential parameters, the ensemble machine learning methods have been rarely developed to estimate the scour around the complex piers. The positive and different point of this study with other studies is that there is no applied an ensemble model to predict the LSCP. In other words, although some models and techniques have been used and suggested for predicting the LSCP, the proposed model, RS-REPTree, has not been yielded for this purpose worldwide. Therefore, the main aim of this study was to use a hybrid intelligence model to predict current-induced local scour at complex piers. The presented model enhances the accuracy of scouring predictions and the understanding of the local scour at the complex pier and its dominant variables.

2. Methodology

2.1. Data Acquisition

In this study, two sets of experimental data on local scour at CPs (LSCP) were used. The first dataset was measured by fourth author of this paper at the National Hydraulic Research Institute of Malaysia, NAHRIM. Five models (CPs) were used and experiments were performed over the whole range of possible pile cap elevations with different geometrical characteristics. More details of the data are presented in Amini et al. [7, 50] and Amini et al. [51]. The second dataset of the experiment works were carried out at Sharif University of Technology, Tehran, Iran by Ataie-Ashtiani et al. [52]. Both datasets were measured under flow and model dimensions so that the sediment size, \( d_{50} \), flow depth, \( h \), and contraction effects on LSCP became insignificant. The flow intensity, \( U/U_c \), was selected so that in all tests the clear water condition was maintained, where \( U \) is mean velocity of the approach flow and \( U_c \) is critical mean velocity for sediment motion.

2.2. Dimensional Analysis

To determine scour depth, \( y_s \), most of the empirical methods use of dimensional analysis, a functional relationship, based on an equivalent pier width, \( b_e \), at CPs, from an existing equation for single piers [7, 9–11]. The \( b_e \) is defined as the diameter of a simple pile for the same flow and sediment characteristics that would produce the same scour depth as the CPs. Depending on the pile cap location (Y) with respect to the undisturbed streambed, \( y_s \) or \( b_e \) is a function of flow and sediment properties and CPs’ geometries. Therefore, a functional relationship for presenting LSCP may be written as Equation (1) using dimensional analysis:

\[
y_s \left( \frac{b_c}{b_e} \right) = f \left( \frac{U}{U_c} \text{ or Fr}, \frac{h}{b_c}, \frac{b_c}{b_{pc}}, \frac{T}{b_p}, \frac{Y}{b_c}, \frac{L_u}{b_c}, \frac{L_f}{b_c}, \frac{K_{sc}}{b_c}, \frac{K_{spc}}{b_c}, \frac{b_{pg}}{b_{pc}}, \frac{S_b}{b_{pg}}, \frac{S_l}{b_{pg}} \right) \tag{1}
\]

where \( b_c \) is the column width; \( b_{pc} \) is the pile cap width; \( h \) is flow depth, \( d_{50} \) is median particle size of the bed sediment, \( U_c \) is critical value of \( U \) associated with initiation of motion of bed sediments, Fr is Froude number, \( T \) is the thickness of the pile cap; \( L_u \) and \( L_f \) are extensions of the pile cap upstream of and sides of the column; \( k_{sc} \) and \( k_{spc} \) are the shape factors for the column and pile cap; \( b_{pg} \) is the pile diameter;
m and n are the number of piles in line and normal with the flow; $S_l$ and $S_b$ are the pile spacing in line and normal with the flow, and $Y$ is pile cap elevation with respect to the undisturbed streambed. A schematic drawing for flow-induced scour around a CP and the corresponding parameters are shown in Figure 1.

![Figure 1](image_url)

**Figure 1.** The various components of the composite pier and the corresponding parameters; (a) Upstream view, (b) Side view, and (c) Plan view.

### 2.3. Empirical Equations

The HEC-18 [9] and FDOT methods [11] use the superposition method to predict scour depth at piers with complex geometry (i.e., column, pile cap, and pile group). The superposition method to calculate scour depth contributions from each component in HEC-18, is expressed as Equation (2):

$$y_s = y_{scol} + y_{spc} + y_{spg}$$

(2)

where $y_{scol}$ is scour of column, $y_{spc}$ is the scour of pile cap, and $y_{spg}$ is the scour of pile group. The FDOT method calculates the equivalents single cylindrical pier that would produce the same scour depth as that complex pier component. Then, the equivalent diameter of the CPs is calculated by adding the equivalent diameters of the CP components and expressed as Equation (3):

$$D_{se} = D_{ecol} + D_{epc} + D_{epg}$$

(3)

where $D_{se}$, $D_{ecol}$, $D_{epc}$, and $D_{epg}$ are equivalent diameters of the CPs, column, pile cap, and pile group, respectively. Finally, the scour depth at CPs can be calculated using the methods presented for scouring calculation at simple piers.

### 2.4. Machine Learning Algorithms

#### 2.4.1. Artificial Neural Networks

The artificial neural networks (ANN), developed based on the neurons, is one of the well-known deep learning algorithms for regression, classification, and pattern recognition challenges [53–56]. A typical ANN architecture (Figure 2) has three layers: (a) input layer, in this case all the predictors of LSCP that also decide the number of neurons in the ANN architecture; (b) hidden layer, the recipient of all the neurons containing a specific weight, and (c) output layer, i.e., the predicted value of LSCP. We have discovered that the number of hidden layers and their neurons can only be decided by trial and error, and may vary case-by-case [57]. The ANN has shown promising results in predicting bridge pier scour depths [58–62].
2.4.2. M5P Model Tree

Although the M5P algorithm was developed in 1992 [63,64] by combining traditional decision tree and linear regression function, resulting in induced trees of regression models, Quinlan [63] reconstructed this algorithm. The M5P is one of the decision tree algorithms that has a branch node and leaf node. The first one indicates a choice between a number of alternatives and another represents a classification or decision issue [65].

Consider $S$ as the set of examples is either associated with a leaf, or some test is chosen that splits $S$ into subsets and this process is recursively conducted to the subsets [63,66]. In the M5P, the split process is done according to the minimizing the intra-subset variation down each branch in the output values. Then, in each node the standard deviation of the expected reduction ($σ_{\text{red}}$) is calculated for the output values of the examples. Eventually, the attribute that maximizes the expected error reduction is selected. The $σ_{\text{red}}$ can be formulated as follows [67]:

$$σ_{\text{red}} = \sigma(S) - \sum_{i=1}^{n} \sigma(S_i) \frac{|S_i|}{|S|}$$

where $S$ is the set of examples corresponding to $S_1$, $S_2$, ..., $S_n$ as the sets that result from splitting of the node according to the chosen attribute [64].

2.4.3. Support Vector Machine

Support vector machine (SVM) was recently used for efficiently predicting scour depth at simple piers [62]. SVM, a nonlinear simplified version of the generalized portrait algorithm, is one of the most known robust machine learning algorithms for both regression and classification problems. It can efficiently handle non-linear relationship between the predictors and the response in a multidimensional space [68–70]. Since SVM is based on the statistical learning theory and was developed using real world datasets, it is advantageous in generalizing the outcomes on the unknown data [68]. In this study, we used support vector regression (SVR), following Vapnik’s method where an alternative $\varepsilon$-loss function is used to minimize the error. The SVM offers four kernel functions including linear, polynomial, radial base, and sigmoid. In the SVR, a linear kernel is used, and users can set the kernel parameters. However, the regularization parameter ‘$C$’ and the size of error in sensitive zone ‘$\varepsilon$’ may vary with the number of features and the context. Mathematically, SVR can be presented as below:

$$f(x, α_i, α_i^*) = \sum_{i=1}^{1} (α_i - α_i^*) M(x_i, x_j) + b$$
where $\alpha_i$ and $\alpha_i^*$ are positive Lagrange multipliers and $M(x_i, x_j)$ denotes a non-linear transformation using linear kernel function of SVM, and ‘b’ stands for the ‘bias’.

2.4.4. REP Tree

The reduced error pruning tree (REPTree) is an ensemble model of decision tree (DT) and reduced error pruning (REP) algorithms, equally efficient for classification and regression problems [71]. The REPTree algorithm forms a decision regression tree by splitting and pruning the regression tree based on the highest information gain ratio (IGR) value [72]. The IGR values were calculated based on the entropy ($E$) function by Equation (6).

$$
IGR(x, S) = \frac{E(S) - \sum_{i=1}^{n} \frac{E(S_i | S_j)}{|S|}}{-\sum_{i=1}^{n} \frac{|S_j| \log_2 |S_j|}{|S|}}
$$

(6)

The IGR considers all the predictors of LSCP from training dataset ($S$) with subset $S_i$: $i = 1, 2, \ldots, n$ in consecutive pruning stages. Since complex decision-trees could lead to over-fitting and the reduced-interpretability of a model, REP helps in decreasing the complexity, by removing leaves and branches of the DT structure [17,71,73,74].

2.4.5. Random Subspace Ensemble Algorithm

The random subspace (RS) is another robust ensemble method used for classification as well as regression challenges [75]. RS distinct itself from other ensemble technique and become advantageous because it trains the model on randomly selected samples of features opposed to whole feature sets, therefore, reduces correlations between estimators [76]. In this process, the first regression of original feature space is performed in $L$ training subsets of $q$ dimensionality. Then, base regression is applied to each of these subsets and a final decision is made on the basis of weighted majority voting [77]. In this technique, first REPTree as a base classifier is selected and, after selecting the optimal parameters, number of seeds, and iterations, the model is run and the dataset trained. In the next step, the RS meta classifier is conducted to hybrid with REPTree base classifier. The RS is created some sub-training and then for each dataset the base classifier is performed and eventually, based on majority voting, the best model is selected as the final outcome. The framework of the LSCP prediction is shown in Figure 3.

2.5. Evaluation and Comparison

2.5.1. Statistical Metrics

In this study, some statistical index-based measures including correlation coefficient ($R$), mean absolute error (MAE), root mean square error (RMSE) and a Taylor diagram were used to evaluate and compare the performance of the models. $R$ is a statistical measure, which represents the percentage of the variance for a dependent variable that’s explained by an independent variable. MAE measures the mean absolute value of each difference. Compared with RMSE, MAE can be given to be a more natural and unambiguous index to measure errors between estimated and actual observed values [78,79]. RMSE has been applied as a regular statistical metric to measure model performance [80]. Lower values of RMSE indicate the better result. It has been widely used different fields around the world [81]. The Taylor diagram graphically shows how the prediction models are matched with observations in terms of correlation, their root-mean-square difference, and the ratio of their variance considered in a single diagram [82]. On this diagram, the mode that is closer to the observation (validation dataset)
has the highest predictive performance [83]. The abovementioned statistical indexes can be calculated by Equations (7)–(9) as below [84]:

\[
R = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}
\] (7)

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |x_i - y_i|
\] (8)

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{N} (x_i - y_i)^2}{N}}
\] (9)

where \(R\) is the correlation coefficient, \(x_i\) and \(y_i\) are measured and predicted values respectively, \(\bar{x}\) and \(\bar{y}\) are the mean of measured and predicted values respectively, and \(N\) is the number of input data.

**Figure 3.** A methodological framework of LSCP prediction.

2.5.2. Non-Parametric Statistical Tests

In order to check the statistical difference between five models to enhance the prediction accuracy of LSPC, Friedman and Wilcox rank tests were used. As a null hypothesis, it was assumed that there was no statistical significance between the performances of all LSCP models at the significance level of \(\alpha = 0.05\). In other words, it attempted to show there was no variation between the models or a single model was no different from its mean. Regarding the P-value, if the null hypothesis is rejected
(\( p < 0.05 \)), it concludes that there is a significant difference between the models \([85]\). The Friedman test only shows the significant differences among all LSCP models but not show as pairwise comparisons. In this case, the Wilcoxon sign-rank test can be used to evaluate the pairwise differences between the LSCP models. Statistically, if the null hypothesis is rejected \( (p < 0.05 \text{ and } z\text{-value} > (-1.96 \text{ and } +1.96) \), there is significant differences between the performance of the models \([72,85]\).

2.6. Sensitivity Analysis

The effect of each factor on the local scour depth is assessed by the sensitivity analysis (SA) technique \([49,86]\). The SA method in this study was computed by the proposed machine learning algorithm named RS-REPTree. In the SA method, each factor is removed individually from the modeling process and then the values of MAE, RMSE, and R metrics are recorded. Eventually, the sensitive factors during the modeling process are known and listed. Judgement is performed based on the effectiveness of each input factor. The higher the values of RMSE and MAE and the lower the R for each factor is, the higher sensitivity of the factor for the modeling process of local scour depth will be \([49]\).

3. Results and Analysis

3.1. Optimal Selection of Modeling Parameters

Theoretically, an accurate model will be obtained when the parameters of the model are correctly selected and optimized. The optimal values are achieved by trial and error using the modification of the parameters. The optimal value of each machine learning parameters is shown in Table 1. In the proposed ensemble model, RS-REPTree, the most important parameters are the number of seeds and iterations during the modeling process. Figure 4 shows the best values for these two parameters of the ensemble model based on the RMSE and R metrics. According to the test/validation dataset, the result indicated that the best values for the number of seed and iteration based on the lowest RMSE metric (0.0181) were 6 and 10, respectively (Figure 4a,b). Additionally, the optimal values for the number of seeds and iterations regarding the R metric (0.945) were also 6 and 10 (Figure 4c,d). Therefore, the proposed ensemble model was constructed by the obtained parameters to enhance the prediction accuracy of LSCP.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>Number of hidden layer: 7; learning rate: 0.3; momentue: 0.2; Number of seed: 3; training time: 500; validation threshold: 20; validation set size: default</td>
</tr>
<tr>
<td>M5P</td>
<td>Build regression tree: True; minimum number of instance: 4</td>
</tr>
<tr>
<td>SVM</td>
<td>C: 0.95; filter type: normalized training data; regOptimizer: RegSMO improved; number of seed: 1; tolerance: 0.001</td>
</tr>
<tr>
<td>REPTree</td>
<td>Maximum depth: −1; minimum number: 2; minimum variance probability: 0.001; number of fold: 2; number of seed: 1</td>
</tr>
<tr>
<td>RS-REPTree</td>
<td>Classifier: REPTree; Number of iteration: 10; number of seed: 6; subspace size: 0.5</td>
</tr>
</tbody>
</table>

3.2. Model Validation and Comparison

The goodness-of-fit and performance of the models using training and testing/validation datasets are shown in Table 2 and Figure 5a–n. The experimental data of LSCP was used to compare obtained models and then their predictive accuracies were evaluated using the statistical metrics as MAE, RMSE and R. It is noteworthy that the goodness-of-fit and performance of the models are judged using training and validation datasets, respectively. However, to compare the power prediction and
performance of the models testing/validation dataset were used. Basically, the number of MAE, RMSE, and R for the FDOT method were 0.058, 0.062, 0.726 and for the HECT-18 were 0.051, 0.064, and 0.620, respectively. In the ANN model, these values were 0.016, 0.021, and 0.907 while in the M5P machine learning model they were 0.017, 0.022, and 0.912, respectively. For the SVM model, the number of MAE, RMSE, and R were 0.016, 0.024, and 0.918 and for the REPTree mode were 0.018, 0.025, and 0.885, respectively. In the proposed ensemble model the values of 0.014, 0.018, and 0.945 were obtained for MAE, RMSE, and R, respectively. Overall, results indicated that, although all machine learning models had higher accuracy than the empirical models, the proposed model, RS-REPTree, well enhanced the accuracy of the based classifier of REPTree. On the other hand, this model had the highest prediction accuracy in predicting LSCP.

Figure 4. Cont.
Figure 4. Determination of the number of optimal values of iteration and seed in the modeling process based on the RMSE and R; (a) number of seed by RMSE, (b) number of iteration using RMSE, (c) number of iteration using R and (d) number of seed using R.

Table 2. Model results and comparison using training and validation datasets.

<table>
<thead>
<tr>
<th>Models</th>
<th>MAE Training</th>
<th>MAE Validation</th>
<th>RMSE Training</th>
<th>RMSE Validation</th>
<th>R Training</th>
<th>R Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDOT</td>
<td>0.045</td>
<td>0.058</td>
<td>0.032</td>
<td>0.062</td>
<td>0.736</td>
<td>0.726</td>
</tr>
<tr>
<td>HEC-18</td>
<td>0.053</td>
<td>0.051</td>
<td>0.067</td>
<td>0.064</td>
<td>0.625</td>
<td>0.620</td>
</tr>
<tr>
<td>ANN</td>
<td>0.012</td>
<td>0.016</td>
<td>0.015</td>
<td>0.021</td>
<td>0.954</td>
<td>0.907</td>
</tr>
<tr>
<td>M5P</td>
<td>0.014</td>
<td>0.017</td>
<td>0.020</td>
<td>0.022</td>
<td>0.943</td>
<td>0.912</td>
</tr>
<tr>
<td>SVM</td>
<td>0.015</td>
<td>0.016</td>
<td>0.020</td>
<td>0.024</td>
<td>0.924</td>
<td>0.918</td>
</tr>
<tr>
<td>REPTree</td>
<td>0.013</td>
<td>0.018</td>
<td>0.021</td>
<td>0.025</td>
<td>0.931</td>
<td>0.885</td>
</tr>
<tr>
<td>RS-REPTree</td>
<td>0.013</td>
<td>0.014</td>
<td>0.019</td>
<td>0.018</td>
<td>0.946</td>
<td>0.945</td>
</tr>
</tbody>
</table>
Figure 5a–n graphically shows the relationship between actual and predicted scour depths (m) based on the training and validation datasets. It was concluded that the empirical models had the lowest prediction because the distance between actual and predicted local scour depth was higher than the other models. The lower the MAE and RMSE and the higher R, the higher the model prediction and lower distance between actual and predicted local scour depths. Although all machine learning models had a reasonable fit of actual and predicted local scour depth, the proposed model showed the best simulating results.

Figure 6 shows the correlation between actual and predicted values of LSCP for empirical and machine learning models. Figure 6 was plotted in SPSS software to show the agreement of the actual and predicted values of local scour depth. The results illustrated that, based on the validation dataset, the values of R for the FDOT, HEC-18, ANN, M5P, SVM, REPTree, and RS-REPTree models were 0.726, 0.620, 0.907, 0.912, 0.918, 0.885, and 0.945, respectively. This implied that the lowest value of R belonged to HEC-18 as an empirical model, and the highest one was obtained for the proposed model. Overall, the result received that all machine learning models outperformed and outclassed the empirical models for predicting the local scour depth; however, the RS-REPTree ensemble model was more powerful than the other models.

Figure 5. Cont.
Figure 5. Cont.
Figure 5. Comparison between actual and predicted scour depth using training dataset; (a) FDOT model, (c) HEC-18, (e) ANNMLP, (g) M5P, (i) SVM, (k) REPTree, (m) RS-REPTree, and Testing dataset; (b) FDOT, (d) HEC-18, (f) ANNMLP, (h) M5P, (j) SVM, (l) REPTree and (n) RS-REPTree.

Figure 6. Cont.
Figure 6. Analysis of correlation between actual and predicted values of LSCP (m) for empirical models and machine learning algorithms.

A Taylor diagram was plotted to further analyze model applicability (Figure 7). This diagram showed that although all machine learning models were close to observed LSCP, the proposed model, RS-REPTree, had the most predictive power compared to other models. The proposed model had a closer correlation (0.946), RMSE (0.018), and SD (0.044) with observed data. The results indicated that the models had good predictive power.
Figure 7. Taylor diagram for displaying the correlation between the observed and predicted LSCP by different machine learning models.

Beside the above-mentioned graphs and tables, the results were assessed based on the boxplot of the models as shown in Figure 8. The box plots were plotted using the observed versus predicted LSCP for machine learning and empirical methods. The spread of measured and predicted values revealed that machine learning models present more accurate predictions in comparison to HEC-18 and FDOT models.

Figure 8. Box plots of the actual and predicted values of local scour depth.
Additionally, to find significant differences among models evaluated in this research the Friedman’s test, a non-parametric test for detecting differences in treatments across multiple attempts [87], was used. Since the Friedman’s test does not show pairwise comparisons, the Wilcoxon test was used to quantify pairwise comparisons among the models. The results of the Wilcoxon test for performance of the RS-REPTree model and other LSCP models are shown in Table 3. The obtained significance value for the Friedman test was 0.000 (<0.05), which indicated that at 95% confidence level there was no evidence to accept the null hypothesis and then it was rejected that there are no differences between the mean of all models. Therefore, a significant difference between the performances of the models to predict the LSCP was indicated. In addition, the Wilcoxon sign rank test, as a pairwise comparison of the performance of the models, was used.

Table 3. Average ranking of the seven local scour depth models using the Friedman’s test.

<table>
<thead>
<tr>
<th>No</th>
<th>Scour Depth Models</th>
<th>Mean Ranks</th>
<th>$\chi^2$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FDOT</td>
<td>6.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>HEC-18</td>
<td>6.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>ANN</td>
<td>3.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>MSP</td>
<td>3.62</td>
<td>158.012</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>SVM</td>
<td>4.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>REPTree</td>
<td>3.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>RS-REPTree</td>
<td>3.38</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4 shows significant differences between empirical methods predictions and actual data using Wilcoxon pairwise rank test. In compare with machine learning models, both HEC-18 and FDOT methods are with more over and under predicted data. It should be noted that avoiding under predicting data, is of most importance for bridge designers. Table 4 shows that, apart from a significant difference between the empirical models and actual experimental data, the empirical models and machine learning models are significantly different. However, no significant differences between machine learning predictions and actual data can be concluded from Table 4. Overall, according to the statistical tests, it can be safely said that the result of machine learning models to predict local scour depth is more reasonable and reliable than the empirical models.

Table 4. Performance of the RS-REPTree model compared to other LSCP models using Wilcoxon signed-rank test (two-tailed).

<table>
<thead>
<tr>
<th>NO</th>
<th>Pairwise Comparison</th>
<th>NND</th>
<th>NPD</th>
<th>z-Value</th>
<th>p-Value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Actual-FDOT</td>
<td>9</td>
<td>65</td>
<td>−6.608</td>
<td>0.000</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>Actual-HEC18</td>
<td>14</td>
<td>68</td>
<td>−6.732</td>
<td>0.000</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>Actual-ANN</td>
<td>39</td>
<td>44</td>
<td>−0.409</td>
<td>0.683</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>Actual-MSP</td>
<td>45</td>
<td>39</td>
<td>−0.085</td>
<td>0.932</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>Actual-SVM</td>
<td>37</td>
<td>38</td>
<td>−0.481</td>
<td>0.631</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>Actual-REPTree</td>
<td>45</td>
<td>39</td>
<td>−0.112</td>
<td>0.911</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>Actual-RSREPTree</td>
<td>41</td>
<td>40</td>
<td>−0.443</td>
<td>0.658</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>HEC18-FDOT</td>
<td>40</td>
<td>24</td>
<td>−0.994</td>
<td>0.320</td>
<td>No</td>
</tr>
<tr>
<td>9</td>
<td>HEC18-ANN</td>
<td>68</td>
<td>14</td>
<td>−6.619</td>
<td>0.000</td>
<td>Yes</td>
</tr>
<tr>
<td>10</td>
<td>HEC18-MSP</td>
<td>74</td>
<td>10</td>
<td>−6.927</td>
<td>0.000</td>
<td>Yes</td>
</tr>
<tr>
<td>11</td>
<td>HEC18-SVM</td>
<td>68</td>
<td>16</td>
<td>−6.442</td>
<td>0.000</td>
<td>Yes</td>
</tr>
<tr>
<td>12</td>
<td>HEC18-REPTree</td>
<td>70</td>
<td>15</td>
<td>−6.806</td>
<td>0.000</td>
<td>Yes</td>
</tr>
<tr>
<td>13</td>
<td>HEC18-RSREPTree</td>
<td>71</td>
<td>13</td>
<td>−6.848</td>
<td>0.000</td>
<td>Yes</td>
</tr>
<tr>
<td>14</td>
<td>FDOT-ANN</td>
<td>78</td>
<td>10</td>
<td>−6.799</td>
<td>0.000</td>
<td>Yes</td>
</tr>
<tr>
<td>15</td>
<td>FDOT-MSP</td>
<td>73</td>
<td>12</td>
<td>−6.765</td>
<td>0.000</td>
<td>Yes</td>
</tr>
<tr>
<td>16</td>
<td>FDOT-SVM</td>
<td>67</td>
<td>18</td>
<td>−6.536</td>
<td>0.000</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Table 4. Cont.

<table>
<thead>
<tr>
<th>NO</th>
<th>Pairwise Comparison</th>
<th>NND</th>
<th>NPD</th>
<th>z-Value</th>
<th>p-Value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>FDOT-REPTree</td>
<td>78</td>
<td>7</td>
<td>-7.072</td>
<td>0.000</td>
<td>Yes</td>
</tr>
<tr>
<td>18</td>
<td>FDOT-RSREPTree</td>
<td>67</td>
<td>13</td>
<td>-6.799</td>
<td>0.000</td>
<td>Yes</td>
</tr>
<tr>
<td>19</td>
<td>ANN-M5P</td>
<td>40</td>
<td>39</td>
<td>-0.364</td>
<td>0.716</td>
<td>No</td>
</tr>
<tr>
<td>20</td>
<td>ANN-SVM</td>
<td>32</td>
<td>50</td>
<td>-1.371</td>
<td>0.170</td>
<td>No</td>
</tr>
<tr>
<td>21</td>
<td>ANN-REPTree</td>
<td>49</td>
<td>32</td>
<td>-0.393</td>
<td>0.694</td>
<td>No</td>
</tr>
<tr>
<td>22</td>
<td>ANN-RSREPTree</td>
<td>37</td>
<td>47</td>
<td>-0.116</td>
<td>0.908</td>
<td>No</td>
</tr>
<tr>
<td>23</td>
<td>M5P-SVM</td>
<td>36</td>
<td>46</td>
<td>-1.318</td>
<td>0.188</td>
<td>No</td>
</tr>
<tr>
<td>24</td>
<td>M5P-REPTree</td>
<td>42</td>
<td>36</td>
<td>-0.416</td>
<td>0.677</td>
<td>No</td>
</tr>
<tr>
<td>25</td>
<td>M5P-RSREPTree</td>
<td>35</td>
<td>49</td>
<td>-0.989</td>
<td>0.323</td>
<td>No</td>
</tr>
<tr>
<td>26</td>
<td>SVM-REPTree</td>
<td>46</td>
<td>39</td>
<td>-0.734</td>
<td>0.463</td>
<td>No</td>
</tr>
<tr>
<td>27</td>
<td>SVM-RSREPTree</td>
<td>47</td>
<td>36</td>
<td>-0.115</td>
<td>0.265</td>
<td>No</td>
</tr>
<tr>
<td>28</td>
<td>RSREPTree-RSREPTree</td>
<td>43</td>
<td>37</td>
<td>-0.187</td>
<td>0.852</td>
<td>No</td>
</tr>
</tbody>
</table>

NND: Number of negative differences; NPD: Number of positive differences; “the standard p value is 0.05”.

3.3. Sensitivity Analysis

The role of each factor on the result of scour depth modeling was assessed by sensitivity analysis of the proposed model (Figure 9). The results are ordered according to the effectiveness of each factor that meaning the factor with removal creates higher RMSE and MAE, as well as the lower R located at the top of the list. The results stated that the pile cap level is the most important factor in the LSCP in a particular model. Moreover, other significant factors were pile cap thickness (T) and width (b<sub>pc</sub>). The rest factors have slight effect for modeling process by the proposed ensemble model (Figure 9).

![Figure 9](image-url)  

**Figure 9.** Sensitivity analysis graphically based on the proposed ensemble model to predict local scour depth.

4. Discussion

The flow disturbances around obstacles such as bridge piers, inserted into an alluvial streambed, induce local scouring which is one of the most common being riverbed scour. Since the late 1950s the estimation of scour at bridges has attracted the attention of many researchers [88]. Unlike the local scour at simple pier, the scour at CPs is a complex phenomenon. However, the accurate estimation of equilibrium scour depth at CPs is vital for safe designing of the bridges. The experimental data were used to compare the most commonly used method for predicting scour depth at CPs. Based on data set analyses, an ensemble model was constructed to improve the accuracy of
local scour depth prediction of REPTree as a base classifier. The statistical measurements indicated that the best values for the number of seeds and iterations were 6 and 10, respectively.

The capabilities and performance of the empirical methods and obtained models in scour prediction were evaluated using statistical tests. Overall, the statistical tests showing the relationship between observed and predicted scour depths indicated that all machine learning models are with higher power prediction than the empirical models. The inaccuracy of empirical methods at CPs scour including HEC-18 and FDOT methods was reported [6,12,52]. In contrast, the superiority of intelligent models to empirical methods for scour predicting was stated by Zounemat-Kermani et al. [48] and Hosseini et al. [49]. The same results can be concluded from the correlation between the observed and predicted values of local scour depth for empirical and machine learning models. The lowest values of R were obtained for HEC-18 and the highest belonged to the RS-REPTree as 0.620 and 0.945, respectively. Moreover, the boxplot and Friedman's test of the models support the above statements.

In the case of detecting the dominant parameters at inducing local scour at CPs, the sensitivity analyses were conducted and the parameters were ordered according to their effectiveness. The results depicted that the pile cap location (Y) in respect to undisturbed streambed is the most important parameter which influence the scouring at CPs. These results are consistent with those reported by [51,52,89]. Furthermore, the pile cap width (bpc), thickness (T) and column width (bc) are with higher influences on LSCP, respectively. These results are in agreement with the findings of Ferraro et al. [90] and Moreno et al. [89]. It should be noted that unlike the simple piers and pile group [91], the role of the parameters at producing LSCP is various versus the pile cap level in respect to undisturbed streambed. Particularly when the pile cap is lower or inside the scour hole, the pile cap prevents scouring. This process continues until flow penetrates below the pile cap and the pile cap becomes undercut. The undercutting of pile caps intensifies the scour depth. As the level of pile cap reached the position for undercutting, apart from column and pile cap, the pile group is exposed to the flow, and contributes towards scouring. As the pile cap level increased, the pile group prevents the scouring and diminishes the LSCP [50,52]. In this case, the LSCP depends on the pile cap and pile group characteristics. It is worth noting that the bridge pier models used to obtain the data in this research were selected so that the sediment size and flow depth effects on LSCP became negligible and flow intensity was in a confined range.

The ensemble models could more decrease the noise and over-fitting problems between the training dataset, resulting in enhancing the accuracy of the model [15,17,92,93]. Basically, the findings depicted that the RS-REPTree ensemble model could well enhance the prediction accuracy of the REPTree as a classifier for the prediction of local scour depth at piers with complex geometry. This finding is in agreement with Cheng and Cao [46] who reported the capability of the IFRIM as a promising tool for civil engineers to estimate local scour at piers with simple geometry.

5. Conclusions

The complexity of the scour mechanisms at piers with non-uniform geometry caused the inaccuracy in the empirical methods presented for scouring prediction at these piers. In this research, the comprehensive datasets were used to evaluate the most commonly used empirical methods and to present a machine learning algorithms approach for predicting LSCP. The typical geometry of non-uniform piers is a complex pier (CP) composed of a column resting on a pile cap supported by a group of piles, which is investigated in this research. The most obvious findings to emerge from this study can be present as:

1. The machine learning algorithms have the powerful capability to predict LSCP and the hybrid models can improve the performance of separate models in predicting LSCP.
2. Computing benchmark algorithms presented in this research have the potential to alter the LSCP prediction in comparison with the most well-known empirical methods, namely HEC-18 and FDOT methods.
3. The state-of-the-art RS-REPTree ensemble model, with the highest accuracy of the REPTree, is proposed as a classifier for the prediction of the LSCP.

4. The pile cap location (Y) was a more sensitive factor for LSCP among other factors based on the availability of data.

Author Contributions: D.T.B., A.S., A.A., H.S., N.A.-A., S.H., S.K.S., B.T.P., and B.B.A. contributed equally to the work. A.S., A.A., H.S., and S.H. collected field data and conducted the analysis. A.S., A.A., H.S., N.A.-A., S.H., S.K.S. and B.T.P. wrote the manuscript. D.T.B., N.A.-A., S.K.S., B.T.P., B.B.A. and P.T.G. provided critical comments in planning this paper and edited the manuscript. All the authors were involved in discussing the results. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>Root Mean Squared Error</td>
</tr>
<tr>
<td>LSCP</td>
<td>Local Scour Depth at Complex Piers</td>
</tr>
<tr>
<td>RS</td>
<td>Random Subspace</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Network</td>
</tr>
<tr>
<td>R</td>
<td>Correlation Coefficient</td>
</tr>
<tr>
<td>d50</td>
<td>Median Sediment Size</td>
</tr>
<tr>
<td>Ys</td>
<td>Scour Depth</td>
</tr>
<tr>
<td>h</td>
<td>Water Depth</td>
</tr>
<tr>
<td>bc</td>
<td>Column Width</td>
</tr>
<tr>
<td>lc</td>
<td>Column Length</td>
</tr>
<tr>
<td>bpc</td>
<td>Pile Cap Width</td>
</tr>
<tr>
<td>lpc</td>
<td>Pile Cap Length</td>
</tr>
<tr>
<td>T</td>
<td>Pile Cap Thickness</td>
</tr>
<tr>
<td>Lu</td>
<td>Extension length of pile cap out from the column face</td>
</tr>
<tr>
<td>Lf</td>
<td>Extension width of pile cap out from the column</td>
</tr>
<tr>
<td>ksc</td>
<td>Shape factor for the column</td>
</tr>
<tr>
<td>kspc</td>
<td>Shape factor for the pile cap</td>
</tr>
<tr>
<td>bpg</td>
<td>Pile diameter</td>
</tr>
<tr>
<td>Fr</td>
<td>Froude number</td>
</tr>
<tr>
<td>m</td>
<td>Number of piles in line with the flow</td>
</tr>
<tr>
<td>n</td>
<td>Number of piles normal with the flow</td>
</tr>
<tr>
<td>Sl</td>
<td>Pile spacing in line with the flow</td>
</tr>
<tr>
<td>Sb</td>
<td>Pile spacing normal with the flow</td>
</tr>
<tr>
<td>Y</td>
<td>Pile cap elevation in respect to undisturbed streamflow</td>
</tr>
<tr>
<td>be</td>
<td>Equivalent width/diameter</td>
</tr>
<tr>
<td>yscol</td>
<td>Column’s scour</td>
</tr>
<tr>
<td>yspc</td>
<td>Pile cap’s scour</td>
</tr>
<tr>
<td>yspg</td>
<td>Scour of pile group</td>
</tr>
<tr>
<td>Des</td>
<td>Equivalent diameters of the complex pier</td>
</tr>
<tr>
<td>Decol</td>
<td>Equivalent diameters of the column</td>
</tr>
<tr>
<td>Depc</td>
<td>Equivalent diameters of the pile cap</td>
</tr>
<tr>
<td>Depg</td>
<td>Equivalent diameters of the pile group</td>
</tr>
<tr>
<td>X</td>
<td>Training dataset</td>
</tr>
<tr>
<td>S</td>
<td>Subset of training dataset</td>
</tr>
<tr>
<td>Uc</td>
<td>Critical velocity for the beginning of sediment motion</td>
</tr>
<tr>
<td>U</td>
<td>Mean approach flow velocity</td>
</tr>
</tbody>
</table>

References


17. Pham, B.T.; Prakash, I.; Singh, S.K.; Shirzadi, A.; Shahabi, H.; Bui, D.T. Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. *Catena* 2019, 175, 203–218. [CrossRef]


78. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geosci. Model. Dev.* 2014, 7, 1247–1250. [CrossRef]


84. Barnston, A.G. Correspondence among the correlation, RMSE, and Heidke forecast verification measures; refinement of the Heidke score. *Weather Forecast.* 1992, 7, 699–709. [CrossRef]


86. Najafzadeh, M.; Rezaie Balf, M.; Rashedi, E. Prediction of maximum scour depth around piers with debris accumulation using EPR, MT, and GEP models. *J. Hydroinform.* 2016, 18, 867–884. [CrossRef]


89. Moreno, M.; Maia, R.; Couto, L. Effects of relative column width and pile-cap elevation on local scour depth around complex piers. *J. Hydraul. Eng.* 2015, 142, 04015051. [CrossRef]


