A Comprehensive Experimental and Computational Investigation on Estimation of Scour Depth at Bridge Abutment: Emerging Ensemble Intelligent Systems

Manish Pandey[1] · Masoud Karbasi[2] · Mehdi Jamei[3,4] · Anurag Malik[5] · Jaan H. Pu[6]

Abstract

Several bridges failed because of scouring and erosion around the bridge elements. Hence, precise prediction of abutment scour is necessary for the safe design of bridges. In this research, experimental and computational investigations have been devoted based on 45 flume experiments carried out at the NIT Warangal, India. Three innovative ensemble-based data intelligence paradigms, namely categorical boosting (CatBoost) in conjunction with extra tree regression (ETR) and K-nearest neighbor (KNN), are used to accurately predict the scour depth around the bridge abutment. A total of 308 series of laboratory data (a wide range of existing abutment scour depth datasets (263 datasets) and 45 flume data) in various sediment and hydraulic conditions were used to develop the models. Four dimensionless variables were used to calculate scour depth: approach densimetric Froude number (F_{d50}), the upstream depth (y) to abutment transverse length ratio (y/L), the abutment transverse length to the sediment mean diameter (L/d_{50}), and the mean velocity to the critical velocity ratio (V/V_{cr}). The Gradient boosting decision tree (GBDT) method selected features with higher importance. Based on the feature selection results, two combinations of input variables (comb1 (all variables as model input) and comb2 (all variables except F_{d50})) were used. The CatBoost model with Comb1 data input (RMSE = 0.1784, R = 0.9685, MAPE = 10.4724) provided better accuracy when compared to other machine learning models.

✉ Masoud Karbasi
m.karbasi@znu.ac.ir

✉ Mehdi Jamei
mehdi.jamei59@gmail.com

Extended author information available on the last page of the article

1 Introduction

Scour is a natural phenomenon that occurs in alluvial streams as a result of the erosive action of flowing water (Oliveto and Hager 2005; Eghlidi et al. 2020). Abutments are constructed near the stream banks for constructing and supporting a bridge. It is generally recognized that abutments are undermined by river-bed erosion and scouring, which are the leading causes of abutment failure (Pandey et al. 2020; Afzal et al. 2021). Natural processes or man-made interactions can cause local scour in streams (Barbhuiya and Dey 2004; Kothyari et al. 2007; Goyal and Ojha 2011). Abutments and other hydraulic structures are often failed by scour (Pandey et al. 2020). Earlier studies on scour around abutments tended to focus on the prediction of maximum, and time-dependent scour depth (Barbhuiya and Dey 2004; Kothyari et al. 2007). The Federal Highway Administration 1973 report illustrates that more than 400 bridges failed due to pier and abutment scour (Pandey et al. 2018). These collapsed bridges show the importance of realistic prediction of scour around the bridge elements (Kumar et al. 2022). Scour around the abutments in natural streams is still a matter of alarm, while numerous researchers and engineers have proposed numerous mathematical and numerical models using laboratory and field studies (Barbhuiya and Dey 2004; Dey and Barbhuiya 2005). Therefore, improving the abutment scour phenomenon and processes is vital to calculate the maximum abutment scour depth at equilibrium scour conditions. (Oliveto and Hager 2005) carried out an experimental study on abutment scour and stated that the minimum required laboratory dimensions to apply Froude number similarity. They also checked the influence of sloping abutments on scour. As long as the limitations of the computational approach are respected, the outcomes of their study could be applied in practice. (Dey and Barbhuiya 2005) completed an experimental flume study on abutment scour and derived a semi-empirical method to calculate the abutment scour around a short vertical wall, semicircular, and 45° wing wall abutments (length/flow depth ≤ 1). By considering the horse-shoe vortex system as the prime agent of abutment scour, they followed the conservation of sediment mass for analyzing the experimental data. The maximum scour depth at equilibrium conditions around abutments is key to the river and bridge engineers (Barbhuiya and Dey 2004).

The scour processes and flow patterns around the abutments are so complicated; thus, it is hard to derive a general empirical relationship (Azamathulla et al. 2010, 2013; Singh et al. 2020). The scour depth around the abutment can be calculated using numerous empirical relationships (Dey and Barbhuiya 2005; Mohammadpour et al. 2017). Each relationship yields good agreements with experimental values just for the specific datasets. Previous studies show the importance of the prediction of abutment scour precisely. Further, insufficient field data would lead to uncertainty of abutment scour equations (Mohammadpour et al. 2016).

With this in view, precise prediction of the abutment is cumbersome and needs to implement robust approaches like gene-expression programming (GEP), artificial neural network (ANN), generalized reduced gradient (GRG), genetic algorithm (GA), evolutionary polynomial regression, modern multi-level ensemble approach and adaptive neuro-fuzzy inference system (ANFIS) to predict abutment scour (Azamathulla et al. 2010; Mohammadpour et al. 2016; Mohammadpour 2017; Aamir and Ahmad 2019; Singh et al. 2022). Azamathulla et al. (Najafzadeh and Azamathulla 2013) applied soft computing approaches to calculate the pipeline scour and illustrated the best results with experimental datasets. (Mohammadpour et al. 2016) applied ANN and ANFIS to

calculate the abutment scour. They stated that the ANN and ANFIS could be successfully used to predict the scour depth at different bridge elements.

In this study, the authors aim to examine the existing maximum scour depth data at equilibrium scour conditions for rectangular wall abutments. Existing abutment scour depth datasets (263 datasets) are collected from (Coleman et al. 2003) and (Dey and Barbhuiya 2005). Further, 45 flume experiments are carried out at the National Institute of Technology in Warangal, India. Previously proposed empirical relationships are also examined to analyze the performance based on available datasets (a total of 308 datasets). Although the applications of robust approaches in maximum scour depth prediction have been of interest to several investigators because of their accuracy and simplicity, no research work has been undertaken to develop a machine learning approach to predict the maximum scour depth at the bridge abutment. However, previous studies on other hydraulic structures scour have shown good agreements with observed values. Considering the significance of the concern of bridge abutment scour, which is mainly responsible for the scour hazards. An effort has been made to develop machine learning based models with a wide range of experimental data. The main objectives of this study are: 1- Application of the CatBoost model as new a machine learning approach to model abutment scour depth using a vast experimental database, 2- Compare developed model with common machine learning approaches such as K nearest neighbor and extra tree models. 3- Compare existing empirical equations with the machine learning models.

2 Methodology

2.1 Maximum Scour Depth Relationships

Numerous studies are available on abutment scour. Most studies have examined the maximum scour at equilibrium scour conditions. At equilibrium conditions, maximum scour depth is influenced by flow properties, sediment characteristics, and abutment geometry (Barbhuiya and Dey 2004; Bressan et al. 2011). The variables that influence the maximum abutment scour depth (d_{se}) at equilibrium conditions in uniform sediment beds are as follows:

$$d_{se} = f\left(V, y, \rho, \upsilon, d_{50}, V_{cr}, \rho_s, L, g\right) \tag{1}$$

where d_{se} is the maximum abutment scour depth, V is time-average flow velocity, y is the approach flow depth, ρ is the mass density of the fluid, υ is the kinematic viscosity of fluid, d_{50} is the median diameter of sediment, V_c is the threshold velocity of sediment, ρ_s is the mass density of sediment, L is the transverse length of the abutment and g is the gravitational acceleration.

For sediment-fluid interaction, Eq. (1) should not contain independent parameters ρ, ρ_s and g (Dey and Barbhuiya 2005). (Dey and Barbhuiya 2005) gave a better representation of above mentioned parameters, and hence, Eq. (1) becomes

$$d_{se} = f\left(F_{d_{50}}, \frac{V}{V_{cr}}, \frac{y}{L}, \frac{L}{d_{50}}\right) \tag{2}$$

where, densimetric Froude number $F_{d_{50}} = \frac{V}{\sqrt{(S-1)g d_{50}}}$, S is the relative density of sediment.

Many researchers derived mathematical relationships that calculate the maximum abutment scour depth using various parameters (Kandasamy and Melville 1998; Melville and Coleman 2000). (Melville and Coleman 2000) proposed an abutment scour relationship on the basis of different empirical factors or K factors which show the influence of flow, sediment, and abutment characteristics. These K-factors can be calculated by the curve fitting method, and the maximum abutment scour depth (d_{se}) expressed in terms of the product of K-factors is given as

$$d_{se} = K_{d_{50}} K_I K_S K_y K_\alpha \tag{3}$$

where d_{se} is the maximum abutment scour depth at equilibrium condition, K_{d50} is the sediment gradation factor, K_I is the flow intensity factor, K_s is the abutment shape factor, K_t is the time factor, K_y is the flow depth–abutment size factor, and K_α is the abutment alignment factor. For vertical wall abutments, K_s and $K_\alpha = 1$.

K factors can be calculated using different empirical equations, given as

$$\left.\begin{array}{l} K_{d_{50}}\left(L/d_{50} \leq 25\right) = 0.57 \log\left(2.24\dfrac{L}{d_{50}}\right) \\[2mm] K_{d_{50}}\left(L/d_{50} > 25\right) = 1 \end{array}\right\} \tag{4}$$

$$\left.\begin{array}{l} K_I = \dfrac{V - \left(V_a - V_{cr}\right)}{V_c}, \, for \, \dfrac{V - \left(V_a - V_{cr}\right)}{V_c r} < 1 \\[3mm] K_I = 1, \, for \, \dfrac{V - \left(V_a - V_{cr}\right)}{V_{cr}} \geq 1 \end{array}\right\} \tag{5}$$

$$\left.\begin{array}{l} K_y(L/y \leq 1) = 2L \\[1mm] K_y(1 < L/y < 25) = 2(yL)^{0.5} \\[1mm] K_y(L/y \geq 25) = 10y \end{array}\right\} \tag{6}$$

where V_a is armor peak velocity.

Other abutment scour depth relationships for vertical-wall abutments under different regimes are given in Table 1.

2.2 Description of Collected Data from Literature and Present Experimental Work

A wide range of existing abutment scour depth datasets (263 datasets) have been collected from (Coleman et al. 2003) and (Dey and Barbhuiya 2005). In addition, 45 flume experiments have been carried out at the NIT Warangal, India. Additional tests were completed in a fixed bed masonry rectangular flume. Flume dimensions were 16.0 m long, 1.0 m wide, and 0.40 m deep. The flume's test section started 8.0 m from its entrance and had dimensions of 3.0 m long, 0.80 m wide, and 0.25 m deep. Uniform sand with a median diameter of 0.27 mm and a geometric standard deviation of 1.17 was used as the sediment bed. We used vertical wall abutment models with transverse lengths viz. 5.0, 6.8, 7.5, 8.4, 9.8, 10.6, 12.5, 15.0, 17.5 and 19.0 cm. All tests were conducted under non-submerged conditions. A valve was fixed into the flume inlet pipe to control the flowrate. At the downstream end

Table 1 Summary of maximum scour depth relationships for vertical–wall abutments

Investigator/s	Equation	Regime condition	Note
(Laursen 1963)	$\frac{d_{se}}{L} = 1.89\left(\frac{y}{L}\right)^{0.5}$	Threshold	
(Liu et al. 1961)	$\frac{d_{se}}{L} = 2015\left(\frac{y}{L}\right)^{0.6} F_r^{0.33}$	Live-bed	F_r is approach Froude number
(Gill 1972)	$\frac{d_{se}}{L} = 8.375\left(\frac{y}{L}\right)^{0.75}\left(\frac{d_{50}}{L}\right)^{0.25}\left(\frac{B}{B-L}\right)^{0.86} - \frac{y}{L}$	Threshold	B is the effective channel width
(Sturm and Janjua 1994)	$\frac{d_{se}}{L} = 7.7\frac{y}{L}\left(\frac{1}{M} \cdot \frac{V}{V_{cr}} - 0.35\right)$	Clearwater and live bed	M is the discharge contraction ratio
(Kandasamy and Melville 1998)	$\frac{d_{se}}{L} = K_S K_y \left(\frac{y}{L}\right)^m$	Clearwater and live bed	
(Dey and Barbhuiya 2005)	$\frac{d_{se}}{L} = 5.857\left(\frac{y}{L}\right)^{0.128}\left(\frac{V_{cr}}{(S-1)gL}\right)^{0.314}\left(\frac{d_{50}}{L}\right)^{0.167}$	Clearwater	S is the specific gravity

of the flume, a tail gate and a pre-calibrated rectangular notch were fixed to maintain the flow depth and measure the flowrate, respectively. The experiments were performed until they reached an equilibrium scour, i.e., no change in scour geometry over time. All tests in this study were carried out for 20 h. Experimentally, it was observed that the maximum scour depth (d_{se}) at equilibrium condition is located at the upstream nose of the abutment. The Vernier point gauge was used to determine the maximum abutment scour depth under equilibrium conditions. Due to the fact that all experiments were conducted in clear-water scour, the threshold velocity ratio was always less than one.

The vertical-wall abutment model was fixed in the test section at the right side of the flume prior to the start of the experiment (as can be seen in Fig. 1), i.e., located 9.5 m from the flume entrance. The sand bed was leveled perfectly with the bed slope and then covered with a 3 mm acrylic sheet to avoid unwanted scouring around the abutment. We achieved desired flow conditions using the inlet valve and flume tail gate. The acrylic sheet was sensibly removed after getting the desired flow conditions. Table 2 summarizes several parameters that influence the maximum abutment scour depth.
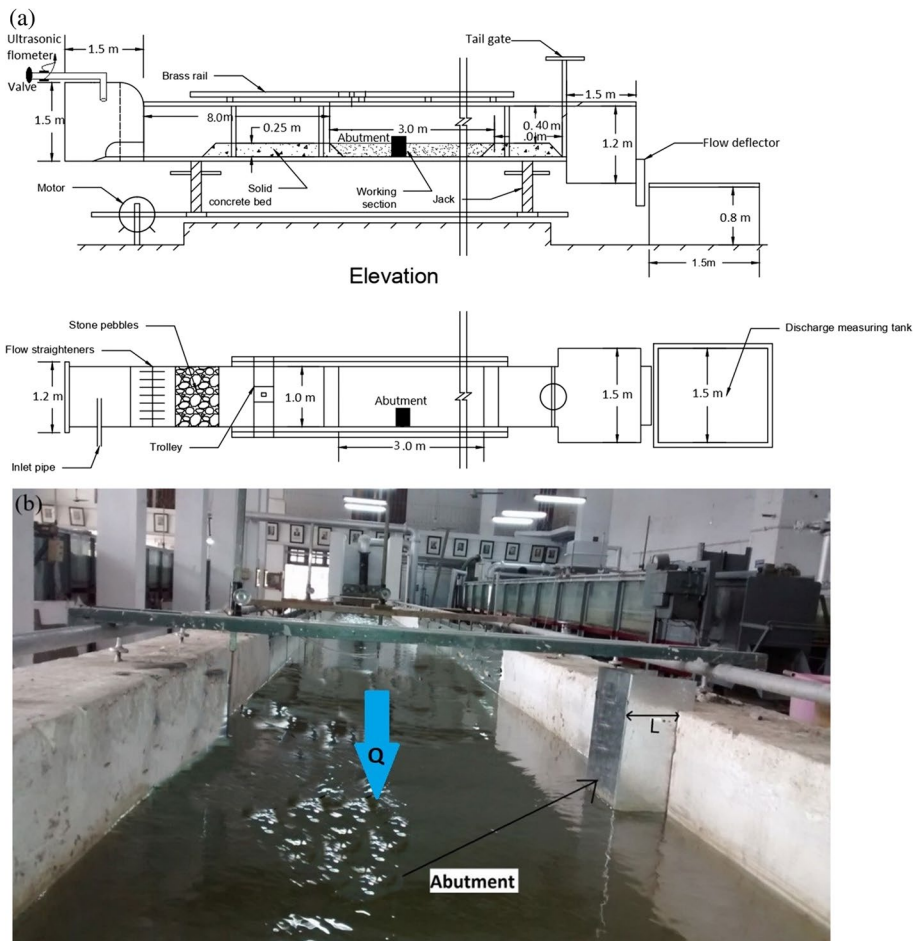

Fig. 1 **a** Flume layout, **b** photometric view of running the experiment

Table 2 Summary of influencing parameters for all datasets

Parameters	Present study	(Dey and Barbhuiya 2005)	(Coleman et al. 2003)
V (m/s)	0.16–0.24	0.18–0.67	0.17–1.54
l (m)	0.05–0.19	0.05–0.30	0.05–0.60
d_{50}(mm)	0.22	0.26–3.10	0.60–3.30
y (m)	0.1–0.13	0.05–0.60	0.048–0.60

Table 3 summarizes the statistical analysis of all data utilized in this investigation. According to the values of kurtosis and skewness presented in Table 3, the densimetric Froude number (F_{d50}) has an approximately normal distribution. However, other dimensionless input variables do not follow the normal distribution. The dimensionless output parameter (d_{se}/L) also approximately follows the normal distribution.

Correlation and regression analyses should be utilized to determine which variables influence the target variable. Using Pearson correlation coefficients, Fig. 2 depicts the relationship between the dimensionless abutment scour depth (d_{se}/L) and the input components. Correlation coefficients with a positive value suggest a direct association between two variables, while those with a negative value imply an inverse relationship. According to Fig. 2, the ratio of depth to transverse length of the abutment (y/L) with a correlation coefficient of $+0.75$ is the most effective input variable in predicting scour. The densimetric Froude number (F_{d50}) with a correlation coefficient of $+0.51$ is the second effective variable. The lowest correlation coefficient $+0.4$ belongs to the mean velocity to critical velocity ratio (V/V_c). The ratio of transverse length of the abutment to the diameter of sediment particles with a correlation coefficient of -0.45 has the opposite effect on scour depth, and with increasing it, the amount of scour decreases.

2.3 Gradient Boosting Decision Tree (GBDT) for Feature Selection

Many machine learning algorithms have been proposed and utilized to solve classification and regression problems over the year. But, the gradient boosting decision tree (GBDT) is one of the most popular algorithms for handling the classification and regression issues based on weak decision trees integration (Friedman 2002). In other words, the GBDT model is an ensemble of decision trees that assimilate a series of weak base learners with

Table 3 Descriptive statistics for all variables

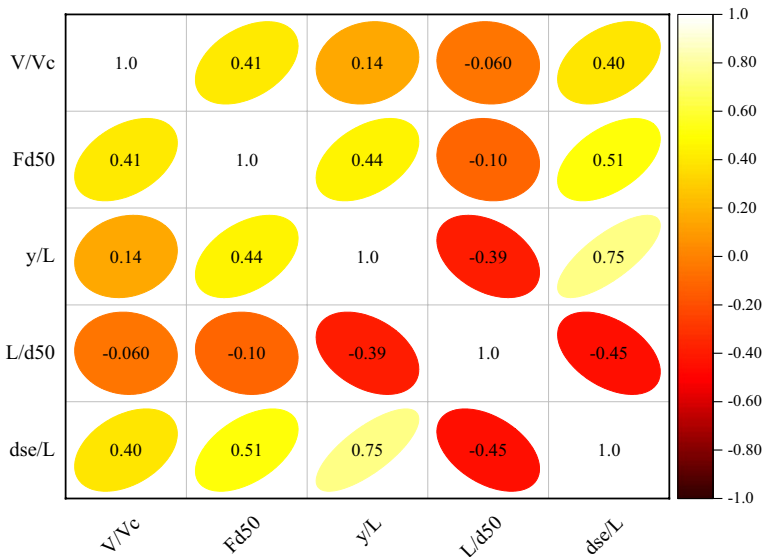| Statistics	V/V_c	F_{d50}	y/L	L/D_{50}	d_{se}/L
Mean	0.869	2.732	1.370	455.77	1.266
Std. Deviation	0.306	0.542	1.144	920.646	0.781
Skewness	4.323	0.201	0.897	4.309	0.893
Kurtosis	28.284	-0.202	-0.149	19.827	0.779
Minimum	0.460	1.467	0.007	8.333	0.781
25% Percentile	0.729	2.343	0.371	81.818	0.673
Median	0.880	1.724	1.000	166.667	1.090
75% Percentile	0.950	3.074	2.000	375	1.772
Maximum	3.512	4.069	4.166	6388	4.080

Fig. 2 Correlogram of input and target variables

many leaf nodes and avoid the overfitting problem (Tao et al. 2022). The amount of error in each node is measured using the weak learners, and a test function is utilized for splitting the node (Fan et al. 2018). The comprehensive background of GBDT can be obtained from Friedman (2002). In this study, the GBDT algorithm is exploited for feature selection and to predict the abutment scour depth (ASD).

2.4 Machine Learning Approaches

2.4.1 Categorical Boosting (CatBoost)

CatBoost is a new machine learning algorithm that was exposed by Prokhorenkova et al. (2018) for dealing the categorical features. It is a subset of the gradient boosting decision tree (GBDT) family but is different in working style. CatBoost is more powerful than other machine learning algorithms, i.e., XGBoost (extreme gradient boosting) and LightGBM (gradient boosting machine)(Ke et al. 2017), in handling complex and noisy data. Recently, the CatBoost algorithm has been widely used in hydrological modeling like reference evapotranspiration estimation (Bian et al. 2020), pan-evaporation estimation (Dong et al. 2021), and prediction of flash flood susceptibility (Saber et al. 2021). The enhancement of CatBoost comprises the following aspects:

1. Manage categorical features during the training period instead of pre-processing period. In training, the complete dataset is permitted by CatBoost. The Greedy target-based statistics (Greedy TBS) method is used for handling categorical features with the least information loss. Precisely, for each sample, a random permutation of the dataset was performed by CatBoost to calculate an average label value for the sample with the same

category value positioned before the given one in the permutation. Assume a dataset of observations $D = \{X_i, Y_i\} i = 1, \dots, n$ and if a permutation is $\theta = \left(\sigma_1, \sigma_2, \dots, \sigma_n\right)_n^T$, it is changed with (Prokhorenkova et al. 2018):

$$x_{\sigma_{p,k}} = \frac{\sum_{j=1}^{p-1} \left[x_{\sigma j,k} = x_{\sigma_{p,k}}\right] \times Y_{\sigma_j} + \beta \times P}{\sum_{j=1}^{p-1} \left[x_{\sigma_{j,k}} = x_{\sigma_{p,k}}\right] + \beta} \tag{7}$$

Here, β = weight of prior, P = prior value. In the dataset, the prior is the average label value, which helps in reducing the low-frequency category noise.

2. Feature combinations. CatBoost integrates all categorical features and their combinations in the current tree with all categorical features in the dataset using a greedy method.
3. Unbiased boosting with categorical features. CatBoost used an ordered boosting method to overcome the gradient bias (Prokhorenkova et al. 2018) and improve the generalization ability of the model.
4. Fast scorer. CatBoost uses oblivious trees as base predictions since they use the same splitting criterion across the tree's levels. These trees are less prone to overfitting and have a more balanced growth pattern. Each leaf index in oblivious trees is represented as a binary vector with a length equal to the tree's depth. This rule is employed to compute model predictions in CatBoost model evaluators because all binaries comprise float, statistics, and one-hot encoded features. Figure 3 shows the architecture of the CatBoost algorithm.

2.4.2 K-Nearest Neighbor (KNN)

K-nearest neighbor (KNN) is a non-parametric regression technique that was first proposed by Fix and Hodges (1951) for optimizing classification and prediction problems (Karlsson and Yakowitz 1987). The background history of the KNN exposes its effective application in hydrology (Sikorska-Senoner and Quilty 2021). In KNN, the independent variables (or predictors) are the input for the prediction objective.

2.4.3 Extra Tree Regression (ETR)

Geurts et al. (2006) proposed the idea of extra tree regression (ETR), which is a new ensemble machine learning model to perform regression or classification tasks based on many united decision trees (DT). A classical top-down procedure is used to construct the ETR model (Geurts et al. 2006). Many applications of the ETR model have been found in different fields (Heddam et al. 2020; Seyyedattar et al. 2020; Asadollah et al. 2021). The ETR is a highly randomized version of random forest (RF) with two main differences. The K (the number of features randomly nominated at each node), and n_{min} (the minimum sample size for splitting a node) are the two main parameters of the ETR, which avoids the overfitting and enhance the prediction accuracy of the model (Asadollah et al. 2021). In this research, ETR model was developed in scikit-library of Python programming for ASD prediction by tuning its parameters.
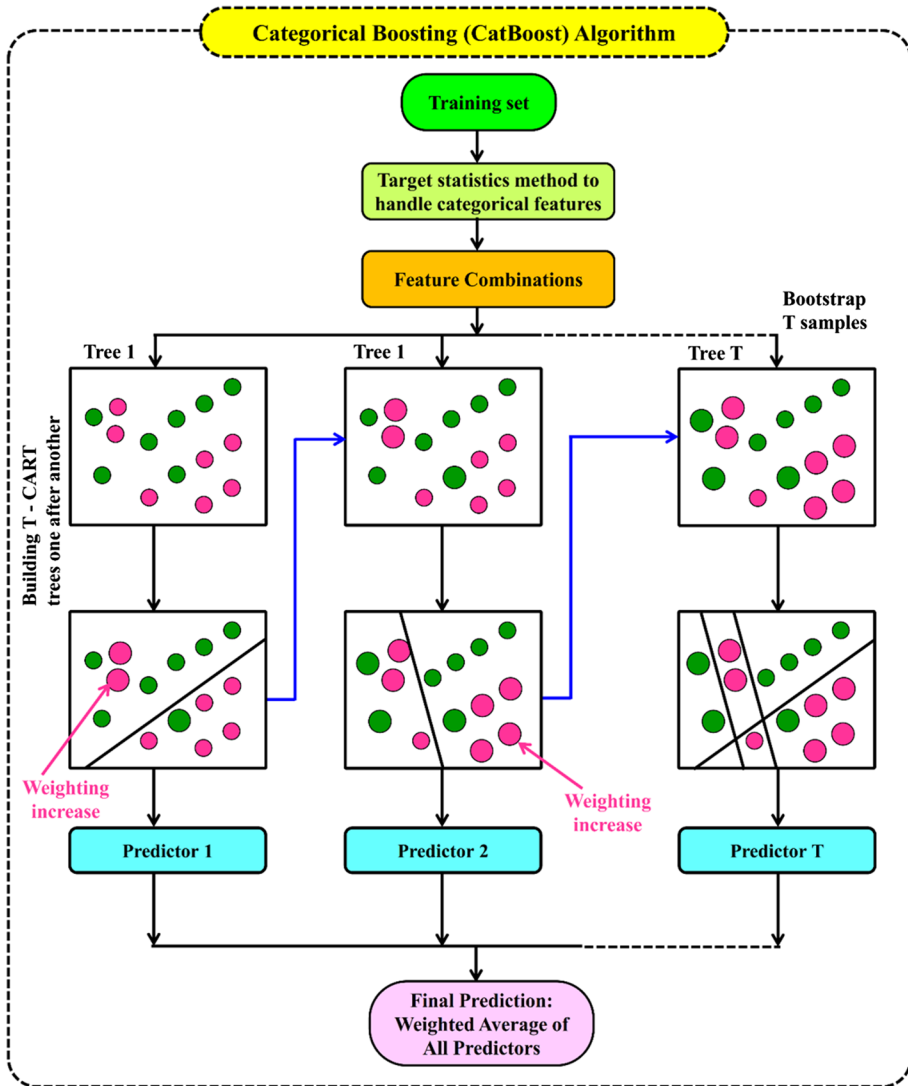
Fig. 3 The flow diagram of the CatBoost model

2.5 Performance Indicators

he efficacy of the applied machine learning paradigms i.e., KNN, Extra Tree, and CatBoost, for predicting the abutment scour depth (ASD) was evaluated by employing the five different performance indicators, including MAPE (mean absolute percent error), RMSE (root mean square error), R (coefficient of correlation), I_A (Willmott agreement index) (Willmott 1982), and $U_{95\%}$ (uncertainty coefficient with 95% confidence level) (Patino and Ferreira 2015). The mathematical formulas of R, RMSE, MAPE, I_A, and $U_{95\%}$ indicators are listed as follows:

$$R = \frac{\sum_{i=1}^{N} \left(ASD_{meas,i} - \overline{ASD_{meas}}\right)\left(ASD_{pred,i} - \overline{ASD_{pred}}\right)}{\sqrt{\sum_{i=1}^{N}(ASD_{meas,i} - \overline{ASD_{meas}})^2 \sum_{i=1}^{N}(ASD_{pred,i} - \overline{ASD_{pred}})^2}} \qquad (8)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(ASD_{meas,i} - ASD_{pred,i})^2} \qquad (9)$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{ASD_{meas,i} - ASD_{pred,i}}{ASD_{meas,i}}\right| \qquad (10)$$

$$I_A = 1 - \left[\frac{\sum_{i=1}^{N}\left(ASD_{pred,i} - ASD_{meas,i}\right)^2}{\sum_{i=1}^{N}\left(\left|ASD_{pred,i} - \overline{ASD_{meas}}\right| + \left|ASD_{meas,i} - \overline{ASD_{meas}}\right|\right)^2}\right] \qquad (11)$$

$$U_{95} = 1.96\sqrt{SD^2 + RMSE^2} \qquad (12)$$

Here, $ASD_{meas,i}$ and $ASD_{pred,i}$ are measured and predicted abutment scour depth (ASD) values for i^{th} data points, values; $\overline{ASD_{meas}}$ and $\overline{ASD_{pred}}$ are means of measured and predicted ASD values, SD is the standard deviation, and N is the total number of observations.

2.6 Model Development and Configuration

The data were normalized using the following equation to equalize the data scale:

$$x_{normal} = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (13)$$

where x_{max} and x_{min} denote the maximum and minimum values of the dataset used to generate the prediction models, respectively. Then 70% of the data for training and 30% of the data for evaluating the models' performance were considered test data sets.

The CatBoost model was used to predict scour depth around the bridge abutment in the present study. Two powerful models, including Extra Tree Regression (ETR) and KNN models, were used to compare the performance of the CatBoost model. The Gradient Boosting Decision Tree (GBDT) method was used to select the effective features in scouring prediction. Figure 4 shows the importance of each of the input variables in the scour estimation. According to Fig. 7, the y/L ratio is the most effective factor in predicting scouring, and F_{d50} is the least important compared to other input variables. As a result, according to the feature selection results, two combinations, namely: comb1 (all variables $(V/V_c, F_{d50}, y/L, L/D_{50})$) and comb2 (all variables except F_{d50} $(V/V_c, y/L, L/D_{50})$), were considered.

The proper adjustment of model parameters is one of the most important aspects of using machine learning models. Fine-tuning the parameters leads to higher accuracy. The grid search method was used to find the optimal value of machine learning model parameters. All models are implemented in a PC with an Intel-core i7-10750H 2.6 GHz processor and 32 GB of RAM.

Fig. 4 The importance of input variable based on GBDT feature selection
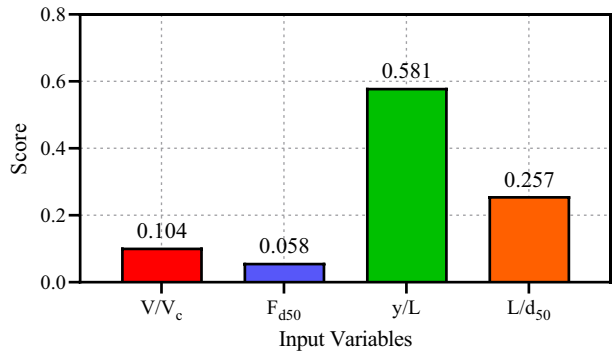
Extra Tree Regression (ETR) and KNN and the feature selection algorithm were developed in the Scikit-Learn package (Pedregosa et al. 2011), and the CatBoost package (Dorogush et al. 2018) in Python was used to implement the CatBoost model. In the CatBoost model, the four main learning rate (learning_rate), tree depth (depth), number of iterations (iterations), and L2 regularization (l2_leaf_reg) parameters must be set. In the present study, the range of learning rates [0.001–0.5], depth [2–20], iterations between [200–2000], and L2 regularization in the range [0.5–1.5] were considered. Table 4 presents the optimal parameters of the CatBoost model for the two input combinations, comb1, and comb2. In the Extra Tree model, there are two important parameters, the number of trees in the forest (n_estimators) and the maximum depth of the tree (max_depth), that need to be adjusted. The maximum depth of the tree was considered between [2–20], and the number of trees in the forest was set between [10–150]. Table 4 presents the optimal parameters of the ETR model for both input combinations. The only parameter of the KNN model is the number of neighbors, which is considered between [1–10], and its optimal values are presented in Table 4 for both input combinations. Figure 5 indicates the schematic flowchart of the present study.

3 Results and Discussion

In this research, a comprehensive ML-based investigation was performed on the normalized scour depth (d_{se}/L) of the abutments in uniform bed based on four dimensionless input features, namely V/V_c, F_{d50}, y/L, and L/D_{50}. The feature selection process was addressed to determine the most significant candidate input combinations using a

Table 4 Model tuning parameters for abutment scour prediction

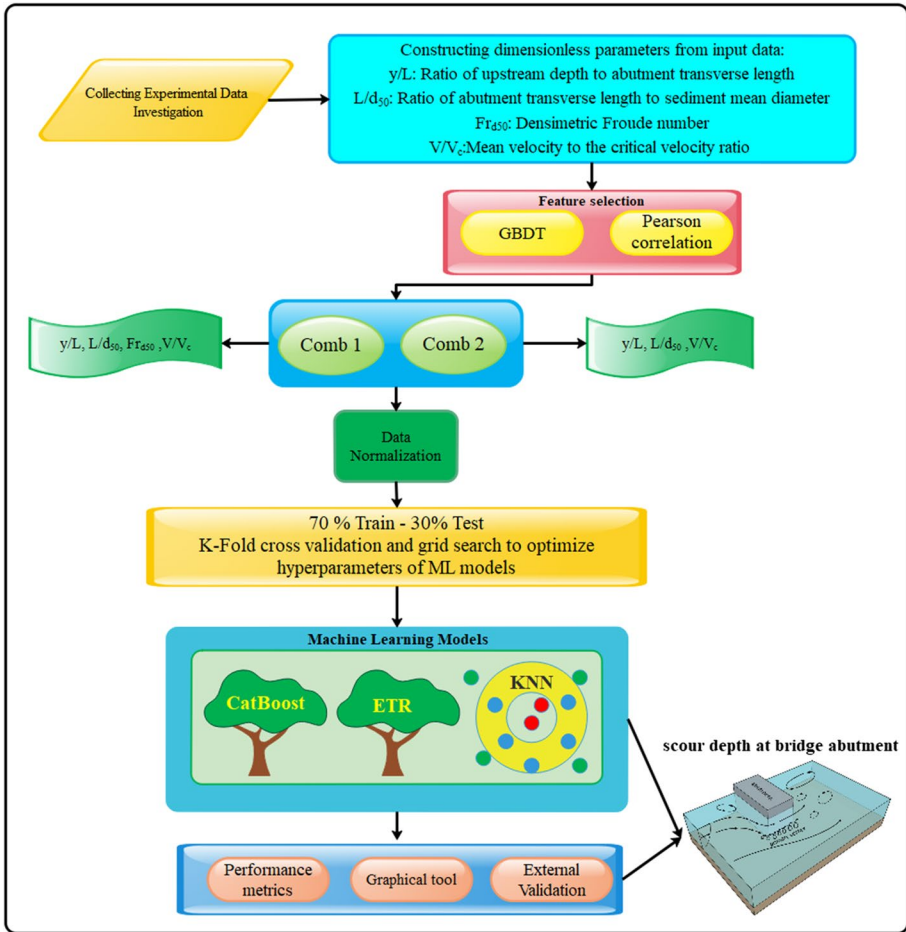| Study site	Models	Tuning parameter models
Comb1	CatBoost	l2_leaf_reg = 1.2, iterations = 1000, learning_rate = 0.15, depth = 5
	Extra Tree	n_estimators = 140, max_depth = 15
	KNN	n_neighbors = 1, weights = ('distance')
Comb2	CatBoost	l2_leaf_reg = 1, iterations = 1000, learning_rate = 0.15, depth = 3
	Extra Tree	max_depth = 15, n_estimators = 140
	KNN	n_neighbors = 2, weights = ('distance')

Fig. 5 The flowchart of the present study

tree-based, namely the Tree decision FS method. The outcomes of pre-processing indicate that two scenarios comprised all features (Comb1) and all features except the F_{d50} (Comb2) were considered to feed the ML model to model the normalized scour depth around abutments. Table 5 summarizes the goodness-of-fit statistics of the simulation of the normalized scour depth at the abutment. According to Table 5, the CatBoost model in Comb 1 (comprised of all features), regarding the best accuracy in training (R = 0.9993, RMSE = 0.0290, MAPE = 0.5069%, and $U_{95\%}$ = 0.0569) and testing (R = 0.9685, RMSE = 0.1784, MAPE = 10.4724%, and $U_{95\%}$ = 0.0612) outperformed the Extra Tree (R = 0.9491, RMSE = 0.2236, MAPE = 11.3301%, and $U_{95\%}$ = 0.0979 for the testing phase) and KNN models (R = 0.9251, RMSE = 0.2778, MAPE = 18.4478%, and $U_{95\%}$ = 0.1503 for the testing phase). Moreover, in the second variant of input combination (i.e., Comb 2), the CatBoost model owing to the highest values of R = 0.9608 and I_A = 0.9790 and least diagnostic metrics (RMSE = 0.1995, MAPE = 12.2251%, and $U_{95\%}$ = 0.0759) in the testing stage had the best performance among three considered models followed by Extra Tree (R = 0.9378, RMSE = 0.2482, MAPE = 12.4341%, and $U_{95\%}$ = 0.1193), and KNN

Table 5 Statistical metrics of the predicted abutment scour depths for Comb 1 and 2

Models		Phase	R	RMSE	MAPE	I_A	$U_{95\%}$
Comb 1	KNN	Training	0.9983	0.0481	0.4290	0.9991	0.0937
		Testing	0.9251	0.2778	18.4478	0.9609	0.1503
	Extra Tree	Training	0.9993	0.0291	0.4642	0.9996	0.0571
		Testing	0.9491	0.2236	11.3301	0.9730	0.0979
	CatBoost	Training	0.9993	0.0290	0.5069	0.9996	0.0569
		Testing	**0.9685**	**0.1784**	**10.4724**	**0.9833**	**0.0612**
Comb 2	KNN	Training	0.9993	0.0303	0.3938	0.9996	0.0595
		Testing	0.9122	0.2936	21.3313	0.9545	0.1701
	Extra Tree	Training	0.9993	0.0299	0.6115	0.9996	0.0588
		Testing	0.9378	0.2482	12.4341	0.9656	0.1193
	CatBoost	Training	0.9989	0.0375	2.4707	0.9994	0.0738
		Testing	**0.9608**	**0.1995**	**12.2251**	**0.9790**	**0.0759**

Bold font: The superior predictive model

(R = 0.9122, RMSE = 0.2936, MAPE = 21.3313%, and $U_{95\%}$ = 0.9545) models, respectively. Overall, regarding the goodness-of-fit statistics reported in Table 5, it can be concluded that the predictive performance of Comb 1, including all input features, is superior to Comb 2.

For validation of the provided model to the estimation of the normalized scour depth at the abutment, several infographic tools, and diagnostic analyses were addressed in the forms of scatter plots, Rug-Histograms density distribution function, trend variation graphs, Taylor diagrams, and the violin plots of residual and relative deviation error.

Figure 6 demonstrates the scatter plots of Comb 1 to compare the predicted and measured values of normalized scour depth at the abutments. According to Fig. 6a, it can be seen that all three models in the training phase of Comb 1 have perfect performance regarding the best fitness with actual values of d_{se}/L. Although in the testing phase, the Catboost model representation with green color (R = 0.9685) led to the best agreement with actual values of d_{se}/L compared with the Extra Tree (R = 0.9491) and KNN (R = 0.9251) approaches. Besides, in Comb 2, (Fig. 6b) the best compatibility between the predicted and measured d_{se}/L is related to both ensemble-based ML models (CatBoost and Etra Tree) and KNN stands on the last rank of accurateness of simulation. A closer comparison of scatter plots indicates that the performance of all three methods in Comb 1 is better than Comb 2 due to the better alignment around the 45° line. In this stage of evaluating the capability of the models, the Taylor diagrams in Fig. 7 are employed to qualitatively assess the accuracy of the representatives of each model in comparison with the actual values, which criterion are the correlation coefficient and standard deviation (Xu et al. 2016). Based on Fig. 7, it can be seen that the representative of the CatBoost method for both combinations is located in the range of 0.95 to 0.99 from a smaller physical distance than other methods with a reference point. Although, the precision of Comb 1 for all three considered models is superior to Comb 2 for estimating the d_{se}/L values. Figure 8 displaces the box plot of residual error in the testing stage, which reveals that the CatBoost models regarding the lowest residual error range in Comb 1 (1.08) and Comb 2 (1.37) yielded the most reliable outcomes than KNN (Comb 1|1.486 and Comb 2|1.575) and Extra Tree (Comb 1|1.634 and Comb 2|1.616). Overall, the diagnostic analyses evident that Comb 1 results in more accurate
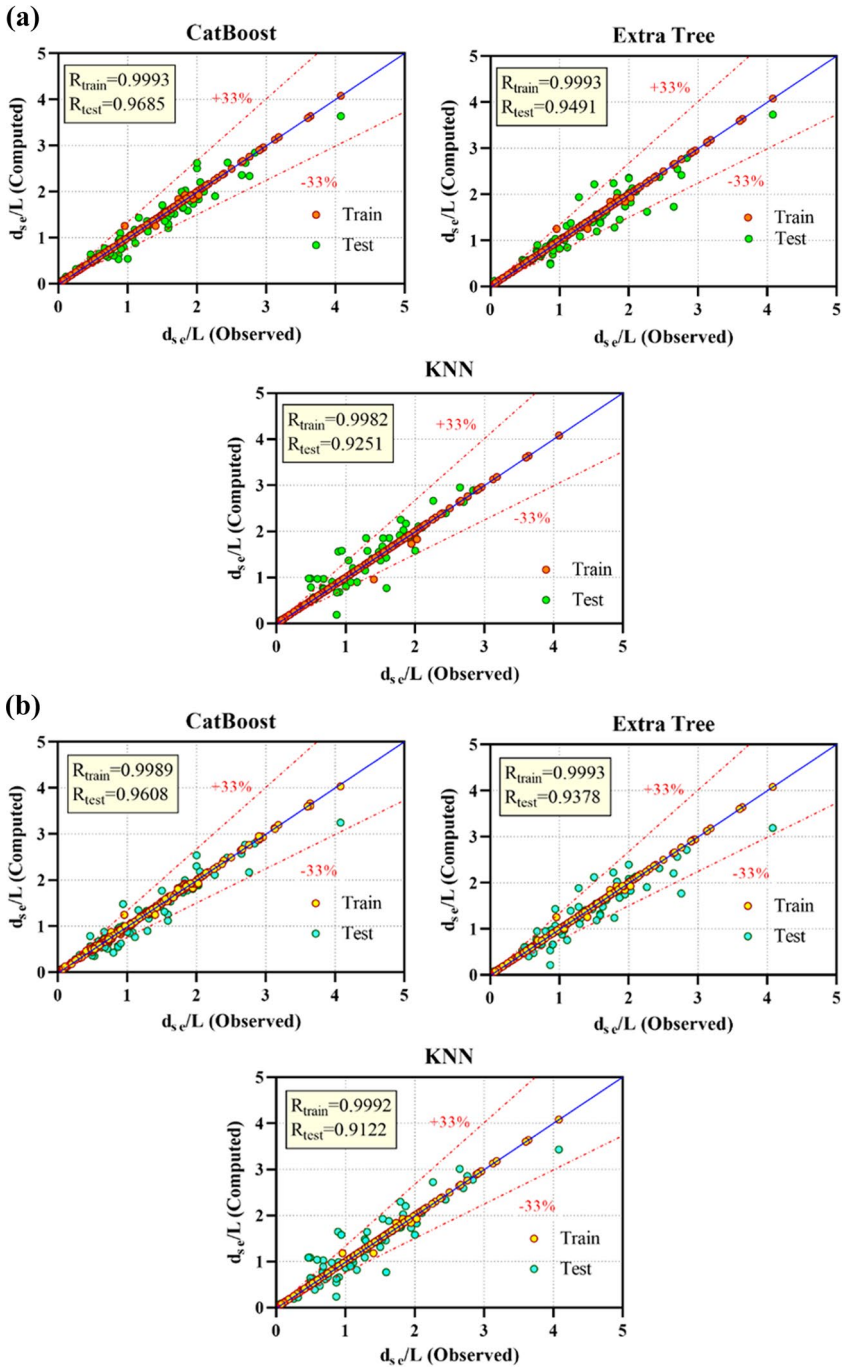
Fig. 6 **a** Scatter plots of computed and observed values of dimensionless scour in comb 1 for the training and test data. **b** Scatter plots of computed and observed values of dimensionless scour in comb 2 for the training and test data
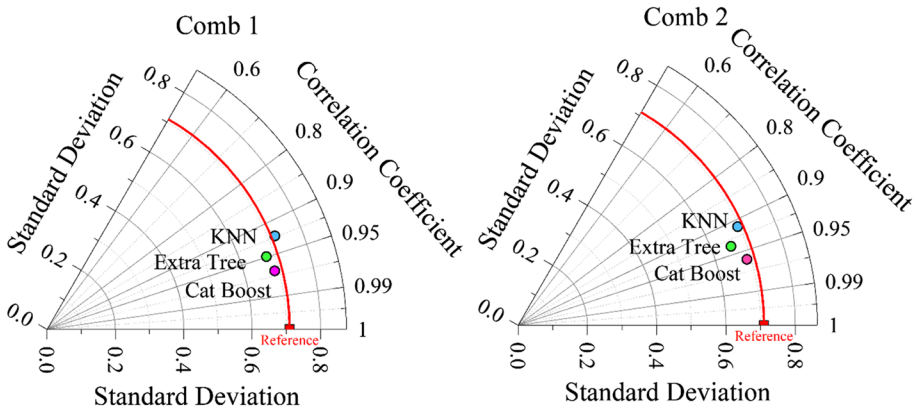
Fig. 7 Taylor diagrams of models in combs 1 and 2

outcomes than Comb 2, and the Extra Tree as the second-best predictive model can be considered a reliable model for precise modeling of the d_{se}/L values. Furthermore, the results presented in this research showed that the novel ensemble-based machine learning methods (CatBoost and Extra Tree) have a good performance in solving significant non-linear scour depth estimation problems at the hydraulic structures, which is also confirmed by previous research.
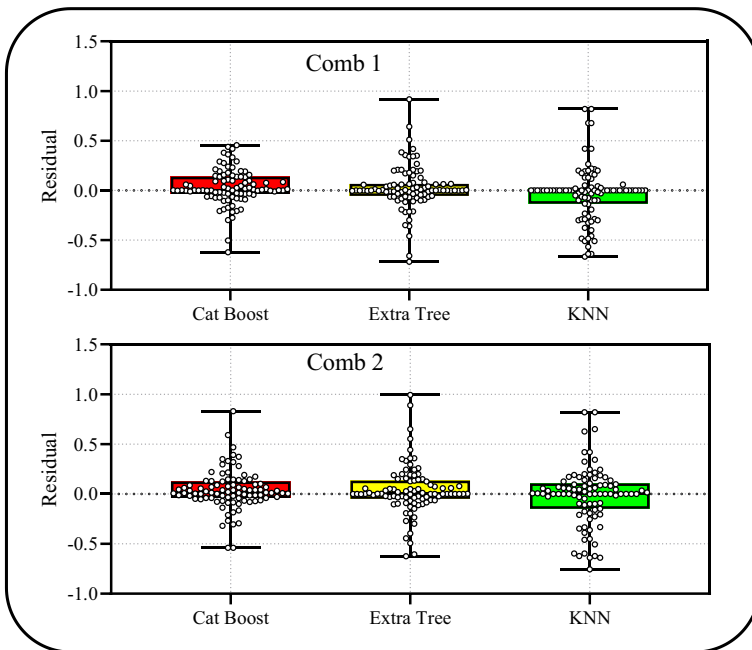
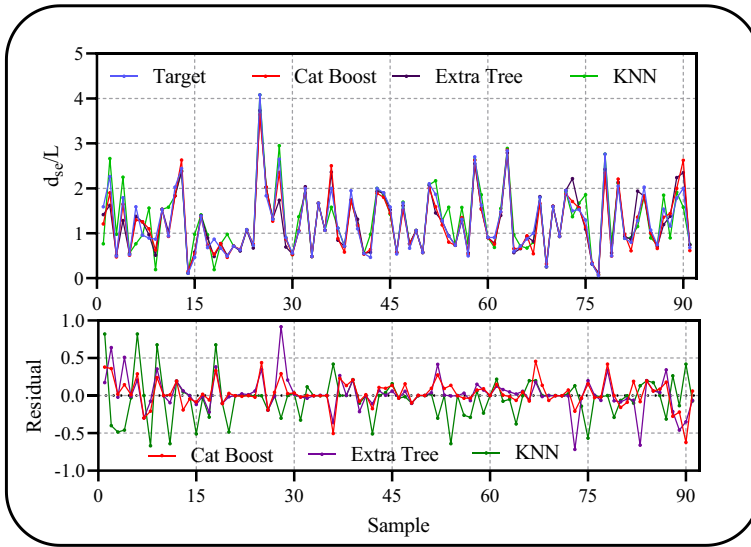**Fig. 8** Box plots of residuals for different models in comb 1 and 2

Fig. 9 Comparison of results and residuals for different models in comb 1

3.1 Extra Discussion and Comparison

Here, the performance of provided ML models in the superior candidate input combination is examined for the prediction of the normalized scour depth at abutments. Figure 9 displaces the trend variation of predicted d_{se}/L values using CatBoost, Extra Tree, and KNN versus the measured d_{se}/L. Best performance in capturing non-linear behavior of the scour depth data points (d_{se}/L) in the testing phase is related to the CatBoost method, and the Extra Tree and KNN methods are in the next ranks, respectively. Also, the residual distribution indicates that the KNN model has the highest oscillation among the three ML approaches. Figure 10 depicted the Rug-Histogram of the predicted normalized scour depth (d_{se}/L) obtained by the CatBoost, Extra Tree, and KNN models vs. the measured d_{se}/L for the superior input combination for whole data points. Here, the density distribution function of the CatBoost model appeared to be more matched with the measured d_{se}/L in comparison with the Extra Tree and KNN models. Also, the greater conformity of the compression band of the predicted and actual data in the CatBoost model ascertains a better performance than the other methods in estimating scour depth at abutments.

3.2 Comparison of Previously Proposed Relationships and Present Approaches

A comparison has also been done using experimental and computed scour depths for previously proposed empirical relationships, viz. Melville and Coleman (Melville and Coleman 2000) and Dey and Barbhuiya (Dey and Barbhuiya 2005) and soft computing approaches viz. CatBoost, ETR, and KNN. Figure 11a illustrates scatter plots between observed and computed values of normalized maximum scour depth around the abutment with the error line bands of ±20%. One can easily identify that novel ensemble-based data-intelligence paradigms, viz. CatBoost, ETR, and KNN predict abutment scour depth more precisely than empirical relationships. Melville and Coleman (Melville and Coleman 2000) equation
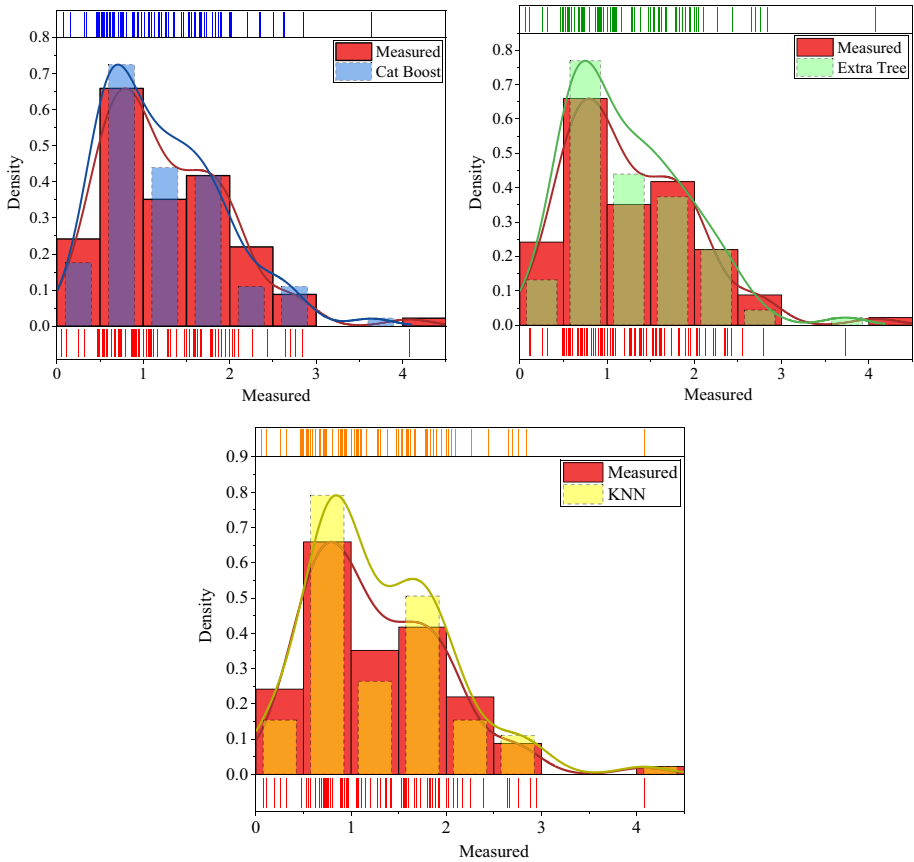
Fig. 10 Comparison of the prediction form of the Rug-Histograms density distribution function

overestimates the abutment scour depth and shows maximum error for used datasets, as can be seen in Fig. 11a. Figure 11b illustrates the variation between the error and total datasets. For CatBoost, ETR, and KNN, approximately 95% of the datasets were found to have less than ± 20% error than whereas more than 60% of datasets for Dey and Barbhuiya (Dey and Barbhuiya 2005) were found to have less than ± 30% error, while only 40% datasets for Melville and Coleman (Melville and Coleman 2000) were found to have less than ± 40% error, as can be seen in Fig. 11b.

4 External Validation

Tropsha et al. (Tropsha et al. 2003) established new criteria for model external validation. These criteria are derived from the known prediction performance of the model. The validation criteria and supporting data for the suggested models are summarized in Table 6. K and K' must be between 0.85 and 1.15, and m and n must be smaller than 0.1 to meet the requirements. For the test dataset, CatBoost's correlation coefficient (R) is 0.968, while for the training dataset, it is 0.999. CatBoost's m and n coefficients ($n = $ -0.064 and m = -0.061
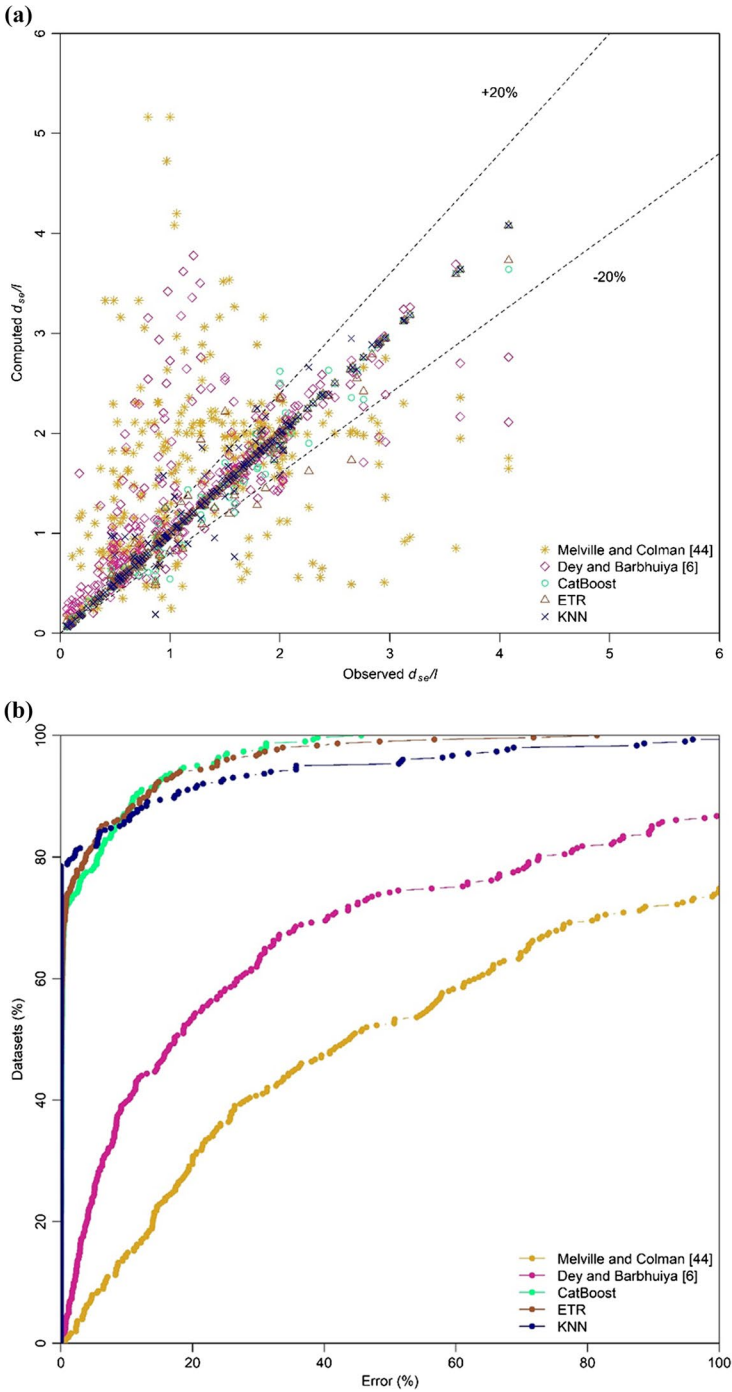
Fig. 11 **a** Comparison between observed and computed normalized abutment scour depth. **b** Variation of percentage error v/s total datasets

Table 6 External validation parameters for the created models

Model	Phase	K	K'	m	n	R
CatBoost	Training	0.999	1.001	-0.001	-0.001	0.999
	Testing	0.966	1.019	-0.061	-0.064	0.968
ETR	Training	0.999	1.001	-0.001	-0.001	0.999
	Testing	0.962	1.014	-0.108	-0.109	0.949
KNN	Training	0.995	1.003	-0.003	-0.003	0.998
	Testing	1.007	0.957	-0.168	-0.159	0.925
Formula		$K = \sum_{i=1}^{n} \frac{T_i \times P_i}{P_i^2}$	$K' = \frac{T_i \times P_i}{P_i^2}$	$m = \frac{R^2 - R_o^2}{R^2}$	$n = \frac{R^2 - R_o^{2\prime}}{R^2}$	Equation 8

Where $R_0^2 = 1 - \sum_{i=1}^{n} P_i^2 (1-K)^2 / \sum_{i=1}^{n} \left(P_i - \overline{P} \right)^2$

$R_0^2 = 1 - \sum_{i=1}^{n} T_i^2 \left(1 - K'\right)^2 / \sum_{i=1}^{n} \left(T_i - \overline{T} \right)^2$

| Condition | | $0.85 < K < 1.15$ | $0.85 < K' < 1.15$ | $m < 0.1$ | $n < 0.1$ | $R > 0.8$ |

for the test dataset) are greater than those of the other models. As seen in Table 6, all three models in the training and testing datasets met all important criteria, demonstrating that the models' predictive quality and significant connection between goal and output are not coincidental.

5 Conclusion

The purpose of this work was to undertake a complete machine learning analysis to predict the scour depth around the bridge abutment. Three machine learning models were created for this purpose: CatBoost, Extra Tree Regression (ETR), and K-nearest neighbor (KNN). The models were developed using 308 samples series of laboratory data (a wide range of existing abutment scour depth datasets (263 datasets) and 45 flume experiments data at the NIT Warangal, India). Four dimensionless parameters including upstream densimetric Froude number (F_{d50}), the upstream depth (y) to abutment transverse length ratio (y/L), the abutment transverse length to the sediment mean diameter ($L/d50$), and the mean velocity to the critical velocity ratio (V/Vc) were considered as the model inputs, and the normalized scour depths (d_{se}/L). Based on the GBDT feature selection method, two combinations: comb1 ($V/V_c, F_{d50}, y/L, L/D_{50}$) and comb2 ($V/V_c, y/L, L/D_{50}$), were selected as input for models. The results of this study showed that the use of input combination 1 (comb1), which includes all input variables, provided more accurate results. Comb1 has provided better results in the test phase in all three models used. The CatBoost model performed best in predicting scour depth in both input combinations 1 and 2 (RMSE=0.1784 and R=0.9685 for comb1 and RMSE=0.1995 and R=0.9608 for comb2). In the ET model, input combination 1 (RMSE=0.2236 and R=0.9491) also performed better than input combination 2 (RMSE=0.2482 and R=0.9378). The ET model performed worse than the CatBoost. The KNN model has the weakest results among the models used (RMSE=0.2778 and R=0.9251 for comb1 and RMSE=0.2936 and R=0.9122 for comb2). Additionally, a comparison of purposed intelligent models to prior empirically-based research demonstrates the superiority of all established machine learning models.

Finally, external validation established that all prediction techniques were constrained to values of $0.85 < K, K' < 1.15$, and $(m, n) < 0.1$. The performance of the created model on test data reveals its ability to generalize effectively.

Authors Contributions Manish Pandey: Conceptualization, formal analysis, writing up the manuscript; **Masoud Karbasi**: Validation, investigation, software, writing up the manuscript; **Mehdi Jamei**: Conceptualization, formal analysis, visualization, writing up the manuscript, Supervision; **Anurag Malik:** Formal analysis, writing up the manuscript and manuscript editing; **Jaan H. Pu**: Investigation, and writing up the manuscript, supervision.

Data Availability Data are available from the corresponding author upon reasonable request.

Declarations

Ethical Approval The manuscript is conducted within the ethical manner advised by the water resource management journal.

Consent to Participate Not applicable.

Consent to Publish The research is scientifically consented to be published.

Competing Interests The authors declare that he has no conflict of interest.

References

Aamir M, Ahmad Z (2019) Estimation of maximum scour depth downstream of an apron under submerged wall jets. J Hydroinformatics 21:523–540

Afzal MS, Holmedal LE, Myrhaug D (2021) Sediment transport in combined wave–current seabed boundary layers due to streaming. J Hydraul Eng 147:4021007

Asadollah SBHS, Sharafati A, Motta D, Yaseen ZM (2021) River water quality index prediction and uncertainty analysis: a comparative study of machine learning models. J Environ Chem Eng 9:104599. https://doi.org/10.1016/j.jece.2020.104599

Azamathulla HM, Cuan YC, Ghani AA, Chang CK (2013) Suspended sediment load prediction of river systems: GEP approach. Arab J Geosci 6:3469–3480. https://doi.org/10.1007/s12517-012-0608-4

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

Barbhuiya AK, Dey S (2004) Local scour at abutments: a review. Sadhana - Acad Proc Eng Sci 29:449–476. https://doi.org/10.1007/BF02703255

Bian C, He H, Yang S, Huang T (2020) State-of-charge sequence estimation of lithium-ion battery based on bidirectional long short-term memory encoder-decoder architecture. J Power Sources 449:227558. https://doi.org/10.1016/j.jpowsour.2019.227558

Bressan F, Ballio F, Armenio V (2011) Turbulence around a scoured bridge abutment. J Turbul 12:1–24. https://doi.org/10.1080/14685248.2010.534797

Coleman SE, Lauchlan CS, Melville BW (2003) Développement de l'affouillement en eau claire aux butées de pont. J Hydraul Res 41:521–531. https://doi.org/10.1080/00221680309499997

Dey S, Barbhuiya AK (2005) Time variation of scour at abutments. J Hydraul Eng 131:11–23. https://doi.org/10.1061/(ASCE)0733-9429(2005)131:1(11)

Dong L, Zeng W, Wu L et al (2021) Estimating the pan evaporation in Northwest China by coupling catboost with bat algorithm. Water 13:256. https://doi.org/10.3390/w13030256

Dorogush AV, Ershov V, Gulin A (2018) CatBoost: gradient boosting with categorical features support. arXiv Prepr arXiv181011363

Eghlidi E, Barani G-A, Qaderi K (2020) Laboratory investigation of stilling basin slope effect on bed scour at downstream of stepped spillway: physical modeling of javeh RCC dam. Water Resour Manag 34:87–100

Fan J, Yue W, Wu L et al (2018) Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. Agric for Meteorol 263:225–241. https://doi.org/10.1016/j.agrformet.2018.08.019

Fix E, Hodges JL (1951) Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. USAF Sch Aviat Med Randolph Field, Texas. https://doi.org/10.2307/1403797

Friedman JH (2002) Stochastic gradient boosting. Comput Stat Data Anal 38:367–378

Geurts P, Ernst D, Wehenkel L (2006) Extremely randomized trees. Mach Learn 63:3–42

Gill MA (1972) Erosion of sand beds around spur dikes. J Hydraul Div 98(9):1587–1602

Goyal MK, Ojha CSP (2011) Estimation of scour downstream of a ski-jump bucket using support vector and M5 model tree. Water Resour Manag 25:2177–2195

Heddam S, Ptak M, Zhu S (2020) Modelling of daily lake surface water temperature from air temperature: Extremely randomized trees (ERT) versus Air2Water, MARS, M5Tree, RF and MLPNN. J Hydrol 588:125130. https://doi.org/10.1016/j.jhydrol.2020.125130

Kandasamy JK, Melville BW (1998) Maximum local scour depth at bridge piers and abutments. J Hydraul Res 36:183–198. https://doi.org/10.1080/00221689809498632

Karlsson M, Yakowitz S (1987) Nearest-neighbor methods for nonparametric rainfall-runoff forecasting. Water Resour Res 23:1300–1308. https://doi.org/10.1029/WR023i007p01300

Ke G, Meng Q, Finley T et al (2017) LightGBM: A highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems. pp 3149–3157

Kothyari UC, Hager WH, Oliveto G (2007) Generalized approach for clear-water scour at bridge foundation elements. J Hydraul Eng 133:1229–1240. https://doi.org/10.1061/(ASCE)0733-9429(2007)133:11(1229)

Kumar L, Afzal MS, Ahmad A (2022) Prediction of water turbidity in a marine environment using machine learning: A case study of Hong Kong. Reg Stud Mar Sci 52:102260

Laursen EM (1963) An Analysis of Relief Bridge Scour. J Hydraul Div 89:93–118. https://doi.org/10.1061/JYCEAJ.0000896

Liu H-K, Chang FM, Skinner MM (1961) Effect of bridge constriction on scour and backwater. CER; 60–22

Melville BW, Coleman SE (2000) Bridge scour. Water Resources Publication

Mohammadpour R (2017) Prediction of local scour around complex piers using GEP and M5-Tree. Arab J Geosci 10. https://doi.org/10.1007/s12517-017-3203-x

Mohammadpour R, Ab. Ghani A, Zakaria NA, Mohammed Ali TA (2017) Predicting scour at river bridge abutments over time. Proc Inst Civ Eng Water Manag 170:15–30. https://doi.org/10.1680/jwama.14.00136

Mohammadpour R, Ghani AA, Vakili M, Sabzevari T (2016) Prediction of temporal scour hazard at bridge abutment. Nat Hazards 80:1891–1911. https://doi.org/10.1007/s11069-015-2044-8

Najafzadeh M, Azamathulla HM (2013) Group method of data handling to predict scour depth around bridge piers. Neural Comput Appl 23:2107–2112. https://doi.org/10.1007/s00521-012-1160-6

Oliveto G, Hager WH (2005) Further results to time-dependent local scour at bridge elements. J Hydraul Eng 131:97–105. https://doi.org/10.1061/(ASCE)0733-9429(2005)131:2(97)

Pandey M, Sharma PK, Ahmad Z, Karna N (2018) Maximum scour depth around bridge pier in gravel bed streams. Nat Hazards 91:819–836. https://doi.org/10.1007/s11069-017-3157-z

Pandey M, Valyrakis M, Qi M et al (2020) Experimental assessment and prediction of temporal scour depth around a spur dike. Int J Sediment Res:1–13. https://doi.org/10.1016/j.ijsrc.2020.03.015

Patino CM, Ferreira JC (2015) Confidence intervals: a useful statistical tool to estimate effect sizes in the real world. J Bras Pneumol. https://doi.org/10.1590/s1806-37562015000000314

Pedregosa F, Varoquaux G, Gramfort A et al (2011) Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830

Prokhorenkova L, Gusev G, Vorobev A et al (2018) Catboost: Unbiased boosting with categorical features. In: Advances in Neural Information Processing Systems. pp 6637–6647

Saber M, Boulmaiz T, Guermoui M et al (2021) Examining LightGBM and CatBoost models for wadi flash flood susceptibility prediction. Geocarto Int:1–26. https://doi.org/10.1080/10106049.2021.1974959

Seyyedattar M, Ghiasi MM, Zendehboudi S, Butt S (2020) Determination of bubble point pressure and oil formation volume factor: extra trees compared with LSSVM-CSA hybrid and ANFIS models. Fuel 269:116834. https://doi.org/10.1016/j.fuel.2019.116834

Sikorska-Senoner AE, Quilty JM (2021) A novel ensemble-based conceptual-data-driven approach for improved streamflow simulations. Environ Model Softw:105094. https://doi.org/10.1016/j.envsoft.2021.105094

Singh RK, Pandey M, Pu JH et al (2020) Experimental study of clear-water contraction scour. Water Supply 20:943–952. https://doi.org/10.2166/ws.2020.014

Singh UK, Jamei M, Karbasi M et al (2022) Application of a modern multi-level ensemble approach for the estimation of critical shear stress in cohesive sediment mixture. J Hydrol 607:127549. https://doi.org/10.1016/j.jhydrol.2022.127549

Sturm TW, Janjua NS (1994) Clear-water scour around abutments in floodplains. J Hydraul Eng 120:956–972

Tao H, Salih S, Oudah AY et al (2022) Development of new computational machine learning models for longitudinal dispersion coefficient determination: case study of natural streams, United States. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-022-18554-y

Tropsha A, Gramatica P, Gombar VK (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. QSAR Comb Sci 22:69–77

Willmott CJ (1982) Some comments on the evaluation of model performance. Bull Am Meteorol Soc 63:1309–1313

Xu Z, Hou Z, Han Y, Guo W (2016) A diagram for evaluating multiple aspects of model performance in simulating vector fields. 4365–4380. https://doi.org/10.5194/gmd-9-4365-2016

Authors and Affiliations

Manish Pandey[1] · Masoud Karbasi[2] · Mehdi Jamei[3,4] · Anurag Malik[5] · Jaan H. Pu[6]

Manish Pandey
manishpandey3aug@gmail.com

Anurag Malik
anuragmalik_swce2014@rediffmail.com

Jaan H. Pu
J.H.Pu1@bradford.ac.uk

[1] Civil Engineering Department, National Institute of Technology Warangal, Hanamkonda, India

[2] Faculty of Agriculture, Water Engineering Department, University of Zanjan, Zanjan, Iran

[3] Faculty of Engineering, Shohadaye Hoveizeh Campus of Technology, Shahid Chamran University of Ahvaz, Dashte Azadegan, Iran

[4] New Era and Development in Civil Engineering Research Group, Scientific Research Center, Al-Ayen University, Thi-Qar, Nasiriyah 64001, Iraq

[5] Punjab Agricultural University, Regional Research Station, Bathinda, Punjab 151001, India

[6] Faculty of Engineering and Informatics, University of Bradford, Bradford, UK