

Article

SHAP-Based Interpretable Ensemble Models for Final Scour Depth Prediction at Collar-Protected Bridge Pier

Nadir Murtaza ¹, Sohail Iqbal ², Muhammad Ali Sikandar ^{1,*} and Mohd Aamir Mumtaz ^{3,*}

¹ Department of Civil Engineering, CECOS University of IT and Emerging Sciences, Peshawar 25000, Pakistan; nadir.murtaza@cecos.edu.pk

² Department of Civil Engineering, Faculty of Science and Technology, Tokyo University of Science, Chiba 278-8510, Japan; sohail.sakhani147@rs.tus.ac.jp

³ Civil Engineering Department, College of Engineering, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia

* Correspondence: msikandar@cecos.edu.pk (M.A.S.); mmaamir@imamu.edu.sa (M.A.M.)

Abstract

Bridge pier scour is one of the primary causes of bridge foundation failure, making accurate prediction of final scour depth essential for the safe and economical design of hydraulic structures. This study develops interpretable ensemble machine learning models for predicting final scour depth around collar-protected circular bridge piers using a laboratory dataset comprising 48 experimental observations. Four dimensionless hydraulic and geometric parameters, namely (b/b_c , b : bridge pier diameter, b_c : collar diameter), (z/d_{50} , z : collar elevation, d_{50} : median diameter of bed particles), (z/b_c), and (U/U_c , U : time average velocity, U_c : critical velocity of bed particles), were employed as model inputs, while the normalized final scour depth (d_{sf}/b_c) was considered as the target variable. Three ensemble learning algorithms, namely XGBoost (XGB), Random Forest (RF), and Extra Trees (ET), were developed and evaluated using training, testing, and 5-fold cross-validation procedures. The predictive performance of the models was assessed using the coefficient of determination (R^2), Kling–Gupta Efficiency (KGE), root mean square error (RMSE), and mean absolute error (MAE). Among the investigated models, XGBoost demonstrated the highest predictive accuracy, achieving an R^2 of 0.987, a KGE of 0.982, and an RMSE of 0.0117 on the training dataset. Furthermore, SHapley Additive exPlanations (SHAP) were employed to interpret the influence of individual input variables, revealing that hydraulic intensity and collar-related parameters exert the greatest influence on equilibrium scour prediction, consistent with established scour mechanics. The proposed framework combines high predictive accuracy with model interpretability, providing a reliable decision-support tool for bridge scour assessment and demonstrating the potential of explainable machine learning to support the design and management of scour protection measures under controlled hydraulic conditions.

Academic Editor: Jueyi Sui

Received: 17 June 2026

Revised: 8 July 2026

Accepted: 10 July 2026

Published: 10 July 2026

Keywords: local scour prediction; collar-protected bridge piers; interpretable machine learning; SHAP analysis; ensemble learning

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Scour remains one of the primary causes of bridge foundation instability and failure, posing serious challenges to structural safety, serviceability, and the long-term resilience of transportation infrastructure [1,2]. Local scour around bridge piers results from the

complex interaction of approaching flow, sediment transport, and pier geometry, producing highly nonlinear hydraulic processes that are difficult to represent using conventional analytical or empirical equations. Most empirical scour prediction formulas were developed from laboratory experiments conducted under specific hydraulic and geometric conditions, restricting their applicability when extrapolated to different flow regimes or sediment characteristics [3–5]. Consequently, improving the reliability of scour prediction has become a major focus of bridge engineering research, particularly within risk-informed approaches for bridge assessment and maintenance [6].

Among the various scour countermeasures, collars have attracted considerable attention because they alter the local flow field by weakening the downflow and suppressing the horseshoe vortex, which is responsible for sediment erosion around bridge piers [7]. Laboratory investigations have consistently demonstrated that appropriately designed collars can substantially reduce final scour depth, scour volume, and scour extent, with reported reductions often exceeding 70–80% under controlled experimental conditions [8]. However, these studies have generally been conducted for limited combinations of flow intensity, collar geometry, sediment properties, and pier dimensions, resulting in predictive relationships that are highly experiment-specific [9,10]. Consequently, generalized models capable of accurately estimating scour around collar-protected bridge piers under varying hydraulic conditions remain limited.

To overcome the limitations of empirical approaches, machine learning (ML) techniques have increasingly been applied to bridge scour prediction because of their ability to capture complex nonlinear relationships among hydraulic, sediment, and structural variables [11–13]. Previous investigations have shown that ensemble learning algorithms generally outperform conventional empirical equations and single learning models owing to their improved generalization capability and robustness [14–17]. Recent studies employing tree-based methods, including Random Forest (RF), XGBoost (XGB), and LightGBM, have reported further improvements in predictive performance through advanced ensemble strategies and hybrid optimization frameworks [18,19]. Nevertheless, most of these investigations have focused on scour around unprotected or conventionally shaped bridge piers. Comparatively few studies have examined collar-protected piers, despite the substantially different flow mechanisms introduced by collars. Furthermore, existing studies often evaluate only one or two algorithms using different datasets and experimental conditions, making direct comparison of modern ensemble techniques under identical hydraulic conditions difficult. This limitation hinders the identification of the most reliable predictive framework for collar-protected bridge pier scour.

Alongside improvements in predictive accuracy, recent research has emphasized the importance of interpreting machine learning models to ensure that their predictions remain consistent with established hydraulic principles. Explainable artificial intelligence (XAI) techniques, particularly SHapley Additive exPlanations (SHAP), provide a quantitative framework for evaluating the contribution of individual input variables to model predictions [20,21]. SHAP has been successfully applied in several hydraulic and water resources applications, including suspended sediment transport, seepage and uplift assessment, runoff prediction, and hydroclimatological modeling [22–24]. More recently, SHAP has also been incorporated into bridge scour prediction models to improve model transparency and assess the consistency between machine learning predictions and established dimensionless hydraulic relationships [25]. The application of interpretable ensemble machine learning to scour prediction around collar-protected circular bridge piers remains largely unexplored. Existing empirical relationships are generally derived using predefined functional forms that may not fully represent the complex nonlinear interactions among hydraulic, sediment, and collar-related parameters. Therefore, this study investigates the capability of interpretable ensemble learning combined with SHAP analysis to model these nonlinear relation-

ships using a carefully controlled experimental dataset while providing physically meaningful explanations of model predictions. Concurrently, risk assessment and resilience strategies for infrastructure underscore the need for transparent and reliable scour prediction models as integral components of strategic asset management, particularly under extreme events, climatic variability, and multi-hazard risk scenarios.

The primary contributions include: (i) developing an interpretable ensemble learning framework using a carefully controlled laboratory dataset; (ii) implementing optimized ensemble learning algorithms capable of robust performance with limited data; and (iii) employing SHAP-based interpretability to elucidate the combined effects of collar geometry, placement, and hydraulic conditions on final scour depth. By demonstrating that high-fidelity, physically consistent predictions are attainable with a relatively small dataset, this research seeks to bridge the gap between experimental collar studies and the development of practical, explainable design tools for scour-resilient bridge foundations.

2. Materials and Methods

2.1. Experimental Setup

The experimental dataset employed in this study was extracted from the laboratory investigation conducted by Pandey et al. [26], which examined equilibrium scour around collar-protected circular bridge piers under clear-water conditions. The experiments were performed in a recirculating rectangular flume measuring 20.0 m in length and 1.0 m in width, with an 8.0 m long test section located 7.0 m downstream from the flume entrance (Figure 1a). The sediment bed consisted of two nearly uniform sands with median grain sizes (d_{50}) of 0.27 mm and 0.40 mm and corresponding geometric standard deviations (σ_g) of 1.23 and 1.20, respectively. Before each experiment, the sediment bed was leveled using a two-dimensional bed profiler to ensure uniform bed conditions.

Two hollow circular cast-iron pier models having diameters of 66 mm and 84 mm were installed vertically at the center of the test section. Three circular collar diameters, corresponding to 1.5b, 2.0b, and 2.3b (where b is the pier diameter), were investigated (Figure 1b). The collars were positioned at four elevations relative to the initial bed level, namely at the bed surface ($z = 0$), one-quarter of the flow depth above the bed, one-half of the flow depth above the bed, and three-quarters of the flow depth above the bed, enabling the evaluation of collar size and elevation on equilibrium scour development.

The discharge was regulated using an ultrasonic flow meter, while the approach flow velocity and three-dimensional velocity components were measured using an Acoustic Doppler Velocimeter (ADV) operating at a sampling frequency of 150 Hz. Maximum final scour depth was measured using a point gauge after each experiment. All tests were conducted under clear-water scour conditions ($U/U_c < 1$), and each experiment was continued until no measurable increase in final scour depth was observed, indicating that equilibrium scour conditions had been achieved. It should be noted that the time required to reach equilibrium scour under clear-water conditions is not constant and depends on several factors, including flow intensity, sediment characteristics, and pier geometry. Melville and Chiew [27] have shown that experiments conducted near the threshold of sediment motion may require several days to achieve true equilibrium, whereas higher flow intensities generally result in substantially shorter equilibrium times. The present study adopts the final scour depths reported in the source experimental investigation, in which each test was continued until no measurable increase in final scour depth was observed. Consequently, the developed machine learning models inherit the equilibrium criterion adopted in the original laboratory study, and their applicability should be interpreted within the experimental conditions investigated.

Two experimental cases were considered: (i) reference tests without a collar to determine the final scour depth around an unprotected pier, and (ii) tests with collars to eval-

uate the influence of collar diameter and elevation on scour reduction. The present study utilizes the resulting 48 experimental observations reported by Pandey et al. [26] to develop and interpret ensemble machine learning models for equilibrium scour prediction.

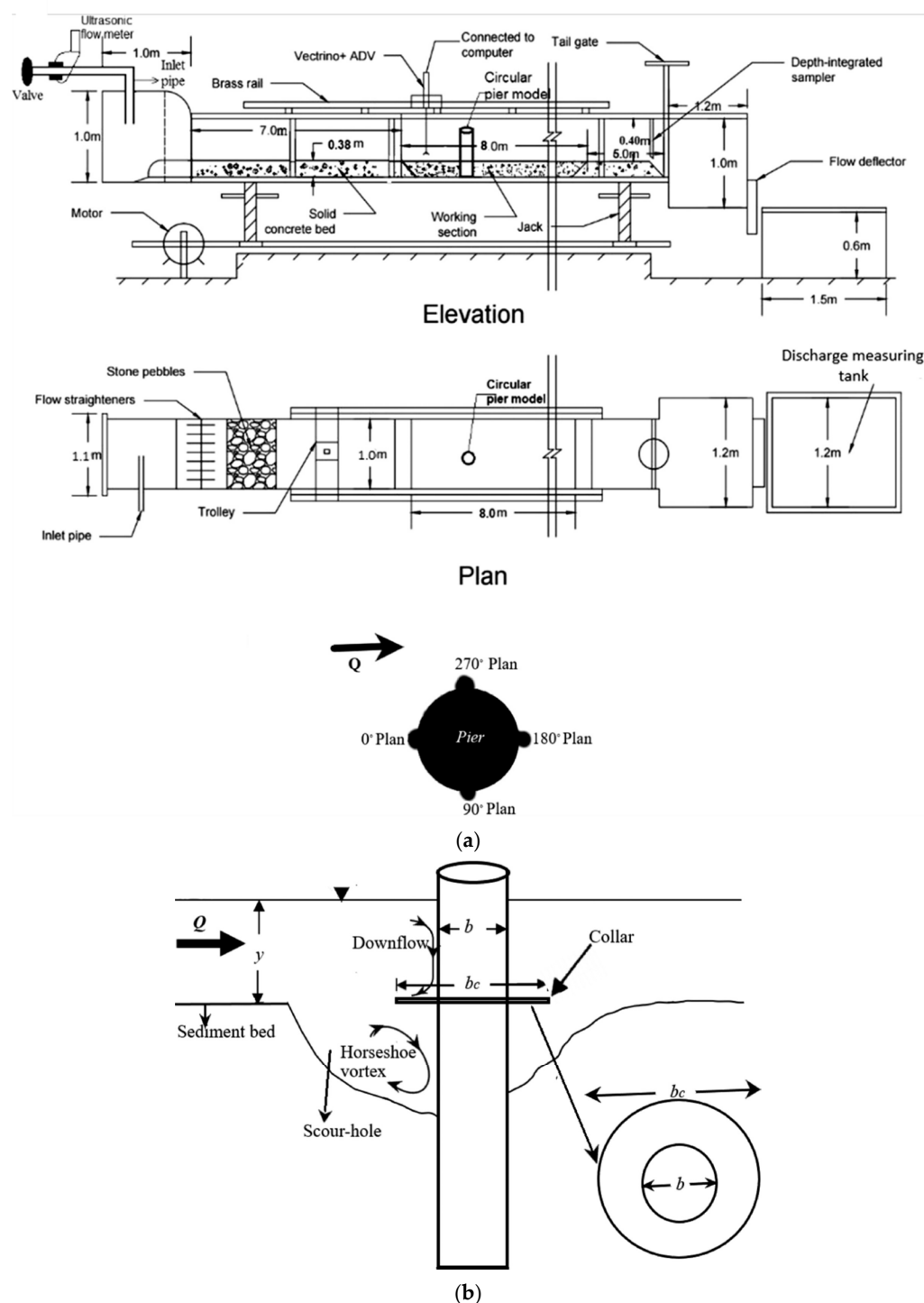


Figure 1. (a) Layout of the present experimental setup with 0–360° plan view around the pier (Pandey et al. [26]), (b) scour around a pier in the presence of collar at bed-level: A definition sketch (Pandey et al. [26]).

2.2. Data Collection

The data set developed in this study is based on the experimental investigation reported by Pandey et al. [26]. The research examined changes in local scour around circular bridge piers with collars installed at various levels and pier diameters in a controlled laboratory environment. Experiments were conducted under low water velocity conditions, specifically within the clear-water scour regime ($U/U_c < 1$). Several hydraulic parameters were systematically measured, including approach flow depth, flow velocity, sediment size, collar diameter, pier diameter, and final scour depth.

A total of forty-eight data points corresponding to collar-protected pier configurations were extracted and compiled from the experimental database for the purpose of training machine learning models. The original experimental program was conducted using two nearly uniform sediment beds with median grain sizes (d_{50}) of 0.27 mm and 0.40 mm, corresponding to geometric standard deviations (σ_g) of 1.23 and 1.20, respectively. These values indicate well-sorted sediments and minimize the influence of sediment gradation on scour development. All experiments were performed under clear-water scour conditions ($U/U_c < 1$), ensuring that sediment transport occurred only in the vicinity of the bridge pier rather than throughout the channel bed.

According to the original experimental investigation, each test was continued until equilibrium scour conditions were reached, defined by the absence of any significant increase in final scour depth with time. Since the present study develops machine learning models using experimentally measured final scour depths reported in the validated laboratory investigation, the reliability of the predictive models depends on the quality and consistency of those experimental measurements. Therefore, the present work adopts the experimentally established final scour depths without modification. The data set encompasses a range of hydraulic conditions and sediment median diameters (d_{50}), including variations in approach velocity (U/U_c), collar diameter to pier diameter ratio (b_c/b), and collar elevations to collar diameter ratio (z/b_c). Consequently, the data set is suitable for both physical interpretation and statistical analysis.

The sample size is relatively small compared to large-scale data-driven studies, which is typical for hydraulic laboratory experiments due to their time-consuming, costly, and physically constrained nature. To address the limitations of a small sample size, appropriate validation schemes were implemented, including cross-validation and ensemble-based learning techniques, both of which are effective when working with limited data. Furthermore, experimentally validated datasets yield reliable measurements with less noise compared to field-collected data. Although the number of observations is limited, the data set remains physically consistent and can be used to develop interpretable and robust predictive models.

Although the dataset originates from a single experimental investigation, it represents a systematically designed laboratory program in which hydraulic and geometric variables were independently varied under carefully controlled clear-water scour conditions. The consistency of the experimental procedures minimizes inter-study variability arising from differences in laboratory facilities, measurement techniques, sediment properties, and testing protocols. Consequently, the dataset provides a reliable benchmark for evaluating the capability of interpretable ensemble learning models to capture nonlinear scour mechanisms. Nevertheless, the predictive relationships developed in this study should be interpreted within the investigated hydraulic conditions, and their broader applicability should be verified using datasets compiled from multiple experimental studies and field measurements. Figure 2 present the overall methodology adopted in this research paper.

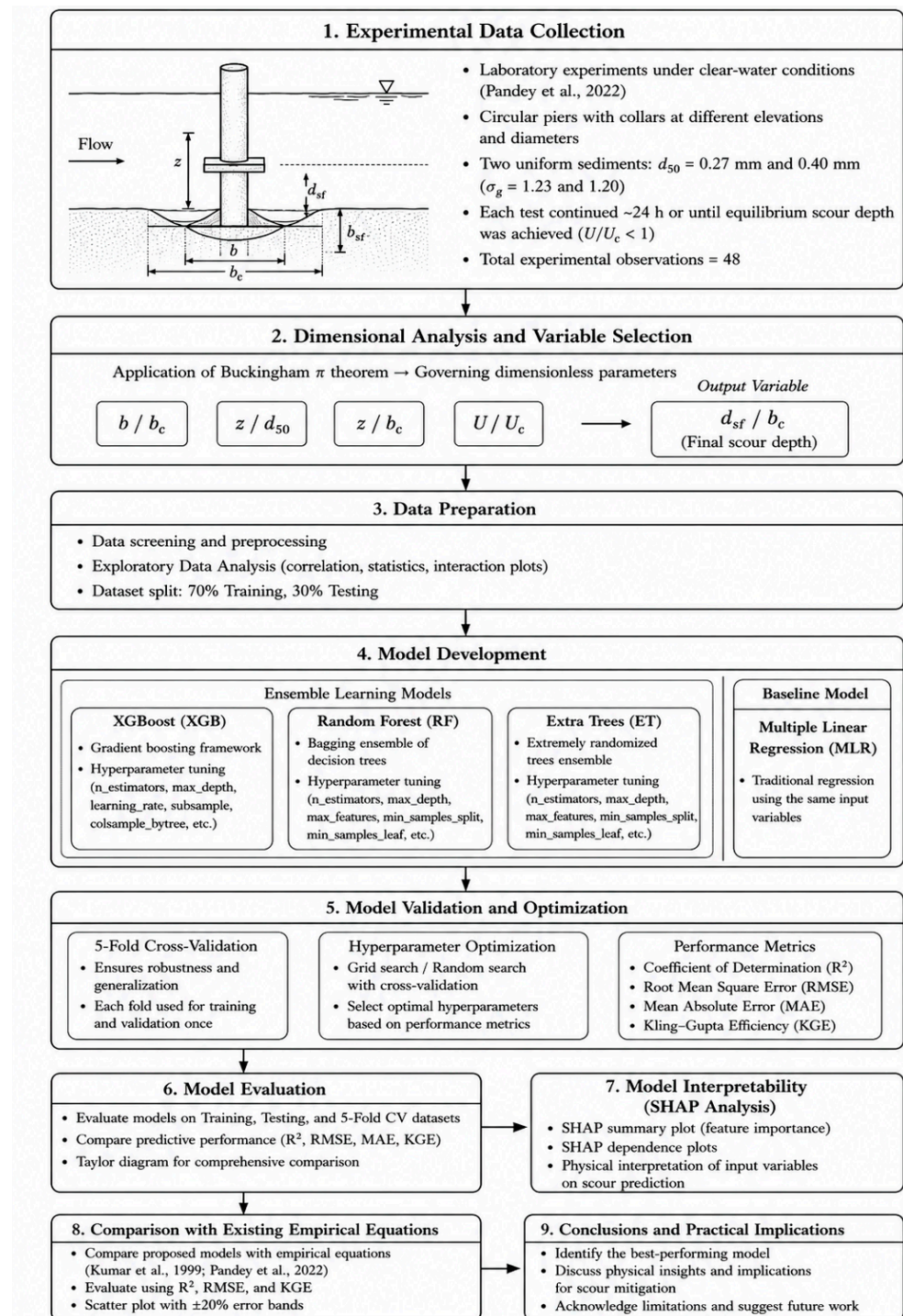


Figure 2. Experimental methodology adopted in this study, illustrating the workflow from data collection, preprocessing and dimensional analysis to ensemble machine learning model development, hyperparameter tuning, model validation using 5-fold cross-validation, performance evaluation, SHAP-based interpretation, regression model development, and comparison with existing empirical equations [26,28].

2.3. Dimensional Analysis

The final scour depth around a collar-protected circular bridge pier is governed by the interaction of hydraulic, geometric, and sediment characteristics. Accordingly, the final scour depth (d_{sf}) may be expressed as;

$$d_{sf} = f(U, \rho, \mu, g, b, b_c, z, y, U_c, d_{50}, \rho_s) \quad (1)$$

where (U) is the approach flow velocity, (ρ) is the fluid density, (μ) is the dynamic viscosity, (g) is the gravitational acceleration, (b) is the pier diameter, (b_c) is the collar diameter, (z) is the collar elevation relative to the initial bed level, (y) is the flow depth, (U_c) denotes the critical flow velocity corresponding to the threshold of sediment motion, (d_{50}) is the median sediment diameter, and (ρ_s) is the sediment density.

Applying the Buckingham π theorem and considering clear-water scour under fully turbulent flow conditions, where the effects of fluid viscosity and density remain approximately constant, the governing functional relationship can be expressed in dimensionless form as

$$\frac{d_{sf}}{b_c} = f\left(\frac{b}{b_c}, \frac{U}{U_c}, \frac{z}{b_c}, \frac{z}{d_{50}}\right) \quad (2)$$

The dimensionless groups (b/b_c), (z/d_{50}), and (z/b_c) characterize the geometric configuration of the collar system, whereas (U/U_c) represents the hydraulic intensity governing sediment entrainment under clear-water conditions. The normalized final scour depth (d_{sf}/b_c) was therefore selected as the response variable, while the four dimensionless parameters served as the input variables for the machine learning models developed in this study.

2.4. Method Adopted and Ensemble Learning Models

A methodological framework was developed to predict the final scour depth around collar-protected circular bridge piers by integrating hydraulic dimensional analysis with ensemble machine learning techniques. Initially, the governing hydraulic and geometric variables were transformed into physically meaningful dimensionless parameters to satisfy hydraulic similarity principles and improve the general applicability of the developed models. The adopted dimensionless variables were selected based on the dimensional analysis presented in Section 2.3. Before model development, exploratory data analysis was conducted to examine the dataset's statistical characteristics. Correlation analysis, interaction plots, and descriptive statistics were employed to evaluate the relationships among the predictor variables and identify potential multicollinearity. Subsequently, sensitivity analysis was conducted to identify the variables exerting the greatest influence on final scour depth.

The dataset was randomly split into training (70%) and test (30%) sets. The training dataset was used to construct the predictive models, whereas the testing dataset was employed for independent model evaluation. In addition, a five-fold cross-validation procedure was adopted to assess model robustness and generalization capability, ensuring that each observation was used for both training and validation. All ensemble learning models, hyperparameter implementation, performance evaluation, and SHAP analyses were independently developed by the authors using self-written Python scripts in the Google Colaboratory (Google Colab: <https://colab.research.google.com/>) environment. The implementation utilized standard Python libraries, including Scikit-learn for the Random Forest and Extra Trees models, the XGBoost library for the Extreme Gradient Boosting model, and the SHAP library for model interpretability. All analyses, model training, hyperparameter configuration, cross-validation, and visualization were performed within the Google Colab environment.

Three ensemble learning algorithms, namely Random Forest (RF), Extra Trees (ET), and Extreme Gradient Boosting (XGBoost), were selected because of their proven ability to model complex nonlinear relationships without requiring predefined functional forms (Table 1) [27]. For comparison, a multiple linear regression model was also developed using the same input variables. The predictive performance of all models was evaluated using the coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Kling–Gupta Efficiency (KGE). The overall methodological framework adopted in this study is illustrated in Figure 2.

To ensure reliable model performance while minimizing overfitting, the hyperparameters of each ensemble learning algorithm were systematically tuned through an iterative trial-and-error process based on five-fold cross-validation. Different combinations of key hyperparameters, including the number of trees, tree depth, learning rate, subsampling ratio, and feature sampling strategy, were evaluated to identify configurations that provided the best balance between predictive accuracy and model generalization. Considering the relatively small experimental dataset, conservative hyperparameter settings were adopted to avoid unnecessarily complex models while maintaining stable predictive performance. Hyperparameter optimization was performed using an iterative trial-and-error procedure guided by the five-fold cross-validation results. Only the principal hyperparameters that strongly influence model complexity and predictive performance (e.g., `n_estimators`, `max_depth`, `learning_rate`, `subsample`, and `colsample_bytree`) were adjusted during model development. The remaining hyperparameters were intentionally kept fixed at their default values provided by the Scikit-learn and XGBoost libraries. These default parameters were not included in the optimization process because the limited dataset (48 observations) did not justify a large hyperparameter search space, which could increase the risk of overfitting while providing limited improvement in predictive performance. Although Random Forest and Extra Trees employed 300 trees with unrestricted maximum depth, overfitting was mitigated through ensemble averaging, independent testing, and cross-validation, as evidenced by the relatively small differences between the training and testing performance metrics. The final optimized hyperparameter values used for model development are summarized in Table 1.

Table 1. Configuration and development details of ensemble machine learning models used for prediction of dimensionless final scour depth (d_{sf}/b_c).

Category	Random Forest (RF)	Extra Trees (ET)	Extreme Gradient Boosting (XGBoost)
Model Type	Bagging ensemble	Randomized bagging ensemble	Gradient boosting ensemble
Learning Strategy	Bootstrap aggregation	Extremely randomized trees	Sequential gradient boosting
Base Learner	Decision Tree Regressor	Decision Tree Regressor	Regression Trees
Dataset Size	48 experimental observations	48 experimental observations	48 experimental observations
Input Variables	(b/b_c), (z/d_{50}), (z/b_c), (U/U_c)	Same	Same
Output Variable	(d_{sf}/b_c)	(d_{sf}/b_c)	(d_{sf}/b_c)
Data pre-processing	Missing values removed; numerical conversion	Same	Same
Training Strategy	70% training, 30% testing	70% training, 30% testing	70% training, 30% testing
Cross-Validation	5-Fold Cross Validation	5-Fold Cross Validation	5-Fold Cross Validation
Hyperparameter Optimization	Iterative trial-and-error using 5-fold CV	Iterative trial-and-error using 5-fold CV	Iterative trial-and-error using 5-fold CV
<code>n_estimators</code>	300 (optimized)	300 (optimized)	300 (optimized)
<code>max_depth</code>	None (fixed at library default)	None (fixed at library default)	4 (optimized)
<code>min_samples_split</code>	2 (fixed at library default)	2 (fixed at library default)	Not specified (library default)
<code>min_samples_leaf</code>	1 (fixed at library default)	1 (fixed at library default)	Not specified (library default)
<code>max_features</code>	$\sqrt{\text{ }}$ (fixed at library default)	$\sqrt{\text{ }}$ (fixed at library default)	Not applicable
Bootstrap Sampling	Enabled (fixed)	Disabled (fixed)	Not applicable

Split Selection	Best split among random feature subset	Random feature and random threshold	Gradient-based optimization
learning_rate	Not applicable	Not applicable	0.05 (optimized)
subsample	Not applicable	Not applicable	0.80 (optimized)
colsample_bytree	Not applicable	Not applicable	0.80 (optimized)
Objective Function	Squared error (default)	Squared error (default)	Squared error regression
Regularization	Tree averaging	Randomization	Learning-rate shrinkage
Random State	42 (fixed)	42 (fixed)	42 (fixed)
Feature Importance	Mean decrease in impurity + SHAP	Mean decrease in impurity + SHAP	Gain-based importance + SHAP
Model Interpretation	SHAP TreeExplainer	SHAP TreeExplainer	SHAP TreeExplainer
Performance Metrics	(R ²), RMSE, MAE, KGE	(R ²), RMSE, MAE, KGE	(R ²), RMSE, MAE, KGE
Programming Language	Python 3	Python 3	Python 3
Software Library	Scikit-learn	Scikit-learn	XGBoost
Development Platform	Google Colab	Google Colab	Google Colab

Note: Hyperparameter optimization was performed using an iterative trial-and-error procedure guided by five-fold cross-validation. Only the principal hyperparameters influencing model complexity and predictive performance (i.e., *n_estimators*, *max_depth*, *learning_rate*, *subsample*, and *colsample_bytree*) were optimized. All remaining hyperparameters were intentionally fixed at their library default values and were not included in the optimization process to reduce model complexity and minimize the risk of overfitting associated with the limited dataset.

2.4.1. Random Forest (RF)

Random Forest (RF) is a supervised ensemble learning algorithm based on the bagging (Bootstrap Aggregating) technique, in which multiple decision trees are independently constructed from bootstrap samples of the training dataset [29]. During the construction of each tree, a random subset of predictor variables is selected at every node split, thereby introducing diversity among the individual trees. The final prediction is obtained by averaging the outputs of all decision trees, thereby substantially reducing prediction variance and improving the model's generalization. Owing to its ability to capture complex nonlinear interactions without requiring predefined mathematical relationships, RF has been widely adopted in hydraulic engineering applications involving highly nonlinear processes such as sediment transport and local scour prediction [30].

In the present study, the Random Forest model was developed using four dimensionless hydraulic parameters, namely b/b_c , z/d_{50} , z/b_c , and U/U_c , as predictor variables, while the normalized final scour depth (d_{sf}/b_c) was considered as the target variable. The dataset was randomly split into training (70%) and test (30%) sets, and model robustness was further evaluated using 5-fold cross-validation. The RF model was implemented with 300 decision trees (*n_estimators* = 300) to ensure stable ensemble predictions while maintaining computational efficiency. Bootstrap sampling was enabled to increase the diversity of the training samples, whereas the maximum tree depth was left unrestricted, allowing each tree to grow until the stopping criteria were satisfied. At each node, a random subset of predictor variables was considered for splitting, thereby reducing the correlation among individual trees and minimizing overfitting. A fixed random state of 42 was employed to ensure reproducibility of the model development. The complete configuration of the RF model is summarized in Table 1.

The selected hyperparameter settings represent a balanced trade-off between prediction accuracy and model complexity, which is particularly important for relatively small experimental datasets. Increasing the number of trees generally improves model stability by reducing prediction variance; however, beyond approximately 300 trees, the improve-

ment becomes marginal while computational cost increases. Similarly, allowing unrestricted tree growth enables the model to capture the nonlinear hydraulic interactions governing scour formation, whereas the ensemble averaging mechanism effectively controls overfitting. After model development, feature importance analysis and SHapley Additive exPlanations (SHAP) were employed to quantify the relative contribution of each hydraulic variable, thereby improving the interpretability of the RF predictions and facilitating their physical interpretation in terms of established scour mechanisms.

2.4.2. Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) is an advanced ensemble learning algorithm based on the gradient boosting framework, in which decision trees are constructed sequentially rather than independently. Each successive tree is trained to minimize the residual errors of the preceding trees by optimizing a differentiable loss function using gradient descent. In addition to gradient boosting, XGBoost incorporates regularization techniques to control model complexity, improve generalization, and reduce overfitting, making it particularly suitable for nonlinear regression problems involving relatively small datasets [31].

For the present study, the XGBoost model was implemented using the same dimensionless input variables and target variable as those adopted for the other ensemble models to ensure a consistent comparison. The model configuration included 300 boosting trees ($n_{\text{estimators}} = 300$) with a maximum tree depth of 4, which provided sufficient model complexity while preventing excessively deep trees that could overfit the limited experimental dataset. A learning rate (η) of 0.05 was adopted to gradually update the model by applying small corrections during each boosting iteration, thereby improving prediction stability. Furthermore, $\text{subsample} = 0.80$ and $\text{colsample_bytree} = 0.80$ were employed so that each tree was trained using 80% of the available observations and predictor variables, respectively. These stochastic sampling strategies reduce the correlation among trees and improve the model's generalization capability. A fixed random state of 42 was specified to ensure reproducibility of the modeling process. The complete hyperparameter settings are summarized in Table 1.

The selected hyperparameters were chosen to achieve an appropriate balance between predictive accuracy and model robustness. A relatively low learning rate, together with a moderate tree depth, enables the model to learn complex hydraulic relationships progressively while minimizing the risk of overfitting. Unlike bagging-based methods, which reduce variance through parallel tree construction, XGBoost improves prediction performance by sequentially correcting residual errors and incorporating regularization during model optimization. This capability allows the algorithm to effectively capture the nonlinear interactions among hydraulic and geometric variables governing final scour depth. The resulting model predictions were subsequently interpreted using SHapley Additive exPlanations (SHAP) to quantify the influence of individual variables and verify the physical consistency of the learned relationships with established scour mechanics.

2.4.3. Extra Trees (ET)

The Extra Trees (ET) algorithm, also known as Extremely Randomized Trees, is a tree-based ensemble learning method that extends the Random Forest approach by introducing a higher degree of randomness during tree construction. Unlike Random Forest, which determines the optimal split among a randomly selected subset of features, ET randomly selects both the predictor variables and the split thresholds for each node. Furthermore, ET constructs trees using the entire training dataset rather than bootstrap samples. This increased randomization reduces the correlation among individual trees, resulting in

lower model variance and improved generalization performance, particularly for complex nonlinear regression problems [32].

In this study, the ET model was developed using 300 decision trees ($n_estimators = 300$) with unrestricted tree depth to enable the algorithm to capture the complex nonlinear interactions among the governing hydraulic variables. Bootstrap sampling was disabled, allowing each tree to be trained using the complete training dataset while introducing randomness through randomly generated split points. The default feature selection strategy was adopted, in which a random subset of predictor variables was selected at each node split, and a fixed random state of 42 was specified to ensure reproducibility. The complete hyperparameter settings adopted for the ET model are summarized in Table 1.

Compared with conventional bagging methods, the higher level of randomization employed by ET often improves predictive stability for relatively small datasets by reducing sensitivity to noise and avoiding overly complex decision boundaries. These characteristics make the algorithm particularly suitable for hydraulic datasets that exhibit nonlinear relationships among flow intensity, sediment properties, and geometric variables. Following model development, the predictive behavior of the ET model was further interpreted using SHapley Additive exPlanations (SHAP), enabling the relative importance and hydraulic significance of each input variable to be evaluated.

2.4.4. Performance Evaluation Metrics

The predictive performance of the developed ensemble learning models was evaluated using four widely adopted statistical indices, namely the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and Kling–Gupta Efficiency (KGE). These metrics provide complementary information regarding the goodness-of-fit, prediction error, and overall agreement between observed and predicted final scour depths.

Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

Where y_i = observed final scour depth, $\hat{y}_i$ = predicted final scour depth, $\bar{y}$ = mean of the observed final scour depths, n = total number of observations. The value of R^2 ranges from 0 to 1, with values closer to 1 indicating better predictive performance.

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

Lower RMSE values indicate higher prediction accuracy.

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

A smaller MAE indicates better agreement between observed and predicted values.

Kling–Gupta Efficiency (KGE)

$$KGE = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (6)$$

$$r = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}} \quad (7)$$

where $\alpha = \frac{\sigma_{\hat{y}}}{\sigma_y}$, $\beta = \frac{\mu_{\hat{y}}}{\mu_y}$, r = Pearson correlation coefficient, α = variability ratio, β = bias ratio, μ_y and $\mu_{\hat{y}}$ = means of observed and predicted values, respectively, and σ_y and $\sigma_{\hat{y}}$

= standard deviations of observed and predicted values, respectively. A KGE value closer to 1 indicates excellent agreement between observed and predicted values by simultaneously considering correlation, variability, and bias.

2.5. Development of the Regression Model

A regression-based approach was employed to develop a predictive equation for estimating final scour depth around collar-protected bridge piers. The analysis utilized 48 experimental observations conducted under various hydraulic and geometric conditions. Based on dimensional analysis and established knowledge of the scour process, four dimensionless variables were selected as independent variables: relative pier-to-collar width (b/b_c), relative collar elevation to sediment size (z/d_{50}), relative collar elevation to collar width (z/b_c), and flow intensity ratio (U/U_c). The dependent variable was defined as the final scour depth (d_{sf}/b_c).

Initially, a simple power law relationship between the dependent and independent variables was adopted, as is common in hydraulic engineering applications. The non-linear multiplicative relationship among the predictor variables was linearized by applying natural logarithm transformations to both the variables and the response. All machine learning models, hyperparameter implementation, and SHAP analyses were independently developed by the authors using self-written Python scripts in the Google Colaboratory (Google Colab) environment. The Scikit-learn library was used to implement the Random Forest and Extra Trees models, while the XGBoost and SHAP Python libraries were used for the Extreme Gradient Boosting model and model interpretability, respectively.

The regression coefficients obtained from the log-transformed model served as the exponents for the dimensionless variables in the resulting Equation 8, while the regression constant corresponded to the exponent of the intercept term. The intercept was exponentiated to yield the coefficient for the original power-law equation. Finally, the developed regression model was used to predict final scour depths, and the predicted values were converted back to the original scale using inverse logarithmic transformation for comparison with the experimental final scour depth measurements.

Although the dimensionless parameters z/d_{50} and z/b_c share a common numerator and may exhibit some degree of correlation, they represent different physical aspects of the scour process. The ratio z/d_{50} characterizes the collar elevation relative to the sediment size, whereas z/b_c represents the collar elevation relative to the collar geometry. Consequently, both variables were retained to preserve their physical significance. The empirical regression equation is intended primarily for comparison with the ensemble learning models, which are less sensitive to correlated predictors because of their tree-based learning structure.

$$\frac{d_{sf}}{b_c} = 13.6228 \left(\frac{b}{b_c}\right)^{3.2016} \times \left(\frac{z}{d_{50}}\right)^{0.0028} \times \left(\frac{z}{b_c}\right)^{0.0028} \times \left(\frac{U}{U_c}\right)^{3.3997} \quad (8)$$

The power-law regression equation was developed using only the training dataset. During data partitioning, observations with zero collar elevation ($z = 0$) were assigned to the independent testing/validation dataset and were therefore excluded from the regression fitting process. Consequently, logarithmic transformation was applied only to positive-valued observations, avoiding the undefined operation $\ln(0)$. Furthermore, the variables z/d_{50} and z/b_c exhibited perfect correlation within the dataset. This relationship was verified against the original experimental data and was found to result from the experimental design, where collar elevation was normalized using two different characteristic length scales while only a limited number of sediment sizes and collar diameters were considered. Both variables were retained in the regression equation to maintain consistency with the dimensionless hydraulic parameters reported in the source experimental study and to facilitate comparison with the developed machine learning models. Never-

theless, the associated multicollinearity should be considered when interpreting the individual regression coefficients, and future studies using larger datasets with a broader range of hydraulic conditions are recommended to improve predictor independence.

The proposed regression equation was formulated using dimensionless variables selected based on well-established hydraulic relationships reported in the bridge scour literature. The relatively small exponents associated with (z/d_{50}) and (z/b_c) indicate that these parameters have a comparatively limited influence on scour prediction within the range of experimental conditions investigated. This finding is consistent with the feature importance and SHAP analyses, which identify hydraulic intensity and collar geometry as the primary factors governing scour development. Consequently, the proposed equation should be regarded as a data-driven predictive relationship applicable to the experimental domain under consideration, rather than a universally applicable physical formulation. Future research incorporating more diverse laboratory and field datasets may further refine the selection and significance of the governing dimensionless parameters through comprehensive dimensional analysis, thereby improving the generalizability and robustness of the proposed model.

2.6. Descriptive Statistics

Descriptive statistical analysis was performed to characterize the distribution and variability of the dimensionless parameters used in the scour prediction models (Table 2). The dataset comprised 48 experimental observations of scour around collar-protected circular bridge piers under clear-water conditions. The relative pier-to-collar width ratio (b/b_c) had a mean of 0.534 and a standard deviation of 0.099, indicating moderate variability in the collar configurations. Both the median and mode were close to 0.50, which suggests a symmetric distribution around the central tendency. The narrow 95% confidence interval (0.505–0.563) further demonstrates consistency in the experimental design regarding collar dimensions.

The collar elevation parameters, z/d_{50} and z/b_c , display identical descriptive characteristics due to their derivation from the same experimental configurations. Both parameters had a mean of 0.135 and a median of 0.11. The mode values of z/d_{50} and z/b_c were significantly low (0), indicating that a significant proportion of the experiments involved collars placed directly on the sediment bed. The low standard deviation (0.112) suggests controlled variations in collar placement. The flow intensity parameter (U/U_c) had a mean of 0.753 and ranged from 0.620 to 0.880, confirming that all experiments were conducted under clear-water scour conditions $(U/U_c < 1)$. The small standard deviation (0.101) indicates limited variability in the approach flow conditions, which reduces potential experimental uncertainty.

The dependent variable, final scour depth (d_{sf}/b_c) , had the highest standard deviation (0.139) among all parameters, indicating substantial variability in scour response. This variation reflects the complex interaction between hydraulic loading and collar geometry. The broad range (0.404–0.949) demonstrates the suitability of the dataset for developing predictive models that capture diverse scour behaviors. Overall, the descriptive statistics confirm that the dataset is physically consistent, sufficiently variable, and appropriate for machine learning-based scour prediction studies.

Table 2. Descriptive Statistics of the Dimensionless Parameters Used for Scour Prediction.

Parameter	N	Mean	Median	Mode	Standard Deviation	Minimum	Maximum	95% Confidence Interval
b/b_c	48	0.534	0.5	0.5	0.099	0.434	0.667	0.505–0.563
z/d_{50}	48	0.135	0.111	0	0.112	0	0.333	0.103–0.167
z/b_c	48	0.135	0.111	0	0.112	0	0.333	0.103–0.167

U/U_c	48	0.753	0.755	0.62	0.101	0.62	0.88	0.724–0.783
d_{sf}/b_c	48	0.638	0.612	0.5	0.139	0.404	0.949	0.598–0.678

3. Results

3.1. Correlation Heatmap of Collar, Hydraulic and Final Scour Depth Parameters

The correlation map presented in Figure 3 provides a comprehensive statistical overview of the relationships among hydraulic, geometric, and scour-related variables considered in this study. This figure is significant as it visually summarizes both the magnitude and direction of correlations between all input variables and the output variable used for final scour depth analysis prior to the development of ensemble machine learning (ML) models. The results indicate that certain hydraulic and geometric parameters exhibit strong positive or negative correlations with final scour depth, highlighting the complex and non-linear nature of local scour around collar-protected circular piers. Generally, parameters associated with flow intensity and hydraulic conditions are positively correlated with final scour depth, indicating that increased flow strength enhances sediment entrainment and vortex formation around the pier.

In contrast, collar-related parameters display negative correlations with equilibrium scour, suggesting that the presence of a collar inhibits scouring at the pier foundation and reduces the intensity of horseshoe vortices. The heatmap further reveals interdependencies among some input variables. In hydraulic engineering, such multicollinearity is common because variables such as flow depth, velocity, pier diameter, sediment properties, and protection dimensions are physically interconnected. This observation is also relevant from a machine learning perspective, as ensemble models such as XGBoost, Random Forest, and Extra Trees can effectively process datasets with nonlinear and correlated variables, unlike traditional regression models.

From a hydraulic engineering standpoint, the figure validates the physical realism of the experimental dataset. The observed trends align with the classical scour mechanism, where higher approach flows increase bed stress and turbulence, resulting in deeper scour holes. Similarly, the reduction in final scour depth due to collar protection supports the effectiveness of collars in disrupting downward flow and mitigating horseshoe vortex intensity around bridge piers. The interpretation of the heatmap also informs feature selection and model interpretability. Parameters with stronger correlations to final scour depth are likely to contribute more significantly to ML prediction performance.

Therefore, the figure serves as an initial diagnostic tool for identifying dominant parameters influencing scour prior to advanced modeling and SHAP interpretation. Overall, the statistical and hydraulic relationships established in this work provide a foundation for connecting the physical scour process with interpretable data-driven insights (see Figure 2). The dataset demonstrates physical validity and sufficient correlation to support the development of reliable, interpretable ensemble ML models for forecasting scour around collar-protected circular piers.

The perfect correlation between z/d_{50} and z/b_c is not due to a labeling or data-processing error but is a consequence of the experimental design adopted by Pandey et al. [26]. Both variables represent collar elevation normalized by different characteristic length scales (median sediment size and collar diameter). Because the dataset contains only two sediment sizes and a limited number of collar diameter configurations, these variables become linearly dependent, resulting in a correlation coefficient of 1.0. The original dataset was verified against the source publication, confirming the consistency of the reported values.

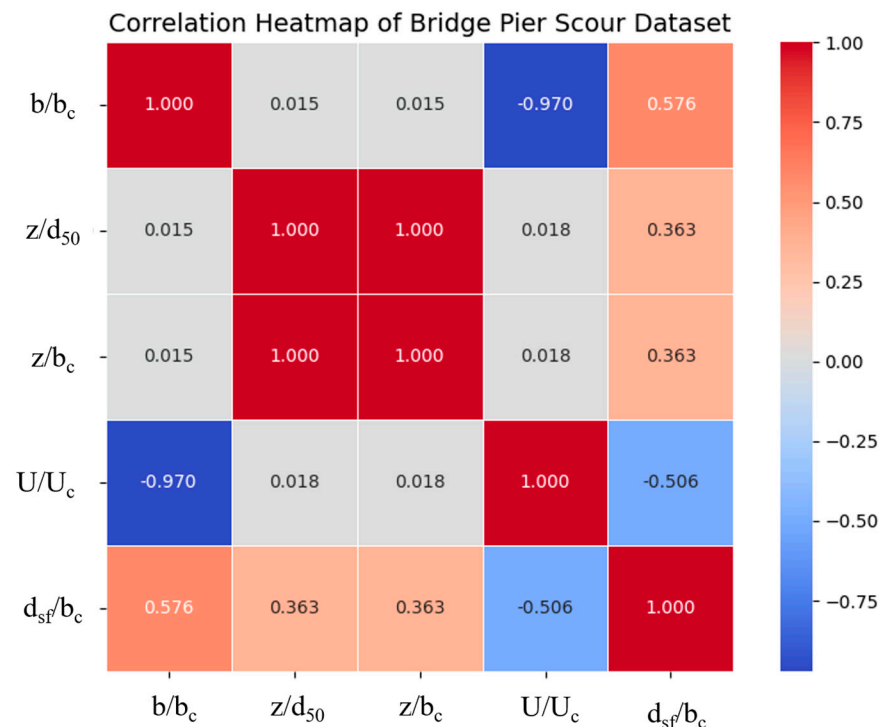


Figure 3. Correlation heatmap illustrating the relationships among the dimensionless input variables (b/b_c , z/d_{50} , z/b_c , and U/U_c) and the output variable (d_{sf}/b_c). The color scale represents the strength and direction of the correlation, providing a preliminary assessment of variable dependence and potential multicollinearity.

3.2. Scatter Pair Plot: Indicating Interaction of Scour Parameters

Figure 4 presents the scatter pair plot of the selected input variables against the final scour depth output variable. This figure is significant as it comprehensively illustrates the distribution, clustering behavior, linearity, and nonlinear interactions between hydraulic and geometric variables influencing equilibrium scour around collar-protected circular piers. The scatter pair plots indicate that the relationships between input parameters and final scour depth are predominantly nonlinear, thereby justifying the use of advanced ensemble machine learning (ML) techniques rather than traditional empirical regression equations. Several combinations display clear trends of increasing final scour depth with higher hydraulic loading conditions, which is attributed to increased vortex formation and sediment entrainment under strong flows.

Additionally, some parameters exhibit patterns beyond simple linearity, with clustering suggesting interactions between hydraulic variables and collar geometry. Local scour in hydraulic engineering is influenced by the combined effects of flow intensity, sediment properties, pier dimensions, and protective measures, rather than by a single parameter. These multiple dependencies are effectively visualized in the pair plot. The presence of several data points within millimeter-scale error bars, yet not coinciding, demonstrates necessary variability in the dataset for effective machine learning model training. Such variation is essential, as diverse data patterns are critical for accurately capturing nonlinear scour behavior using machine learning algorithms such as Random Forest, Extra Trees, and XGBoost. The absence of outliers further indicates that experimental measurements are consistent and physically valid.

The figure also demonstrates the impact of collar protection on the scour process from a hydraulic engineering perspective. Certain scatter distributions reveal lower final scour depth values as collar-related parameters increase under identical hydraulic condi-

tions, validating the effectiveness of collars in reducing vortex strength and bed material erosion around the pier foundation. This observation aligns with the hydraulic principle that collars reduce downward flow acceleration at the pier nose and are expected to inhibit the formation of the horseshoe vortex system responsible for local scour. The inclusion of distributions below the diagonal aids in understanding the statistical distribution of each parameter within the pair plot. A wider spread of values suggests greater experimental variability, which supports the generalization capability of interpretable ML models.

Furthermore, the observed nonlinear interaction patterns indicate that SHAP analysis would be beneficial for elucidating feature importance and model behavior in subsequent analyses. In summary, Figure 4 provides both statistical and hydraulic insights into the complex interaction mechanisms associated with collar-protected circular piers. The figure confirms that the scour process is highly nonlinear and multi-dimensional, and that interpretable ensemble machine learning methods are reliable for accurate scour prediction, even with relatively small datasets.

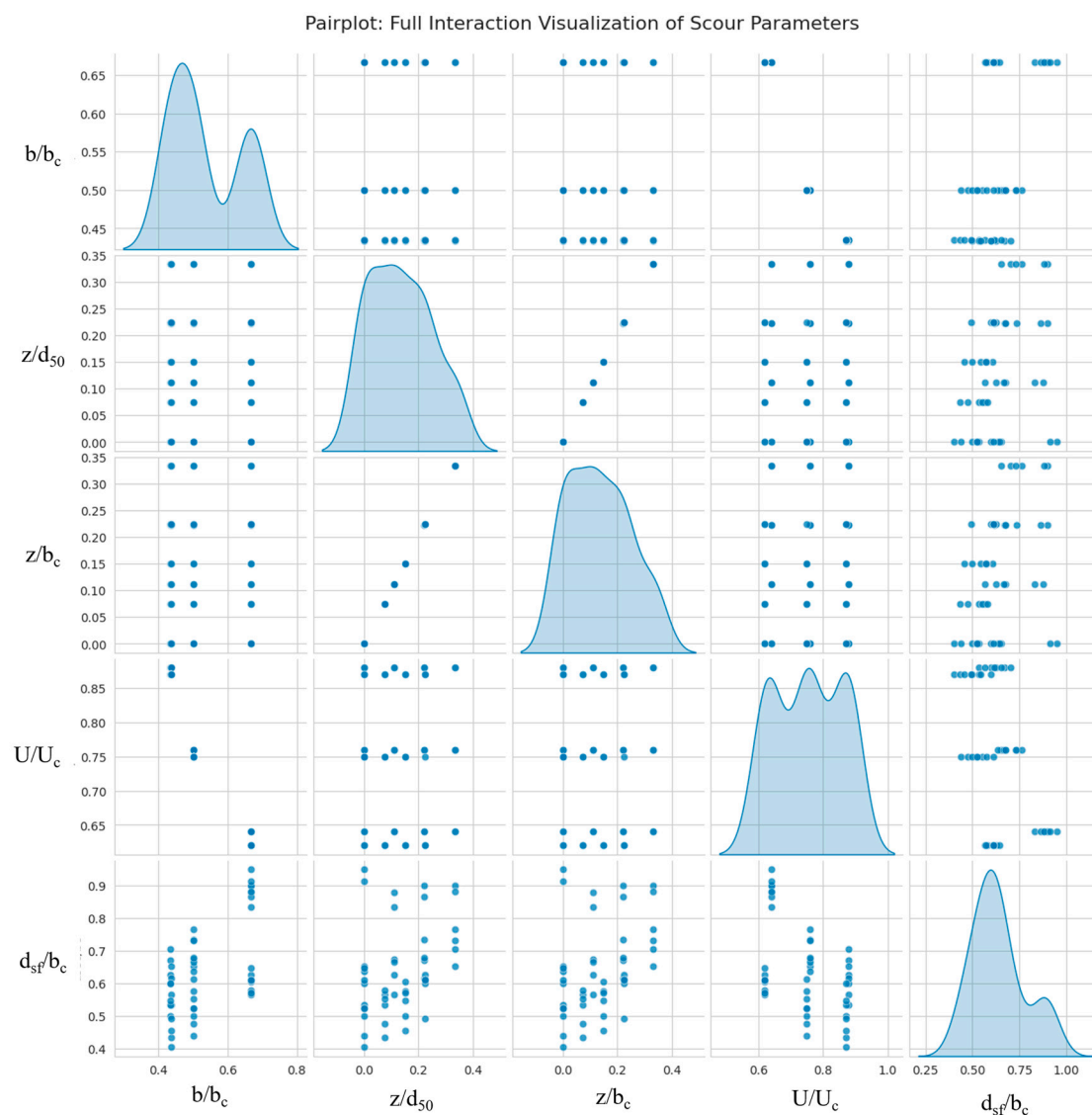


Figure 4. Scatter pair plots showing the interactions among the dimensionless input variables and the normalized final scour depth (d_{sf}/b_c). The plots illustrate the distribution of the experimental data and the pairwise relationships between hydraulic and geometric parameters.

3.3. Hydraulic Interpretation of Input and Output Parameter Relationships Using Scatter Correlation Plots

A hydraulic interpretation of the relationships among input parameters and final scour depth is presented as scatter correlation plots in Figure 5. This analysis is significant because it directly connects the experimental results with the hydraulic mechanisms observed in local scour around collar-protected circular piers. Unlike the correlation heatmap and pair plot, Figure 5 emphasizes the physical interpretation of each hydraulic parameter's effect on scour behavior.

The results indicate that, in general, an increase in hydraulic intensity parameters such as flow velocity and flow depth ratio leads to greater final scour depth. The scatter trends further demonstrate that approach flow conditions influence the shear stress and turbulence generated around the pier, both of which increase with higher approach flow velocity. Consequently, the downward flow and the horseshoe vortex system become more intense, resulting in deeper scour holes and a greater number of entrained sediment particles. These findings are consistent with classical local scour theory in bridge hydraulics.

The addition of collar protection is observed to reduce the likelihood of scour formation. Several scatter relationships demonstrate a decrease in final scour depth as collar-related characteristics increase. These observations suggest that collars perturb the near-bed flow field around the pier, altering the downflow motion responsible for horseshoe vortex formation. Larger collars weaken these vortices, leading to reduced sediment transport capacity near the pier base and limiting final scour depth development.

Another important observation is the presence of nonlinearity in many scatter plots. The increase in scour is not uniform across the entire range of hydraulic variables. At lower flow intensities, scour development is gradual, whereas at higher intensities, beyond a threshold, final scour depth increases more rapidly. This nonlinear response results from the interaction between turbulent generation, sediment mobility, and flow separation around the pier. Therefore, interpretable ensemble machine learning models are more suitable than traditional empirical equations for capturing these nonlinear dependencies.

From a hydraulic engineering perspective, the figure effectively supports the predominant mechanisms influencing local scour in protected pier systems. Both hydraulic loading and collar geometry are observed to jointly control scour variation, as illustrated in the scatter plot. Recognizing these factors is essential in bridge foundation design, as excessive local scour is a primary cause of bridge instability and structural failure during flood events.

Furthermore, the figure demonstrates that the hydraulic trends observed within the limited experimental data enhance confidence in the dataset for developing machine learning models such as Random Forest, Extra Trees, and XGBoost. The trends identified in the scatter plots correspond to physical processes, rather than being solely statistical artifacts.

Therefore, Figure 5 provides a critical link between data-driven analysis and hydraulic interpretation. The figure supports experimental observations, highlights the benefits of collar protection, and identifies nonlinear scour mechanisms that necessitate the use of interpretable ensemble machine learning methods for predicting scour around circular bridge piers.

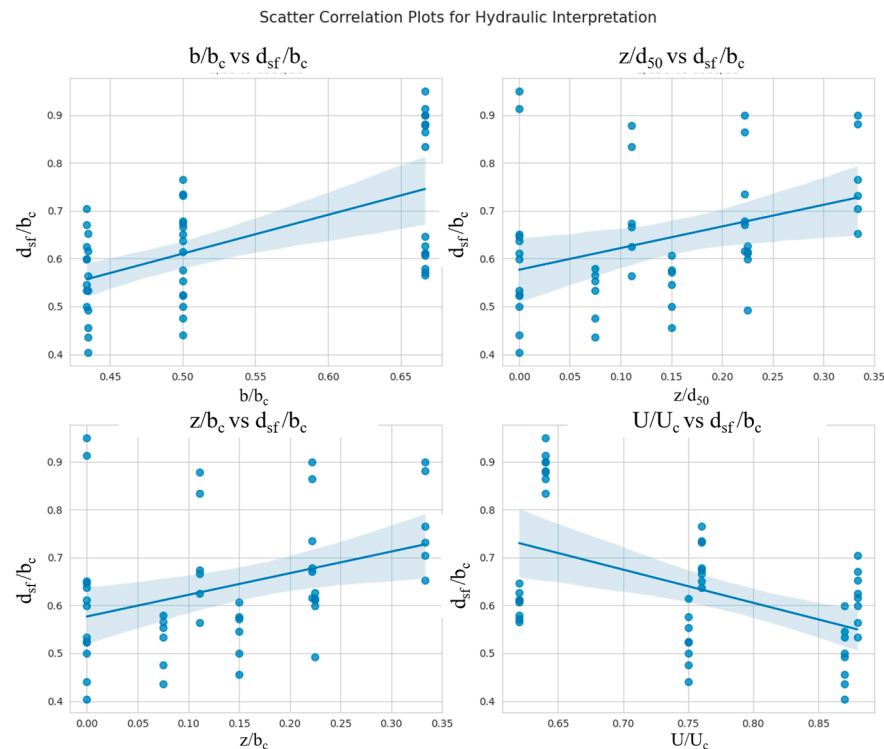


Figure 5. Hydraulic interpretation of the relationships between the dimensionless input variables and the normalized final scour depth (d_{sf}/b_c) using scatter correlation plots. The trends provide insight into the influence of collar geometry, collar elevation, sediment size, and flow intensity on scour development.

3.4. Sensitivity Analysis of Various Input Parameters Inducing Scour Around Collar-Protected Pier

Figure 6 presents the sensitivity analysis of the input variables governing final scour depth around collar-protected circular bridge piers. The analysis quantitatively evaluates the relative contribution of each hydraulic and geometric parameter to the prediction of final scour depth, thereby providing valuable insight into the decision-making process of the developed machine learning models. In addition to assessing predictive performance, sensitivity analysis is an important interpretability tool that helps determine whether model responses are consistent with established hydraulic principles.

The results indicate that the hydraulic intensity-related variables are the most influential predictors of final scour depth. Parameters describing the approach flow conditions exhibit the highest sensitivity, reflecting their fundamental role in governing local scour processes through the interaction of flow turbulence, sediment entrainment, and pier-induced flow separation. The elevated sensitivity associated with these variables demonstrates that variations in hydraulic loading substantially affect the development of horseshoe vortices, bed shear stress, and sediment transport in the vicinity of the pier foundation.

The analysis further reveals that collar-related parameters have a pronounced influence on scour reduction, confirming the effectiveness of collar protection as a scour mitigation measure. The presence of a collar alters the local flow field by obstructing the downward flow and weakening the horseshoe vortex system, thereby reducing sediment entrainment around the pier. Furthermore, the sensitivity of the model to collar dimensions indicates that geometric characteristics of the collar significantly influence scour mitigation performance, emphasizing the importance of optimizing collar geometry during the design of bridge foundations.

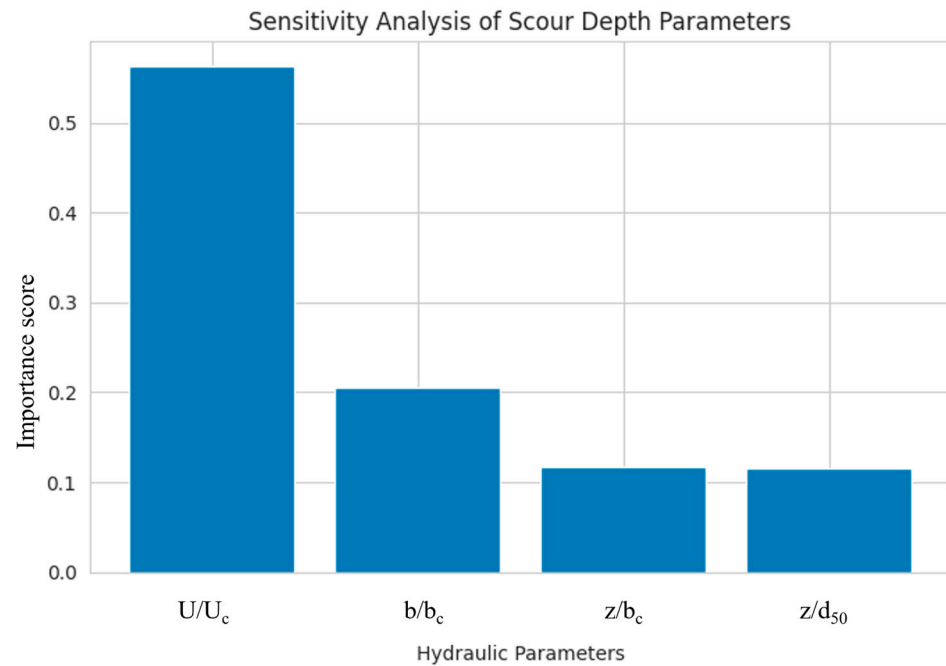


Figure 6. Sensitivity analysis showing the relative importance of the selected dimensionless input variables in predicting the normalized final scour depth around collar-protected circular bridge piers. Higher sensitivity values indicate a greater contribution of the corresponding variable to the prediction model.

The sensitivity analysis also demonstrates that the influence of the investigated variables is not uniform. While several parameters contribute substantially to scour prediction, others exhibit comparatively lower sensitivity, indicating a secondary role under the range of hydraulic conditions considered in this study. Distinguishing between dominant and less influential variables is valuable for experimental design, numerical simulations, and field-scale engineering applications, as it enables researchers and practitioners to prioritize the parameters that exert the greatest control over scour development. Such knowledge also improves the interpretability and robustness of the predictive models by ensuring that their responses are governed by physically meaningful hydraulic relationships rather than spurious statistical correlations.

Although ensemble learning algorithms such as Random Forest, Extra Trees, and XGBoost are primarily developed to maximize predictive accuracy, their practical application in hydraulic engineering requires consistency with the underlying physical mechanisms governing scour processes. The sensitivity results demonstrate that the developed models assign greater importance to variables that are well recognized in scour mechanics, indicating that the models have successfully captured the dominant hydraulic controls instead of relying solely on statistical patterns within the dataset. This agreement between model behavior and established hydraulic theory enhances confidence in the reliability and generalizability of the proposed models.

From an engineering perspective, the sensitivity analysis provides useful insight into the relative importance of the investigated variables for laboratory-scale collar-protected bridge piers under clear-water conditions. These findings may assist future experimental studies and support the development of predictive tools but should not be interpreted as direct guidance for bridge design or field-scale flood risk management without further validation. Identifying the primary controlling variables may help researchers prioritize influential hydraulic and geometric parameters in future laboratory investigations and model development. Since local scour remains one of the leading causes of bridge foun-

dation failure worldwide, understanding the relative importance of the governing parameters is essential for improving the safety, resilience, and long-term sustainability of hydraulic infrastructure. Overall, the sensitivity analysis indicates that the developed ensemble machine learning models capture relationships consistent with the laboratory observations and established hydraulic principles. The results provide insight into the relative influence of the investigated variables within the experimental conditions considered in this study. Additional validation using larger and field-scale datasets is required before extending these findings to practical scour mitigation strategies.

3.5. Five-Fold Cross Validation

The five-fold cross-validation results for the developed ensemble machine learning (ML) models are presented in Figure 7a–c, using three key statistical performance metrics: coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). This figure is particularly significant for evaluating the robustness, stability, and generalizability of the tested ML models across various data subsets. Such evaluation is crucial for obtaining reliable test approximations, especially when working with relatively small hydraulic datasets.

The results demonstrate that the XGBoost model consistently outperformed other models, including Random Forest (RF) and Extra Trees (ET), across all folds. The R^2 values for XGBoost remained relatively high (0.59–0.91) in each fold, indicating strong agreement between predicted and observed final scour depth values (Table 3). Table 3 summarizes the predictive performance of the developed ensemble models for the training, testing, and five-fold cross-validation datasets. To quantify the variability in model performance, the mean and standard deviation of the cross-validation metrics are also reported. Although some variation was observed among the individual folds, particularly in the first fold due to the limited number of experimental observations, the remaining folds exhibited consistent predictive performance. Among the evaluated models, XGBoost achieved the highest mean R^2 , the lowest mean RMSE, and the highest mean KGE, while also maintaining relatively low variability across the five folds. These results indicate that the proposed ensemble models, particularly XGBoost, provide satisfactory robustness and generalization capability for predicting the final scour depth of collar-protected bridge piers within the available experimental dataset. These high R^2 values suggest that the model effectively captures the nonlinear relationships between hydraulic variables, collar-protected piers, and the associated final scour depths.

Similarly, the RMSE (0.047–0.039) and MAE (0.034–0.035) results indicate that the XGBoost model achieved the lowest prediction errors among the tested models. Lower RMSE values reflect reduced magnitude of prediction errors, while lower MAE values demonstrate greater consistency in predictions. These findings indicate that XGBoost effectively learns complex hydraulic interactions and accurately models final scour depth, even when limited experimental data and essential nonlinear relationships are present.

The variation in results across folds demonstrates that the developed models are stable and reliable, exhibiting relatively small fold-wise variation. Consistency is essential in hydraulic engineering applications due to variable flow states and uncertain field environments where scour prediction models are implemented. The overall stability across folds suggests that the algorithm does not overfit the dataset and can generalize effectively to unseen laboratory observations having characteristics similar to those represented in the present dataset.

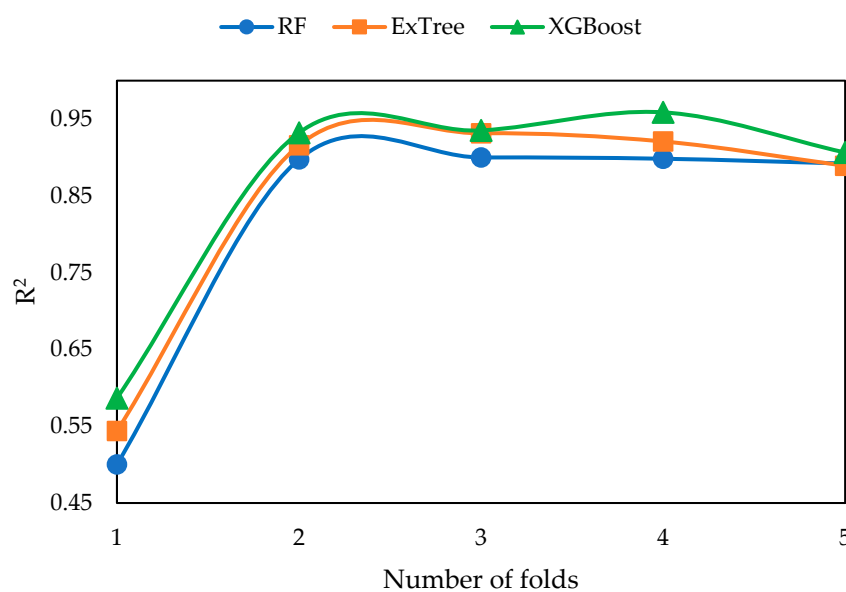
Figure 7a–c illustrates the potential of interpretable ensemble ML approaches for accurately predicting scouring using comparatively small experimental datasets in hydraulic engineering. Traditional scour predictions based on empirical equations often rely on limited laboratory experiments and may not accurately represent flow–sediment interac-

tions. Ensemble ML models provide more reliable predictions while maintaining physical consistency, as confirmed by the cross-validation results.

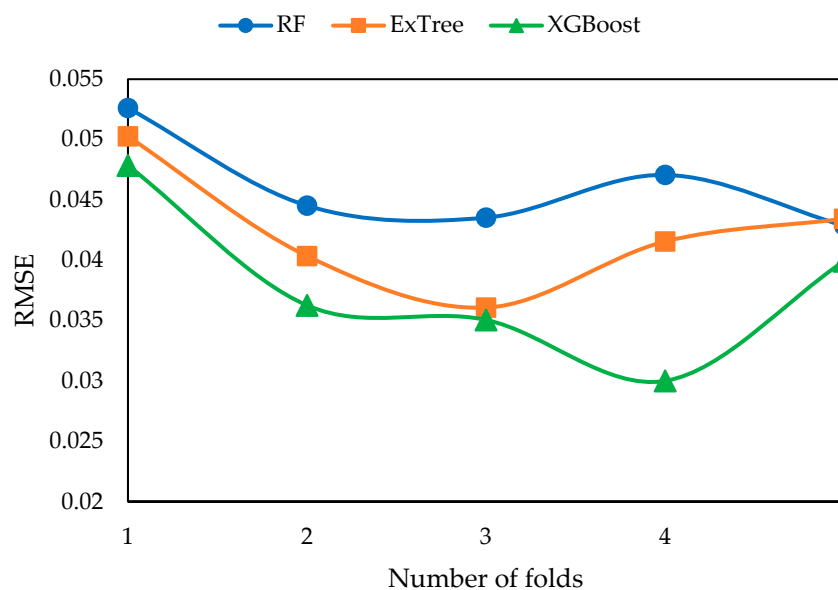
Another observation from Figure 7a–c is the general robustness of the model to experimental variability. Turbulent flow fluctuations, uncertain sediment transport, and vortex interactions around bridge piers contribute to the complexity of local scour phenomena. The stability in performance across folds indicates that the developed models maintain predictive accuracy despite these uncertainties. In summary, Figure 7a–c confirms that the tested ensemble ML models are predictive and generalize effectively. The results indicate that XGBoost is particularly effective in modeling local scour around collar-protected circular piers, where hydraulic behavior is nonlinear. These findings demonstrate the potential of interpretable ensemble machine learning for predicting laboratory-scale final scour depth around collar-protected circular piers. Further validation using independent laboratory and field datasets is required before application to bridge foundation safety assessment.

Table 3. Comparative performance of the XGBoost, Random Forest (RF), and Extra Trees (ET) models on the training, testing, and five-fold cross-validation datasets using different statistical performance metrics.

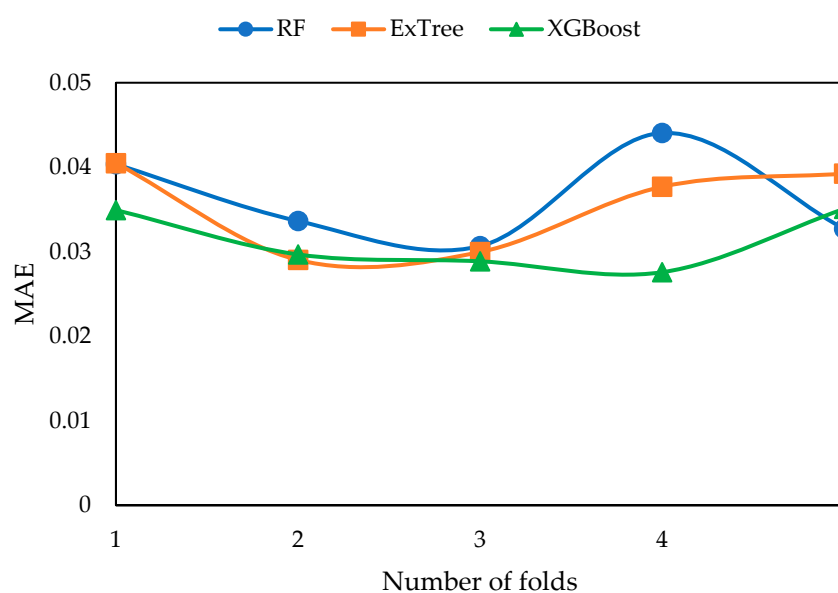
Model	Dataset	R ²	RMSE	KGE
XGBoost	Training	0.987	0.0117	0.982
	Testing	0.976	0.0213	0.955
	5-Fold CV (Mean ± SD)	0.864 ± 0.153	0.0378 ± 0.0065	0.905 ± 0.055
Random Forest	Training	0.96	0.0215	0.93
	Testing	0.942	0.0435	0.889
	5-Fold CV (Mean ± SD)	0.818 ± 0.178	0.0461 ± 0.0039	0.830 ± 0.089
Extra Trees	Training	0.982	0.0191	0.97
	Testing	0.963	0.0403	0.954
	5-Fold CV (Mean ± SD)	0.840 ± 0.172	0.0423 ± 0.0051	0.896 ± 0.052



(a)



(b)



(c)

Figure 7. Results of the five-fold cross-validation used to evaluate the robustness and generalization capability of the tested ensemble machine learning models. The performance metrics across each fold are presented for (a) coefficient of determination (R^2), (b) root mean square error (RMSE), and (c) mean absolute error (MAE).

3.6. Performance Metrics

Figure 8 presents the performance of three key statistical measures, coefficient of determination (R^2), Kling–Gupta efficiency (KGE), and root mean square error (RMSE) for the evaluated ensemble machine learning (ML) models. The figure provides a direct comparison of Random Forest (RF), Extra Trees (ET), and XGBoost in predicting final scour depth for collar-protected circular piers.

The results indicate that the XGBoost model outperforms all other models evaluated. XGBoost achieved an R^2 value of approximately 0.987, a KGE value of 0.982, and an RMSE of nearly (0.0117). These metrics demonstrate the model's strong capability in capturing

local scour processes involving nonlinear hydraulic interactions. The high R^2 and KGE values indicate strong agreement between observed and predicted final scour depths, while the low RMSE reflects minimal prediction error.

The Extra Trees (ET) model also demonstrated strong performance, with a KGE of approximately 0.97 and an RMSE of (0.01914). Notably, ET achieved the moderate RMSE among the models, suggesting medium precision in prediction. These results highlight the potential of tree-based ensemble methods for generalizing nonlinear, multivariate data related to scour.

The Random Forest (RF) model performed comparatively less well, with R^2 approximately 0.96, KGE approximately 0.93, and RMSE around (0.0215). Although RF achieved reasonable accuracy, its higher error rate indicates reduced efficiency in modeling complex hydraulic relationships relative to the ET and XGBoost models.

These results are significant for hydraulic engineering, as local scour prediction involves complex mechanisms of sediment interaction. Traditional empirical equations often fail to capture these nonlinear processes with sufficient accuracy. The ensemble ML models demonstrate strong predictive capability for the laboratory dataset considered in this study.

The figure further demonstrates that interpretable ensemble ML models can achieve satisfactory predictive performance for the relatively small laboratory dataset investigated in this study. This capability is particularly valuable in hydraulic engineering studies where large-scale experimental data collection is challenging and costly. The improved performance of XGBoost and ET indicates that boosting and randomized tree learning methods are well suited for scour modeling applications.

Additionally, the combined use of R^2 , KGE, and RMSE provides a comprehensive assessment of model reliability. R^2 quantifies the strength of correlation, KGE accounts for both bias and accuracy, and RMSE measures the magnitude of prediction error. The consistently high values across all metrics confirm the robustness and practical applicability of the developed models. In summary, Figure 8 demonstrates that the XGBoost and Extra Trees models can accurately and effectively predict scour at collar-protected circular piers. These models provide a promising framework for laboratory-scale scour prediction and may support future research on data-driven approaches for bridge scour assessment. However, additional validation using larger experimental and field datasets is necessary before practical engineering application. Figure 9 demonstrate the relationship between measured and predicted values of the final scour depth.

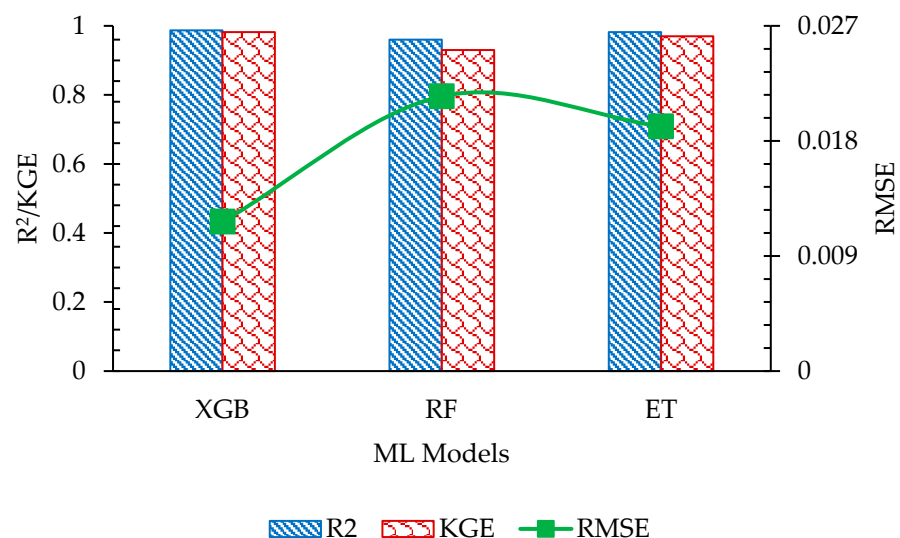


Figure 8. Comparison of the predictive performance of the tested ensemble machine learning models using the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error

(MAE), and Kling–Gupta Efficiency (KGE). Higher R^2 and KGE values together with lower RMSE and MAE indicate superior predictive performance.

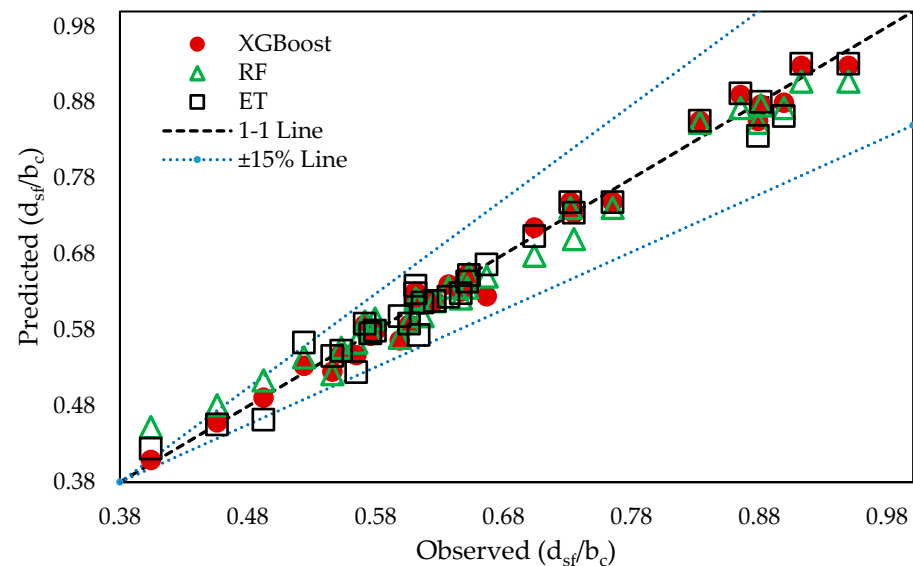


Figure 9. Comparison between the observed and predicted normalized final scour depth (d_{sf}/b_c) obtained using the XGBoost, Random Forest, and Extra Trees models. The 1:1 line and the $\pm 15\%$ error bounds are included to assess the prediction accuracy and agreement with the experimental observations.

3.7. SHAP Analysis

The ensemble machine learning (ML) models developed in this study, including XGBoost, Random Forest (RF), and Extra Trees (ET), were analyzed using SHAP (SHapley Additive exPlanations) as shown in Figure 10a–c. This analysis is essential because it enables quantification of the contribution of each input parameter to the prediction of final scour depth. In hydraulic engineering applications, interpretability is crucial; predictive capability must be complemented by a physical understanding of the dominant hydraulic mechanisms involved. The SHAP analysis was performed using the TreeExplainer algorithm, which is specifically designed for tree-based ensemble models such as Random Forest, Extra Trees, and XGBoost. TreeExplainer efficiently computes exact SHAP values by exploiting the tree structure, thereby quantifying the contribution of each input variable to the predicted final scour depth for individual observations and the overall model.

The magnitude of the SHAP results demonstrates that parameters related to hydraulic intensity significantly influence scour prediction across all three models. Flow condition parameters consistently exhibit large SHAP values, reinforcing that the primary mechanisms affecting local scour development around bridge piers are associated with flow-induced turbulence and bed shear stress. Positive SHAP values generally correspond to increased final scour depths, while negative SHAP values indicate decreased final scour depths due to protective or stabilizing parameters.

Among the models tested, XGBoost exhibits the most structured and focused performance, with fewer random fluctuations and superior capability in handling feature interactions. This observation aligns with the results in Figure 7, where XGBoost achieved a high prediction accuracy ($R^2 = 0.98$) and a low RMSE (0.0117). Furthermore, the SHAP interpretation confirms that the SHAP values produced by the model are physically meaningful rather than being mere statistical artifacts.

The same figure also illustrates the strong negative SHAP values associated with collar-related parameters on final scour depths. This result provides hydraulic validation for

the effectiveness of collars in controlling local scour by reducing downflow motion and diminishing the strength of the pier horseshoe vortex at the base. Experimental observations and classical hydraulic scour protection theory are thus supported, as scour predictions decrease with increasing collar dimensions.

Another major objective of applying SHAP is to identify parameter interaction behavior. The analysis reveals that certain variables contribute differently to scour depending on hydraulic conditions, highlighting the nonlinearity and coupling inherent in scour processes. Such complex behavior cannot be fully captured by current empirical equations, further underscoring the value of interpretable ensemble ML approaches in hydraulic engineering research.

From a hydraulic engineering perspective, the figure provides insight into the physical reliability of the developed ML framework. Advanced ML techniques are often criticized for their ‘black box’ nature. However, this limitation can be addressed through SHAP analysis, which offers clear explanations of each parameter’s impact on scour prediction. Such transparency significantly enhances confidence in applying ML models to real-world bridge foundation design and scour risk assessment.

The SHAP analysis provides an interpretable representation of the relationships learned by the ensemble models, and the observed feature importance is consistent with established hydraulic principles. However, owing to the correlation among some predictors, these results should be interpreted as model-based explanations rather than direct confirmation of the underlying physical mechanisms. The dominance of hydraulic intensity-related variables reflects their direct control over the strength of the downflow and horseshoe vortex system, which governs bed shear stress, sediment entrainment, and the subsequent development of the scour hole. As flow intensity increases, stronger vortical structures promote greater sediment removal around the pier, leading to increased final scour depth. Similarly, the importance of collar-related variables confirms the hydraulic role of collars in modifying the near-pier flow field by interrupting the downward flow and weakening the horseshoe vortex, thereby reducing sediment transport capacity and limiting scour development. In contrast, the relatively smaller SHAP contributions associated with sediment-related variables indicate that, within the investigated experimental conditions, their influence is secondary to that of hydraulic loading and collar geometry. This behavior is primarily attributed to the relatively narrow range of sediment characteristics represented in the dataset and should not be interpreted as a general indication that sediment properties are insignificant under all hydraulic conditions. Overall, the agreement between the SHAP-derived feature importance and the well-established physical processes of local scour demonstrates that the developed XGBoost model captures hydraulically meaningful relationships rather than merely statistical correlations, thereby enhancing the interpretability and engineering reliability of the proposed prediction framework.

Additionally, the consistent ranking of parameters across XGBoost, RF, and ET models indicates stable hydraulic relationships within the dataset. This consistency suggests that the models have learned underlying physical scour mechanisms rather than fitting to random noise present in the smaller dataset.

Although SHAP provides a transparent interpretation of the relative contribution of individual predictors to the ensemble model outputs, the results should not be interpreted as direct evidence of causal hydraulic mechanisms. In particular, the dimensionless variables z/d_{50} and z/b_c exhibit strong correlation because they originate from the same collar-elevation measurements used in the experimental design. Consequently, part of the feature contribution may be shared between these correlated predictors. Therefore, the SHAP analysis is interpreted as explaining the predictive behavior of the trained models rather than establishing independent physical causality. Nevertheless, the overall SHAP

trends remain consistent with established hydraulic understanding, indicating that collar elevation and collar geometry are influential factors controlling final scour depth.

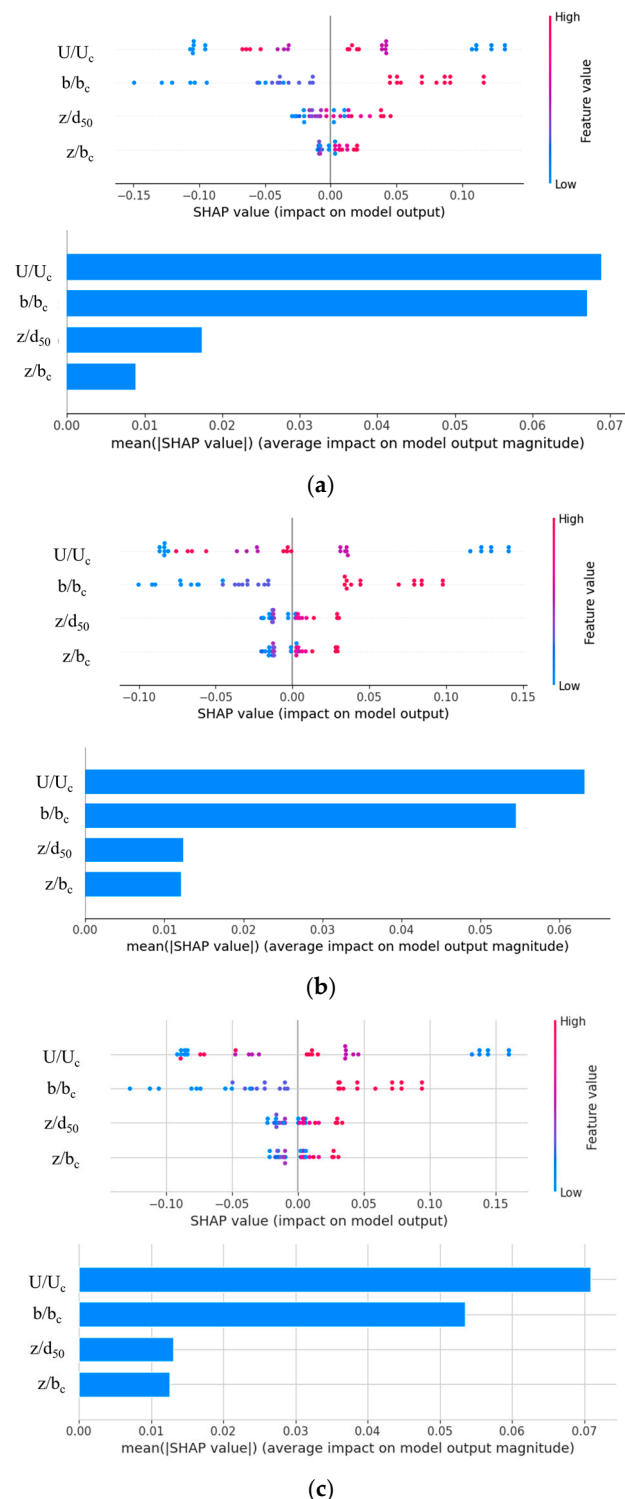


Figure 10. SHAP (SHapley Additive exPlanations) summary plots for the tested ensemble machine learning models: (a) XGBoost, (b) Random Forest, and (c) Extra Trees. The plots illustrate the relative importance of each input variable and its positive or negative contribution to the prediction of normalized final scour depth.

3.8. Prediction by Developed Equation

The regression equation was evaluated using four widely recognized performance indicators: coefficient of determination (R^2), Kling–Gupta Efficiency (KGE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). Comparison of observed and predicted relative final scour depths (d_{sf}/b_c) showed that the regression model achieved an R^2 value of 0.757, indicating that 75.7% of the variability in observed final scour depths was explained by the proposed dimensionless equation (Figure 11). This result demonstrates that the regression model effectively captured the primary hydraulic and geometric mechanisms governing equilibrium scour development around collar-protected circular piers. The remaining unexplained variance suggests that certain nonlinear interactions inherent in the scour process were not fully represented by the power-law formulation.

The RMSE value of 0.0678 reflects a small average prediction error between measured and estimated dimensionless scour depths. Similarly, the MAE value of 0.0547 indicates that the equation deviated from observed values by approximately 5.5% on average in dimensionless terms. The close agreement between RMSE and MAE suggests that model predictions were not significantly affected by extreme outliers and that errors were distributed relatively evenly across the dataset.

The regression model produced a KGE value of 0.697, indicating satisfactory overall performance in reproducing the correlation structure, variability, and mean characteristics of the observed data. Although this KGE value was lower than those achieved by the ensemble machine learning models developed in this study, it demonstrates acceptable predictive capability for a simple and interpretable regression equation. These results indicate that the proposed dimensionless regression equation serves as a practical and physically meaningful tool for preliminary estimation of final scour depth around collar-protected bridge piers. More advanced machine learning approaches may provide improved predictive accuracy under complex hydraulic conditions.

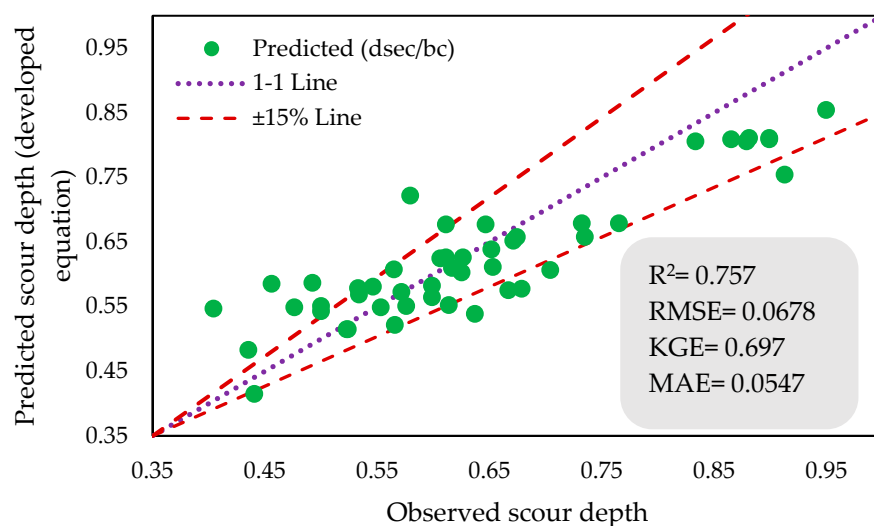


Figure 11. Comparison between the observed and predicted normalized final scour depth (d_{sec}/b_c) obtained using the proposed dimensionless regression equation. The statistical performance metrics (R^2 , RMSE, MAE, and KGE) and the $\pm 15\%$ error bounds are included to evaluate the predictive capability of the developed equation.

3.9. Comparison with Existing Empirical Equations

To demonstrate the effectiveness of the proposed models, their predictive performance was compared with two widely used empirical equations developed by Kumar et al. [28] and Pandey et al. [26]. For consistency with the response variable adopted in the

present study, predictions from both empirical equations were converted to the dimensionless final scour depth (d_{sf}/b_c) and evaluated using the same 48 experimental observations (Figure 12). As summarized in Table 4, the empirical equation proposed by Kumar et al. [28] produced reasonable agreement with the measured data ($R^2 = 0.848$, $RMSE = 0.0535$, and $KGE = 0.818$), whereas the equation developed by Pandey et al. [26] showed comparatively lower predictive accuracy ($R^2 = 0.516$, $RMSE = 0.0956$, and $KGE = 0.579$). In contrast, the proposed regression equation and the ensemble machine learning models achieved substantially higher predictive performance. Among the machine learning models, XGBoost produced the highest overall accuracy, followed by Extra Trees and Random Forest, indicating that tree-based ensemble learning is more effective in capturing the complex nonlinear interactions between hydraulic, sediment, and geometric parameters than conventional empirical formulations. These results demonstrate the advantages of the proposed data-driven framework and provide a direct benchmark against established empirical scour prediction equations. These results demonstrate that the proposed machine learning framework not only outperforms existing empirical equations but also provides greater robustness across the investigated hydraulic conditions.

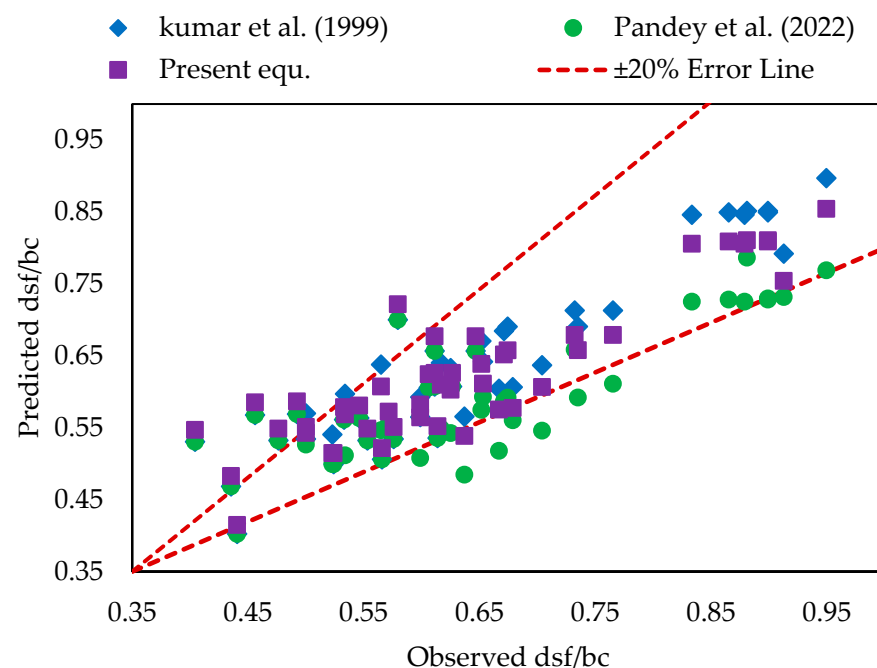


Figure 12. Comparison between the observed and predicted normalized final scour depths (d_{sf}/b_c) obtained using the empirical equations of Kumar et al. [28] and Pandey et al. [26], together with the proposed regression equation. The dashed red lines represent the $\pm 20\%$ error bounds used to assess the prediction accuracy of the different empirical formulations.

Table 4. Comparison of the proposed regression and ensemble machine learning models with existing empirical equations for predicting dimensionless final scour depth (d_{sf}/b_c).

Model	Training R^2	Testing R^2	Training RMSE	Testing RMSE	Training KGE	Testing KGE
XGBoost	0.987	0.976	0.0117	0.0213	0.982	0.955
Random Forest	0.96	0.942	0.0215	0.0435	0.93	0.889
Extra Trees	0.982	0.963	0.0191	0.0403	0.97	0.954
Present Equation	0.886	0.475	0.0628	0.1073	0.746	0.675
Kumar et al. [28]	0.897	0.498	0.0387	0.0946	0.864	0.834
Pandey et al. [26]	0.826	0.465	0.0843	0.1314	0.672	0.65

3.10. Taylor Diagram

The Taylor diagram provides a comprehensive graphical comparison of the predictive performance of the developed ensemble learning models by simultaneously evaluating the agreement between simulated and measured final scour depths in terms of standard deviation, correlation, and the centered root-mean-square difference. As shown in Figure 13, all three ensemble learning models are located close to the reference point representing the measured data, indicating that they successfully reproduced the statistical characteristics of the experimental observations. Among the evaluated models, XGBoost is closest to the reference point and exhibits the smallest centered root-mean-square difference, demonstrating the best overall agreement with the measured final scour depths. The Extra Trees (ET) model also shows excellent predictive capability and closely matches the observed variability, although its deviation from the reference point is slightly greater than that of XGBoost. The Random Forest (RF) model provides satisfactory predictions but is located comparatively farther from the reference point, indicating a relatively larger prediction error and a slight underestimation of the variability in the observed final scour depths. Overall, the Taylor diagram confirms the superior predictive performance and robustness of the XGBoost model, while also demonstrating that all three ensemble learning approaches are capable of accurately representing the laboratory clear-water scour data around collar-protected circular bridge piers.

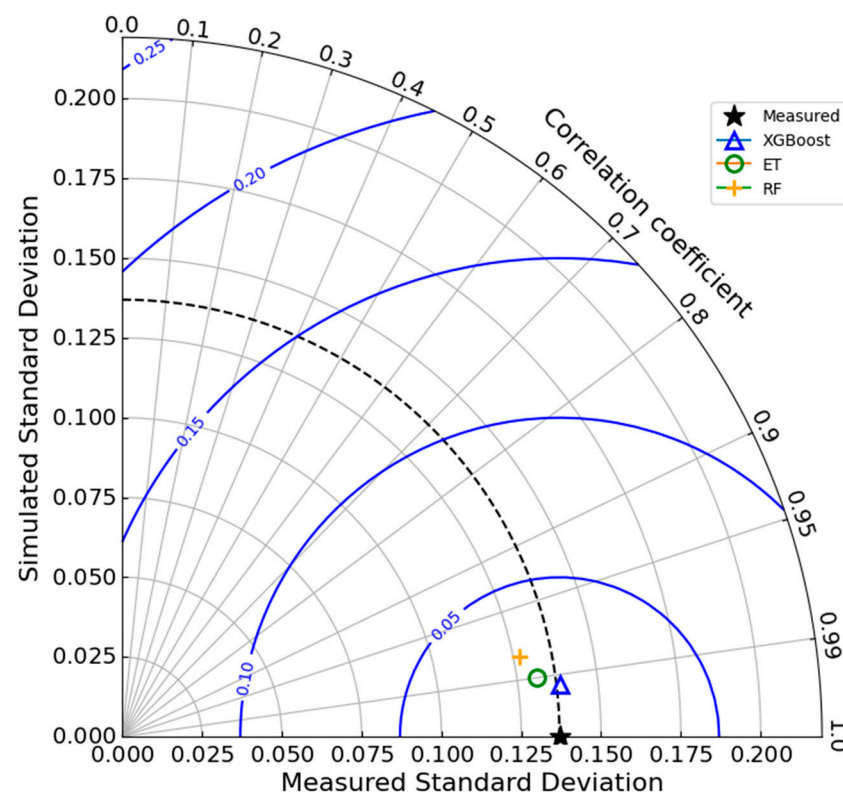


Figure 13. Taylor diagram illustrating the comparative performance of the XGBoost, Extra Trees (ET), and Random Forest (RF) models in predicting the dimensionless final scour depth (d_{sf}/b_c), based on the agreement between predicted and measured experimental observations in terms of standard deviation, correlation, and centered root mean square difference.

4. Discussion

The results demonstrate that interpretable ensemble machine learning models can accurately predict local scour around collar-protected circular bridge piers, even when trained on a relatively small experimental dataset. The strong performance of the devel-

oped models, particularly XGBoost, underscores the effectiveness of boosting-based algorithms in modeling scour due to their ability to capture complex nonlinear interactions inherent in scour processes. Managing local scour requires balancing hydraulic loading, sediment mobility, pier geometry, and protective measures. Conventional empirical equations often oversimplify these relationships and fail to represent the coupled effects among governing parameters. Consequently, the present findings align with recent studies, indicating that ensemble machine learning tools offer a more flexible and accurate alternative for scour prediction.

Correlation and scatter analyses indicate that parameters related to hydraulic intensity have a strong positive correlation with final scour depth, while collar-related parameters exhibit a strong negative correlation. These results are consistent with classical scour theory, which attributes scour formation to downflow and horseshoe vortices generated at the upstream side of bridge piers. Increased flow intensity leads to higher bed shear stress and greater turbulence, intensifying sediment entrainment and the formation of scour holes. Pandey et al. [26] observed that increasing approach velocity results in greater final scour depth, regardless of collar protection.

The findings further validate the effectiveness of collars as scour countermeasures. Larger collars, or those positioned at optimal elevations above the sediment bed, are associated with reduced final scour depth [33,34], as indicated by negative correlations and SHAP values for collar-related variables. This result aligns with numerous experimental studies that demonstrate significant scour reduction through the use of collars. Moghanloo et al. [35] and Memar et al. [36] reported that collars counteract the downflow component of the flow approaching the pier nose, thereby reducing the intensity of horseshoe vortices. More recently, Kassem et al. [37] found that optimized collar geometry significantly reduces final scour depth under clear-water conditions. The present results extend these findings by quantitatively demonstrating, through SHAP analysis, that collar characteristics play a prominent role in reducing scour height within predictive machine learning models.

A notable outcome of this study is that XGBoost outperformed both Random Forest and Extra Trees. While all ensemble algorithms demonstrated strong performance, XGBoost achieved the highest R^2 and KGE values and the lowest prediction errors. Previous scour studies have reported similar findings. For example, Choi et al. [38] and Khan and Ismael [39] demonstrated that boosting-based methods outperform regression-based and individual machine learning methods due to their sequential approach to reducing residual errors. The present study confirms this advantage, even with a dataset comprising only 48 experimental observations. Attention should be given to the effectiveness of ensemble learning with small sample sizes. Hydraulic experiments are typically labor-intensive, time-consuming, and costly, which limits the availability of large datasets compared to other machine learning applications [40]. Under these constraints, the reliability of data-driven approaches may be called into question. However, recent research in hydraulic engineering has shown that tree-based ensemble methods can be dependable when supported by robust validation procedures [41,42]. The developed models did not overfit the training data, as evidenced by stable performance across all folds in the five-fold cross-validation conducted in this study. This finding is especially significant, as it demonstrates the feasibility of using interpretable machine learning methods in contexts where large datasets are not available. The principal contribution of this study lies in demonstrating that interpretable ensemble learning can effectively model nonlinear scour mechanisms within a controlled experimental domain while establishing a methodological framework that can be extended and validated using larger and more diverse datasets in future investigations.

The developed models established credibility by demonstrating consistency with established hydraulic principles, as further supported by SHAP analysis. The ranking of parameter importance was similar across XGBoost, Random Forest, and Extra Trees, suggesting that the models primarily captured statistical rather than physical relationships. Explainable artificial intelligence (XAI) is increasingly emphasized in engineering applications, where interpretability is essential for widespread adoption. Recent studies in hydraulics, has shown that SHAP analysis enables interpretation of complex machine learning models without compromising predictive performance [43]. The present investigation extends this perspective by demonstrating that explainability tools can clarify the combined effects of hydraulic loading and collar geometry on final scour depth.

These results have practical implications for the design of bridge foundations and the management of scour risk. Local scour remains a leading cause of bridge failure globally, highlighting the need for reliable predictive tools to support decision-making during design, inspection, and maintenance. The findings indicate that interpretable ensemble machine learning models can complement existing empirical methods by enhancing predictive accuracy while maintaining consistency with physical laws. Additionally, identifying the most influential variables in scour containment can assist engineers in prioritizing key factors during bridge evaluation. In summary, this study demonstrates that interpretable ensemble machine learning constitutes an effective approach for predicting scour around collar-protected bridge piers. The proposed model integrates predictive capability with interpretability, facilitating its application in real-world decision-making for the design of resilient bridge infrastructure.

Practical Implications and Research Limitations

The findings of this research have several potential applications in bridge engineering and scour risk management. Local scour is a primary factor contributing to bridge foundation instability and failure during flood events. Accurate estimation of final scour depth is therefore critical for ensuring the safety and resilience of transportation infrastructure, and the development of reliable predictive tools is essential to achieve this goal. The machine learning model developed in this study utilizes an interpretable ensemble approach, providing researchers with an alternative to the traditional empirical equations, which are often limited to experimental conditions and may not adequately capture complex hydraulic interactions. The performance of XGBoost and Extra Trees demonstrates that ensemble learning methods can serve as a viable supplement for the design and evaluation of collar-protected bridge piers. Additionally, SHAP analysis enables the identification of the relative importance of hydraulic and collar-related parameters influencing scour, thereby enhancing model transparency. This approach assists engineers in determining optimal collar dimensions, prioritizing variables during bridge inspections, and focusing efforts on improving scour mitigation strategies. The proposed framework could also be integrated into decision support systems for bridge asset management, particularly under increasingly uncertain hydrological conditions associated with climate change.

Despite these promising results, several limitations should be acknowledged. The developed models were trained on a relatively small dataset comprising 48 laboratory observations collected under clear-water scour conditions. Although rigorous cross-validation procedures were employed, the limited sample size may restrict the generalizability of the findings. Furthermore, the database was constructed using experiments involving only circular piers with collar protection, which raises uncertainty regarding the model's applicability to other protection methods, pier geometries, sediment characteristics, and flow regimes. Future research should incorporate more extensive data from multiple laboratory studies and field measurements to improve the model's transferability. Integrating uncertainty quantification methodologies and physics-informed machine

learning techniques may further enhance the reliability and practical applicability of scour prediction models in engineering practice.

The findings of this study should be interpreted in the context of certain data-related limitations. The proposed machine learning models were developed using a controlled laboratory dataset obtained from a single experimental campaign. While this ensured a high degree of data consistency and experimental reliability, it limited the diversity of hydraulic and geometric conditions represented in the model training data. Consequently, although the developed models demonstrated excellent predictive accuracy within the investigated domain, their applicability to a wider range of bridge scour conditions requires further validation. Future research should incorporate larger, multi-source experimental datasets from different laboratories, complemented by prototype field measurements, to enhance the robustness, reliability, and generalizability of the proposed models across diverse hydraulic environments.

5. Conclusions

This research examined the applicability of interpretable ensemble machine learning methods for predicting final scour depth around collar-protected circular bridge piers, utilizing a limited experimental dataset of 48 laboratory observations. The study addressed the shortcomings of conventional empirical equations, which frequently fail to capture the complex interactions among hydraulic conditions, sediment properties, and scour mitigation measures. By combining predictive accuracy with model interpretability, this work contributes to the development of transparent data-driven approaches for laboratory-scale bridge scour assessment.

Three ensemble learning algorithms, XGBoost (XGB), Random Forest (RF), and Extra Trees (ET), were developed and evaluated using five-fold cross-validation. The results indicated that all models accurately predicted final scour depth, despite the limited dataset size. XGBoost consistently achieved the highest coefficient of determination and Kling–Gupta Efficiency, as well as the lowest prediction errors. This superior performance underscores the effectiveness of boosting-based algorithms in capturing the nonlinear relationships present in scour processes. These results demonstrate that reliable machine learning models can be developed for predicting laboratory-scale final scour depth, even with limited experimental data, when appropriate validation strategies are implemented.

The use of SHAP analysis facilitated interpretation of the developed models and offered insights into the relative importance of governing variables. Parameters related to hydraulic intensity had the most significant positive impact on scour development, while collar-related parameters reduced final scour depth, confirming the effectiveness of collar systems as mitigation measures. The consistent parameter rankings across all ensemble models suggest that the developed models learned relationships consistent with the laboratory observations. However, because some predictors are correlated, the SHAP results should be interpreted as explaining model behavior rather than demonstrating the underlying physical mechanisms. This level of transparency enhances confidence in the interpretation of machine learning predictions within the investigated laboratory conditions and may support future research on explainable data-driven scour prediction.

A dimensionless regression equation was also developed to offer a simple and interpretable alternative for scour estimation. This equation demonstrated acceptable predictive capability, with R^2 , KGE, RMSE, and MAE values of 0.757, 0.697, 0.0678, and 0.0547, respectively. Although its performance was lower than that of the ensemble learning models, the regression equation remains a useful and interpretable alternative for preliminary laboratory-scale scour estimation and comparative analysis where simplicity is preferred.

The proposed XGBoost-based framework provides an accurate and interpretable tool for predicting final scour depth under laboratory clear-water conditions for collar-pro-

tected circular bridge piers within the range of the investigated experimental data. Although the developed models demonstrate excellent predictive capability, their applicability to prototype bridge piers, complex field conditions, and different scour environments should be treated with caution. Future investigations should further evaluate the developed models using larger datasets obtained under different sediment gradations, longer-duration experiments, varying hydraulic conditions, and field-scale measurements before broader engineering application is considered. Additionally, incorporating uncertainty quantification and physics-informed machine learning techniques may further enhance the robustness and applicability of scour prediction models in engineering practice.

Author Contributions: Conceptualization, N.M., S.I., M.A.M. and M.A.S.; methodology, N.M. and S.I.; software, M.A.M. and M.A.S.; validation, M.A.S., M.A.M. and N.M.; formal analysis, N.M., S.I., and M.A.S.; investigation, N.M. and M.A.S.; resources, M.A.M. and S.I.; data curation, S.I., M.A.S. and M.A.M.; writing—original draft preparation, N.M. and M.A.M.; writing—review and editing, N.M., S.I., M.A.S., and M.A.M.; visualization, N.M., S.I., and M.A.M.; supervision, M.A.M. and M.A.S.; funding acquisition, M.A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-DDRSP 2602).

Data Availability Statement: The dataset used in this study was obtained from the research conducted by Pandey et al. [26].

Acknowledgments: During the preparation of this review paper, generative AI tools (ChatGPT-5.5) and Premium Grammarly were used to improve language proficiency. The author (s) take full responsibility for the content of the paper upon publication.

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

References

1. Briaud, J.-L.; Gardoni, P.; Yao, C. Statistical, Risk, and Reliability Analyses of Bridge Scour. *J. Geotech. Geoenvironmental Eng.* **2014**, *140*, 04013011. [https://doi.org/10.1061/\(ASCE\)GT.1943-5606.0000989](https://doi.org/10.1061/(ASCE)GT.1943-5606.0000989).
2. Shahriar, A.R.; Ortiz, A.C.; Montoya, B.M.; Gabr, M.A. Bridge Pier Scour: An overview of factors affecting the phenomenon and comparative evaluation of selected models. *Transp. Geotech.* **2021**, *28*, 100549. <https://doi.org/10.1016/j.trgeo.2021.100549>.
3. Pasha, G.A.; Murtaza, N.; Hamidifar, H. Collar systems for bridge pier scour protection. *Innov. Infrastruct. Solut.* **2025**, *10*, 557. <https://doi.org/10.1007/s41062-025-02394-z>.
4. Pizarro, A.; Manfreda, S.; Tubaldi, E. The Science behind Scour at Bridge Foundations: A Review. *Water* **2020**, *12*, 374. <https://doi.org/10.3390/w12020374>.
5. Salauddin, M.; Shaffrey, D.; Habib, M.A. Data-driven approaches in predicting scour depths at a vertical seawall on a permeable shingle foreshore. *J. Coast. Conserv.* **2023**, *27*, 18. <https://doi.org/10.1007/s11852-023-00948-w>.
6. Tabasi, N.; Fereshtehpour, M.; Roghani, B. A review of flood risk assessment frameworks and the development of hierarchical structures for risk components. *Discov. Water* **2025**, *5*, 10. <https://doi.org/10.1007/s43832-025-00193-2>.
7. Khan, Z.U.; Ahmed, A.; Valyrakis, M.; Pasha, G.A.; Farooq, R.; Murtaza, N.; Khan, D. Effectiveness of Collars and Hooked-Collars in Mitigating Scour around Different Abutment Shapes. *Water* **2024**, *16*, 2550. <https://doi.org/10.3390/w16172550>.
8. Tang, Z.; Melville, B.; Shamseldin, A.; Guan, D.; Singhal, N.; Yao, Z. Experimental study of collar protection for local scour reduction around offshore wind turbine monopile foundations. *Coast. Eng.* **2023**, *183*, 104324. <https://doi.org/10.1016/j.coastaleng.2023.104324>.
9. Affandy, E.; Pallu, M.S.; Maricar, F.; Bakri, B. An Experimental Study on Local Scour Protection for Round Slotted Bridge Piers with Collar Variations. *Eng. Technol. Appl. Sci. Res.* **2025**, *15*, 29685–29695. <https://doi.org/10.48084/etasr.14582>.
10. Liu, Z.; Zhang, S.; Wu, C.; Shen, D.; Zhang, J.; Song, Z. Experimental study on scour protection method based on EICP with anti-permeation covers. *Case Stud. Constr. Mater.* **2025**, *22*, e04270. <https://doi.org/10.1016/j.cscm.2025.e04270>.

11. Murtaza, N.; Pasha, G.A. Ensemble Learning-Driven Flood Risk Management Using Hybrid Defense Systems. *AI* **2025**, *7*, 2. <https://doi.org/10.3390/ai7010002>.
12. Murtaza, N.; Khan, D.; Rezzoug, A.; Khan, Z.U.; Benzougagh, B.; Khedher, K.M. Scour depth prediction around bridge abutments: A comprehensive review of artificial intelligence and hybrid models. *Phys. Fluids* **2025**, *37*, 021306. <https://doi.org/10.1063/5.0244974>.
13. Khan, Z.U.; Khan, D.; Murtaza, N.; Pasha, G.A.; Alotaibi, S.; Rezzoug, A.; Benzougagh, B.; Khedher, K.M. Advanced Prediction Models for Scouring Around Bridge Abutments: A Comparative Study of Empirical and AI Techniques. *Water* **2024**, *16*, 3082. <https://doi.org/10.3390/w16213082>.
14. Eini, N.; Janizadeh, S.; Bateni, S.M.; Jun, C.; Heggy, E.; Kirs, M. Predicting final scour depth around non-circular bridge piers with shallow foundations using hybrid explainable machine learning methods. *Results Eng.* **2024**, *24*, 103492. <https://doi.org/10.1016/j.rineng.2024.103492>.
15. Chou, J.-S.; Nguyen, N.-M. Scour depth prediction at bridge piers using metaheuristics-optimized stacking system. *Autom. Constr.* **2022**, *140*, 104297. <https://doi.org/10.1016/j.autcon.2022.104297>.
16. Pasha, G.A.; Ghumman, A.R.; Khalid, M.A.; Murtaza, N.; Ahmed, A.; Alodah, A.; AlSalamah, I.S.; Shafiquzzaman, M.D.; Al-Dawodi, H.; Haider, H. Evaluating Various Techniques for Total Sediment Concentration Prediction: Case of Marala Ravi Link Canal in Pakistan. *Adv. Civ. Eng.* **2025**, *2025*, 6656721. <https://doi.org/10.1155/adce/6656721>.
17. Parisouj, P.; Jun, C.; Bateni, S.M.; Heggy, E.; Band, S.S. Machine learning models coupled with empirical mode decomposition for simulating monthly and yearly streamflows: A case study of three watersheds in Ontario, Canada. *Eng. Appl. Comput. Fluid Mech.* **2023**, *17*, 2242445. <https://doi.org/10.1080/19942060.2023.2242445>.
18. Zuhairi, A.H.; Yakub, F.; Omar, M.; Sharifuddin, M.; Razak, K.A.; Faruq, A.; Wijaya, A.A. Performance Analysis of Tree-Based Ensemble Machine Learning Model for Flood Forecasting in Tropical Regions. *IEEE Access* **2025**, *13*, 200840–200860. <https://doi.org/10.1109/ACCESS.2025.3632033>.
19. Reina-Jiménez, P.; Martínez-Ballesteros, M.; Riquelme, J.C. Tree-based vs. deep learning in time series: A prescriptive framework for performance-driven model selection. *Appl. Soft Comput.* **2026**, *195*, 114983. <https://doi.org/10.1016/j.asoc.2026.114983>.
20. Rezzoug, A.; Khan, A.; Murtaza, N.; Rizvi, S.A.S. Predicting flood energy reduction in vegetated open channel: Comparative assessment of hybrid artificial intelligence techniques. *Eng. Appl. Artif. Intell.* **2025**, *159*, 111756. <https://doi.org/10.1016/j.engappai.2025.111756>.
21. Murtaza, N.; Akbar, Z.; Alrowais, R.; Iqbal, S.; Pasha, G.A.; Alquraish, M.; Bashir, M.T. Machine Learning and SHAP-Based Prediction of Tip Velocity Around Spur Dikes Using a Small-Scale Experimental Dataset. *Water* **2025**, *18*, 26. <https://doi.org/10.3390/w18010026>.
22. Ajmal, M.; Khan, T.; Kim, T.-W. A CN-Based Ensembled Hydrological Model for Enhanced Watershed Runoff Prediction. *Water* **2016**, *8*, 20. <https://doi.org/10.3390/w8010020>.
23. Fan, J.; Li, R.; Zhao, M.; Pan, X. A BiLSTM-Based Hybrid Ensemble Approach for Forecasting Suspended Sediment Concentrations: Application to the Upper Yellow River. *Land* **2025**, *14*, 1199. <https://doi.org/10.3390/land14061199>.
24. Shojaezadeh, S.A.; Al-Wardy, M.; Nikoo, M.R. Suspended sediment load modeling using Hydro-Climate variables and Machine learning. *J. Hydrol.* **2024**, *633*, 130948. <https://doi.org/10.1016/j.jhydrol.2024.130948>.
25. Abioye, S.O.; Babatunde, Y.O.; Abikoye, O.A.; Shaibu, A.N.; Bankole, B.J. Optimized machine learning algorithms with SHAP analysis for predicting compressive strength in high-performance concrete. *AI Civ. Eng.* **2025**, *4*, 16. <https://doi.org/10.1007/s43503-025-00061-x>.
26. Pandey, M.; Pu, J.H.; Pourshahbaz, H.; Khan, M.A. Reduction of scour around circular piers using collars. *J. Flood Risk Manag.* **2022**, *15*, e12812. <https://doi.org/10.1111/jfr3.12812>.
27. Melville, B.W.; Chiew, Y.-M. Time Scale for Local Scour at Bridge Piers. *J. Hydraul. Eng.* **1999**, *125*, 59–65. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1999\)125:1\(59\)](https://doi.org/10.1061/(ASCE)0733-9429(1999)125:1(59)).
28. Kumar, V.; Raju, K.G.R.; Vittal, N. Reduction of Local Scour around Bridge Piers Using Slots and Collars. *J. Hydraul. Eng.* **1999**, *125*, 1302–1305. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1999\)125:12\(1302\)](https://doi.org/10.1061/(ASCE)0733-9429(1999)125:12(1302)).
29. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. <https://doi.org/10.1023/A:1010933404324>.
30. Tyralis, H.; Papacharalampous, G.; Langousis, A. A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources. *Water* **2019**, *11*, 910. <https://doi.org/10.3390/w11050910>.
31. Niazkar, M.; Menapace, A.; Brentan, B.; Piraei, R.; Jimenez, D.; Dhawan, P.; Righetti, M. Applications of XGBoost in water resources engineering: A systematic literature review (Dec 2018–May 2023). *Environ. Model. Softw.* **2024**, *174*, 105971. <https://doi.org/10.1016/j.envsoft.2024.105971>.

32. Mame, M.; Huang, S.; Li, C.; Zhou, J. Application of Extra-Trees Regression and Tree-Structured Parzen Estimators Optimization Algorithm to Predict Blast-Induced Mean Fragmentation Size in Open-Pit Mines. *Appl. Sci.* **2025**, *15*, 8363. <https://doi.org/10.3390/app15158363>.
33. Gupta, L.K.; Pandey, M.; Raj, P.A. Impact of airfoil collar on scour reduction around the bridge pier. *Ocean Eng.* **2023**, *290*, 116271. <https://doi.org/10.1016/j.oceaneng.2023.116271>.
34. Shahhosseini, M.; Yue, S.; Xu, J.; Zhang, M.; Yu, G. Scour depth at a pile with a collar at bed level in clear-water current. *Ocean Eng.* **2025**, *320*, 120319. <https://doi.org/10.1016/j.oceaneng.2025.120319>.
35. Moghanloo, M.; Vaghefi, M.; Ghodsian, M.; Kisi, O. The influence of collar parameters on local scour mechanism around the circular pier at the bend. *Appl. Water Sci.* **2024**, *14*, 194. <https://doi.org/10.1007/s13201-024-02257-5>.
36. Memar, S.; Zounemat-Kermani, M.; Beheshti, A.; Rahimpour, M.; De Cesare, G.; Schleiss, A.J. Influence of collars on reduction in scour depth at two piers in a tandem configuration. *Acta Geophys.* **2020**, *68*, 229–242. <https://doi.org/10.1007/s11600-019-00393-0>.
37. Kassem, H.; El-Masry, A.A.; Diab, R. Influence of collar's shape on scour hole geometry at circular pier. *Ocean Eng.* **2023**, *287*, 115791. <https://doi.org/10.1016/j.oceaneng.2023.115791>.
38. Choi, J.; Kim, J.; Kwon, S.; Kim, T. Probabilistic Prediction of Local Scour at Bridge Piers with Interpretable Machine Learning. *Water* **2025**, *17*, 3574. <https://doi.org/10.3390/w17243574>.
39. Khan, A.; Ismael, D. Interpretable machine learning for bridge-pier scour prediction and flood resilience. *Front. Built Environ.* **2026**, *11*, 1731114. <https://doi.org/10.3389/fbuil.2025.1731114>.
40. Cai, M.; Wang, H.; Lv, X.; Liu, Z.; Chen, Y. Deep learning-based defect identification in hydraulic structures: A comprehensive review. *KSCE J. Civ. Eng.* **2026**, *30*, 100410. <https://doi.org/10.1016/j.kscej.2025.100410>.
41. Ghobadi, F.; Kang, D. Application of Machine Learning in Water Resources Management: A Systematic Literature Review. *Water* **2023**, *15*, 620. <https://doi.org/10.3390/w15040620>.
42. Ghosh, S.; Borse, D.; Padilam, L.R.; Jose, D.M.; Kondapalli, S.; Sawicz, K.A.; Chinnayakanahalli, K.; Chowdhary, H. *Application of XGBoost in Flood Modeling*; Springer Nature: Singapore, 2025; pp. 217–233. https://doi.org/10.1007/978-981-97-9168-2_16.
43. Lamane, H.; Mouhir, L.; Moussadek, R.; Baghdad, B.; Kisi, O.; El Bilali, A. Interpreting machine learning models based on SHAP values in predicting suspended sediment concentration. *Int. J. Sediment Res.* **2025**, *40*, 91–107. <https://doi.org/10.1016/j.ijsrc.2024.10.002>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.