Numerical Methods in Civil Engineering, 10-3 (2026) 12-24

Numerical Methods in Civil Engineering

Journal Homepage: https://nmce.kntu.ac.ir/

An Interpretable Machine Learning Approach for Predicting Bearing Capacity of Driven Piles Using SHAP

Seyed Emad Miri*, Seyed Amin Mousavi Shaarbaf*, Hamid Mohammadnezhad**

ARTICLE INFO

RESEARCH PAPER

Article history:
Received:
October 2025
Revised:
December 2025
Accepted:
February 2026

Keywords:
Driven pile
Bearing capacity
Flap Number
Machine learning
Interpretability
SHAP

Abstract:

Determination of the ultimate pile bearing capacity is still one of the major concerns in geotechnical engineering due to the complex interaction between soil and structure. This study employs interpretable machine learning models to provide precise predictions of pile capacities while identifying the role of the key design variables. A detailed data set of 100 steel and concrete piles is evaluated by including eight important design variables: effective pile length, cross-sectional area, Flap number, drained cohesion, drained soil friction angle, effective unit weight of soil, pile–soil friction angle, and pile material. Prediction models for the pile capacities are established using the Random Forest, XGBoost, CatBoost, and Extra Trees algorithms, which are validated through a strict 5-fold cross-validation. The results show that the Extra Trees algorithm is the most stable and has the highest predictive capability, with a coefficient of determination (R^2) of 0.95 ± 0.03 and RMSE of 1806 ± 999. Furthermore, SHapley Additive exPlanations (SHAP) analysis is performed to calculate the importance of the design parameters, indicating that effective pile length, cross-sectional area, and Flap number are the major contributing factors. This reveals that the proposed unique combination of Flap number with cutting-edge machine learning analysis is an accurate, clear, and viable process for pile capacities.

1. Introduction

Piles are employed as foundation elements to transfer structural loads to soil strata. Their efficiency in functioning directly affects the performance of the superstructure built over them. The bearing capacity of piles is one of the most important geotechnical aspects for their design and performance analysis from a geotechnical perspective, and the prediction of this capacity is a significant and complex problem in the discipline of geotechnical engineering [1].

Accurate estimation of pile bearing capacity is a critical element in deep foundation design. It is common practice to use a unified bearing capacity equation for piles under similar geotechnical conditions; however, this approach should be used cautiously. Also, due to the incomplete understanding and inherent complexity involved in pile–soil interaction, many proposed methods face limitations when it decades, several methods, including empirical, numerical, and analytical methods, have been suggested for estimating pile bearing capacity [2–5]. Yet, the accuracy and consistency of the estimated capacity are still of great concern. Most of the existing methods simplify the problem by making assumptions about the factors that affect pile capacity. Because of the complexity of the pile–soil behavior, none of the suggested methods offer notable consistency in the accurate prediction of pile behavior. For example, the estimated pile capacity using the well-known Meyerhof formula may differ from values predicted by semi-empirical methods [6]. Also, some of the simplifying assumptions in these methods may impose certain limitations. Thus, a more direct and reliable method of pile capacity determination is through pile static load testing. However, the high cost and time requirements are the major limitations of performing static load tests.

Dynamic tests are faster and cheaper than static load tests. However, several PDA tests are typically required for each project. With the expense of each dynamic load test,

* *BSc, Faculty of Civil, Water and Environmental Engineering, Shahid Beheshti University, Tehran, Iran.*
** *Corresponding author: Assistant Professor, Faculty of Civil, Water and Environmental Engineering, Shahid Beheshti University, Tehran, Iran.*
Email: h_mohammadnezhad@sbu.ac.ir

https://doi.org/10.61882/NMCE.2510.1102

reducing the number of required tests would be advantageous and could significantly reduce overall project expenses. Therefore, alternative approaches that can overcome the limitations of existing methods while still yielding sufficient accuracy for pile bearing capacity prediction are required.

During the recent few years, the application area of data-driven approaches in geotechnical engineering has expanded significantly, thereby revealing their flexibility in dealing with complex issues. Various reviews have emphasized the usefulness of machine learning approaches for modeling the dynamic properties of soils, with a considerable success in identifying the shear-wave velocity [7, 8] as well as the small-strain shear modulus of sandy soils [9]. In underground construction techniques, these approaches have been utilized successfully for predicting the stability of rock tunnel headings [10] and the behavior of shallow tunnels [11]. In addition, novel applications have appeared in the area of soil improvement techniques [12], the prediction of bearing capacity of shallow foundations [13], and the prediction of settlement of shallow foundations using CPT records [14]. Successful applications of such approaches in various areas of geotechnical engineering clearly indicate the great potential of machine learning approaches in dealing with non-linear engineering problems.

Due to the large number of parameters that influence the bearing capacity of the pile, efficient computational methods are required for the analysis of complex interactions among various parameters. In the last two decades, there has been a growing use of non-aprioristic methods such as artificial intelligence and machine learning techniques for this purpose. These methods are an improvement over previously used techniques because of the ability to determine the nonlinear relations among various parameters. In this context, many researchers have investigated Artificial Neural Networks (ANNs) for predicting pile bearing capacity [15–19]. Notably, Goh [20, 21] has presented an ANN model for the estimation of driven pile bearing capacity in cohesionless soils. From their study, pile length, pile width, pile cross-sectional area, and hammer properties such as weight and drop height are the primary input parameters in pile capacity estimation. From their study, ANNs are capable of capturing the complex and nonlinear interactions among the factors influencing the performance of driven piles. They can accurately model pile behavior under complex and varying loads, offering tremendous advances in pile performance prediction.

Syed Jamal Arbi et al. developed optimized machine learning models to predict pile-bearing capacity in layered soils. They investigated algorithms including Random Forest, Support Vector Machine, and XGBoost, refining model performance using random search and grid search optimization techniques. The results showed that XGBoost achieved the highest accuracy, demonstrating the effectiveness of tree-based models for capturing geotechnical variability in pile data [22].

Nguyen et al. applied a hybrid machine learning approach combining Adaptive Neuro-Fuzzy Inference System (ANFIS) and Genetic Algorithm (GA) to predict the total bearing capacity of driven piles. Using a dataset of 95 Pile Driving Analyzer (PDA) tests from Vietnam, their GA-ANFIS model demonstrated high accuracy, achieving R^2 values of 0.976 and 0.925 for training and testing datasets, respectively, indicating its effectiveness for rapid and reliable pile capacity prediction [23].

Kardani et al. used particle swarm optimization to tune hyperparameters and estimate the bearing capacity of piles in cohesionless soils, employing six machine learning algorithms, namely Decision Tree, k-Nearest Neighbor, Multilayer Perceptron Artificial Neural Network, Random Forest, Support Vector Regressor, and Extreme Gradient Boosting. The models considered key input parameters such as soil effective stress and pile geometric properties to capture the dominant factors influencing pile behavior. The results demonstrated that the optimized models significantly improved prediction accuracy, with Extreme Gradient Boosting exhibiting the best performance (R^2 = 0.96). Sensitivity analysis revealed effective stress as the most influential factor. This approach offers a more accurate and efficient alternative to traditional empirical methods for pile capacity estimation [24].

In recent years, the development of interpretable models and the growing emphasis on model explainability have sparked significant interest across the area of geotechnical engineering. The usage of SHapley Additive exPlanations (SHAP) has witnessed a substantial increase across research within geotechnical engineering, to explain machine learning models and determine the relative importance of soil and structural parameters. For instance, SHAP was utilized to explain models that predict residual friction angle [25], settlement [26], and slope stability [27], and generated significant insights pertaining to the effects of key geotechnical parameters.

Within the area of pile foundations, various recent studies have applied SHAP to determine the model sensitivity with respect to pile capacity, and thus emphasized the overwhelming dominance of geometric and soil-related parameters. For example, Li et al. applied SHAP with the aid of a Back Propagation (BP) neural network to determine the key contributors to pile bearing capacity, including undrained shear strength and average effective vertical stress [28]. Similarly, Khan et al. applied SHAP with the aid of a gene expression programming model and demonstrated that pile tip elevation was found to be the key factor that contributed to the prediction of pile capacity [29].

Table 1: Statistical Summary of the Dataset

	A(m^2)	L(m)	C′ (kN/m^2)	ϕ′ (°)	γ′ (kN/m^3)	δ (°)	Fn	Qm (kN)
Min	0.07	14.20	0.00	0.00	5.38	10.14	14.00	540.00
Max	1.59	98.00	148.00	36.62	13.49	17.00	2291.00	52100.00
Mean	0.43	27.11	32.37	25.58	10.20	13.68	494.99	5133.12
SD	0.4656	18.60	32.84	9.65	1.85	1.70	602.32	9290.15

Most previous studies mainly focused on the SPT value. However, there is a need for interpretable models that use direct pile–soil interaction parameters. This study fills this gap by using the standardized Flap number to predict the bearing capacity of 100 steel and concrete piles. It applies advanced tree-based algorithms combined with eXplainable Artificial Intelligence (XAI) techniques, specifically SHAP. This method avoids the "black-box" issue of traditional AI. It provides both high accuracy and a clear view of the factors affecting pile behavior, offering a transparent framework for geotechnical design.

2. Database

In this study, data from 100 concrete and steel samples from Iran, the USA, Canada, and India were employed [30]. The database consists of pile geometric parameters, physical parameters of the soil, Flap number, hammer impact energy, and results of static load tests. From this dataset, specific parameters were selected as model inputs, primarily divided into two categories that dominate pile bearing capacity: pile-related parameters and soil-related parameters. The drained cohesion, drained friction angle, and effective unit weight are employed to describe the parameters of the soil for the case of this database. The geometry of the pile is described by the pile cross-sectional area and the effective length. Additionally, the pile–soil friction angle and the material of piles are considered to be pile material-related parameters. Besides the selected parameters, the Flap number was incorporated to take into account loading conditions and other unknown variables. All of the piles were deployed by driving, and the dataset includes a wide variety of concrete and steel piles of different lengths, ranging from 15-meter concrete piles in sand to 98-meter offshore steel piles [30]. Table 1 summarizes the descriptive statistics of all input and output variables. The statistical characteristics of the dataset, including the distribution of parameters and their relation to pile bearing capacity, are shown in Fig. 1.

The qualitative characteristics of the compiled dataset were investigated to evaluate both its representativeness and generalizability. As shown in Figure 2(a), steel and concrete piles constitute 55% and 45% of the dataset, respectively, indicating a relatively balanced distribution of pile types. In addition, data from multiple geographical locations were included to capture geological variability. As illustrated in Figure 2(b), although 81% of the compiled dataset is representative of Iran, test data from other locations, including the United States, India, and Canada, accounting for 12%, 4%, and 3% of the compiled dataset, respectively, is also considered.

In the determination of the input parameters, several aspects were considered. The interpreted bearing capacity was defined as the failure load by adhering to the approach described by Eslami [31]. For cases where the failure load was not clearly defined, Hansen's 80% criterion [32] was followed. Mean values were recorded for parameters that vary along the length of the active pile, such as drained cohesion, soil friction angle, effective unit weight, and pile–soil friction angle. The effective unit weight of the soil deposits lying below the water table was computed according to the advice of Bowles [33]. Static load test results were adopted due to their acknowledged precision, and as these tests typically last several hours, during their conduct, the soil was adequately drained, and thus the consolidated drained (CD) conditions were assumed during calculations. For different pile–soil friction angles, recommended values described by Bowles [33] for various piles and rock or soil material types were utilized. To allow for a comparison of Flap numbers for different types of hammers, the number of hammer blows was normalized by the relative hammer energy. For these purposes, the choice of Kobe 35 was made as the reference point, following the work of Tomlinson and Woodward [34]. Following the findings of Prakash and Sharma [35], the Flap number was defined as the product of the hammer relative energy and the number of blows. Finally, following the approach laid out by Fellenius [36], the gross cross-sectional area of H-piles and open-ended steel pipe piles was considered, due to the generation of soil plugs.

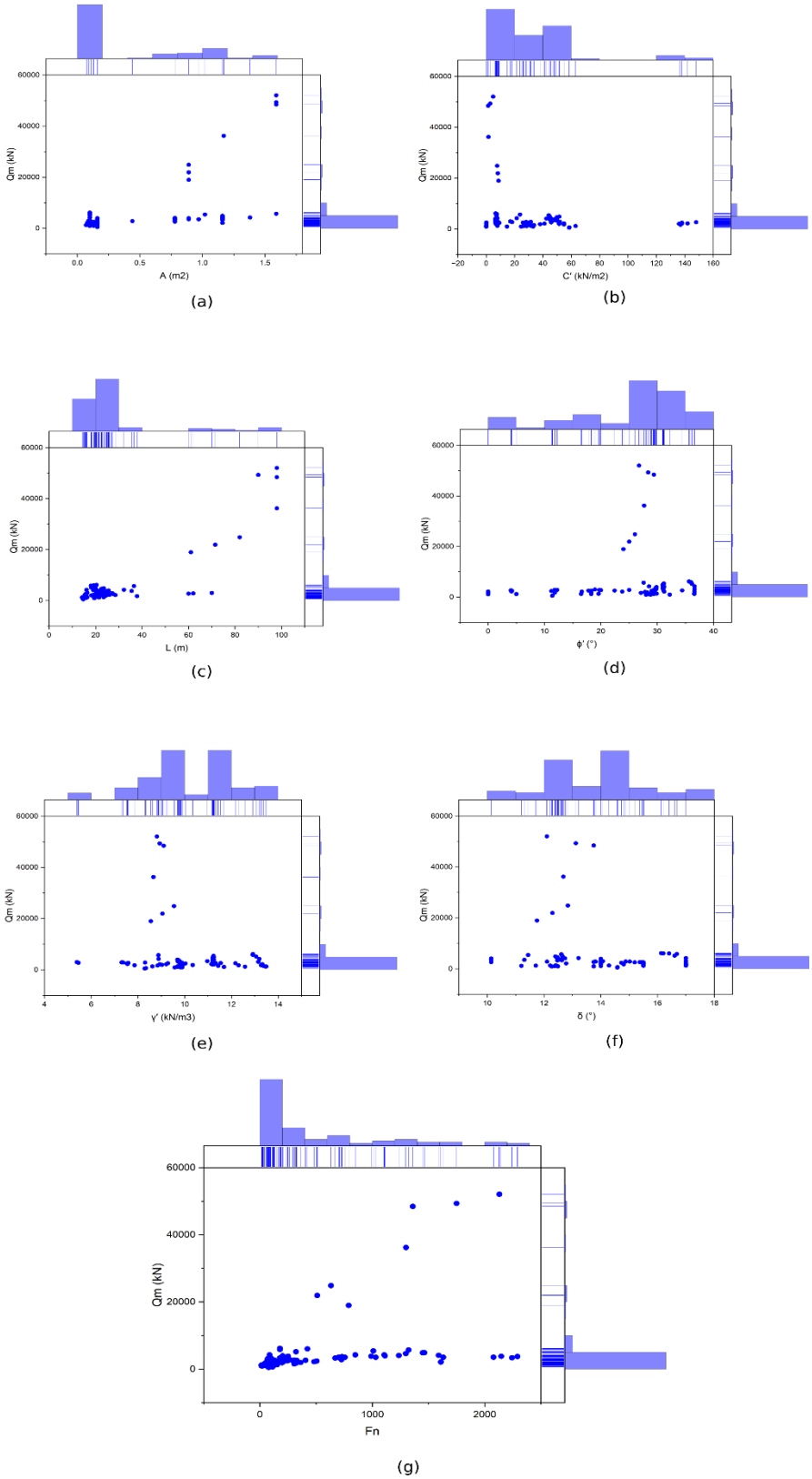


Fig. 1: Density contour plots of input variables with respect to the output parameter: (a) pile cross-sectional area, (b) drained cohesion, (c) effective pile length, (d) drained soil friction angle, (e) effective unit weight of soil, (f) pile–soil friction angle, and (g) Flap number.

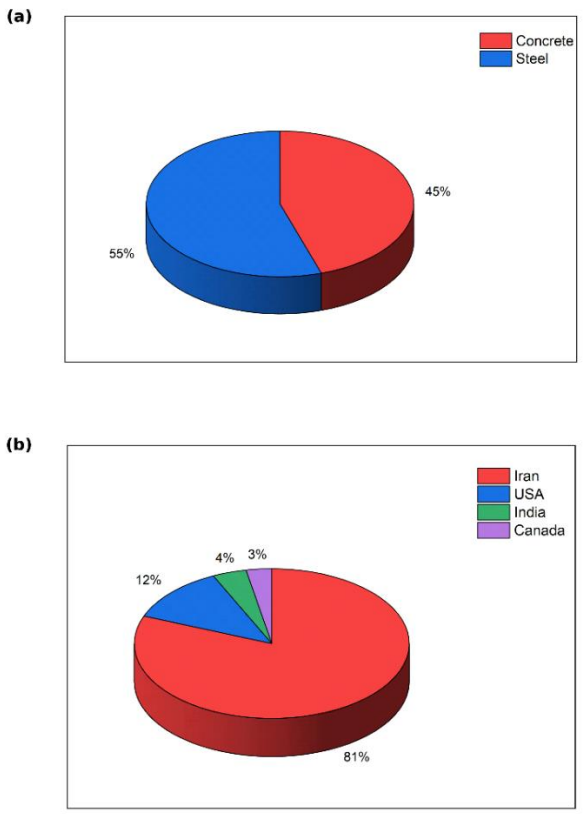


Fig. 2: Statistical distribution of the compiled pile load test database: (a) distribution of pile material types, and (b) geographical origin of the test sites

Fig. 3 is a correlation matrix representing the correlation of seven variables. Each value in the matrix is the Pearson correlation coefficient between the corresponding pair of variables. Near +1 or −1 values indicate strong linear correlation, and near-zero values indicate weak or no linear correlation. Colors represent visually the strength as well as the direction of correlations, with red representing positive correlation and

blue representing negative correlation. Analysis of this matrix shows the degree of correlation of each of the input variables to the output. Effective pile length and effective pile cross-sectional area correlate most positively with the output, having coefficients of 0.87 and 0.56, respectively. The Flap number also shows a moderate positive correlation with a coefficient of 0.45. In contrast to this, soil-related parameters like drained cohesion, drained friction angle, effective unit weight, and pile–soil friction angle have lower correlation with the output.

While the correlation matrix presented in Fig. 3 provides an initial overview of the pairwise relationships, a closer look has been taken regarding multicollinearity among the input parameters by a detailed VIF analysis. The VIF quantifies the extent to which the variance of an estimated regression coefficient increases due to correlation with the other predictors. Normally, VIF values greater than 10 are considered to indicate high multicollinearity, while VIF values below 5 do not imply any problems in terms of correlations.

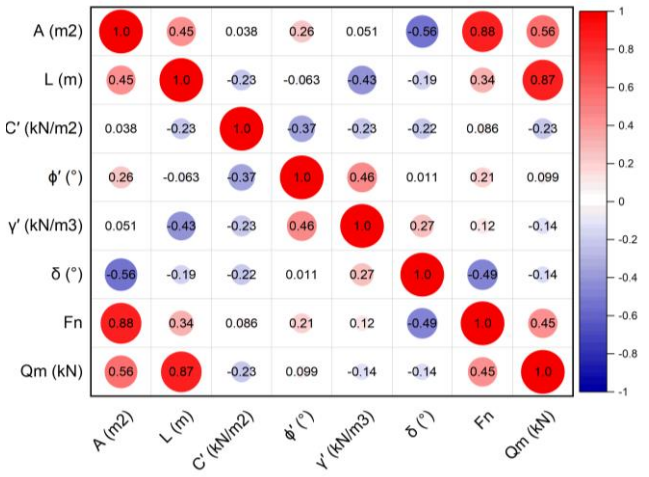


Fig. 3: Correlation matrix of the parameters

The VIF values for the majority of the parameters are under the conservative threshold of 5 (summary provided in Table 2), proving minimal collinearity. The cross-sectional area has a medium VIF value of 6.30 due to its geometrical relationship with the other physical property data, but this is still far below the critical threshold value of 10. Thus, no strong multicollinearity exists that can disrupt the learning process, and justification is made for all selected input variables to be included in the predictive models.

Table. 2: Multicollinearity analysis using the Variance Inflation Factor (VIF)

Parameter	Symbol	VIF Value	Status
Cross-sectional Area	A	6.30	Moderate
Flap number	Fn	4.61	Low
Effective Pile Length	L	2.42	Low
Effective Unit Weight	γ′	2.35	Low
Pile-Soil Friction Angle	δ	1.89	Low
Internal Friction Angle	ϕ′	1.61	Low
Drained Cohesion	C′	1.58	Low
Pile Material	-	1.46	Low

3. Data Preprocessing

The first step in applying machine learning methods is data preprocessing, which involves cleaning and transforming the data to make it suitable for the models. Preprocessing real-world datasets usually entails handling missing data and encoding categorical variables. Below is a detailed explanation of the preprocessing steps applied to the dataset in this research.

3.1 Removal of Invalid or Missing Data

Invalid or missing data in datasets can adversely affect the performance of machine learning models. Removing such data often improves model performance and

generalizability. For this case and research, however, no data were removed due to data completeness and consistency.

3.2 Encoding of Categorical Data

One-hot encoding was applied to the categorical variable "Pile Material" (steel and concrete), converting it into binary indicators (0 and 1) compatible with machine learning models and allowing them to evaluate the effect of pile material on bearing capacity.

4. Methodology

4.1 Machine Learning Algorithms

4.1.1 XGBoost (Extreme Gradient Boosting)

The XGBoost Regressor is an advanced machine learning algorithm uniquely designed for regression tasks, commonly known as "Extreme Gradient Boosting." The model is an ensemble comprising outputs from various decision trees to provide a more accurate and consistent final prediction. XGBoost Regressor is known for its high performance and efficient handling of large, complex datasets. This algorithm has one outstanding characteristic: it handles missing data excellently. The algorithm can learn by itself to handle missing values during training and thus reduces manual data cleaning and extensive preprocessing work significantly [37].

4.1.2 Random Forest

Random Forest Regressor is a machine learning regression algorithm. It is an ensemble method where the output of multiple decision trees is combined to produce a more accurate and robust outcome. Each decision tree is trained by selecting subsets of training features and data randomly, thus imparting diversity among the trees. The final prediction is then obtained by averaging all the predictions given by trees in a forest. Random Forest Regressor has one of its core advantages of being able to prevent overfitting compared to a decision tree [38].

4.1.3 CatBoost

CatBoost Regressor is a machine learning model belonging to the gradient boosting family. CatBoost Regressor is utilized to perform significantly better with categorical features than the standard gradient boosting algorithm. CatBoost performs both ordered boosting and random permutations on categorical variables to avoid overfitting and provide improved generalization. It is mostly renowned for its high speed, precision, and strong performance in processing complex datasets with numerical and categorical features [39].

4.1.4 Extra Trees Regressor

The algorithm resembles a Random Forest but introduces greater randomness during tree construction. Feature and split point selection are entirely random, promoting tree diversity and enhancing model performance. By training an ensemble of decision trees on random data subsets and averaging their predictions, the algorithm achieves fast training and reduces overfitting, particularly with noisy or high-dimensional data [40].

4.2 K-Fold Cross-Validation

In order to rigorously test the generalization performance of the proposed machine learning models, and given the limited dataset size of 100 samples, this study employed the K-Fold Cross-Validation technique. Unlike the traditional hold-out method, where a fixed portion of data is put aside for testing, and which may give spurious results due to the randomness of the split, cross-validation ensures that every single data point will be used for both phases: training and validation. This offers the full test of the robustness and predictive consistency of the model.

In this work, a 5-fold cross-validation scheme was implemented. The actual process involves randomly partitioning the entire dataset into five mutually exclusive and equal-sized subsets called folds. Subsequently, the modeling procedure is iterated five times. In each iteration, one specific fold is isolated as a test set to validate the model using unseen data, while the remaining four folds are combined and used as the training set. In order to ensure that these experiments are reproducible and that the outcome of shuffling for data distribution across runs remains identical, a fixed random seed with the value of 13 was used.

The final performance for each algorithm will be the arithmetic mean of the evaluation metrics calculated over the five iterations. It reduces the impact of any potential outliers that may exist in any single partition and will yield a robust estimate of the capability of the model in real-world scenarios. In other words, all the values presented within the Results section will represent the aggregated mean performances derived from these five rounds of validation.

4.3 Shapley Additive Explanations (SHAP)

Grounded in Lloyd Shapley's cooperative game theory, the SHAP (SHapley Additive exPlanations) method provides a comprehensive XAI framework for interpreting feature contributions in machine learning model predictions. Developed by [41], SHAP values fairly allocate the contribution of each feature to a specific prediction by averaging its marginal contribution across all possible feature combinations. This can be mathematically expressed as shown in Equation (1):

$$\varphi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} v(S \cup \{i\} - v(S)) \quad (1)$$

Where φ_i is the SHAP value for feature *i*, *S* is a subset of all features *N* excluding *i*, and *v(S)* is the value function representing the prediction for subset S.

SHAP values quantify the contribution of each feature to a model prediction, offering a straightforward and intuitive explanation for a particular prediction. These explanations enhance model explainability, highlight relevant features, expose biases, and build trust in complex models. Relatively and equally weighting every feature, SHAP promotes transparency and accountability of machine learning systems across domains where model decisions need to be understandable.

5. Model Evaluation

Model performances are evaluated using metrics such as the coefficient of determination (R^2), mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). A higher R^2 indicates superior model performance, whereas lower values for MAE, MSE, RMSE, and MAPE signify better predictive accuracy. Mathematical definitions of these evaluation measures are explained separately in Equations (2–6). These parameters quantify the performance of the machine learning models by comparing the estimated pile bearing capacity with the values determined from the static load test.

$$R^2 = 1 - \frac{\Sigma(y - \hat{y})^2}{\Sigma(y - \bar{y})^2} \quad (2)$$

$$\text{MAPE} = \frac{100\%}{n}\Sigma\left|\frac{y - \hat{y}}{y}\right| \quad (3)$$

$$MAE = \frac{1}{n} \times \Sigma|y - \hat{y}| \quad (4)$$

$$MSE = \frac{1}{n} \times \Sigma(y - \hat{y})^2 \quad (5)$$

$$RMSE = \sqrt{MSE} \quad (6)$$

6. Result and Discussion

To ensure reproducibility and enable fair comparison among algorithms, a systematic search for hyperparameter tuning using Grid Search with cross-validation among 5 folds was conducted. The search space for every model contained sets of hyperparameters that would result in an optimal value for validating the coefficient of determination and, at the same time, reduce the standard deviation. The strategy was employed as a method to reduce the risk of overfitting, given that there were limited numbers within the dataset. Moreover, there was a global random number seed of 13. These search ranges and respective optimal hyperparameters are listed in Table 3.

Table 3: Hyperparameter search spaces and optimal settings for the developed models.

Model	Hyperparameter	Search Range	Optimal Value
XGBoost	Booster	[gbrtree, dart]	dart
	reg_lambda	[0,1,5,10]	10
	subsample	[0.4,0.6,0.8,1]	0.4
	colsample_bytree	[0.4,0.6,0.8,1]	0.4
	learning_rate	[0.01,0.05,0.1]	0.05
	n_estimators	[100,300,500]	300
	max_depth	[3,5,7]	3
Random Forest	max_features	[sqrt,log2,None]	sqrt
	n_estimators	[100,200,500]	100
	max_depth	[None,10,20,30]	30
Extra Trees	max_features	[sqrt,log2,None]	None
	n_estimators	[100,200,500]	100
	min_sample_split	[2,5,10]	2
CatBoost	learning_rate	[0.01,0.05,0.1]	0.01
	Depth	[4,6,8,10]	4
	l2_leaf_reg	[1,3,5,10]	1
	iterations	[500,1000,2000]	1000

The prediction accuracy of the four developed machine learning models, namely XGBoost, Random Forest, Extra Trees, and CatBoost, on the compiled datasets was measured using the 5-fold cross-validation method. The average value of these metrics would help obtain an unbiased result about generalizability. The result is shown below in Table 4.

Table 4: Performance evaluation of the developed machine learning models using 5-fold cross-validation (Mean ± Standard Deviation).

Model	R^2	MAPE	MAE	MSE	RMSE
XGBoost	0.72 ± 0.06	42.82 ± 22.35	2114 ± 734	$2.61\times10^7 \pm 1.38\times10^7$	4830 ± 1663
Random Forest	0.84 ± 0.11	49.19 ± 34.01	1704 ± 807	$1.26\times10^7 \pm 1.17\times10^7$	3192 ± 1541
CatBoost	0.91 ± 0.07	28.93 ± 10.98	1253 ± 620	$7.44\times10^6 \pm 8.02\times10^6$	2444 ± 1211
Extra Trees	0.95 ± 0.03	25.28 ± 7.90	1003 ± 475	$4.26\times10^6 \pm 4.32\times10^6$	1806 ± 999

As suggested by the result obtained from the quantitative analysis, it was realized that among all the models, the Extra Trees algorithm worked best as a predictive model for pile bearing capacity. It obtained the highest accuracy on all criteria and had the lowest prediction error with a Root Mean Square Error value of 1806kN and a Mean Absolute Percentage Error value of 25.28%, and a Coefficient of Determination value of 0.95. Its outstanding performance can be ascribed to its highly random character, which

effectively reduced variance and minimized the risk of overfitting.

Following the Extra Trees algorithm, CatBoost performed well with high prediction accuracy and retained the second rank. Conversely, the accuracy and stability of the Random Forest and XGBoost models were relatively low. XGBoost obtained an R^2 value of 0.72 with more variability in its values. Although techniques for regularizing were incorporated, the presence of outliers in the small data set still acted as a constraint on the gradient boosting algorithm. It can be understood from these findings that ensemble learning with high randomness works as a better generalizer compared to traditional gradient boosting methods for the geotechnical task at hand.

The predictive accuracy and calibration of the developed models were checked by plotting the predicted values against the actual measurements for all four models in Fig. 4. These scatter plots can act as calibration curves, with good calibration shown by points closely following the 1:1 diagonal line.

As is seen in Fig. 4a and Fig. 4b, the XGBoost and Random Forest models tend to capture the underlying trend rather well, but dispersion clearly increases at higher capacities of piles. This would therefore mean that beyond capacity values of 5000 kN, these models have larger deviations from the reference line. In contrast, Fig. 4c presents results indicating that CatBoost has better stability since data points remain closer to the diagonal axis compared with the other applied algorithms. Of the algorithms considered, the Extra Trees model, as shown in Fig. 4d, exhibited the best agreement between the predicted and observed values. The points in this model are closer to the 1:1 line for most of the range. This visual comparison would, therefore, suggest that for the dataset considered, the algorithm Extra Trees has a highly consistent and robust predictive capability compared to the other methods.

Residual plot analysis further assessed the Extra Trees model's reliability. Fig. 5a depicts residuals plotted against predicted capacity, along with 95% prediction intervals ($\pm 2\sigma$). It should be evident that no systematic pattern is followed by residuals, which are distributed uniformly around zero, with the vast majority falling within those dashed bounds, indicating accuracy and an unbiased prediction. Moreover, as an additional check on the variability of model accuracy as a function of different geometric arrangements, plot analysis of residuals with Pile Length and Cross-sectional Area, which are primarily influencing factors are shown in Fig. 5b and Fig. 5c. The uniform distribution of residuals around the predicted values for such factors will reveal that there are no systematically low or high predictions concerning the pile geometry, greater or smaller.

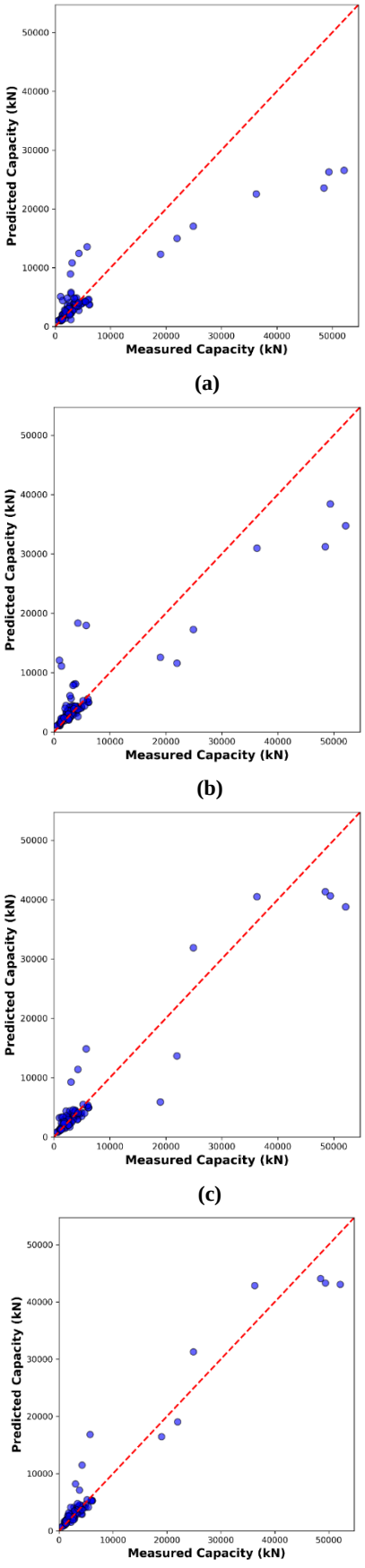
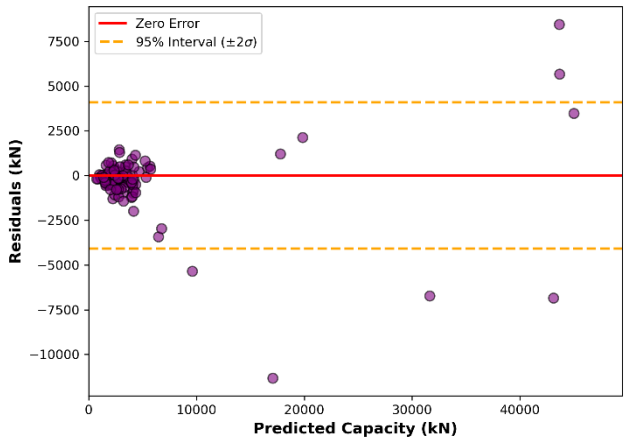
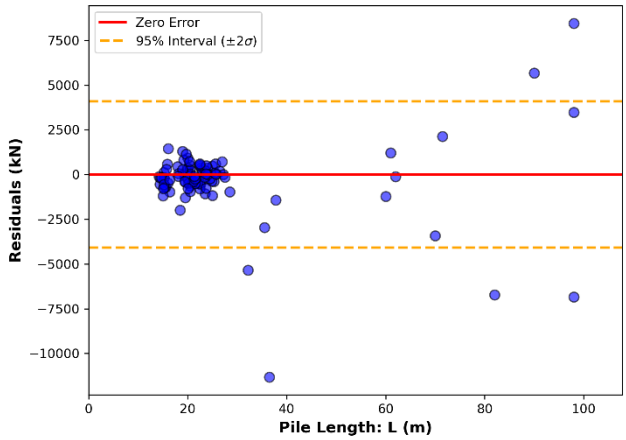


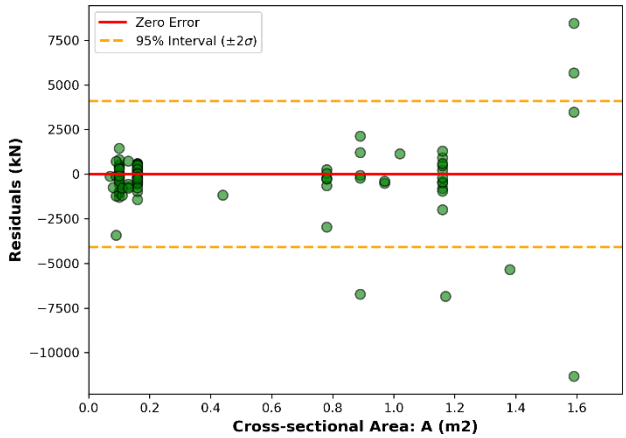
Fig. 4: Predicted versus measured pile bearing capacities for the developed machine learning models: (a) XGBoost, (b) Random Forest, (c) CatBoost, and (d) Extra Trees.



(a)



(b)



(c)

Fig. 5: Residual analysis for the Extra Trees model: (a) Residuals vs. Predicted Capacity; (b) Residuals vs. Pile Length; and (c) Residuals vs. Cross-sectional Area.

The data is divided into three clear-cut capacity bands: Low (< 5000 kN), Medium (5000–15000 kN), and High (> 15000 kN) to clearly highlight model performance for different load magnitudes. Quantitative error metrics for each band are provided in Table 5.

Table. 5: Quantitative error analysis of the Extra Trees model across different capacity bands

Capacity band	Range (kN)	No. of Sample	RMSE (kN)	MAPE (%)	Mean Bias (kN)
Low	<5000	86	972	25.30	-318
Medium	5000-15000	7	4324	36.54	-1160
High	> 15000	7	5526	13.84	1056

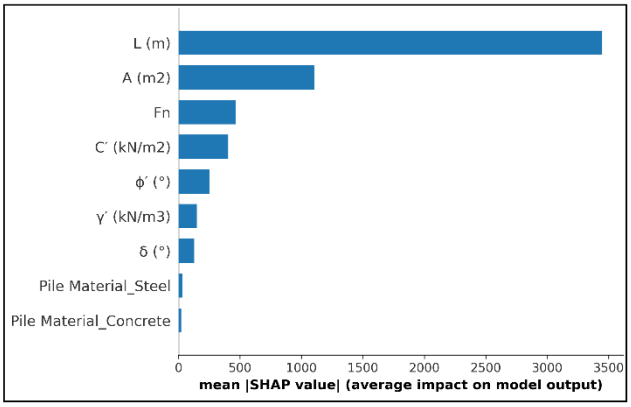
As expected based on the distribution of the dataset, 86% of the samples are within the low-capacity range. Here, the model shows consistent performance with an MAPE magnitude of about 25.30%. In the medium-capacity band, a brief trend of increasing relative error is seen due to the sparse data density in this transition zone. Notably, the Extra Trees algorithm gives remarkably good relative accuracy in the high-capacity band (>15 MN), which shows the lowest MAPE of 13.84%. Though the absolute error (RMSE) necessarily scales with the magnitude of the target variable, the relative precision of estimation significantly increases for larger piles. This aside, the mean bias is 1056 kN in the high-capacity category, implying that the model tends to somewhat underpredict the ultimate bearing capacity of massive piles. From the point of view of geotechnical engineering, this conservative performance is highly desirable as it remains on the side of safety for critical high-load structures.

While ensemble models such as Extra Trees are highly accurate predictors, they have often been described as black box algorithms. In order to provide a balance between predictive performance and engineering interpretability, the SHapley Additive exPlanations framework was utilized. SHAP assigns an importance value to each feature for every individual prediction, which indicates how much, according to the model, the feature pushes the output away from the baseline value. To ensure the robustness of this analysis, SHAP values were computed using aggregated results of the 5-fold cross-validation on the held-out test data. The outcome of this global sensitivity analysis is depicted in Fig. 6.

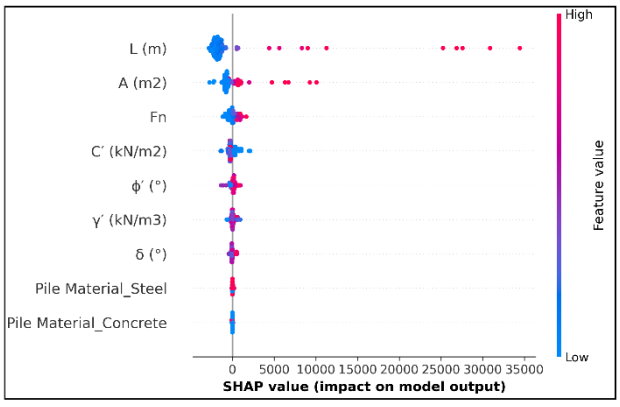
Fig. 6a presents the feature importance ranking based on the mean absolute SHAP value. This panel clearly shows that geometric properties of the pile are dominant for bearing capacity. Effective Pile Length (L) comes out as the most valuable predictor, followed by Cross-sectional Area (A) and Flap number (Fn), which also have a relatively significant impact. This is in exact agreement with geotechnical theory, as the pile resistance is directly proportional to the pile length, while the toe resistance is related to cross-sectional area. After geometry, other geotechnical parameters, specifically drained cohesion (C′), drained internal friction angle (φ′), effective unit weight (γ′), and pile soil interface friction angle (δ), have a much lower

global effect. Their steep falling importance values reveal that for the dataset considered herein, pile geometry variations dominate the impact of soil property variations. Finally, the Pile Material has negligible influence.

These relationships are further investigated using the SHAP summary (beeswarm) plot in Fig. 6b. Each dot in the visualization corresponds to one individual data point. The color of each dot signifies the feature value: red stands for high, and blue represents low. The x-axis is used to locate an impact on the model output, with the right-hand side indicating a positive impact and the left side showing a negative one. There are distinct patterns from this analysis for the top features. L, A, and Fn have a clearly positive correlation with bearing capacity. High values of this feature (red dots) cluster on the right-hand side, meaning that enlarged pile dimensions or an increase in Flap number lead to higher predicted capacity. Drained cohesion (C'), though, has an inverse relationship whereby higher values tend to yield lower SHAP values. For the remaining parameters, including φ', γ', δ, and Pile Material, the SHAP values exhibit no distinct patterns.



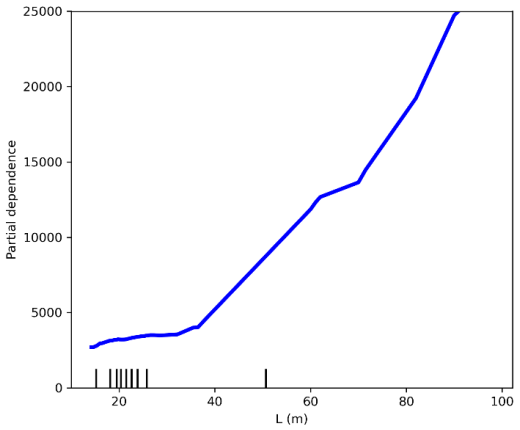
(a)



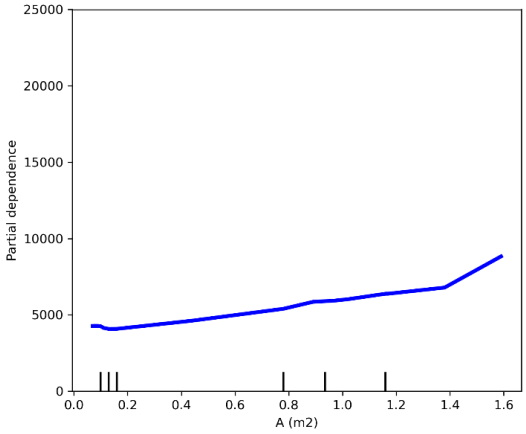
(b)

Fig. 6: SHAP analysis of input parameters for pile bearing capacity prediction: (a) SHAP bar plot showing feature importance, and (b) SHAP beeswarm plot illustrating the distribution of feature impacts.

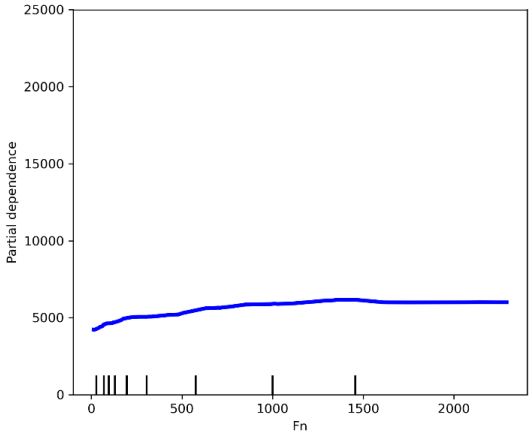
To complement the global feature importance, Partial Dependence Plots were generated. These plots, presented in Fig. 7, focus on the marginal effect of the three most influential features, namely Effective Pile Length (L), Cross-sectional Area (A), and Flap number (Fn), on the model output while averaging out the effects of all other variables.



(a)



(b)



(c)

Fig. 7. Partial Dependence Plots (PDPs) showing the marginal effect of key parameters on pile bearing capacity: (a) Effective Pile Length (L); (b) Cross-sectional Area (A); and (c) Flap number (Fn).

Fig. 7a presents the bearing capacity as a function of the Effective Pile Length (L). The curve represents a rather strong, steep, and near-linear positive relationship. The capacity increases significantly with an increase in the length of the pile. The very steep slope of the relationship evidences, furthermore, from a visual perspective, that pile length is the most important variable responsible for the total bearing capacity, and this is through the mobilization of shaft resistance.

In contrast, Fig. 7b and Fig. 7c show the dependencies of Cross-sectional Area (A) and Flap number (Fn), respectively. These features are positively correlated with capacity due to the geotechnical effect that the greater area increases toe resistance, while the more flaps increase anchorage. However, their slope is far gentler compared to the pile length. The discrepancy in the slopes suggests that while the Area and Flap number are indeed valuable contributors, their marginal effect on the total variation of capacity is considerably less dramatic than that provided by the pile length.

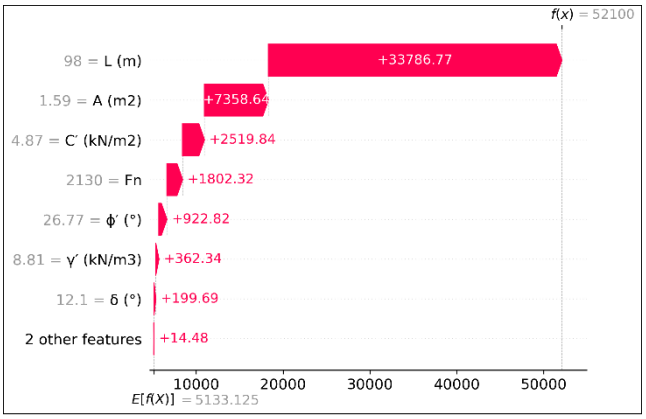
Although global analysis highlights overall trends, engineering demands an understanding of the process followed for making decisions on specific designs. To improve these efforts, SHAP waterfall plots were generated for two different scenarios: a high-capacity pile and a low-capacity pile. Figures 8a and 8b display these waterfall plots, which clearly explain feature contribution towards making the prediction deviate from the average value for a reference prediction (E[f(x)]) to arrive at the predicted result (f(x)).

Fig. 8a shows an explanation for local values for piles with maximum capacity within the dataset. According to the waterfall plot, prediction begins at an average value. Notably, large forces due to positive contributions from Effective Pile Length (L) and Cross-sectional Area (A), shown as red sections, result in a substantial increase in the predicted capacity. It clearly confirms that, based on large geometry parameters, the Extra Tree model predicted high capacity. Subsequently added forces due to parameters like C' and φ' result in an enormous predicted value.

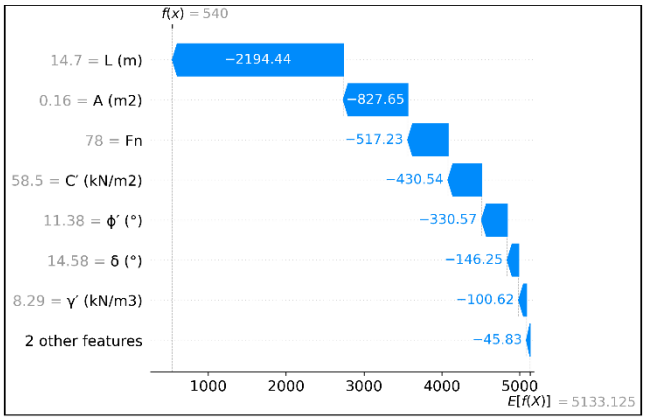
Fig. 8b shows an analysis for piles with low capacity. According to the waterfall plot, prediction begins at an average value. Negative contributions from low geometry parameters (L and A), shown as blue sections, result in a substantial downward adjustment, confirming that low geometry values drive the predicted capacity down.

7. Conclusion

The prediction of the ultimate bearing capacity of driven piles on the basis of the Flap number and explainable machine learning methods plays a significant role in geotechnical engineering. This method achieves high accuracy as it considers the most influential factors while estimating the pile capacity to produce more reliable outcomes. In addition to that, it proves to be cost- as well as time-effective in relation to other methods.



(a)



(b)

Fig. 8. SHAP Waterfall plots showing local feature contributions for: **(a)** A representative high-capacity pile; and **(b)** A representative low-capacity pile. Red bars indicate positive contributions increasing the capacity, while blue bars indicate negative contributions.

In this study, machine learning algorithms were utilized for developing models that predict the ultimate bearing capacity of driven piles using important input variables such as effective pile length, cross-sectional area, Flap number, drained cohesion, drained soil friction angle, effective unit weight, pile-soil friction angle, and pile material. Of the models that were compared: Random Forest, XGBoost, CatBoost, and Extra Trees, it was found that the latter Extra Trees approach is more robust with an average R^2 value of 0.95 ± 0.03 for a stringent 5-fold cross-validation test. This was further confirmed by analysis of residuals, indicating that almost all predictions lay in their respective 95% confidence levels.

SHAP analysis showed that the core factors affecting the proposed models were the effective pile length(L) and cross-sectional area (A), with contributions made by the Flap number (Fn). In particular, it was found that there was a constant relationship in the Partial Dependence plot related

to both strength and pile length, ensuring that the proposed models were interpretable and exhibited properties of a white-box model.

However, it should be kept in mind that the study is limited by the extent of the data set, as well as the fact that it is limited to a narrow category of soil and pile types. Future studies could potentially explore other data sets with a wider number of soil and pile types.

References

[1] Lee, I. M., & Lee, J. H. (1996). Prediction of pile bearing capacity using artificial neural networks. Computers and geotechnics, 18(3), 189-200.

[2] Meyerhof, G. G. (1976). Bearing capacity and settlement of pile foundations. Journal of the Geotechnical Engineering Division, 102(3), 197-228.

[3] Coyle, H. M., & Castello, R. R. (1981). New design correlations for piles in sand. Journal of the Geotechnical Engineering Division, 107(7), 965-986.

[4] Masouleh, S. F., & Fakharian, K. (2008). Application of a continuum numerical model for pile driving analysis and comparison with a real case. Computers and Geotechnics, 35(3), 406-418.

[5] Momeni, E., Maizir, H., Gofar, N., & Nazir, R. (2013). Comparative study on prediction of axial bearing capacity of driven piles in granular materials. Jurnal Teknologi, 61(3), 15-20.

[6] Momeni, E. (2012). Axial bearing capacity of piles and modelling of distribution of skin resistance with depth (Doctoral dissertation, Universiti Teknologi Malaysia).

[7] Payan, M., Asadi, P., Jamaldar, A., Salimi, M., Ranjbar, P. Z., Armaghani, D. J., ... & Sheng, D. (2025). Artificial intelligence-based predictive models for shear wave velocity of soils: A comprehensive review. Engineering Applications of Artificial Intelligence, 155, 111095.

[8] Ghorbanzadeh, S., Armaghani, D. J., Salimi, M., & Payan, M. (2026). Predictive models for dynamic properties of soils using machine learning approaches: A comprehensive review. Engineering Applications of Artificial Intelligence, 163, 113014.

[9] Khodkari, N., Hamidian, P., Khodkari, H., Payan, M., & Behnood, A. (2024). Predicting the small strain shear modulus of sands and sand-fines binary mixtures using machine learning algorithms. *Transportation Geotechnics*, *44*, 101172.

[10] Ngamkhanong, C., Keawsawasvong, S., Jearsiripongkul, T., Cabangon, L. T., Payan, M., Sangjinda, K., ... & Thongchom, C. (2022). Data-driven prediction of stability of rock tunnel heading: an application of machine learning models. Infrastructures, 7(11), 148.

[11] Dashtgoli, D. S., Sadeghian, R., Ardakani, A. R. M., Mohammadnezhad, H., Giustiniani, M., Busetti, M., & Cherubini, C. (2024). Predictive modeling of shallow tunnel behavior: Leveraging machine learning for maximum convergence displacement estimation. *Transportation Geotechnics*, *47*, 101284.

[12] Sangdeh, M. K., Salimi, M., Khansar, H. H., Dokaneh, M., Ranjbar, P. Z., Payan, M., & Arabani, M. (2024). Predicting the precipitated calcium carbonate and unconfined compressive strength of bio-mediated sands through robust hybrid optimization algorithms. *Transportation Geotechnics*, *46*, 101235.

[13] Mohammadnezhad, H., & Eslami, S. (2024). Machine learning models for predicting the bearing capacity of shallow foundations: A Comparative study and sensitivity analysis. *Numerical Methods in Civil Engineering*, *9*(2), 40-54.

[14] MolaAbasi, H., Khajeh, A., Chenari, R. J., & Payan, M. (2022). A framework to predict the load-settlement behavior of shallow foundations in a range of soils from silty clays to sands using CPT records. Soft Computing, 26(7), 3545-3560.

[15] Moayedi, H., & Armaghani, D. J. (2017). Optimizing an ANN model with ICA for estimating bearing capacity of driven pile in cohesionless soil. *Engineering With Computers*, *34*(2), 347–356. https://doi.org/10.1007/s00366-017-0545-7

[16] Nazir, R., & Momeni, E. (2013). Prediction of axial bearing capacity of spread foundations in cohesionless soils using artificial neural network. Proc. GEOCON, 747-757.

[17] Shaik, S., Krishna, K. S. R., Abbas, M., Ahmed, M., & Mavaluru, D. (2018). Applying several soft computing techniques for prediction of bearing capacity of driven piles. *Engineering With Computers*, *35*(4), 1463–1474. https://doi.org/10.1007/s00366-018-0674-7

[18] Jesus, J. N., Maria do Socorro, C., & de SM Gadéa, A. (2022). PREDICTING LOAD CAPACITY OF PRECAST CONCRETE PILES USING SPT AND ARTIFICIAL NEURAL NETWORK. In XLIII Ibero-Latin American Congress on Computational Methods in Engineering (Vol. 4, No. 04).

[19] Momeni, E., Nazir, R., Armaghani, D. J., & Maizir, H. (2014). Prediction of pile bearing capacity using a hybrid genetic algorithm-based ANN. Measurement, 57, 122-131.

[20] Goh, A. T. (1995). Back-propagation neural networks for modeling complex systems. Artificial intelligence in engineering, 9(3), 143-151.

[21] Goh, A. T. (1996). Pile driving records reanalyzed using neural networks. Journal of Geotechnical Engineering, 122(6), 492-495.

[22] Arbi, S. J., Hassan, W., Khalid, U., Ijaz, N., Maqsood, Z., & Haider, A. (2025). Optimized machine learning-based enhanced modeling of pile bearing capacity in layered soils using random and grid search techniques. Earth Science Informatics, 18(4), 1-21.

[23] Nguyen, D. D., Nguyen, H. P., Vu, D. Q., Prakash, I., & Pham, B. T. (2023). Using GA-ANFIS machine learning model for forecasting the load bearing capacity of driven

piles. Journal of Science and Transport Technology, 3(2), 26-33.

[24] Kardani, N., Zhou, A., Nazem, M., & Shen, S. L. (2020). Estimation of bearing capacity of piles in cohesionless soil using optimised machine learning approaches. Geotechnical and Geological Engineering, 38(2), 2271-2291.

[25] Ankah, M. L. Y., Adjei-Yeboah, S., Ziggah, Y. Y., & Asare, E. N. (2025). Advanced hybrid machine learning models with explainable AI for predicting residual friction angle in clay soils. Scientific Reports, 15(1), 25868.

[26] Kannangara, K. P. M., Zhou, W., Ding, Z., & Hong, Z. (2022). Investigation of feature contribution to shield tunneling-induced settlement using Shapley additive explanations method. Journal of Rock Mechanics and Geotechnical Engineering, 14(4), 1052-1063.

[27] Waris, K. A., Rahaman, M. a. U., & Basha, B. M. (2025). Detection and prediction of slope stability in unsaturated finite slopes using interpretable machine learning. Modeling Earth Systems and Environment, 11(4). https://doi.org/10.1007/s40808-025-02454-4

[28] Li, S., Hai, M., Zhang, Q., Zhou, B., Wang, M., & Zhao, Z. (2025). Study on an interpretable prediction model for pile bearing capacity based on SHAP and BP neural networks. Scientific Reports, 15(1), 28134.

[29] Khan, A., Khan, M., Khan, W. A., Afridi, M. A., Naseem, K. A., & Noreen, A. (2025). Predicting pile bearing capacity using gene expression programming with SHapley Additive exPlanation interpretation. Discover Civil Engineering, 2(1), 58.

[30] Milad, F., Kamal, T., Nader, H., & Erman, O. E. (2015). New method for predicting the ultimate bearing capacity of driven piles by using Flap number. KSCE Journal of Civil Engineering, 19, 611-620.

[31] Eslami, A. (1997). Bearing capacity of piles from cone penetration test data. University of Ottawa (Canada).

[32] Hansen, J. B. (1963). Discussion of "Hyperbolic stress-strain response: Cohesive soils". Journal of the Soil Mechanics and Foundations Division, 89(4), 241-242.

[33] Bowles, J. E., & Guo, Y. (1996). Foundation analysis and design (Vol. 5, p. 127). New York: McGraw-hill.

[34] Tomlinson, M. J., & Woodward, J. (1994). Pile design and construction practice. Chapman and Hall.

[35] Prakash, S., & Sharma, H. D. (1991). Pile foundations in engineering practice. John Wiley & Sons.

[36] Fellenius, B. H. (1989). Predicted and observed axial behavior of piles. In Proceedings of the Symposium on Driven Pile Bearing Capacity Prediction (pp. 75–115). American Society of Civil Engineers

[37] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 785-794).

[38] Breiman, L. (2001). Random forests. Machine learning, 45, 5-32

[39] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. Neural Information Processing Systems, 31, 6639–6649. https://papers.nips.cc/paper/7898-catboost-unbiased-boosting-with-categorical-features.pdf

[40] Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. Machine learning, 63, 3-42.

[41] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.