

Comparative Evaluation of Random Forest and XGBoost for Bearing Capacity Prediction On Slopes

Rahmathiyah Sari Amanda¹⁾, Lindung Zalbuin Mase^{2)*}, Aidil Fitriansyah³⁾, Rena Misliniyati⁴⁾,
Muharram Nur Fikri⁵⁾

¹⁾Departement of Civil Engineering, Faculty of Engineering, University of Bengkulu, Bengkulu 38371, Indonesia.

^{2,3,4,5)} Departement Geotechnical Hazard Research Unit, Faculty of Engineering, University of Bengkulu, Bengkulu 3871, Indonesia

¹⁾g1b022054.rahmathiyahsari@mhs.unib.ac.id, ^{2)*}lmase@unib.ac.id, ³⁾ aidil.fitriansyah@unib.ac.id,

⁴⁾renamisliniyati@unib.ac.id, ⁵⁾ m.nurfikri@unib.ac.id

Submitted : Jun 18, 2026 | **Accepted** : Sep 8, 2026 | **Published** : Oct 1, 2026

Abstract: Predicting the ultimate bearing capacity of shallow foundations is crucial in geotechnical planning because it is directly related to the safety and stability of the structure. This is especially true for foundations located at the edge of a slope. However, conventional analytical methods often fail to capture the complex and nonlinear relationships between soil and foundation parameters. This study evaluates the performance of Random Forest (RF) and Extreme Gradient Boosting (XGBoost) ensemble machine learning models for predicting the ultimate bearing capacity (q_{ult}) of shallow foundations in non-cohesive soils on slopes, with emphasis on methodological rigor and physical model interpretability. The dataset consists of 393 secondary data points obtained from numerical modeling based on the Finite Element Method (FEM) in a previous study. Input variables include slope angle ($\sin \beta$), foundation width (B), foundation depth (Df), and soil angle of internal friction ($\tan \phi$), while ultimate bearing capacity (q_{ult}) is used as the output variable. Model performance was evaluated using the coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the A20 index. Both models achieved a test R^2 of 0.97. XGBoost demonstrated superior generalization with a test RMSE of 7.130 kN/m² and MAPE of 8.833%, while Random Forest achieved a higher A20 index of 88.136% compared to 86.44% for XGBoost.

Keywords: Random Forest, XGBoost, machine learning, ultimate bearing capacity, shallow foundations

INTRODUCTION

The ultimate bearing capacity of shallow foundations is a fundamental parameter in geotechnical design, governing the safety and structural integrity of civil infrastructure. When foundations are situated on or near slopes, the asymmetric ground geometry fundamentally alters the failure mechanism relative to level-ground conditions by reducing the passive resistance zone on the slope side and shifting the critical failure surface, thereby reducing the effective bearing capacity and increasing its sensitivity to variations in soil strength and foundation geometry (El-emam et al., 2023). Conventional analytical methods address this condition through simplified reduction factors applied to standard bearing capacity equations. While these methods remain widely adopted in geotechnical practice due to their computational simplicity, they are grounded in idealized assumptions including homogeneous soil profiles, planar failure surfaces, and symmetric loading conditions that frequently diverge from actual field conditions, particularly on sloped terrain. As a result, conventional methods have consistently demonstrated limited predictive accuracy on cohesionless soil datasets, with coefficients of determination typically between 0.68 and 0.82 and mean absolute percentage errors exceeding 19% (El Gendy, 2025), underscoring the need for more accurate and data-driven prediction approaches.

The emergence of machine learning has offered a powerful alternative to conventional approaches for bearing capacity prediction, with ensemble-based methods demonstrating particularly consistent performance

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

across geotechnical prediction tasks. El Gendy (2025) compared eleven ML and five deep learning models for predicting the bearing capacity of shallow foundations on cohesionless soil under level-ground conditions, finding that ensemble methods including XGBoost, Random Forest, and CatBoost achieved R^2 values above 0.988 and MAPE below 5.07%, substantially outperforming conventional equations. A further study evaluating six ML algorithms under identical experimental conditions reported R^2 values above 0.92 during training and above 0.83 during testing on a dataset of 169 experimental observations, confirming the reliability advantages of ensemble methods over simpler baseline algorithms. A comprehensive review of artificial intelligence applications for shallow foundation bearing capacity evaluation similarly identified ensemble learning as the dominant high-performing family across diverse geotechnical datasets and input configurations (Khajehzadeh & Keawsawasvong, 2024). Random Forest (Breiman, 2001) achieves this performance through bagging aggregating predictions from a large number of independently trained decision trees while XGBoost (Chen & Guestrin, 2016) employs sequential gradient boosting with built-in L1 and L2 regularization, which constrains individual tree complexity during training and promotes generalization to unseen data. Both algorithms are robust to multicollinearity among input variables, a property of particular relevance to geotechnical datasets where foundation geometry and soil parameters can exhibit moderate intercorrelation.

Despite this progress, three important gaps persist in the literature. First, the overwhelming majority of ML studies on shallow foundation bearing capacity have focused exclusively on level-ground conditions; no prior study has applied ensemble machine learning models to the prediction of bearing capacity of shallow foundations specifically on slopes, where slope geometry constitutes an additional and dominant input variable that fundamentally changes the physical problem relative to level-ground prediction. Research addressing foundations on slopes has relied primarily on conventional analytical and numerical approaches including finite element parametric studies such as El-emam et al. (2023) without extending to data-driven machine learning prediction. Second, feature importance and model interpretability analyses in geotechnical ML studies have been conducted almost exclusively for level-ground datasets, where foundation width and soil friction angle are typically reported as dominant predictors. Whether this ranking holds under slope-conditioned configurations where the asymmetric failure mechanism may cause slope angle and foundation depth to overtake conventional soil strength parameters in governing bearing capacity has not been examined through a data-driven lens. Third, data integrity practices including the systematic identification and removal of duplicate records prior to training-testing partitioning are rarely reported in ML studies based on finite element parametric datasets, despite having demonstrable consequences for the reliability of reported evaluation metrics.

This study addresses all three gaps through a single integrated investigation. Using a dataset of 352 unique observations derived from the finite element parametric study of El-emam et al. (2023), and following rigorous data integrity verification, Random Forest and XGBoost models are developed and validated through grid search hyperparameter optimization combined with 10-fold cross-validation. The trained models are evaluated using four complementary metrics R^2 , RMSE, MAPE, and the A20 index on an independent hold-out test set to ensure unbiased performance assessment. Feature importance analysis is subsequently conducted across both models to identify which geotechnical parameters most strongly govern the prediction of q_{ult} under slope-conditioned configurations. The results reveal that foundation depth (Df) and slope angle ($\sin \beta$) consistently dominate the prediction in both algorithms, while soil unit weight (γ) a parameter explicitly incorporated in conventional bearing capacity formulations contributes negligibly ($\leq 1.1\%$), suggesting that on sloped ground, geometric factors govern the failure mechanism to an extent not differentiated by conventional analytical approaches. This combination of methodological rigor, slope-specific context, and physically interpretable variable analysis constitutes the principal contribution of this study to the growing body of machine learning applications in geotechnical foundation engineering.

LITERATURE REVIEW

Research on the prediction of ultimate bearing capacity (qult) of shallow foundations has evolved substantially over the past two decades, driven by the recognized limitations of classical analytical formulations and the growing availability of computational resources. Prior studies have demonstrated that machine learning methods can achieve high predictive accuracy for the bearing capacity of shallow foundations (El Gendy, 2025; Liu & Liang, 2024). However, these studies have predominantly focused on level-ground conditions, leaving the influence of slope geometry on bearing capacity prediction largely unexamined. Conversely, research specifically addressing shallow foundations on slopes, such as the finite element parametric study by El-emam et al. (2023) from which the dataset in this study is derived, has relied on conventional analytical or numerical approaches rather than data-driven machine learning techniques. Consequently, while machine learning has proven effective for flat-ground scenarios and numerical methods have been established for sloped conditions, no prior study has combined these two directions by applying ensemble machine learning models specifically Random Forest and XGBoost to predict the bearing capacity of shallow foundations on slopes, supported by a complete methodological validation

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

framework comprising data integrity verification, cross-validation, hyperparameter optimization, and model interpretation through feature importance analysis.

Beyond this methodological gap, the present study offers a substantive contribution at the level of physical interpretation. Classical bearing capacity theories, including the Terzaghi and Meyerhof formulations, treat soil unit weight (γ) as a parameter of comparable significance to soil strength and foundation geometry through its direct contribution in the N_γ term. However, the feature importance analysis conducted in this study confirmed consistently across two structurally different ensemble algorithms, one based on bagging and the other on boosting indicates that foundation depth (D_f) and slope angle ($\sin \beta$) are the dominant predictors of q_{ult} , while unit weight (γ) contributes negligibly ($\leq 1.1\%$) to the prediction in either model. This data-driven finding suggests that on sloped ground, the asymmetric failure mechanism induced by slope geometry and foundation embedment governs bearing behavior to an extent that overshadows the conventional role of soil unit weight, a nuance that classical analytical formulas do not differentiate and that has not been explicitly demonstrated through a data-driven approach in prior slope-related bearing capacity studies.

This study therefore addresses both gaps by evaluating the performance of Random Forest and XGBoost ensemble machine learning models in predicting the ultimate bearing capacity of shallow foundations on sandy slopes, with the goal of providing predictions that are not only accurate and reliable but also physically interpretable for practical geotechnical design.

Machine Learning Applications

In geotechnical engineering have grown rapidly in the last decade. Harle and Wankhade conducted a comprehensive review of ML techniques in geotechnical predictive modelling and concluded that ensemble methods particularly RF and gradient boosting variants consistently outperform single-model approaches in terms of accuracy and robustness. The performed a systematic review and meta-analysis of artificial neural network applications in geotechnical engineering and confirmed that data-driven models can reliably capture nonlinear soil–structure interaction patterns that are intractable for conventional equations (Moayed et al., 2020).

Random Forest

Random Forest is an ensemble learning algorithm that combines multiple decision trees using bootstrap aggregation (bagging) to improve prediction accuracy and reduce overfitting (Breiman, 2001). Random Forest was introduced as an ensemble method that constructs multiple decision trees on bootstrapped subsets of the training data and aggregates their predictions through majority voting (classification) or averaging (regression). The method is robust against overfitting due to the combination of bagging and random feature selection at each split node. In geotechnical contexts, RF has been applied successfully to predict pile bearing capacity, soil liquefaction susceptibility, and foundation settlement.

XGBoost

XGBoost is a gradient boosting framework developed to improve computational efficiency and predictive performance while incorporating regularization to reduce overfitting (Chen & Guestrin, 2016). The incorporation of L1 and L2 regularization terms in the objective function explicitly penalizes model complexity, promoting generalization to unseen data. Previous studies have shown that integrating XGBoost with other machine learning techniques can significantly improve the prediction of ultimate bearing capacity (q_{ult}) of shallow foundations, outperforming both standalone models and traditional analytical approaches (Liu & Liang, 2024).

Research Gap

The expanding use of machine learning in geotechnical engineering has brought significant progress, yet notable limitations persist. A substantial portion of existing studies concentrate on individual ML algorithms or narrowly scoped datasets, which constrains the broader applicability and generalization of the resulting models (Baghbani et al., 2022; Park et al., 2025). Moreover, systematic comparisons across multiple ML algorithms under consistent experimental conditions remain relatively rare in the literature (Khajehzadeh & Keawsawasvong, 2024; El Gendy, 2025). Beyond methodological gaps, the majority of published research has centered on shallow foundations situated on flat terrain, leaving the behavior of foundations on sloped ground comparatively underexplored (Zhang et al., 2023; Zhou et al., 2023). Given that slope geometry fundamentally alters soil stress distribution and failure mechanisms, this constitutes a critical and unresolved research gap (Zhang et al., 2023; Zhou et al., 2023). To address these shortcomings, the present study proposes a structured comparative evaluation of Random Forest and XGBoost models trained on a numerically derived dataset, with the dual objective of achieving high predictive accuracy and offering a practical, dependable tool for estimating the ultimate bearing capacity of shallow foundations under sloped ground conditions.

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

METHOD

This study evaluates the performance of the Random Forest and XGBoost algorithms in predicting the ultimate bearing capacity of shallow foundations in non-cohesive soils through the stages of data collection, preprocessing, modeling, and evaluation

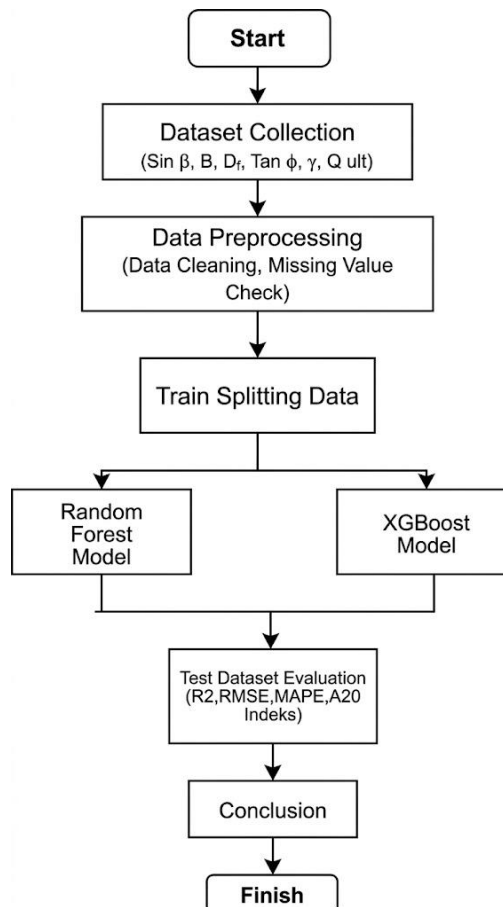


Figure 1. Research Methodology

Dataset Collection

The dataset used consists of 393 simulation results obtained from previous research on shallow foundations in non-cohesive soil near slopes. The input variables used include foundation width (B), foundation depth (D_f), internal friction angle of the soil ($\tan \varphi$), and slope angle ($\sin \beta$), while ultimate bearing capacity (q_{ult}) is used as the output variable (Pantelidis, 2024)

Table 1. Descriptive statistics dataset of research variables

Variables Parameter	N	Mean	Std Dev	Min	Q1	Median	Q3	Max
$\sin \beta$	393	0.394	0.169	0.174	0.174	0.342	0.5	0.643
B	393	1.625	0.415	1	1.50	1.5	2	2
D_f	393	1.247	0.432	1	1	1	1	2
$\tan \varphi$	393	0.329	0.116	0.176	0.176	0.306	0.46	0.466
γ	393	19.005	1.632	17.0	17.0	19.0	21.0	21.0
q_{ult}	393	166.089	80.279	43.04	103.1	147.025	211.26	418.26

Table 1 presents the descriptive statistics for a dataset consisting of 393 data points. The input variables include $\sin \beta$, B , D_f , and $\tan \varphi$, while q_{ult} is used as the output variable. The q_{ult} values range from 43.04 to 418.26, with a mean of 166.09, indicating sufficient data diversity to support the training and evaluation of machine learning models.

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Data Processing

Before model development, the dataset undergoes preprocessing to ensure data quality and consistency. This stage includes data validation, variable normalization, and dataset splitting. Normalization standardizes the scales across variables, thereby improving the effectiveness of the machine learning model training process. Next, 15% of the dataset was used for testing, and the remaining 85% was used for training. Given the small size of the dataset, this ratio was selected so that most of the data could be used as efficiently as possible during the model training phase without sacrificing the availability of suitably representative test data to assess prediction performance. This is consistent with standard procedures in civil engineering machine learning research, where data split ratios usually vary from 70:30 to 90:10 based on the quantity of samples available. Of the total 393 data points, 334 are used for model training, while the remaining 59 are used to test the model's ability to predict data that was not used during the training process.

Cross-Validation and Hyperparameter Optimization

To minimize the risk of overfitting and to ensure that the reported model performance reflects genuine generalization capability, hyperparameter optimization was performed for both Random Forest and XGBoost using grid search combined with 10-fold cross-validation (Kohavi, 1995; Bergstra & Bengio, 2012). For Random Forest, the search space comprised the number of trees ($n_{\text{estimators}}$: 50, 100, 200), maximum tree depth (max_depth : 10, 20, 30, or unrestricted), minimum samples required to split an internal node (min_samples_split : 2, 10, 20), minimum samples required at a leaf node (min_samples_leaf : 1, 5, 10), and the use of bootstrap sampling (True/False), yielding 216 candidate combinations. For XGBoost, the search space comprised the number of boosting rounds ($n_{\text{estimators}}$: 50, 100, 200), maximum tree depth (max_depth : 3, 6, 10), learning rate (0.01, 0.1, 0.2), row subsampling ratio (subsample : 0.8, 1.0), and column subsampling ratio per tree (colsample_bytree : 0.8, 1.0), yielding 108 candidate combinations (Chen & Guestrin, 2016).

Machine Learning Models

This study uses Random Forest (RF) and Extreme Gradient Boosting (XGBoost) to predict the ultimate bearing capacity of shallow foundations. These algorithms were chosen for their ability to model the complex, nonlinear relationship between geotechnical parameters and bearing capacity.

Random Forest

Random Forest is an ensemble-based machine learning method that combines multiple decision trees to produce more stable and accurate predictions. Each tree is built from a different data sample, and the predictions from all the trees are combined to obtain the final prediction. This approach helps reduce the risk of overfitting and improves the model's ability to capture patterns in the data. Due to these capabilities, Random Forest has been widely used in various geotechnical studies to predict soil parameters and foundation behavior (Breiman, 2001).

XGBoost

XGBoost is an ensemble learning algorithm developed from the gradient boosting method. This algorithm builds decision trees incrementally, where each new tree is used to correct prediction errors generated by the previous tree. Through this mechanism, XGBoost can produce more accurate and efficient models for handling complex nonlinear relationships. Additionally, applying regularization techniques in XGBoost improves the model's generalization and reduces the likelihood of overfitting. Therefore, this algorithm is widely used in various predictive studies in engineering and geotechnical engineering (Moayedi et al., 2020); (Liu & Liang, 2024); (Pantelidis, 2024).

Model Evaluation

Model performance was evaluated using four statistical indicators: the coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the a_{20} index. These four parameters were used to compare the predictive capabilities of the Random Forest and XGBoost models in estimating the ultimate bearing capacity of shallow foundations, while also identifying the model with the highest accuracy. It should be noted that the a_{20} index is a relatively new technical indicator that has been widely applied in civil engineering research (Asteris et al., 2021). In general, a model's predictive performance is considered better as the R^2 value approaches one, the RMSE and MAPE values approach zero, and the a_{20} index value approaches one. The mathematical equations used to calculate R^2 and RMSE are formulated as follows.

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Root Mean Square Error (RMSE)

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

Mean Absolute Percentage Error (MAPE).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (2)$$

Coefficient of Determination (R^2)

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

A20 Index

$$A20 = \frac{m_{20}}{n} \times 100 \quad (4)$$

where : m_{20} is the number of predictions satisfying the condition

$$0.8 \leq \frac{\hat{y}_i}{y_i} \leq 1.2 \quad (5)$$

RESULT**Cross-Validation and Hyperparameter Optimization Results**

Grid search identified the optimal Random Forest configuration as $n_estimators = 200$, $max_depth = unrestricted$, $min_samples_split = 2$, $min_samples_leaf = 1$, and $bootstrap = True$. For XGBoost, the optimal configuration was $n_estimators = 100$, $max_depth = 3$, $learning_rate = 0.2$, $subsample = 1.0$, and $colsample_bytree = 1.0$. The relatively shallow optimal tree depth identified for XGBoost ($max_depth = 3$) compared to the unrestricted depth selected for Random Forest reflects the fundamentally different mechanisms underlying the two algorithms: XGBoost builds shallow trees sequentially while correcting residual errors through boosting, whereas Random Forest relies on deep, independent trees aggregated through bagging (Chen & Guestrin, 2016; Breiman, 2001).

Model Performance Evaluation

This section presents the performance evaluation of the developed machine learning models, including Random Forest and XGBoost. The evaluation is based on several statistical metrics, namely R^2 , RMSE, MAPE, and A20 index, to assess the accuracy and reliability of each model.

Table 2. Comparison of model performance

Model Evaluation	Random Forest		XGBoost	
	Training	Testing	Training	Testing
RMSE	6.699	13,349	13.107	7,130
R^2	0.99	0.97	0.99	0.97
MAPE (%)	3.353	8,928	3,934	8,833
A20 Indeks (%)	94,611	88,136	93,712	86,44

Table 2 presents the performance metrics of both models during the training and testing phases. During training, the Random Forest model achieved an R^2 of 0.99, RMSE of 6.699 kN/m², MAPE of 3.353%, and A20 of 94.611%. On the independent test set, Random Forest achieved an R^2 of 0.97, RMSE of 13.349 kN/m², MAPE of 8.928%, and A20 of 88.136%. XGBoost achieved a training R^2 of 0.99, RMSE of 13.107 kN/m², MAPE of 3.934%, and A20 of 93.712%, with test performance of $R^2 = 0.97$, RMSE = 7.130 kN/m², MAPE = 8.833%, and A20 = 86.44%.

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Visual Evaluation of Model Predictions

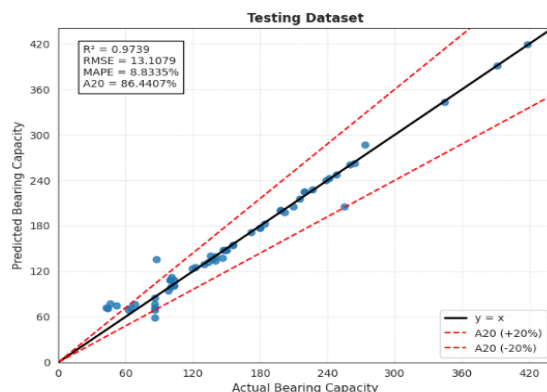


Figure 1. Performance evaluation Testing dataset of the XGBoost model based on the relationship between the actual and predicted values of ultimate bearing capacity

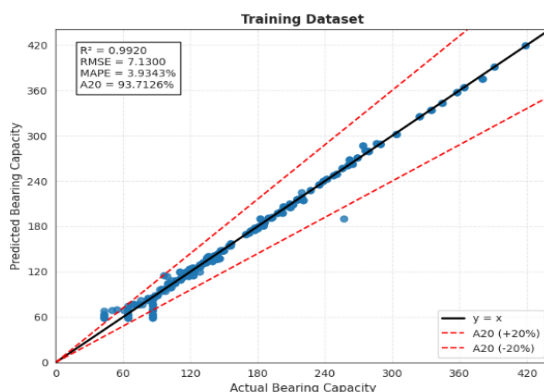


Figure 2. Performance evaluation Training dataset of the XGBoost model based on the relationship between the actual and predicted values of ultimate bearing capacity

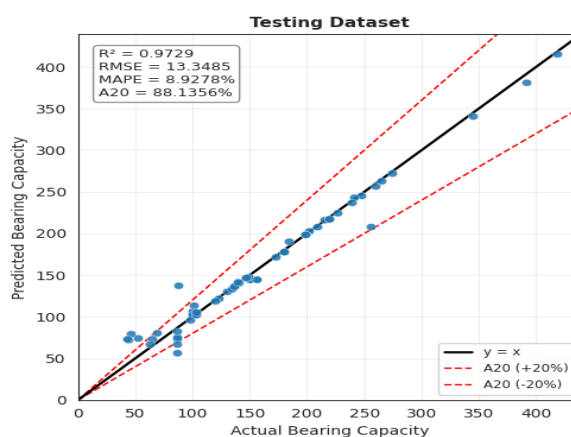


Figure 3. Performance evaluation Testing dataset of the Random Forest model based on the relationship between the actual and predicted values of ultimate bearing capacity

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

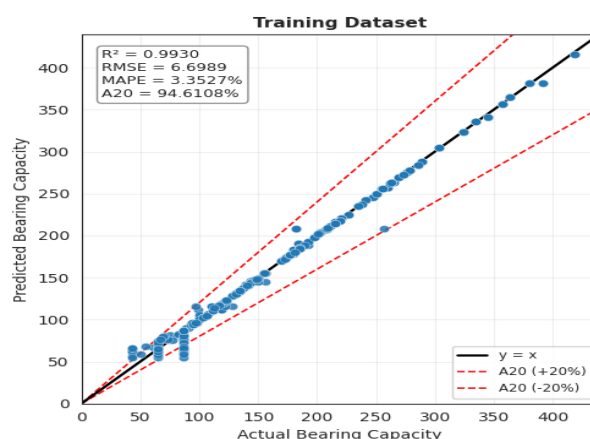


Figure 4. Performance evaluation Training dataset of the Random Forest model based on the relationship between the actual and predicted values of ultimate bearing capacity

A Visual Evaluation of Model Predictions is used to evaluate the agreement between the model's predicted ultimate bearing capacity values and the actual values. The closer the distribution of points is to the line of equality ($y = x$), the better the model's ability to represent the relationship between the input and output variables. Additionally, the $\pm 20\%$ tolerance limit indicated by the A20 line is used to determine the proportion of predicted results that still fall within an acceptable error range. It is worth noting that the q_{ult} values in the dataset ranged from 43.04 to 418.26 kN/m², with a mean of 166.09 kN/m² and a standard deviation of 80.28 kN/m², reflecting considerable variability in the ultimate bearing capacity across different geotechnical configurations.

Based on Figure 1, the XGBoost model on the testing dataset shows that most predicted points are distributed around the line of equality ($y = x$). The coefficient of determination (R^2) of 0.97 indicates that the model explains 97% of the variation in the q_{ult} test data. The RMSE of 7.130 kN/m² corresponds to a relative error of approximately 4.3% of the mean q_{ult} , which is considered highly acceptable in geotechnical prediction. The MAPE of 8.833% confirms that the predicted q_{ult} values deviate on average by less than 9% from the actual values, and the A20 index of 86.44% indicates that predictions within $\pm 20\%$ of the actual values constitute 86.44% of the total test data points.

Based on Figure 2, the XGBoost model on the training dataset achieved an R^2 of 0.99, an RMSE of 13.107 kN/m², a MAPE of 3.934%, and an A20 index of 93.712%. The notably lower MAPE on training data (3.934%) compared to testing data (8.833%) indicates that the model learned the q_{ult} patterns effectively during training. The lower RMSE on the testing dataset (7.130 kN/m²) relative to the training dataset (13.107 kN/m²) further suggests strong generalization capability rather than overfitting a behavior directly attributable to XGBoost's built-in L1 and L2 regularization mechanism, which constrains tree complexity during training and promotes stable out-of-sample performance (Chen & Guestrin, 2016).

Meanwhile, Figure 3 presents the performance of the Random Forest model on the testing dataset. The R^2 value of 0.97 indicates that the model maintains good predictive ability on unseen q_{ult} data. The RMSE of 13.349 kN/m² — approximately 8.0% of the mean q_{ult} is notably higher than that of XGBoost (7.130 kN/m²), indicating larger average absolute prediction errors. The MAPE of 8.928% and A20 index of 88.136% are comparable to XGBoost on the test set, suggesting that while both models capture the general trend of q_{ult} well, Random Forest exhibits slightly larger absolute errors on unseen data.

Figure 4 presents the Random Forest model's performance on the training dataset, where $R^2 = 0.99$, RMSE = 6.699 kN/m², MAPE = 3.353%, and A20 = 94.611%. The gap between Random Forest's training RMSE (6.699 kN/m²) and testing RMSE (13.349 kN/m²), a difference of 6.650 kN/m², is more pronounced compared to XGBoost, where the testing RMSE (7.130 kN/m²) is in fact lower than the training RMSE (13.107 kN/m²). This pattern indicates that Random Forest exhibits a relatively higher degree of overfitting on this dataset, whereas XGBoost demonstrates stronger generalization to unseen data a characteristic consistent with its regularization framework.

When compared visually, both models show closely aligned prediction patterns with relatively small data dispersion relative to the line of equality. The XGBoost model shows a slightly tighter distribution around the line $y = x$, particularly in the test data, consistent with its lower test RMSE. These results align with the quantitative values presented in Table 2, where XGBoost yields a lower prediction error than Random Forest on the q_{ult} test data in terms of RMSE and MAPE, while Random Forest achieves a higher A20 index on the test set (88.136% vs

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

86.44%). Both models can therefore be applied to predict the ultimate bearing capacity of shallow foundations on slopes, although XGBoost demonstrates marginally superior overall generalization performance on the dataset used in this study.

Feature Importance Analysis

Feature importance analysis was conducted for both tuned models to examine the relative contribution of each input variable to the prediction of q_{ult} , based on the mean decrease in impurity for Random Forest (Breiman, 2001) and gain-based importance for XGBoost (Chen & Guestrin, 2016)

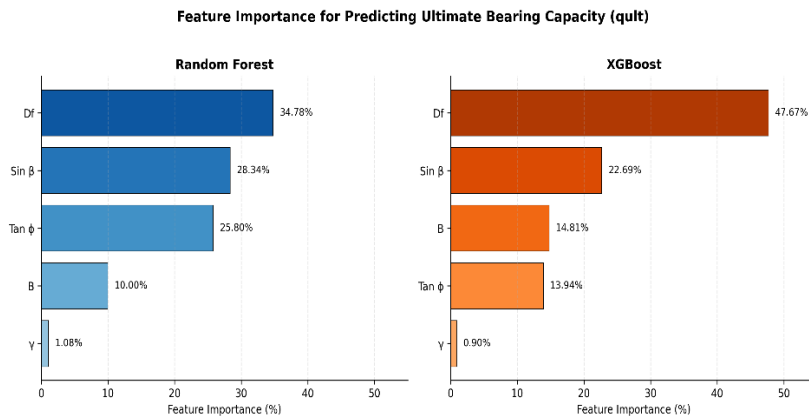


Figure 5. Performance of feature Importance Analysis

For Random Forest, foundation depth (Df) was identified as the most influential variable, contributing 34.78% of total importance, followed by slope angle ($\sin \beta$) at 28.34%, internal friction angle ($\tan \phi$) at 25.80%, foundation width (B) at 10.00%, and soil unit weight (γ) at 1.08%. For XGBoost, (Df) similarly ranked first with a notably higher contribution of 47.67%, followed by $\sin \beta$ at 22.69%, foundation width (B) at 14.81%, $\tan \phi$ at 13.94%, and γ at 0.90%.

DISCUSSIONS

The results presented in the previous section demonstrate that both Random Forest and XGBoost are capable of predicting the ultimate bearing capacity (q_{ult}) of shallow foundations on slopes with high accuracy. Both models achieved identical test R^2 values of 0.97 and comparable MAPE values below 9% on the independent test set, a level of accuracy consistent with findings from recent ensemble-based ML studies on shallow foundation bearing capacity, where R^2 above 0.97 and MAPE below 10% have been reported as characteristic of well-trained ensemble models applied to numerical simulation datasets. More broadly, both models substantially outperform conventional analytical methods, which have been reported to yield R^2 between 0.68 and 0.82 and MAPE exceeding 19% on comparable cohesionless soil datasets, reinforcing the practical case for adopting data-driven approaches in the pre-design stage of geotechnical foundation engineering on slopes.

A closer examination of the performance metrics reveals meaningful differences in how the two algorithms achieve comparable accuracy. XGBoost achieved a notably lower test RMSE (7.130 kN/m²) compared to Random Forest (13.349 kN/m²), corresponding to relative errors of approximately 4.3% and 8.0% of the mean q_{ult} (166.09 kN/m²), respectively. Notably, XGBoost's test RMSE (7.130 kN/m²) is lower than its own training RMSE (13.107 kN/m²) — a pattern directly attributable to its built-in L1 and L2 regularization mechanism, which constrains individual tree complexity during training and systematically promotes generalization to unseen data. Random Forest, in contrast, shows the more conventional pattern in which training RMSE (6.699 kN/m²) is lower than test RMSE (13.349 kN/m²), with a gap of 6.650 kN/m² indicating a higher degree of sensitivity to the specific training partition on this dataset. Despite this difference in generalization behavior, Random Forest achieved a higher A20 index on the test set (88.136% vs 86.44%), meaning a greater proportion of its predictions fell within the $\pm 20\%$ tolerance bound relative to actual q_{ult} values. This tradeoff — XGBoost superior in RMSE and MAPE, Random Forest superior in A20 — suggests that neither model universally dominates the other across all evaluation dimensions, and that the selection of an appropriate model should be guided by the metric most relevant to the intended engineering application.

The feature importance results extend the contribution of this study beyond predictive accuracy into the domain of physical interpretation. Foundation depth (Df) and slope angle ($\sin \beta$) were consistently identified as the two dominant predictors of q_{ult} across both models, collectively accounting for 63.12% of total importance in Random Forest and 70.36% in XGBoost. This finding is physically consistent with the governing mechanism of bearing

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

capacity on sloped terrain: the slope boundary asymmetrically reduces the confining stress on the downslope side, shortening the failure surface and diminishing load-bearing resistance — an effect that intensifies with both increasing slope inclination and foundation embedment depth. Conventional analytical formulations treat all parameters with fixed analytical roles, but the data-driven results of this study empirically confirm that geometric parameters (D_f and $\sin \beta$) govern the failure mechanism to an extent disproportionate to what classical equations implicitly assume.

By contrast, soil unit weight (γ) received the lowest importance in both models only 1.08% in Random Forest and 0.90% in XGBoost contributions so small as to be practically negligible within the parametric range of this dataset. In classical bearing capacity formulations, γ enters directly through the $N\gamma$ term and is treated as a parameter of comparable physical significance to friction angle and foundation geometry. The data-driven results here suggest that, on sloped ground within the ranges examined, the influence of γ on q_{ult} is marginal relative to geometric factors. This finding carries a practical implication: uncertainty in soil unit weight is unlikely to be the dominant source of variability in q_{ult} for this class of slope-foundation problem, a consideration that may inform how uncertainty budgets are allocated in reliability-based design.

The secondary importance rankings differ between the two models Random Forest places $\tan \phi$ third (25.80%) and B fourth (10.00%), while XGBoost reverses this order (B at 14.81%, $\tan \phi$ at 13.94%). This divergence is a known characteristic of comparing impurity-based importance in bagging algorithms against gain-based importance in boosting algorithms, which quantify variable contribution through fundamentally different computational mechanisms and can yield different secondary rankings on the same dataset. The agreement between two structurally different algorithms on the dominant variables (D_f and $\sin \beta$) and the negligible variable (γ) strengthens the physical interpretation: findings that emerge consistently from models with different internal mechanics are more robust than conclusions drawn from a single model alone.

From a practical standpoint, both models are suitable as supplementary computational tools in the pre-design stage of shallow foundation engineering on slopes, where rapid and reliable q_{ult} estimation across a range of geotechnical configurations is needed before committing to more time-intensive numerical analysis. The feature importance findings additionally provide actionable design guidance: given that D_f and $\sin \beta$ collectively explain more than 63% of the variation in q_{ult} in both models, these two parameters should receive priority attention in design optimization and parametric sensitivity studies for foundations on slopes within the soil and geometric ranges examined in this study.

CONCLUSION

This study evaluated the performance of Random Forest and XGBoost for predicting the ultimate bearing capacity (q_{ult}) of shallow foundations in non-cohesive soils on slopes, using a dataset of 393 data points derived from validated finite element numerical simulations. Hyperparameter optimization was performed through grid search combined with 10-fold cross-validation.

Both models demonstrated strong predictive performance. On the independent test set, XGBoost achieved $R^2 = 0.97$, RMSE = 7.130 kN/m², MAPE = 8.833%, and A20 = 86.44%, while Random Forest achieved $R^2 = 0.97$, RMSE = 13.349 kN/m², MAPE = 8.928%, and A20 = 88.136%. XGBoost demonstrated superior generalization in terms of RMSE and MAPE, with its test RMSE lower than its training RMSE — a direct consequence of its L1 and L2 regularization framework. Random Forest achieved a marginally higher A20 index on the test set, indicating a greater proportion of predictions within the $\pm 20\%$ tolerance bound. Both models substantially outperform conventional analytical methods reported in the literature, which typically yield R^2 between 0.68–0.82 and MAPE exceeding 19% on comparable datasets.

Feature importance analysis consistently identified foundation depth (D_f) and slope angle ($\sin \beta$) as the two most influential predictors of q_{ult} across both models, collectively accounting for 63.12% of total importance in Random Forest and 70.36% in XGBoost. Soil unit weight (γ) received negligible importance in both models (1.08% and 0.90%), suggesting that, within the parametric range examined, geometric factors governing asymmetric slope failure geometry dominate the bearing capacity mechanism to an extent that renders γ a minor contributor — a finding with potential implications for uncertainty allocation in reliability-based design of slope foundations.

Both models are recommended as practical supplementary tools for rapid q_{ult} estimation at the pre-design stage of shallow foundation engineering on slopes in non-cohesive soils, within the parametric bounds of the dataset used. XGBoost is preferred where minimizing absolute prediction error (RMSE) is the primary criterion, while Random Forest may be preferred where the proportion of predictions within practical tolerance limits (A20 index) is the priority.

Limitations and Future Work

Should expand beyond the limitations of the current study. The dataset is derived entirely from FEM numerical simulation, which while internally consistent but not fully reflect field variability. Future studies should incorporate field-measured bearing capacity data, extend the soil type range to include cohesive soils, add

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

geometric parameters not captured here such as foundation setback distance from the slope crest, and investigate advanced hyperparameter optimization methods including Bayesian optimization and automated machine learning frameworks to further improve model generalizability and robustness across broader geotechnical conditions.

REFERENCES

- El-emam, M., El Berizi, Y., Mabrouk, A. B., & Tabsh, S. W. (2023). Bearing capacity of strip footing on top of slope: Numerical parametric study. *Ain Shams Engineering Journal*, 14(11), 102522. <https://doi.org/10.1016/j.asej.2023.102522>
- Ismail, A., & Kwon, S. (2023). Development of mathematically motivated hybrid soft computing models for improved predictions of ultimate bearing capacity of shallow foundations. *Journal of Rock Mechanics and Geotechnical Engineering*, 15(3), 747–759. <https://doi.org/10.1016/j.jrmge.2022.04.005>
- Hamidi, A., & Abbeche, K. (2019). Bearing capacity of strip footing built on geogrid-reinforced sand over soft clay slope and subjected to a vertical load. *Geotechnical and Geological Engineering*, 19(1), 23–32.
- Harle, S. M., & Wankhade, R. L. (2025). Machine learning techniques for predictive modelling in geotechnical engineering: A succinct review. *Discover Civil Engineering*. <https://doi.org/10.1007/s44290-025-00224-w>
- As, M., & Subhiyakto, E. R. (2025). Performance comparison of Random Forest, SVM, and XGBoost algorithms with SMOTE for stunting prediction. *Journal of Information Systems and Informatics*, 9(4), 1163–1169.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Çınar, Z. M., Nuhu, A. A., Zeeshan, Q., & Korhan, O. (2020). Machine learning in predictive maintenance towards sustainable smart manufacturing in Industry 4.0. *Sustainability*, 12(19), 8211.
- Ma, L., Han, W., & Li, J. (2025). Machine learning based learning user behavior classification and recognition algorithm under multimodal data.
- Moayedi, H., Mosallanezhad, M., Jusoh, W. A. W., Muazu, M. A., & others. (2020). A systematic review and meta-analysis of artificial neural network application in geotechnical engineering: Theory and applications. *Neural Computing and Applications*, 32(2), 495–518. <https://doi.org/10.1007/s00521-019-04109-9>
- Liu, Y., & Liang, Y. (2024). Integrated machine learning for modeling bearing capacity of shallow foundations. *Scientific Reports*. <https://doi.org/10.1038/s41598-024-58534-5>
- Baghbani, A., Choudhury, T., Costa, S., & Reiner, J. (2022). Application of artificial intelligence in geotechnical engineering: A state-of-the-art review. *Earth-Science Reviews*, 228, 103991. <https://doi.org/10.1016/j.earscirev.2022.103991>
- Park, K., et al. (2025). Methodology for generating diverse geotechnical datasets using Monte Carlo simulation and genetic algorithms. *Computer-Aided Civil and Infrastructure Engineering*. <https://doi.org/10.1111/mice.70106>
- Khajehzadeh, M., & Keawsawasvong, S. (2024). Artificial intelligence for bearing capacity evaluation of shallow foundation: An overview. *Geotechnical and Geological Engineering*, 42, 5401–5424. <https://doi.org/10.1007/s10706-024-02863-9>
- El Gendy, M. (2025). A comparative machine and deep learning approach for predicting ultimate bearing capacity of shallow foundations in cohesionless soil. *Scientific Reports*. <https://doi.org/10.1038/s41598-025-22812-7>
- Zhang, W., Gu, X., & Ou, Q. (2023). Bearing capacity and failure mechanism of strip footings lying on slopes subjected to various rainfall patterns and intensities. *Geological Journal*. <https://doi.org/10.1002/gj.4882>
- Zhou, H., Hu, Q., Yu, X., Zheng, G., Liu, X., Xu, H., & Tian, K. (2023). Quantitative bearing capacity assessment of strip footings adjacent to two-layered slopes considering spatial soil variability. *Acta Geotechnica*. <https://doi.org/10.1007/s11440-023-01875-8>
- Pantelidis, L. (2024). Bearing capacity of shallow foundations: A focus on the depth factors in combination with the respective N-factors. *Arabian Journal of Geosciences*. <https://doi.org/10.1007/s12517-024-11976-7>
- Asteris, P. G., Skentou, A. D., Bardhan, A., Samui, P., & Pilakoutas, K. (2021). Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models. *Cement and Concrete Research*, 145, 106449. <https://doi.org/10.1016/j.cemconres.2021.106449>
- Díaz, E., & Spagnoli, G. (2024). A super-learner machine learning model for a global prediction of compression index in clays. *Applied Clay Science*, 249, 107239. <https://doi.org/10.1016/j.clay.2023.107239>

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

- Ismail, A., & Kwon, S. (2023). *Development of mathematically motivated hybrid soft computing models for improved predictions of ultimate bearing capacity of shallow foundations*. Journal of Rock Mechanics and Geotechnical Engineering, 15(3), 747–759. <https://doi.org/10.1016/j.jrmge.2022.04.005>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>