Published in AJSE, Vol: 24, Issue:3
Received on 14th February 2025
Revised on 26th May 2026
Published on 31th July 2026

AIUB JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 1608 - 3679 (print) 2520 - 4890 (Online)
https://doi.org/10.53799/507kbm95

Predictive Modelling and Slope Stability Using Numerical Simulations and Machine Learning Techniques

Radha Tomar and Smita Tung

***Abstract*— The current study combines numerical modelling and machine learning to identify the stability of applications in nail reinforced slope study. PLAXIS LE was used to develop different slopes having different soil properties including various values for cohesion (5, 10, 15 kPa), angle of internal friction (20°, 25°, 30°), unit weight (17, 18, 19 kN/m³), and slope angle (30°, 35°, 40°, 45°, 50°, 60°, 70°). Safety Factors (FOS) prediction models such as Random Forest (RF), Linear Regression (LR), and K-Nearest Neighbors (KNN) have been developed using the parameters included in the study. The Random Forest model has shown a superior performance among the other models with the lowest Mean Absolute Error (MAE: 0.053) and Mean Squared Error (MSE: 0.006), taking into consideration the highest value of R^2 (0.957) and Adjusted R^2 (0.951) to indicate a better predictive accuracy. With R^2 values of 0.903 and 0.920, respectively, Linear Regression and KNN also showed considerable strength of results. The results mentioned above show the bright future of machine learning models with Random Forest in predicting slope stability and contribute to refining nail reinforcement strategies. It shall also provide an input for developing cost-effective and robust slope rehabilitation measures in a geotechnically unfriendly environment.**

***Index Terms*— PLAXIS LE, linear Regression, Random Forest, K-Nearest Neighbors, Soil Nailing, Numerical Analysis.**

I. Introduction

This research is about applying a predictive model and optimization for slope stability using numerical simulations coupled into machine learning techniques. The numerical simulations using PLAXIS LE software consider different soil properties with varying values on cohesion, angle of internal friction, unit weight, and the slope angle, modelling slopes with various combinations[1]. Then this dataset will be used for prediction the Factor of Safety (FOS) through machine learning models. Thus, the overall objective of this research is to understand how various soil properties affect slope stability and to enhance the predictive accuracy of the machine learning models such as Random Forest, Linear Regression, K-Nearest Neighbors (KNN)[2]. The earliest study on the application of decision trees and Random Forest model for predicting the FOS of slopes was conducted by Bui et al. during 2015[3]. The authors also presented a study on the K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) in Goh et al. (2017), which are models for the prediction of slope stability in sandy soils [4]. Their study proved KNN superiority over SVM on some occasions because KNN simple pattern and ability to shape local structure between input variables. Li et al. (2016) studied the performance of Linear regression, KNN, and Random Forest for stability prediction of slopes [5]. Xu et al. (2013) have shown that, though requiring a large volume of data for training, ANNs can demonstrate superior prediction accuracy in slope stability in soils that have very variable properties. SVM is considered for slope stability analysis [6]. For example, Hong et al. (2014) used an SVM model to predict slope failure in sandy soils and found that SVM can indeed predict slope statically even with a limited data collection. A few attempts at developing hybrid machine learning models have been made in the past few years, which combine different algorithms. Yang et al. (2020) included KNN and ANNs in their hybrid model for predicting slope stability and claimed superior performance of the hybrid model[7]. Albayrak et al. (2020) reported that combined Random Forest with SVM gave the best results in predicting slope stabilities of different environments[8]. According to Sharma et al. (2022), CNNs are equally practical prediction devices, wherein slope failures could be determined and even more sophisticated patterns within huge data sets compared to traditional modelling[9]. Employing machine learning techniques to numerical simulations is a powerful benefit that is ordinarily not possible through the traditional approaches. As a result of the advancement in the technology of machine learning, for example, Random Forest, KNN, and Linear Regression, that work on big data coming out of the numerical simulation, one can detect the main elements that have influenced the slope stability and the Factor of Safety can be predicted with high quality[10]. These models are beneficial additionally to the assessment of the optimizing of slope stability as they can adjust their influence automatically according to the soil properties, slope geometry, or external conditions coming out of the soil testing[11].

The use of PLAXIS LE simulation with Random Forest,

Radha Tomar, *Research Scholar, Department of Civil Engineering, GLA University, Mathura, India*
Smita Tung, *Assistant Professor, Department of Civil Engineering, GLA University, Mathura, India*

Linear Regression, and KNN mechanisms for slope stability analysis from passing through different stages that were previous researchers is undoubtedly so very much in clearing up the prediction of slope stability[12]. Resolve the influence of soil properties on the slope stability by developing optimized models that will make it possible to have a more precise and reliable outcome in the prediction of Factor of Safety[13].

II. Study Area

Kullu district, located in northern Himachal Pradesh, is part of the western Himalayas and features steep slopes along with a variety of soil types, such as sandy loams, silts, and clays that are vulnerable to erosion. Given its high risk of landslides, this study utilizes numerical simulations and machine learning techniques to forecast slope stability and enhance stabilization strategies.

III. Methodology

This research combines numerical simulations with machine learning techniques to forecast the stability of slopes reinforced with soil nails. The goal is to create a predictive model that utilizes algorithms such as Linear Regression, KNN, and Random Forest to estimate the Factor of Safety (FOS). This estimation will take into account essential soil and geometric parameters, while also considering different soil conditions and reinforcement designs.

A. Data Collection

The data for this study was generated using numerical simulations on PLAXIS LE software. The key parameters considered for the simulations are listed in Table 1.

Table 1
Various Soil Properties Used for the Analysis

Soil Properties	Value
Slope Angle	30°, 35°, 40°, 45°, 50°, 60°, 70°
Cohesion	5 kPa, 10 kPa, 15 kPa
Unit Weight	17 kN/m³, 18 kN/m³, 19 kN/m³
Angle of Internal Friction	20°, 25°, 30°

B. Numerical Modelling (PLAXIS LE)

PLAXIS LE was utilized to simulate slope stability across different soil conditions and reinforcement methods. The table below details the characteristics of the soil nails employed in this study.

The numerical simulations were conducted to evaluate the Factor of Safety (FOS) for each slope configuration using the Limit Equilibrium Method (LEM) incorporated within PLAXIS LE. The analysis considered the contribution of soil nail reinforcement to slope stability, enabling a comprehensive assessment of the stability behavior under varying geotechnical and geometric conditions.

Table 2
Properties of Soil Nails

S.No.	Parameters	Values	Units
1	Bond Strength	100	kN/m
2	Plate Capacity	100	kN
3	Tensile Capacity	160	kN
4	Out of Plane Spacing	1	m
5	Length of Soil Nail	8	m
4	Inclination of Soil Nail	20	Degree

Representative failure slip surfaces generated using PLAXIS LE are shown in Figures 1 and 2. The results confirm the development of critical failure zones and validate the numerical modelling setup prior to the machine learning analysis.

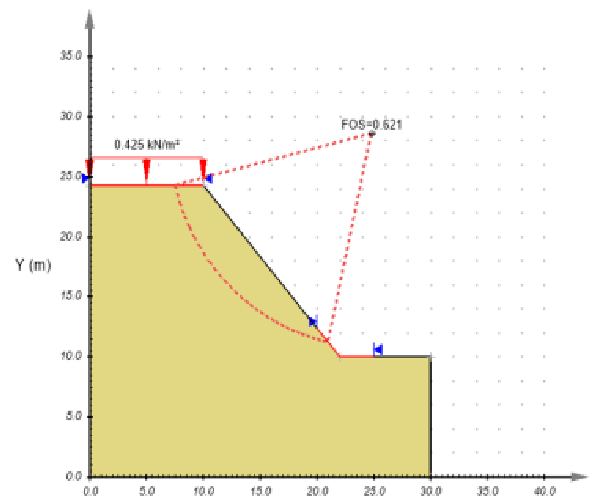

Fig. 1. Critical slip surface obtained from PLAXIS LE analysis for the selected slope configuration.

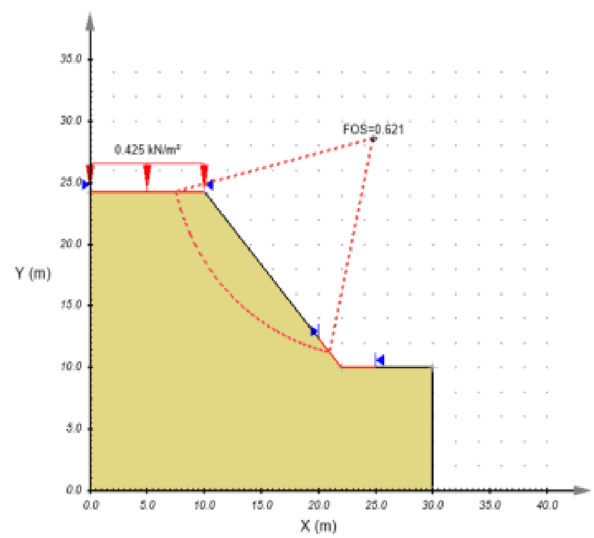

Fig. 2. Failure mechanism and corresponding slip surface generated through PLAXIS LE analysis.

IV. Machine Learning Models

This study utilizes three machine learning models Linear Regression, KNN, and Random Forest to forecast the Factor of Safety (FOS) in slope stability. These models are adept at capturing the intricate relationships among soil properties, slope angle, and FOS (Ahmad et al., 2023). The dataset derived from PLAXIS LE simulations was divided into 80% for training and 20% for testing. The models were trained to understand the input-output relationships and were assessed using various metrics, including MAE, MSE, R^2, Adjusted R^2, AIC, and SBC. The objective is to determine the most effective predictor for FOS, which will support slope stability analysis and the optimization of soil nail reinforcement.

V. Verification of Numerical Modelling Results

A comparative examination of the analysis's reliability was carried out using the finite element method to determine the strength reduction factor. Both deterministic and probabilistic linear equation modelling (LEM) showed a strong correlation with the strength reduction factor derived from the finite element method (FEM). The findings suggested that the stability of the slopes under study is primarily affected by the overall slope angle and its steepness. This research employed PLAXIS 2D for the finite element method and PLAXIS LE for the Limit Equilibrium method to analyze and compare slope stability. Table 3 illustrates the variation in the factor of safety between the LEM and FEM methods.

TABLE 3
THE FACTOR OF SAFETY VALUE FROM LEM AND FEM.

Nail Inclination	15°	20°
Factor of safety from LEM	1.451	1.51
Critical FOS by FEM	1.445	1.51

The LEM results compared well with finite element-model-based approach (FEM), as shown in Fig 3. The FEM yielded a minimum global safety factor (FS) compared to LEM and it evaluates the slopes' stability condition better. The Factor of Safety based using LEM is 1.451 and 1.51 at NI=15° and NI=20° respectively and the critical strength reduction factor-based FEM is 1.445 and 1.51 at NI=15° and NI=20° respectively, indicating that the slopes are critical to marginally stable. The Factor of safety obtained by LEM showed a good correlation with the factor of safety based on

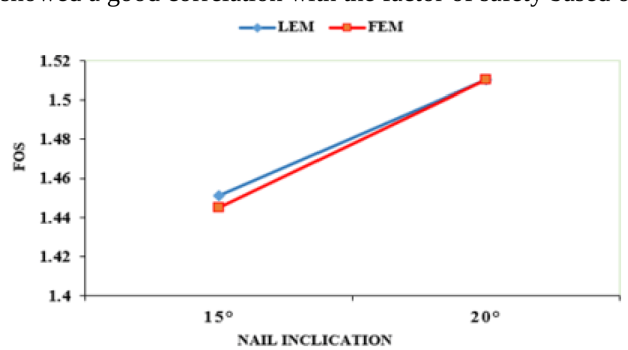

Fig. 3. Comparison between factor of safety from LEM and FEM.

FEM-SRF. The percentage difference between the Factor of Safety based on LEM and the factor of safety based on FEM-SRF is about 1.01%.

VI. Summary of Data Statics

TABLE 4: TABLE OF DATA STATICS

Variable	Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	Mean	Std. deviation
FOS	189	0	189	0.416	1.850	1.241	0.361
Cohesion	189	0	189	5.000	15.000	10.033	4.103
Unit Weight	189	0	189	17.000	19.000	18.026	0.808
Angle of Internal Friction	189	0	189	20.000	30.000	25.000	4.123
Slope Angle	189	0	189	30.000	70.000	46.921	12.895

The summary statistics provide an overview of input parameters and Factor of Safety (FOS) values for slope stability assessment. The FOS ranges from 0.416 to 1.850, with a mean of 1.241 and a standard deviation of 0.361, indicating both stable and unstable slopes. Cohesion varies from 5 to 15 kPa, internal friction angle from 20° to 30°, and unit weight from 17 to 19 kN/m³, with minor variability. Slope angles show the highest variation (30° to 70°), significantly influencing stability. These findings highlight how soil properties and slope geometry collectively impact FOS, emphasizing the need for reinforcement in unstable slopes.

VII. Results and Discussion

A. Numerical Modelling Results

a. Effect of Slope Angle on FOS

The results of the numerical analysis indicate that both the slope angle and height have a notable influence on slope stability. The factor of safety (FOS) decreases as both the height and angle of the slope increase. The highest FOS is observed for a slope angle of 30º, while the lowest FOS is associated with a slope angle of 70º.

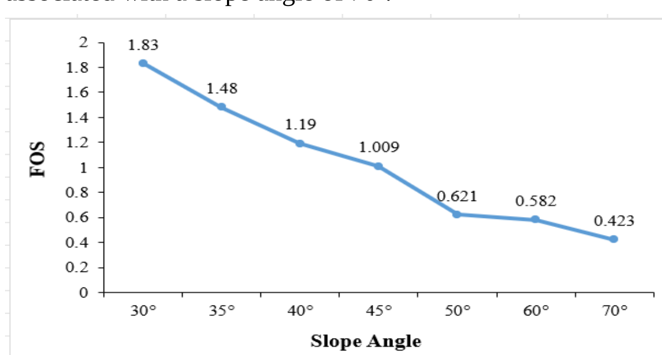

Fig. 4. Slope Angle v/s FOS Graph

b. Effect of Cohesion of Soil on FOS of Nailed Slope

The cohesion of soil plays a significant role in the stability of slopes reinforced with nails. It is observed that Higher cohesion in the soil generally leads to an increase in the Factor of Safety because cohesive soils have better internal strength and are less prone to sliding compared to less cohesive soils. Cohesion contributes to the overall stability of the slope by providing resistance against shear forces. If cohesion is more, it helps in holding the soil particles together, reducing the risk of slope failure. The stabilizing effect of cohesive soil can enhance the effectiveness of soil nail reinforcement, thereby increasing the Factor of Safety.

As illustrated in Figures 5, the cohesion of the soil, plays a crucial role in determining slope stability. For a unit weight of 17 kN/m³ and an angle of internal friction (φ) of 20º, an increase in the cohesion of soil from 5 kPa to 15 kPa leads to an increase in the factor of safety (FOS) from 0.987 to 1.508. Similarly, for a unit weight of 18 kN/m³ and an angle of internal friction (φ) of 30º, an increase in the cohesion of soil from 5 kPa to 15 kPa leads to an increase in the factor of safety (FOS) from 1.503 to 1.712. The cohesion of soil influences the Interaction of the soil and nails in high cohesive soils; the nails may experience greater bond strength due to the cohesive forces acting between the soil particles and the nails. As a result, the nails can provide more effective reinforcement in case of more cohesion, leading to higher FOS values compared to less cohesive soils.

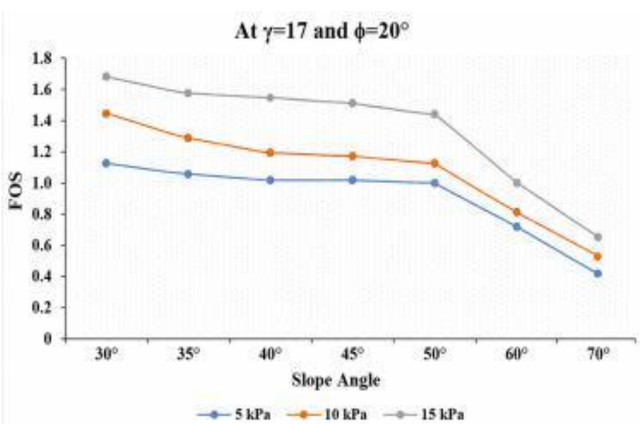

Fig. 5. Graph displaying effect of Cohesion on FOS at ϒ=17 kN/m3 and φ=20°

When designing a nailed slope in high cohesion soil, it is important to consider the cohesive strength of the soil along with other factors such as slope geometry, nail spacing, and groundwater conditions.

c. Effect of Unit Weight of Soil on FOS of Nailed Slope

The unit weight of soil plays a crucial role in the stability of slopes reinforced with nails. It is observed that higher unit weight of the soil generally leads to a decrease in the Factor of Safety. This is because denser soils exert greater downward forces on the slope, increasing the potential for instability. The added weight of the soil can induce higher shear stresses along the slope surface, making it more susceptible to sliding or failure. When analyzing the stability of a reinforced slope, the weight of the soil as one of the driving forces acting on the slope needs to be considered. The increased unit weight of soil results in higher driving forces that must be balanced by resisting forces, such as the shear strength provided by soil nails and the cohesion of the soil. As illustrated in Figures 6, the density of the soil, as measured by its unit weight, plays a crucial role in determining slope stability. For a cohesion (c) of 10 kPa and an angle of internal friction (φ) of 25º, an increase in the unit weight of soil from 17 kN/m³ to 19 kN/m³ leads to a decrease in the factor of safety (FOS) from 1.367 to 1.336.

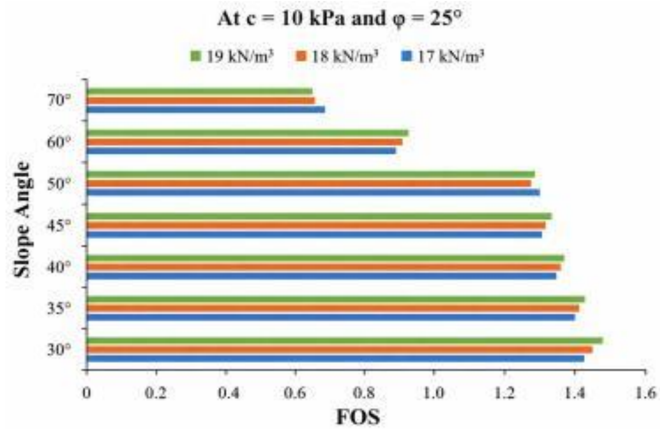

Fig. 6. Graph displaying effect of Unit Weight on FOS at C=10 kPa and φ=25°

The unit weight of soil affects the load-carrying capacity of the soil nails. In soils with higher unit weight, the nails may need to withstand higher loads due to the increased weight of the soil they are supporting. As a result, the design of soil nails, including their diameter, length, and spacing, may need to be adjusted to accommodate the higher unit weight of the soil and ensure adequate reinforcement. When designing nailed slopes unit weight of soil must be accounted. This includes conducting slope stability analyses that consider the unit weight of soil and its distribution along the slope.

d. Effect of Angle of Internal Friction of Soil on FOS of Nailed Slope

The angle of shearing resistance is a critical parameter that significantly influences the stability of nailed slopes. The angle of internal friction (φ) represents the resistance of soil particles to shear deformation. Soils possessing greater friction angles generally exhibit greater shear strength and are more stable than soils with lower angles of internal friction. As the angle of internal friction increases, the shear strength of the soil also increases.

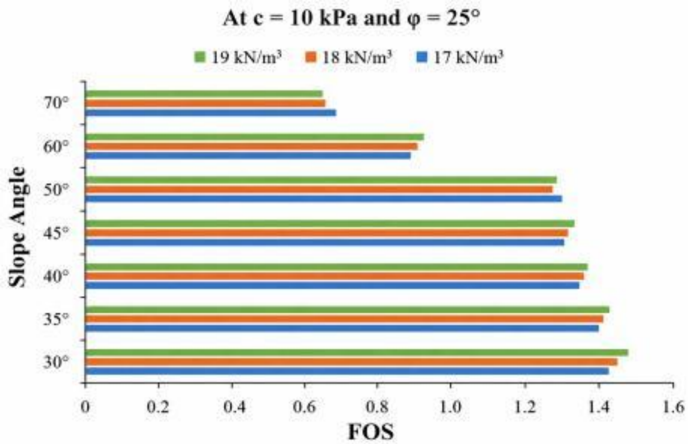

Fig. 7. Graph displaying effect of Angle of Internal Friction on FOS at C=5 kPa and ϒ=17 kN/m3

It is observed that slopes composed of soils with higher angles of internal friction have higher Factor of Safety values compared to slopes with lower angles of internal friction, assuming all other factors remain constant (Figure 7). Nailed slopes are reinforced slopes where soil nails are installed to enhance stability. The friction angle of the soil affects the mobilization of shear resistance along the failure surface of the slope. Higher angles of internal friction result in greater resistance to shear failure, thereby increasing the stability of the reinforced slope. Soils with lower angles of internal friction may require more extensive reinforcement to achieve a satisfactory Factor of Safety. Soils with higher angles of internal friction generally leads to more stable slopes, while soils with lower angles may require additional reinforcement to achieve satisfactory stability.

As depicted in Figure 7, the angle of internal friction of soil has a significant impact on slope stability. For instance, when the slope angle and friction angle of soil are 45° and 20° respectively, the safety factor is 0.987. However, with an increase in the friction angle to 30°, the FOS rises to 1.15. Additionally, the diagram highlights that the safety factor is more strongly influenced by the slope angle in granular materials with higher friction angles compared to fine-grained materials. For example, a change in slope angle from 45° to 70° with a friction angle of 30° leads to a reduction in the safety factor from 1.15 to 0.732.

Cohesion, internal friction angle, unit weight, and slope configurations have been some of the many soil parameters whereby the stability of slopes is affected.

Since, however, these properties interact in complex and nonlinear ways, traditional numerical modelling alone would not effectively make predictions regarding slope stability. To consider this complexity, machine learning techniques XGBoost, SVM, and LASSO Regression were used to analyses the complex relationships between soil properties and Factor of Safety (FOS). Besides enabling slope stability prediction, it also gives insights into the relevant roles of different soil properties to stability outcomes.

The performance of such machine learning models is analyzed using various metrics, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), among many others, on a real ground. The metrics are essential because they are needed to measure the quantitative accuracy and reliability of predictions in what would provide a benchmark of comparison for determining the best technique: also applied would be information-theoretic criteria, such as the Akaike's Information Criterion (AIC) and Schwartz's Bayesian Criterion (SBC), to ascertain model simplicity and goodness of fit.

B. Results of Machine Learning Models

a. Comparison of Machine Learning Models Based on Performance Metrics

The performance of three machine learning models Linear Regression, Random Forests, and K-Nearest Neighbors (KNN) was compared using various evaluation metrics. The comparison of various evaluation metrics is shown in Table 5.

TABLE 5

THE COMPARISON OF VARIOUS EVALUATION METRICS

PERFORMANCE METRICS	LINEAR REGRESSION	RANDOM FORESTS	K NEAREST NEIGHBORS
MAE	0.091	0.036	0.064
MSE	0.012	0.002	0.006
R^2	0.894	0.979	0.945
ADJUSTED R^2	0.882	0.977	0.939
AKAIKE'S AIC	-159.463	-221.174	-184.549
SCHWARZ'S SBC	-152.913	-214.623	-177.998

Mean Absolute Error (MAE): MAE evaluates the average absolute difference between the predicted and actual values, making it a simple measure of prediction accuracy.

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

where,

y_i = actual value, $\hat{y}_i$ = predicted value and n = number of data points

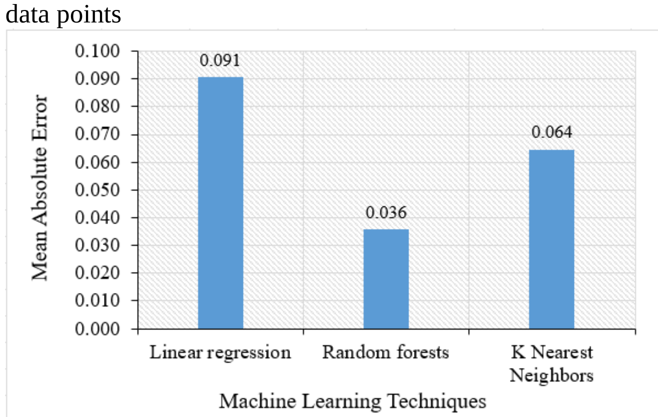

Fig. 8: Comparison of Mean Absolute Error (MAE) of various ML Techniques

A lower MAE indicates better predictive accuracy. Among the three models, Random Forests perform the best with the smallest MAE of 0.036, reflecting the highest accuracy. K-Nearest Neighbors (KNN) comes next with an MAE of 0.064, showing moderate accuracy. Linear Regression has the highest MAE of 0.091, indicating it is the least accurate model.

Mean Squared Error (MSE): MSE measures the average squared differences between predictions and actual values, penalizing larger errors more heavily.

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

where,

y_i = actual value, $\hat{y}_i$ = predicted value and n = number of data points

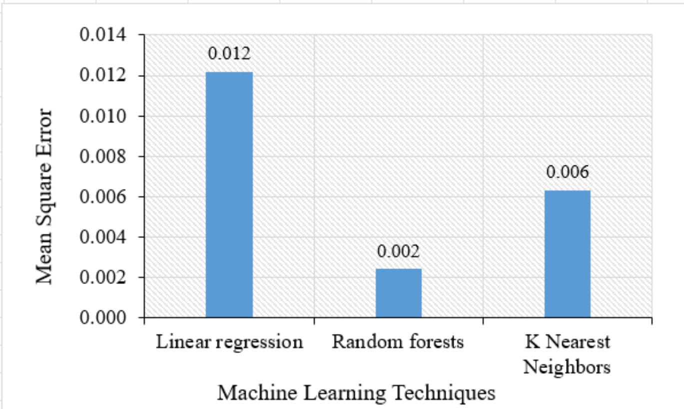

Fig. 9: Comparison of Mean Square Error (MSE) of various ML Techniques

Mean Squared Error is useful for assessing how well the model avoids large errors. Lower values of MSE are preferred. Random Forests again show superior performance with the smallest MSE of 0.002, followed by KNN with 0.006. Linear Regression lags behind with the largest MSE of 0.012, highlighting its weaker predictive capability.

Coefficient of Determination (R^2): The R^2 metric represents the proportion of variance in the target variable that the model can explain.

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

Where:

y_i = actual value, $\hat{y}_i$= predicted value and $\bar{y}$ = mean of the actual values

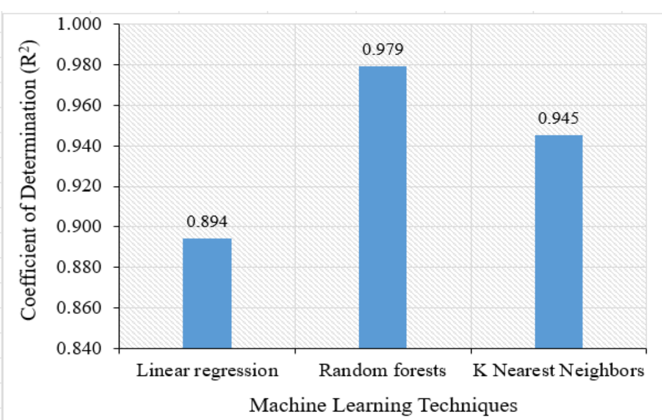

Fig. 10: Comparison of R^2 of various ML Techniques

value of 1 indicates perfect prediction; a 0 stands for the absence of model explanation for variance. Negative R^2 values indicate a poor fit. Random Forests achieve the highest R^2 of 0.979, explaining nearly 98% of the variance, indicating its strong predictive power. KNN performs moderately well with an R^2 of 0.945, while Linear Regression has the lowest R^2 of 0.894, making it the weakest in capturing variance.

Adjusted R^2: Adjusted R^2 accounts for model complexity and is a refined version of R^2. It penalizes the model for including too many predictors that don't improve the model's predictive power.

$$Adjusted\ R^2 = 1 - \left(\frac{(1-R^2)(n-1)}{n-p-1}\right)$$

Where: R^2 = coefficient of determination, n = number of data points and p = number of predictors (independent variables)

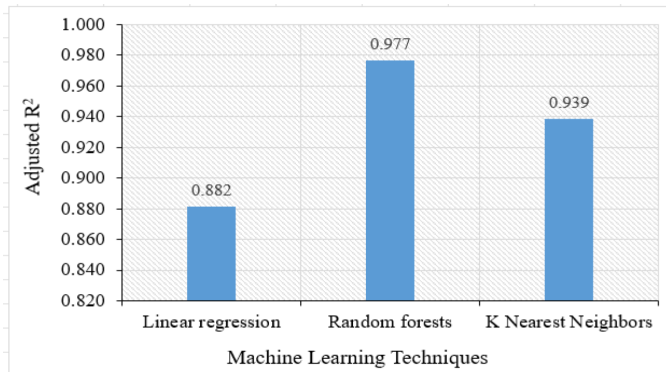

Fig. 11: Comparison of Adjusted R^2 of various ML Techniques

A higher Adjusted R^2 indicates a better model. Random Forests maintain their lead with the highest adjusted R^2 of 0.977, showing excellent model performance while accounting for predictors. KNN follows with an adjusted R^2 of 0.939, while Linear Regression, with the lowest adjusted R^2 of 0.882, reflects its relatively poor performance.

Akaike Information Criterion (AIC): AIC evaluates model quality by balancing fit and complexity, with smaller values being better. Lower AIC values indicate a better model, with simpler models being preferred if they perform similarly to more complex models.

$$AIC = 2k - 2\ln(L)$$

Where:

k = number of parameters in the model and L = likelihood of the model

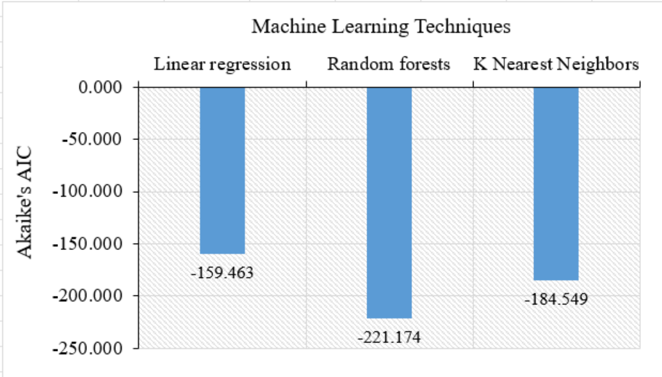

Fig. 12: Comparison of Akaike's AIC of various ML Techniques

Random Forests outperform with the smallest AIC value of -221.174, indicating the best balance. KNN ranks second with an AIC of -184.549, while Linear Regression, with the highest AIC of -159.463, is the least optimal model in terms of performance and complexity.

Schwarz's Bayesian Criterion (SBC/BIC): Similar to AIC, BIC penalizes model complexity more strongly. Lower BIC values suggest a better model.

$$BIC = \ln(n)k - 2\ln(L)$$

Where: n = number of data points, k = number of parameters and L = likelihood of the model

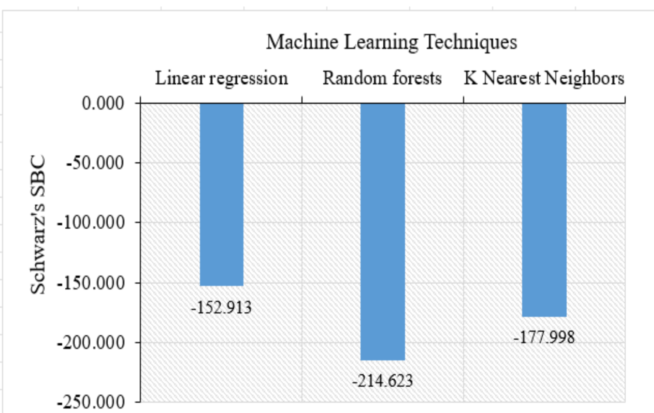

Fig. 13: Comparison of Schwarz's SBC of various ML Techniques

Random Forests again perform best with the smallest BIC of -214.623, indicating its effectiveness in balancing model complexity and performance. KNN, with a BIC of -177.998, performs better than Linear Regression, which has the highest BIC of -152.913, confirming its relative inefficiency.

In overall, Random Forests consistently emerge as the best-performing model across all metrics, providing the most accurate and reliable predictions. KNN offers moderate performance, better than Linear Regression but not as robust as Random Forests. Linear Regression, while straightforward, is the least effective in terms of prediction accuracy and model quality.

The proposed Random Forest model demonstrated improved predictive capability compared to traditional empirical approaches commonly adopted in soil-nailed slope design. Conventional empirical methods, such as FHWA-based approaches, generally rely on simplified assumptions and limited parameter interactions. In contrast, the Random Forest model effectively captured the nonlinear relationships between soil properties, slope geometry, and reinforcement conditions, resulting in higher prediction accuracy and improved generalization capability[14].

Model Performance Comparison Using Predicted vs Actual FOS Plots: Figures 14, 15, and 16 illustrate the comparison between predicted and actual Factor of Safety (FOS) for the Linear Regression, Random Forest, and KNN models. The green points represent the training data, while the blue points show the validation data. The dashed line (y = x) signifies perfect prediction, with points that are closer to this line indicating greater accuracy.

Linear Regression: The scatter plot for Linear Regression shows a moderate alignment of points around the diagonal line, indicating reasonably accurate predictions. However, some points deviate further from the line, especially at extreme values. This is reflected in the performance metrics: the R^2 of 0.894 indicates that the model explains 89.4% of the variance in FOS, but the higher MSE (0.012) and MAE (0.091) compared to other models indicate relatively larger prediction errors. Linear Regression has the highest AIC (-159.463) and SBC (-152.913), suggesting it is less efficient than the other methods.

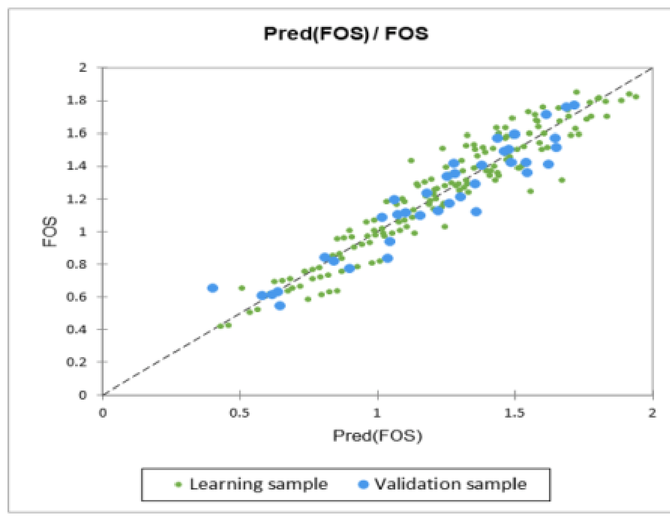

Fig. 14: Predicted vs. Actual FOS for Linear Regression

Random Forest Graph: The Random Forest scatter plot shows points closely clustered around the diagonal line, indicating high accuracy and strong predictive performance. The R^2 of 0.979 indicates that the model explains 97.9% of the variance in FOS, and the MSE (0.002) and MAE (0.036) are the lowest among the three methods, confirming minimal prediction errors. Additionally, Random Forest has the lowest AIC (-221.174) and SBC (-214.623), highlighting its superior efficiency and robustness.

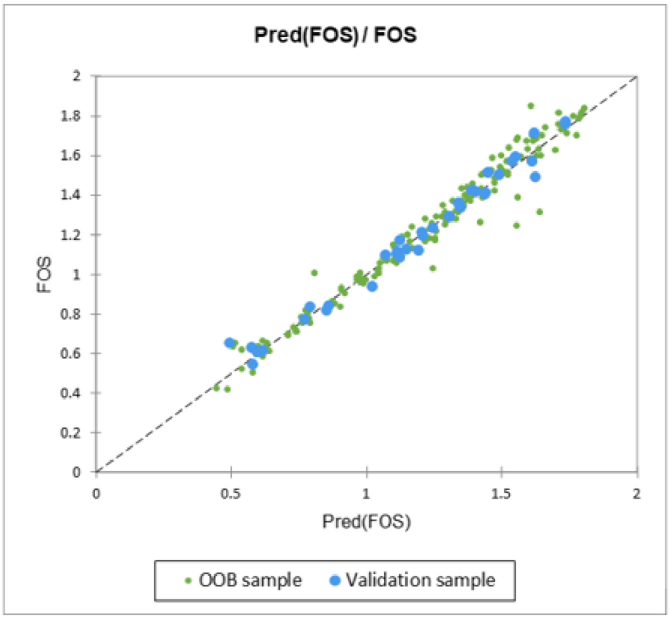

Fig. 15: Predicted vs. Actual FOS for Random Forest

K-Nearest Neighbors (KNN): The scatter plot for KNN shows points that are closer to the diagonal than Linear Regression but not as tightly aligned as Random Forest. The R^2 value of 0.945 indicates that the model explains 94.5% of the variance in FOS, which is better than Linear Regression but slightly worse than Random Forest. The MAE (0.064) and MSE (0.006) are also intermediate, showing that the errors are smaller than Linear Regression but higher than Random Forest. The AIC (-184.549) and SBC (-177.998) suggest better model efficiency compared to Linear Regression but not as good as Random Forest. The visual comparison of these scatter plots confirms that Random Forest is the most accurate and reliable model for predicting the Factor of Safety (FOS).

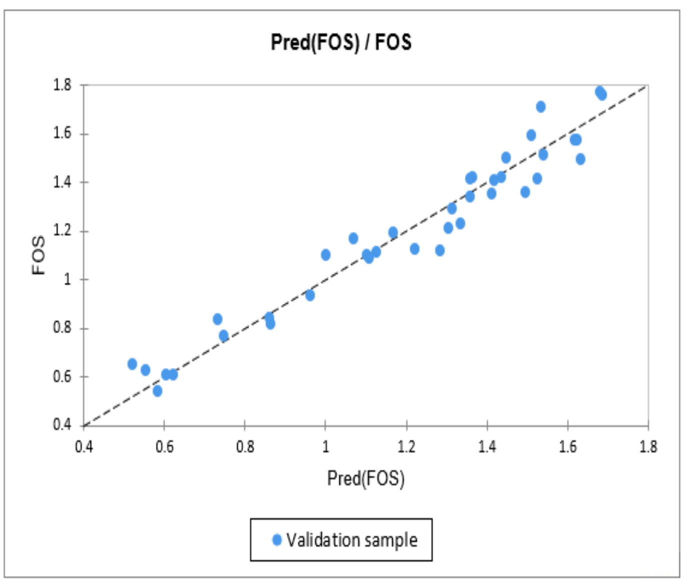

Fig. 16: Predicted vs. Actual FOS for K-Nearest Neighbors

TABLE 6

MACHINE LEARNING MODEL COMPARISON

Aspect	Linear Regression	Random Forest	K-Nearest Neighbors (KNN)
Prediction Accuracy	Moderate	High	Moderate
Handling Nonlinear Data	Weak	Strong	Moderate
Generalization	Moderate	Strong	Moderate
Spread of Data Points	Higher spread	Minimal spread	Moderate spread
Overfitting	Less prone	Well-balanced	Prone to overfitting
Key Strength	Simplicity and interpretability	Handles complexity and outliers	Adapts to local patterns

Sensitivity Analysis and Feature Importance

Standardized Coefficients graph for Linear Regression: This bar chart represents the standardized coefficients of the variables affecting the Factor of Safety (FOS) in the linear regression model. The standardized coefficients indicate the relative impact of each independent variable on the FOS, with their respective 95% confidence intervals. The positive coefficient for cohesion suggests that an increase in cohesion significantly enhances slope stability, resulting in a higher FOS. Angle of Internal Friction has the highest positive impact on FOS among all the factors, indicating that slopes with higher angles of internal friction exhibit greater stability and the near-zero coefficient for unit weight implies a negligible effect on FOS within the given range of values in this study while the slope angle exhibits the largest negative influence on FOS, signifying that steeper slopes are considerably less stable. This analysis highlights the dominant role of internal friction and slope angle in determining slope stability, providing a clear understanding of the critical factors to focus on in slope design and reinforcement. The chart highlights the most significant predictors in the model while identifying variables with minimal impact through coefficients close to zero. Despite having an R^2 of 0.894, which indicates that the Linear Regression model explains a substantial portion of the variance in FOS, its higher MAE (0.091) and MSE (0.012) reflect relatively larger prediction errors compared to more complex methods like Random Forest and KNN. However, the standardized coefficient chart offers a distinct advantage in interpretability, providing clear insights into the direct relationships between variables and FOS, which may not be as easily discerned in non-linear models.

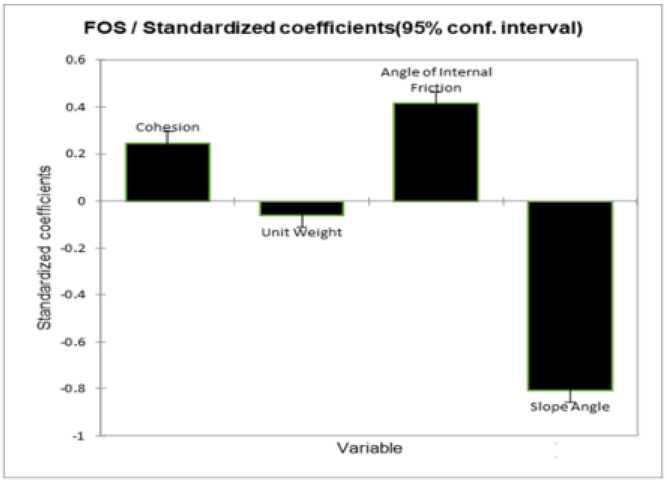

Fig.17: Standardized Coefficients of Variables Influencing

FOS (95% Confidence Intervals) using Linear Regression variable Importance Graph for Linear Regression: This graph depicts the variable importance analysis for the Random Forest model used in predicting the Factor of Safety (FOS) for nail-reinforced slopes. The importance of each variable is quantified using the "Mean Increase Error," which reflects the contribution of the variable to the predictive accuracy of the model. Angle of Internal Friction exhibits the highest importance, with a mean increase error close to 45. This indicates that changes in the angle of internal friction significantly impact the FOS, emphasizing its critical role in slope stability.

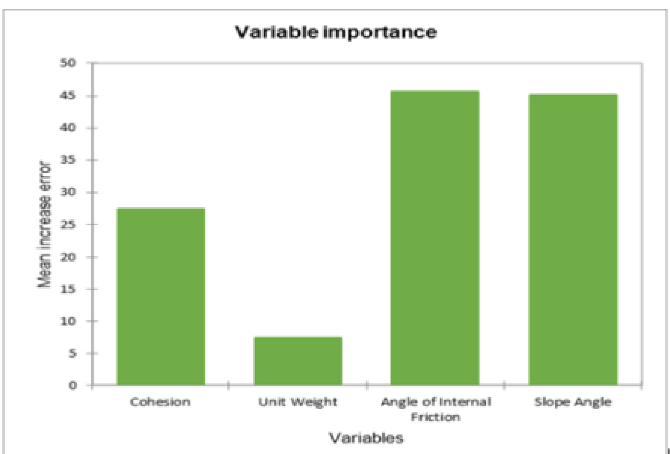

Fig.18: Variable Importance Based on Mean Increase in Error Using Random Forest

The slope angle also shows high importance. This result aligns with geotechnical principles, as steeper slopes are generally less stable and require more accurate predictions for effective stabilization while Cohesion contributes moderately to the model's predictive accuracy, with a mean increase error of approximately 30. This aligns with its role in providing shear strength to soil, which influences slope stability and the unit weight of the soil shows the least importance, suggesting that variations in this parameter have a comparatively smaller effect on the FOS in the context of nail-reinforced slopes. The graph shows that slope angle and the angle of internal friction are the most influential parameters in predicting slope stability. These insights are valuable for optimizing nail reinforcement strategies, as they guide practitioners to prioritize these factors during slope design and stabilization efforts. The Random Forest model effectively captures these relationships, demonstrating its robustness and accuracy in slope stability analysis.

Analysis of Variance (ANOVA): The Analysis of Variance (ANOVA) table summarizes the statistical evaluation of the linear regression model used to predict the Factor of Safety (FOS).

TABLE 7

TABLE OF ANOVA (ANALYSIS OF VARIANCE)

SOURCE	DF	SUM OF SQUARES	MEAN SQUARES	F	PR > F	P-VALUES SIGNIFICATION CODES
MODEL	4	17.757	4.439	364.353	<0.0001	***
ERROR	146	1.779	0.012			
CORRECTED TOTAL	150	19.535				

The model includes four predictors, and its performance is assessed by comparing the variability explained by the model to the residual error. The Sum of Squares for the model is 17.757, indicating that the majority of the total variability (19.535) in FOS is explained by the predictors. The Mean Square for the model is 4.439, while for the error it is 0.012, resulting in a high F-statistic value of 364.353. The associated p-value is less than 0.0001, which is highly significant, confirming that the overall model is statistically significant at a 99.9% confidence level. This indicates that the predictors collectively have a strong and significant influence on the FOS. The low error sum of squares (1.779) further suggests that the model provides a good fit to the data.

Conclusion: Random Forest emerged as the best-performing model, achieving a Mean Absolute Error (MAE) of 0.036, a Mean Squared Error (MSE) of 0.002, and the highest R^2 value of 0.979. Its Adjusted R^2 was also excellent at 0.977, indicating superior accuracy and the ability to explain variability in the Factor of Safety (FOS). The Akaike Information Criterion (AIC) and Schwarz's Bayesian Criterion (SBC) values were the lowest at -221.174 and -214.623, respectively, further validating its efficiency. The K-Nearest Neighbors (KNN) model also performed well, with an MAE of 0.064, an MSE of 0.006, and an R^2 of 0.945. The Adjusted R^2 was 0.939, suggesting reliable predictive performance, though slightly lower than Random Forest. Linear Regression showed reasonable performance but was outperformed by the other models. It produced an MAE of 0.091, an MSE of 0.012, and an R^2 of 0.894. The Adjusted R^2 was 0.882, which, while acceptable, indicates less accuracy compared to the other two models. The AIC and SBC values were also higher at -159.463 and -152.913, respectively. Overall, Random Forest outperformed both Linear Regression and KNN across all metrics, proving to be the most effective model for predicting slope stability. The findings highlight its ability to handle complex relationships between input parameters, making it a valuable tool for geotechnical analysis. Although the proposed Random Forest model demonstrated excellent agreement with numerical simulation results, future studies should validate the developed model using real-world landslide case histories and published experimental datasets to further assess its practical applicability under field conditions.

Future Scope: Future studies may incorporate larger datasets generated using Latin Hypercube Sampling or probabilistic parameter distributions to further improve the robustness and generalization capability of the machine learning models.

Author Contributions: Radha Tomar: Conceptualization, numerical modelling, machine learning analysis, manuscript writing, data interpretation.
Smita Tung: Supervision, methodology review, manuscript editing, validation, and final approval of manuscript.

Acknowledgement: I would like to appreciate and extend my sincere gratitude to Department of Civil Engineering, GLA University for providing the research facilities needed to complete this project.
Conflict of Interest: The authors have no conflicts of interest to declare. All co-authors have seen and agree with the contents of the manuscript and there is no financial interest to report.

AI Tool Usage Acknowledgment: Artificial Intelligence (AI)-based language assistance tools were utilized for grammatical refinement, language improvement, and manuscript editing. All technical analysis, numerical simulations, interpretations, and final scientific content were verified and approved by the authors.

Funding: No Funding is received for this project

References:

[1] M. I. Khan and S. Wang, “Slope stability analysis to develop correlations between different soil parameters and factor of safety using regression analysis,” *Polish J. Environ. Stud.*, vol. 30, no. 5, pp. 4021–4030, 2021, doi: 10.15244/pjoes/131203.

[2] F. Benayoun, D. Boumezerane, and S. R. Bekkouche, “Techniques for Optimizing Parameters of Soil Nailed Vertical Cut,” vol. 16, no. 1, pp. 131–145, 2021, doi: 10.1515/sspjce-2021-0010.

[3] S. Lin, H. Zheng, C. Han, B. Han, and W. Li, “Evaluation and prediction of slope stability using machine learning approaches,” *Front. Struct. Civ. Eng.*, vol. 15, no. 4, pp. 821–833, 2021, doi: 10.1007/s11709-021-0742-8.

[4] A. Burman, S. P. Acharya, R. R. Sahay, and D. Maity, “A comparative study of slope stability analysis using traditional limit equilibrium method and finite element method,” *Asian J. Civ. Eng.*, vol. 16, no. 4, pp. 467–492, 2015.

[5] H. Lin, W. Zhong, W. Xiong, and W. Tang, “Erratum: Slope stability analysis using limit equilibrium method in nonlinear criterion (The Scientific World Journal (2014) 2014 (206062)),” *Sci. World J.*, vol. 2015, 2015, doi: 10.1155/2015/419636.

[6] D. ping Deng, L. Li, and L. heng Zhao, “Limit equilibrium analysis for stability of soil nailed slope and optimum design of soil nailing parameters,” *J. Cent. South Univ.*, vol. 24, no. 11, pp. 2496–2503, 2017, doi: 10.1007/s11771-017-3662-y.

[7] Y. A. Nanehkaran *et al.*, “Comparative Analysis for Slope Stability by Using Machine Learning Methods,” *Appl. Sci.*, vol. 13, no. 3, pp. 1–14, 2023, doi: 10.3390/app13031555.

[8] S. K. Singh and D. Chakravarty, “Assessment of Slope Stability Using Classification and Regression Algorithms Subjected To Internal and External Factors,” *Arch. Min. Sci.*, vol. 68, no. 1, pp. 87–102, 2023, doi: 10.24425/ams.2023.144319.

[9] A. Chakraborty and D. Goswami, “Prediction of slope stability using multiple linear regression (MLR) and artificial neural network (ANN),” *Arab. J. Geosci.*, vol. 10, no. 17, pp. 1–11, 2017, doi: 10.1007/s12517-017-3167-x.

[10] T. E. Elahi, M. A. Islam, and M. S. Islam, “Parametric Assessment of Soil Nailing on the Stability of Slopes Using Numerical Approach,” *Geotechnics*, vol. 2, no. 3, pp. 615–634, 2022, doi: 10.3390/geotechnics2030030.

[11] D. T. Bui, H. Moayedi, M. Gör, A. Jaafari, and L. K. Foong, “Predicting slope stability failure through machine learning paradigms,” *ISPRS Int. J. Geo-Information*, vol. 8, no. 9, 2019, doi: 10.3390/ijgi8090395.

[12] X. Feng, S. Li, C. Yuan, P. Zeng, and Y. Sun, “Prediction of Slope Stability using Naive Bayes Classifier,” vol. 22, pp. 941–950, 2018, doi: 10.1007/s12205-018-1337-3.

[13] S. Suman, S. Z. Khan, S. K. Das, and S. K. Chand, “Slope stability analysis using artificial intelligence techniques,” *Nat. Hazards*, vol. 84, no. 2, pp. 727–748, 2016, doi: 10.1007/s11069-016-2454-2.

[14] “Soil Nail Walls Reference Manual,” no. 132085, 2015.

Radha Tomar received her B. Tech degree in Civil Engineering and completed her M.Tech degree in Structural Engineering from GLA University. She is currently pursuing a Ph.D. in Geotechnical Engineering in the Department of Civil Engineering at GLA University. Her research interests include slope stability analysis, geotechnical engineering, numerical modelling, and machine learning applications in civil engineering.

Dr. Smita Tung received her M.Tech, and Ph.D. degrees in Geotechnical Engineering from Jadavpur University. She is currently serving as an Assistant Professor in the Department of Civil Engineering at GLA University. Her research interests include geotechnical engineering, slope stabilization, numerical modelling, and soil reinforcement techniques.