International Journal of Civil and Structural Engineering Research ISSN 2348-7607 (Online)
Vol. 12, Issue 2, pp: (53-58), Month: October 2024 - March 2025, Available at: www.researchpublish.com

Utilizing Boosting Machine Learning Techniques for Slope Stability Prediction

Saurabh Kumar Anuragi

Department of Civil Engineering

Maulana Azad National Institute of Technology, Bhopal, India

DOI: https://doi.org/10.5281/zenodo.14274474

Published Date: 04-December-2024

Abstract: **The complex nature of slope engineering presents challenges in accurately predicting slope stability with traditional methods. Identifying the appropriate techniques for slope stability prediction is essential in mitigating the risks associated with slope failures. This study conducts a thorough analysis of two boosting machine learning models: Adaboost and LightGBM. By evaluating a wide range of hyperparameters, the research aims to discover the optimal settings for each model, ultimately leading to effective solutions.Six potentially relevant features were identified as key indicators for prediction: height (H), pore water ratio (ru), unit weight (Ɣ), cohesion (c), slope angle (β), and angle of internal friction (ϕ). The models were assessed using evaluation indicators such as AUC and accuracy, revealing that LightGBM significantly outperformed the Adaboost model, achieving an impressive AUC of 0.878 and an accuracy of 0.803. Furthermore, real-world engineering examples illustrate the effectiveness of LightGBM as a predictive tool for slope stability. Its enhanced capacity and efficiency in deformation prediction positions it as a leading instrument for accurate forecasting in this field. To deepen the understanding of these models, a comprehensive analysis of parameter sensitivity was also conducted, highlighting the most significant characteristics contributing to reliable slope stability predictions.**

Keywords: **adaboost, lightgbm, slope stability, finite element, machine learning.**

I. INTRODUCTION

The complexities inherent in the physical state of soil pose significant challenges in accurately estimating slope stability. The increasing frequency of slope failures, which result in considerable economic and social repercussions, has garnered the attention of both researchers and engineers. To mitigate or prevent such adverse outcomes, it is essential to conduct comprehensive slope stability analyses and to implement effective stabilization measures. A deeper understanding of the mechanisms that contribute to slope failure is critical for successfully addressing these challenges. Slope engineering is characterized by its complexity, nonlinearity, dynamism, and inherent uncertainties. Various geological and engineering factors—such as unpredictability, fuzziness, and variability—significantly influence slope stability. It is important to recognize that the relationship between slope stability and its influencing factors is predominantly non-linear. Current trends in slope stability research indicate a shift from traditional deterministic approaches towards a more holistic understanding of the uncertainties arising from the diverse range of slope characteristics. Traditional methods, including the limit equilibrium method, discontinuous deformation analysis, and finite element method, often fall short in accuracy due to the intricate mechanisms that affect slope stability [1, 2]. Nonetheless, ongoing efforts in numerical and analytical modelling aim to reduce potential losses by enabling precise predictions, thereby facilitating the implementation of appropriate preventive actions.

In recent years, advancements in computational techniques have led researchers to increasingly employ machine learning as a robust alternative for slope stability analysis. These methods evaluate slope stability by analyzing features such as slope geometry and material properties, producing significant results. For instance, Lin et al. [3] conducted a comprehensive comparative study of 11 machine learning models, focusing on six critical slope factors. Samui [4] explored the application of support vector machines, utilizing them to predict the factor of safety as a regression model while also classifying the slope status. Similarly, Cheng et al. [5] integrated the K-nearest neighbor method with a Bayesian framework to enhance slope stability predictions. Fattahi [6] adapted three models of the adaptive neuro-fuzzy inference system (ANFIS), namely the subtractive clustering method (SCM), grid partitioning (GP), and fuzzy c-means clustering (FCM), for accurate factor of safety predictions. Hoang et al. [7] undertook a comparative analysis utilizing advanced machine learning methods, including least squares support vector machines (LSSVM), radial basis function neural networks (RBFNN), and extreme learning machines (ELM), for slope stability evaluation. Das et al. [8] applied a differential evolution neural network to carry out slope stability analysis, developing both classification and regression models. Additionally, Manouchehrian et al. [9] created a regression model for predicting slope stability using genetic algorithms (GA). Erzin et al. [10] conducted a comparative study aimed at predicting the factor of safety (FOS) of homogeneous finite slopes through multiple regression (MR) and artificial neural networks (ANN). Qi et al. [11] proposed and compared six artificial intelligence approaches, including logistic regression (LR), random forest (RF), support vector machine (SVM), gradient boosting machine (GBM), decision tree (DT), and multi-layer perceptron neural network (MLPNN), integrating the firefly algorithm (FA) for hyper-parameter tuning. Karir et al. [12] investigated various machine learning models, such as gradient boosting, extreme gradient boosting (XGB), support vector regressors (SVR), random forests (RF), and artificial neural networks (ANN) for predicting factors of safety.

Since, there are limited studies that focus on boosting machine learning techniques for slope stability prediction. Therefore, it is necessary to explore additional techniques that are better suited for analyzing nonlinear slope behavior. Additionally, there is no comprehensive comparison of classifier boosting algorithms for predicting slope stability. To improve the accuracy of predicting nonlinear slope behavior and establish a simple model that can be widely utilized, it is essential to continue exploring boosting algorithms that are better tailored for analyzing nonlinear slope behavior.

II. MATERIALS AND METHODOLOGY

A. Dataset Preprocessing and Visualization

In developing a classification model for slope stability, it is imperative to identify and select features that have a significant influence on stability outcomes. This process involves employing strategic feature selection principles that mitigate dimensionality-related challenges and enhance the model's efficiency. By focusing on essential features, we can minimize computational complexity and ensure that the model highlights the most critical factors affecting slope stability, thereby improving overall predictive performance. The features of particular importance in this analysis include pore water ratio (ru), height (H), unit weight (Υ), cohesion (c), slope angle (β), and angle of internal friction (ϕ). This study analyzes a dataset consisting of 444 slope stability cases to predict slope status, classified as either stable (1) or unstable (0) as shown in Figure 1. To ensure the integrity of the analysis, the dataset is normalized to eliminate discrepancies related to scale, units, and distributions. This normalization process is essential for enhancing model accuracy and its ability to generalize effectively to new and unseen data.

$$y_{normalization} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where, y is a normalized input parameter, x is the original input parameter, x_{max} is the maximum parameter and x_{min} is the minimum parameter. The distribution and variability of each input variable on slope status are shown in Figure 2.

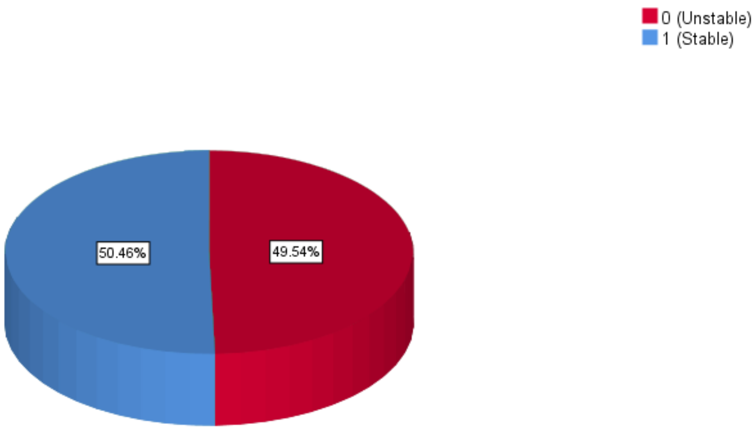


Figure 1: Dataset Pie Chart

Figure 2: Correlation Matrix of Dataset

The violin plots illustrated in Figure 3 provide a comprehensive overview of the distribution and density of the dataset across various categories. In a violin plot, the width at any given point indicates the density of the data, with thicker sections representing areas of high density and thinner sections indicating lower density. The median value of the data is shown by the horizontal line within each violin. Upon analyzing these plots, it is evident that the variables Ύ, ϕ, β, and ru exhibit a wide distribution pattern, which is reflected in the diverse shapes of their corresponding violins. This suggests that the data points for these variables are quite dispersed. Conversely, the variables c and H demonstrate a more densely clustered distribution, as evidenced by their narrower violin shapes, highlighting a higher frequency of data points at specific values. This comparative analysis of the violin plots enhances our understanding of the underlying data distribution in the dataset.

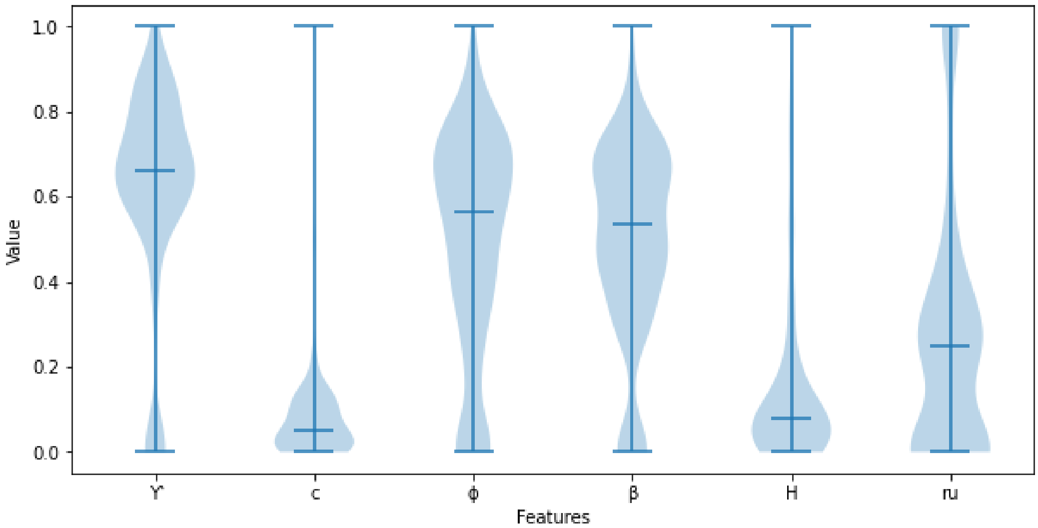


Figure 3: Violin Plots showing distribution of slope cases

B. Model Development and Optimization

This study examines the use of two boosting machine learning algorithms, Adaboost and LightGBM, for slope stability classification. To evaluate model performance, the dataset is divided into a training set (70% or 311 cases) and a testing set (30% or 133 cases). The training set is used to train the model and fine-tune hyperparameters, while the testing set assesses the model's ability to generalize to new data. Various hyperparameter combinations are explored to identify the optimal settings, which are then applied for predictions on unseen data. The model's performance is evaluated using metrics such as Area Under the Curve (AUC), Accuracy, and Sensitivity. The AUC provides a comprehensive view of predictive capacity, while high sensitivity indicates effectiveness at detecting positive instances. Hyperparameter optimization results and predictions are detailed in Table 1.

TABLE I: OPTIMAL HYPERPARAMETER SETTINGS OF EACH MODEL

Model	Hyperparameters	Optimal Hyperparameters	AUC	Accuracy	Sensitivity
AdaBoost	learning_rate = [0.1, 0.01]	0.01	0.762	0.706	0.781
	n_estimators = [50,100,200,300,400,500]	200			
LightGBM	learning_rate = [0.1, 0.01]	0.01	0.878	0.803	0.828
	n_estimators = [50,100,200,300,400,500]	400			

III. RESULTS AND DISCUSSION

This study employs AdaBoost and LightGBM classifiers to evaluate their performance in predicting slope stability. It is important to note that the performance of a classifier is significantly influenced by the Area Under the Curve (AUC), where a value of 1.0 indicates optimal performance. The ROC curves of the classification models (Figure 4) reveal that the AUC for AdaBoost is 0.762, while LightGBM has an AUC of 0.878. The differences in AUC values among the classifiers are attributed to the variations in their underlying algorithms, model complexity, and their effectiveness in capturing the relationships between the features and the target variable. The results indicate that LightGBM demonstrates superior AUC values, suggesting better discriminatory ability and overall performance compared to AdaBoost. Figure 5 presents the confusion matrix for both classification models, showing that the total number of misclassifications for AdaBoost is 32, while for LightGBM, it is 28. Sensitivity analysis, illustrated in Figure 6, further highlights the disparities in performance among the classifiers. From Figure 6, it can be inferred that the LightGBM model exhibits a high sensitivity with a score of 0.828, whereas the AdaBoost model has a comparable sensitivity score of 0.781. These variations emphasize the importance of selecting the appropriate classifier based on the specific characteristics of the dataset and the problem being addressed in order to achieve optimal classification results. Overall, it is evident from the results that the LightGBM classification model outperforms the AdaBoost classifier due to its superior ability to discriminate and better rank positive samples, despite having lower sensitivity.

IV. CONCLUSION

This study presents a comparative analysis of two boosting machine learning classifiers, specifically Adaboost and LightGBM, in assessing the stability of 444 slope cases. The analysis utilizes six features—H, ru, β, c, ϒ, and ϕ—for the prediction and generalization of classification models. Based on the evaluation of receiver operating characteristic (ROC) curves, the LightGBM classifier exhibited a significantly higher area under the curve (AUC) compared to the Adaboost classifier. This finding indicates that LightGBM demonstrates superior overall performance in distinguishing between stable and unstable slopes. Furthermore, LightGBM's model showcased high sensitivity in its predictions, effectively identifying true positives while minimizing false negatives. This attribute is particularly critical in applications where overlooking a positive case can lead to serious consequences. The results highlight the complexity involved in evaluating slope stability and emphasize the necessity of considering a comprehensive range of characteristics rather than relying solely on individual metrics for precise forecasting. As such, boosting classifiers like LightGBM represent a compelling alternative for slope stability predictions, especially in scenarios that demand interpretability and the management of imbalanced datasets.

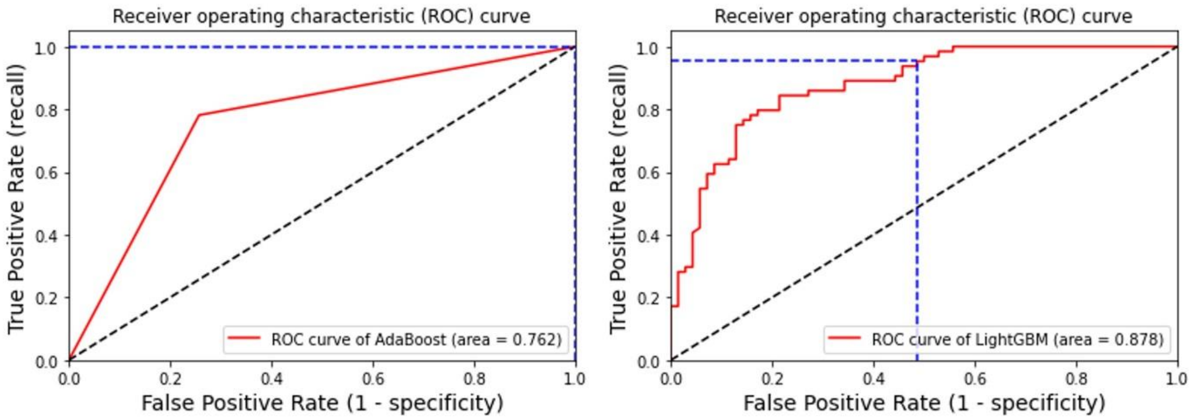
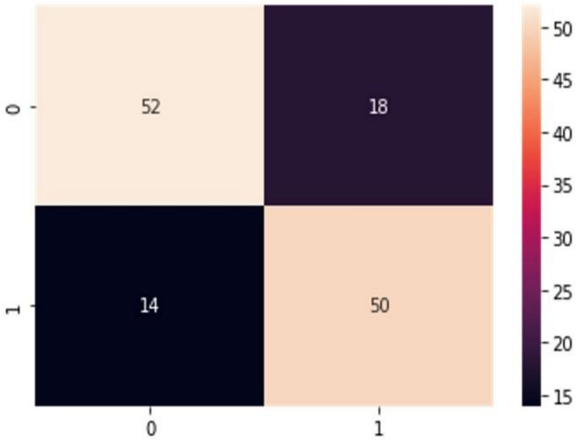
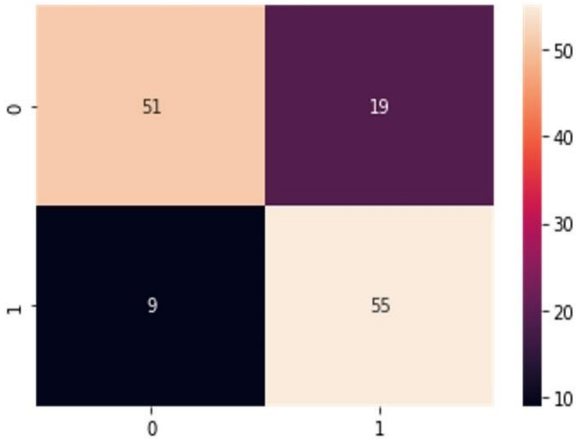


Figure 4: ROC Curves of of classification models on testing dataset



(1) AdaBoost



(2) LightGBM

Figure 5: Confusion Matrix of classification models

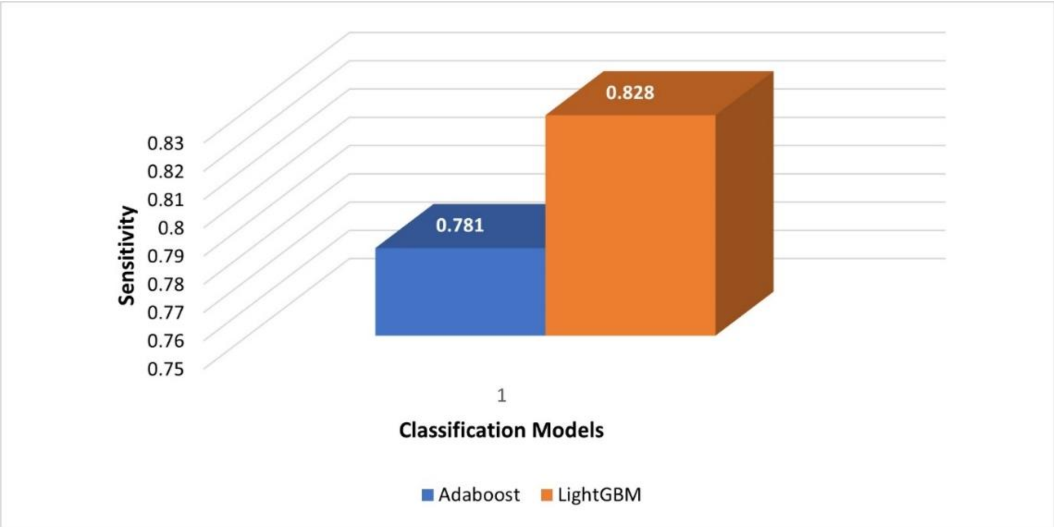


Figure 6: Sensitivity of classification models

REFERENCES

[1] Hammouri, N. A., Malkawi, A. I. H., & Yamin, M. M. Stability analysis of slopes using the finite element method and limiting equilibrium approach. *Bulletin of Engineering Geology and the Environment*, *67*, 471-478, 2008.

[2] Duncan, J. M. State of the art: limit equilibrium and finite-element analysis of slopes. *Journal of Geotechnical engineering*, *122*(7), 577-596, 1996.

[3] Lin, S., Zheng, H., Han, C., Han, B., & Li, W. Evaluation and prediction of slope stability using machine learning approaches. *Frontiers of Structural and Civil Engineering*, *15*(4), 821-833, 2021.

[4] Samui, P. Slope stability analysis: a support vector machine approach. *Environmental Geology*, *56*, 255-267, 2008.

[5] Cheng, M. Y., & Hoang, N. D. Slope collapse prediction using Bayesian framework with k-nearest neighbor density estimation: case study in Taiwan. *Journal of Computing in Civil Engineering*, *30*(1), 04014116, 2016.

[6] Fattahi, H. Prediction of slope stability using adaptive neuro-fuzzy inference system based on clustering methods. *Journal of Mining and Environment*, *8*(2), 163-177, 2017.

[7] Hoang, N. D., & Bui, D. T. Slope stability evaluation using radial basis function neural network, least squares support vector machines, and extreme learning machine. In *Handbook of neural computation* (pp. 333-344). Academic Press, 2017.

[8] Das, S. K., Biswal, R. K., Sivakugan, N., & Das, B. Classification of slopes and prediction of factor of safety using differential evolution neural networks. *Environmental Earth Sciences*, *64*, 201-210, 2011.

[9] Manouchehrian, A., Gholamnejad, J., & Sharifzadeh, M. Development of a model for analysis of slope stability for circular mode failure using genetic algorithm. *Environmental Earth Sciences*, *71*, 1267-1277, 2014.

[10] Erzin, Y., & Cetin, T. The prediction of the critical factor of safety of homogeneous finite slopes using neural networks and multiple regressions. *Computers & Geosciences*, *51*, 305-313, 2013.

[11] Qi, C., & Tang, X. Slope stability prediction using integrated metaheuristic and machine learning approaches: A comparative study. *Computers & Industrial Engineering*, *118*, 112-122, 2018.

[12] Karir, D., Ray, A., Bharati, A. K., Chaturvedi, U., Rai, R., & Khandelwal, M. Stability prediction of a natural and man-made slope using various machine learning algorithms. *Transportation Geotechnics*, *34*, 100745, 2022.