SSRG International Journal of Civil Engineering
ISSN: 2348-8352/ https://doi.org/10.14445/23488352/IJCE-V11I5P116

Volume 11 Issue 5, 168-180, May 2024

Original Article

Comparative Analysis of Ensemble Learning Approaches for Slope Stability Prediction

Saurabh Kumar Anuragi[1], D. Kishan[2], S.K. Saritha[3]

[1,2,3]*Maulana Azad National Institute of Technology, Madhya Pradesh, India.*

[1]*Corresponding Author : saurabh.1399@gmail.com*

Received: 15 March 2024 Revised: 17 April 2024 Accepted: 16 May 2024 Published: 31 May 2024

Abstract *- Due to the complex nature of slope engineering, accurately predicting slope stability using traditional techniques can be difficult. It is, therefore, crucial to identify the correct technique for slope stability prediction in order to prevent disasters caused by slope failures. This study provides a comprehensive analysis of three ensemble models: Random Forest (RF), CatBoost, and Stacking. The models were evaluated for a wide range of hyperparameters to find the optimal settings for each model, resulting in the best solution. Six potentially relevant features, including height (H), pore water ratio (ru), unit weight (Υ), cohesion (c), slope angle (β) and angle of internal friction (ϕ), were selected as prediction indicators. The generalization ability of classification models is enhanced by using a 5-fold CV. Evaluation indicators such as AUC and accuracy were analyzed, and Stacking was found to outperform the other ensemble models with the highest AUC of 0.898 and accuracy of 0.854. The analysis of engineering examples shows that Stacking is a highly effective tool for predicting slope stability due to its ability to enhance capacity and efficiency in deformation prediction models. This makes it the most accurate tool available for forecasting slope stability. In addition, a comprehensive analysis of parameter sensitivity was conducted to determine the most significant characteristics for predicting slope stability.*

Keywords *- Logistic regression, CatBoost, Slope stability, Random Forest, Hyperparameters, Optimization, Finite Element Method, Limit Equilibrium.*

1. Introduction

Due to complexities in the physical state of the soil, the accurate estimation of slope stability is a difficult problem. The increasing slope failures that led to enormous economic and social losses were brought attention of researchers and engineers. In order to mitigate or prevent such damages, it is imperative to conduct a thorough slope stability analysis and implement appropriate stabilization measures. A deeper comprehension of the mechanisms contributing to slope failure is essential for effectively eradicating such events. Slope engineering is a system that is complicated, non-linear, dynamic, and not without uncertainty. Various geological and engineering aspects, including unpredictability, fuzziness, and variability, as well as other uncertain qualities, have a comprehensive impact on its stability.

It is important to note that the relationship between slope stability and the elements that influence it is significantly non-linear. A common tendency in slope stability research is a shift away from traditional deterministic notions and toward a more complete recognition of the uncertainty caused by the broad range of slope parameters. Traditional approaches such as the limit equilibrium method [1–3], discontinuous deformation analysis [4, 5], and finite element method [6–9] are extensive and inaccurate due to the complicated mechanism that influences slope stability. However, efforts are made to minimize the losses by numerical and analytical modelling such that appropriate actions can be taken by making accurate predictions.

In recent years, due to advancements in computational techniques, many researchers have started employing machine learning techniques as an alternative method for slope stability analysis. These techniques evaluate slope stability based on parameters such as slope geometry and slope material properties, giving remarkable results. Lin et al. [10] carried out a comparative study of 11 ML models considering six slope factors. Samui [11] investigated the use of support vector machines to predict the factor of safety as a regression model and slope status as a classification model. Cheng et al. [12] utilized the K-Nearest Neighbour integrated with the Bayesian framework for slope stability prediction. Fattahi [13] adapted three neuro-fuzzy inference system (ANFIS) models, including the Subtractive Clustering Method (SCM), Grid Partitioning (GP) and Fuzzy C-means Clustering Method (FCM) for the prediction of FOS.

Hoang et al. [14] conducted a comparative study of slope stability prediction using advanced machine learning methods, including Least Squares Support Vector Machines (LSSVM), Radial Basis Function Neural Networks (RBFNN) and Extreme Learning Machine (ELM). Das et al. [15] applied a differential evolution neural network for slope stability analysis, developing both classification and regression models. Manouchehrian et al. [16] developed a regression model for the prediction of slope stability using a Genetic Algorithm (GA). A comparison study was carried out by Erzin et al. [17] to predict the Factor of Safety (FOS) of homogeneous finite slopes by utilizing Multiple

Regression (MR) and Artificial Neural Network (ANN). Qi et al. [18] proposed and compared six Artificial Intelligence (AI) approaches, including Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), Gradient Boosting Machine (GBM), Decision Tree (DT) and Multi-Layer Perceptron Neural Network (MLPNN) for slope stability prediction integrated with Firefly Algorithm (FA) for the hyper-parameter optimization. Karir et al. [19] studied various ML models, including gradient boosting, Extreme Gradient Boost (XBG), Support Vector Regressor (SVR), Random Forest (RF) and Artificial Neural Network (ANN) for factor of safety prediction. Recent studies have documented numerous hybrid models used for analyzing slope stability, including different types of data [20–25].

All of the machine learning models that were stated above help us better understand slope behavior and the complexity of slope. Nevertheless, the complexity of the problem continues to vary for the same data type, and this is because each model has its own set of limits.

The purpose of slope stability analysis is to generate better prediction results for machine learning algorithms that are more recent and robust. Because of this, it is essential to discover powerful and high-accuracy ensemble learning algorithms in order to achieve results that are superior to those obtained by standalone algorithms in terms of slope stability. The combination of multiple ML models results in the formation of a "strong learner" that is more comprehensive through the process of ensemble learning.

It is possible to generate more accurate prediction results with ensemble learning, which also can improve generalization performance and broader application applicability [26-28]. Kardani et al. [29] applied a hybrid stacking ensemble approach, including an Artificial Bee Colony (ABC) algorithm for enhancing the prediction of slope stability. Wang et al. [30] developed a highly effective reliability analysis method using Extreme Gradient Boosting (XGBoost) to assess the probability of slope failure in earth dams. Zhang et al. [31] developed a model for the prediction of Factor of Safety (FOS) against basal heave for deep-braced excavations using Random Forest Regressor (RFR) and Extreme Gradient Boosting (XGBoost).

The results above suggest that ensemble learning algorithms present a promising method for predicting slope stability. However, there is a scarcity of studies that concentrate on ensemble algorithm classifiers for this particular application. Therefore, it is imperative to investigate additional ensemble classifiers that are better suited for examining non-linear slope behavior.

Furthermore, there is no comprehensive evaluation of classifier ensemble algorithms for predicting slope stability. To enhance the accuracy of forecasting non-linear slope behavior and develop a straightforward model that can be extensively adopted, it is crucial to continue exploring ensemble algorithms that are more effectively designed for analyzing non-linear slope behavior.

Therefore, the objective of this study is to conduct a comparative analysis among different ensemble learning classifiers with the specific goal of predicting slope stability. The study will explore and evaluate the performance of Random Forest (RF), CatBoost and Stacking Ensemble Learning classifiers. These particular ensemble learning classifiers have been chosen due to their increasing popularity and application within engineering disciplines.

Despite their widespread use, there remains a gap in the literature regarding a thorough comparison of these algorithms for slope stability prediction. Hence, this research seeks to address this gap by providing a detailed assessment of their effectiveness and suitability. This study is outlined as follows: Section 2 provides a concise introduction to the ensemble learning classifiers. Section 3 introduces the dataset of slopes and the techniques used to categorize their stability. Section 4 presents the outcomes and analysis derived from the performance criteria. The study's result is presented in Section 5.

2. Ensemble Learning Approaches

Ensemble learning is a machine learning technique that integrates predictions from numerous independent models (learners) to improve overall prediction performance, as shown in Figure 1. Instead of relying on a single model, ensemble techniques create more accurate predictions by using the diversity and complementary qualities of numerous models.

The fundamental idea of ensemble learning is that when a set of weak learners is joined, it can generate a strong learner that outperforms any individual model in the ensemble. In addition to decision trees, linear models, neural networks, and any other sort of model that is capable of learning from data, these weak learners can also be neural networks.

Ensemble learning has three topologies based on how base learners are combined: parallel, serial, and hybrid. RF represents the parallel structure in which basic learners function independently, and their predictions are combined via a parallel method. The serial structure, represented by boosting (CatBoost), trains base learners progressively, with each succeeding learner focusing on rectifying the errors of the prior ones.

Stacking represents a hybrid structure that includes aspects of both parallel and serial architectures in order to utilize the strengths of several base learners. In this study, some common classifiers and ensemble methods were employed, which were effectively implemented in various fields of geotechnical engineering, delivering excellent results [18, 26, 32, 33].

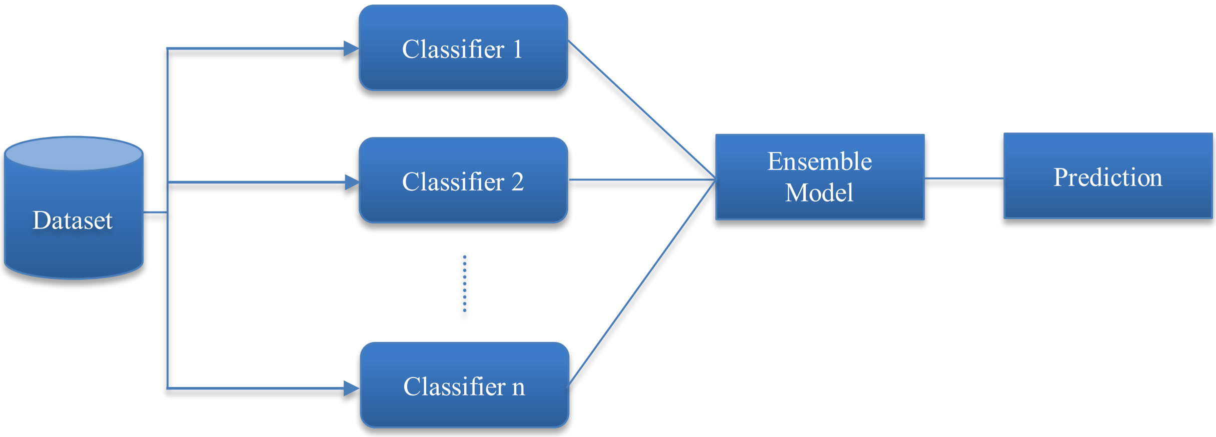

Fig. 1 Schematic diagram of ensemble learning

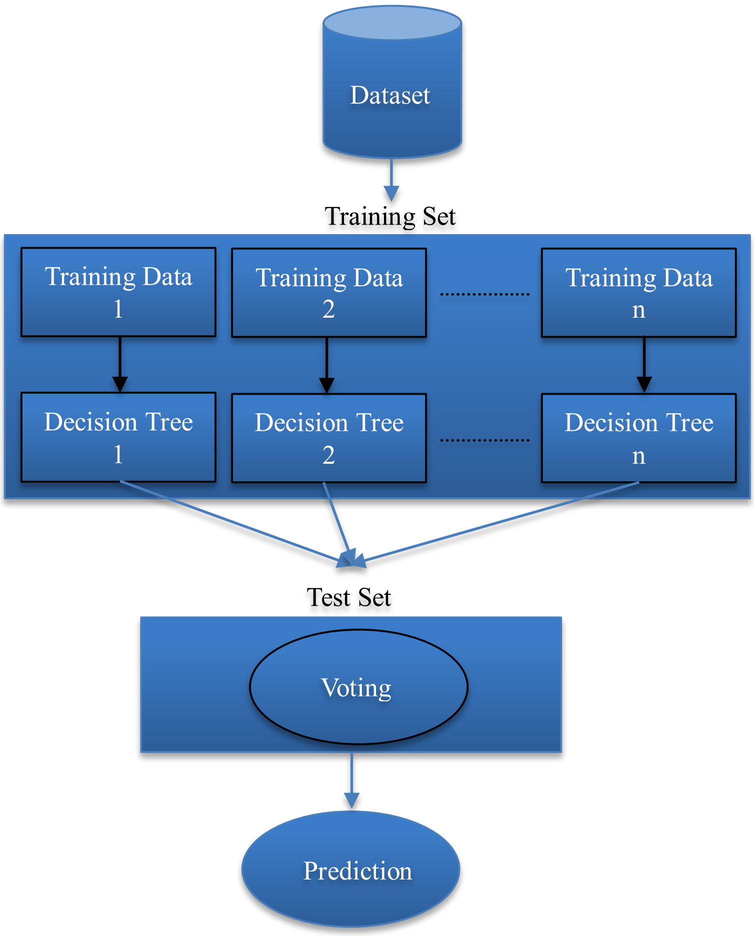

Fig. 2 Schematic diagram of random forest

2.1. Random Forest

The Random Forest (RF) algorithm is a type of ensemble method based on bagging and decision trees used in machine learning. The structure diagram of a random forest is depicted in Figure 2. With a bagging approach, the technique first extracts m multiples of the training data, after which multiple decision trees are built on distinct subsets of the training data. Their predictions are combined to form a final prediction based on the performance (based on the score derived from the number of votes in the classification tree) of multiple decision trees [34–36]. Every decision tree in the forest is constructed by utilizing a random subset of the independent variables and a random subset of the observations. This minimizes the correlation between the trees while increasing their prediction effectiveness. The main advantage of RF is its ability to handle a large number of input variables and complex interactions between them. Another benefit of the random forest technique is variable significance estimation. This is accomplished by assessing the loss in prediction accuracy as a variable is randomly permuted in the data. Therefore, RF has been one of the highly accurate and most robust algorithms in many studies. [37-40].

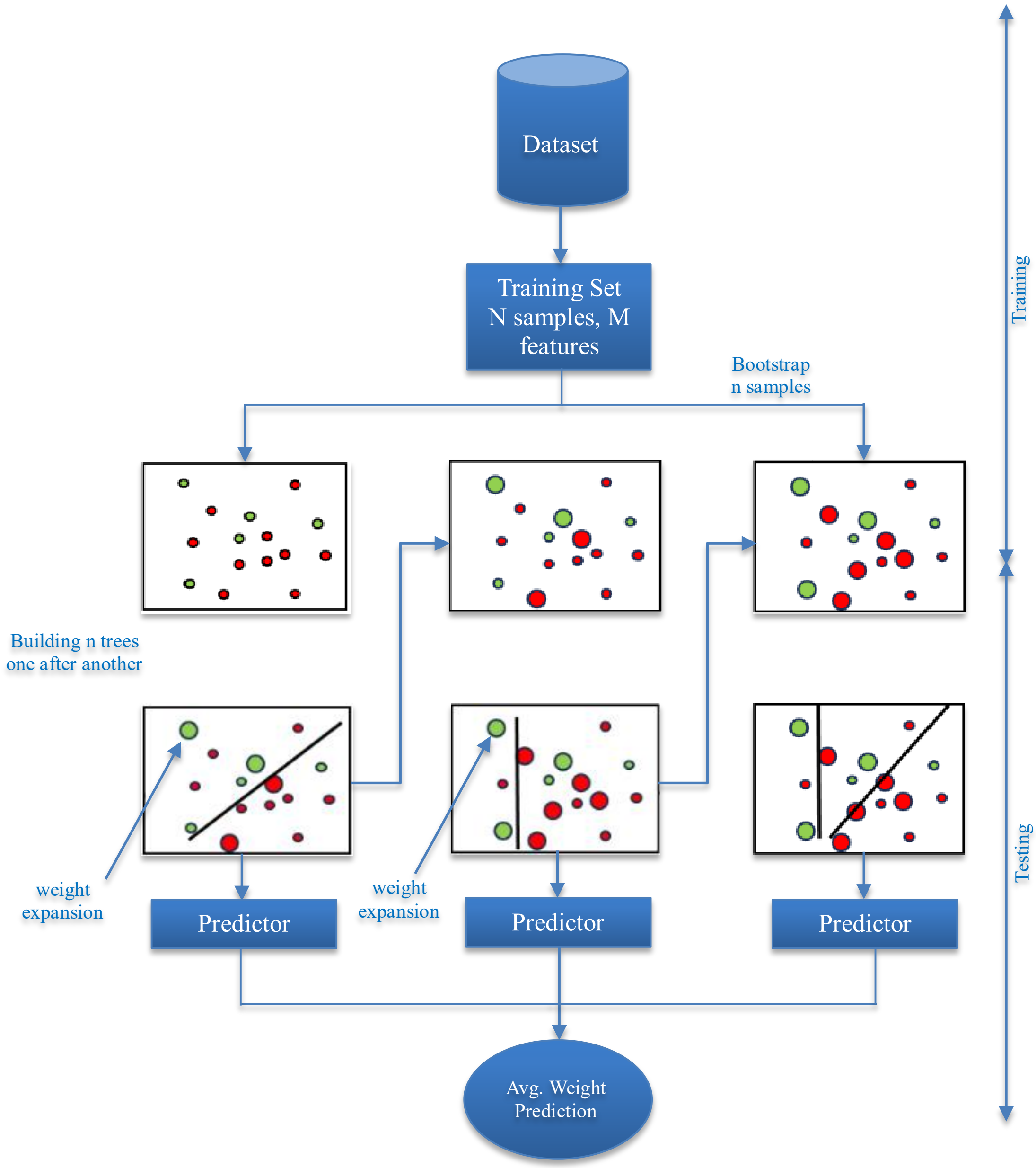

Fig. 3 Schematic diagram of CatBoost

2.2. CatBoost

CatBoost, also known as "Categorical Boosting", is a new Gradient Boosting Decision Tree (GBDT) algorithm specifically designed for handling categorical variables in datasets [41, 42]. Compared to other gradient boosting algorithms (Figure 3), CatBoost differs in many ways. Firstly, it has built-in support for categorical variables, eliminating the need for preprocessing like one-hot encoding. This enhances efficiency during the training process. Secondly, it utilizes the gradient boosting algorithm to construct an ensemble of decision trees sequentially, with each tree correcting the errors of its predecessors. Additionally, CatBoost employs various optimizations, such as ordered boosting, to improve training speed and memory efficiency. It also incorporates regularization techniques like L2 regularization to prevent overfitting during training. Moreover, CatBoost provides built-in support for cross-validation, simplifying the evaluation of model performance and hyperparameter tuning. It is highly regarded for its high performance across a wide range of datasets, particularly those containing both numerical and categorical features. Recently, CatBoost has been widely used in several fields, such as finance [43], health care [44], and academics [45] and has been applied to other forms of data, including time series data [46]. CatBoost is remarkable for its handling of categorical variables, in which the original variable is replaced with a set of binary features corresponding to each category. The approach, as highlighted by [41], provides an advantage by utilizing random combinations to estimate leaf values while selecting tree structures. This technique successfully reduces overfitting, which is common in conventional gradient boosting algorithms. It is worth mentioning that CatBoost utilizes binary decision trees as its fundamental predictor, which greatly enhances its strong performance across various domains and datasets [47].

2.3. Stacking

Stacking, or stacked generalization [48], is an effective ensemble learning method that combines the predictions from multiple base models using a meta-learner or higher-level model. The primary goal of stacking is to learn how to effectively combine the predictions of the basic models to increase overall performance. The stacking algorithm offers a simple structure, high performance, and great classification capabilities. The stacking structure, with the first layer consisting of base models and the second layer comprising the meta-model. Stacking offers enhanced non-linear expression capabilities compared to individual prediction models. This is achieved by utilizing the predicted values from each model in the first layer as input features for the subsequent layer, resulting in a reduction of generalization error. Subsequently, a meta-learner, also referred to as a "blender" or "stacker," is developed to combine these predictions into final predictions. The meta-learner can encompass any machine learning algorithm, such as linear regression or neural networks. To maximize combination weights and minimize error, the meta-learner is trained on a validation set with base model predictions. Finally, the trained base models provide predictions on the test set, which are fed back into the trained meta-learner to produce the final predictions.

The stacking algorithm in this study uses Logistic Regression (LR), Support Vector Classifier (SVC), CatBoost, Random Forest (RF) and K-Nearest Neighbor (KNN) as primary learners (Base Models) with Logistic Regression (LR), functioning as the secondary learner (Meta-Model). Stacking is a flexible method that improves the accuracy of predictions by combining high-level and low-level models.

3. Materials and Methodology

3.1. Data Preprocessing and Visualization

When building a classification model for slope stability, it is crucial to select the features that have a significant impact on slope stability. This involves following certain feature selection principles. To train the model efficiently and prevent dimensionality-related problems, we first select the essential features from the available feature set. The use of this strategic selection helps reduce the likelihood of experiencing issues that are related to high-dimensional data. Second, the learning process's complexity is reduced by eliminating features that are irrelevant to succeeding learning stages. This not only improves the efficiency of the computations but also ensures that the model emphasizes the most important features of slope stability, which in turn improves the model's overall prediction performance. Applying these principles carefully in feature selection is crucial for creating a strong classification model designed for the complexities of slope stability evaluation. At present, the magnitude of slope angle (β), pore water ratio (ru), height (H), unit weight (ϒ), cohesion (c) and angle of internal friction (ϕ) have extensive use in slope stability prediction. This study uses 444 slope stability cases for the prediction of slope status [stable (1) or unstable (0)] collected from [49], modelled as a classification problem (Figure 4). The dataset is normalized using Equation (1) prior to conducting the analysis. It helps to remove the effects of different scales, units, and distributions, which can lead to biased model training and reduced accuracy. By scaling the data to a common range, we can improve the model's ability to generalize and make accurate predictions on new, unseen data.

$$y_{\text{normalization}} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

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 5. The violin plots of the dataset are shown in Figure 6. The violin plot displays the distribution and density of a dataset across different categories or groups. The width of the violin at any point indicates the density of data at that point. The thicker parts of the violin represent regions of high density, while the thinner parts represent regions of low density. The horizontal line inside the violin represents the median value of the data. By examining the violin plots, the variables ϒ, ϕ, β, and ru have a wide distribution pattern, as seen from the spread of the violin plot shapes. This indicates that the data points for these variables are widely dispersed. On the other hand, the variables c and H exhibit a densely clustered distribution with a higher frequency of data points at certain values, as seen from the narrow violin shapes.

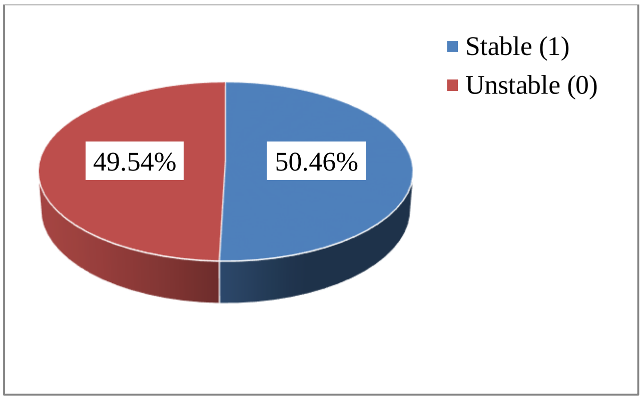

Fig. 4 Dataset pie chart

For maximum classification accuracy, it is ideal that each value of every feature on the diagram is associated with only one class label, either stable or unstable. Figure 7 illustrates the classification of slope stability using various indicators. The graphic illustrates cases where a single indicator value corresponds to both slope classifications. One possible explanation for this behavior is that the data do not exhibit linear separability, which makes it difficult to establish distinct bounds for the feature values.

3.2. Model Development and Optimization

This study investigates the applicability of three ensemble learning algorithms (Stacking, CatBoost and RF) in slope stability classification. When dealing with supervised classification problems, it is essential to evaluate the performance of classification models on new data to determine their capacity for generalization. To accomplish this, the dataset is usually split into two subsets: the training set, which contains most of the data, is used to train the model and optimize hyperparameters, and the testing set, which is a smaller subset of the dataset, is exclusively used to assess the model's capacity to generalize to new, unseen instances.

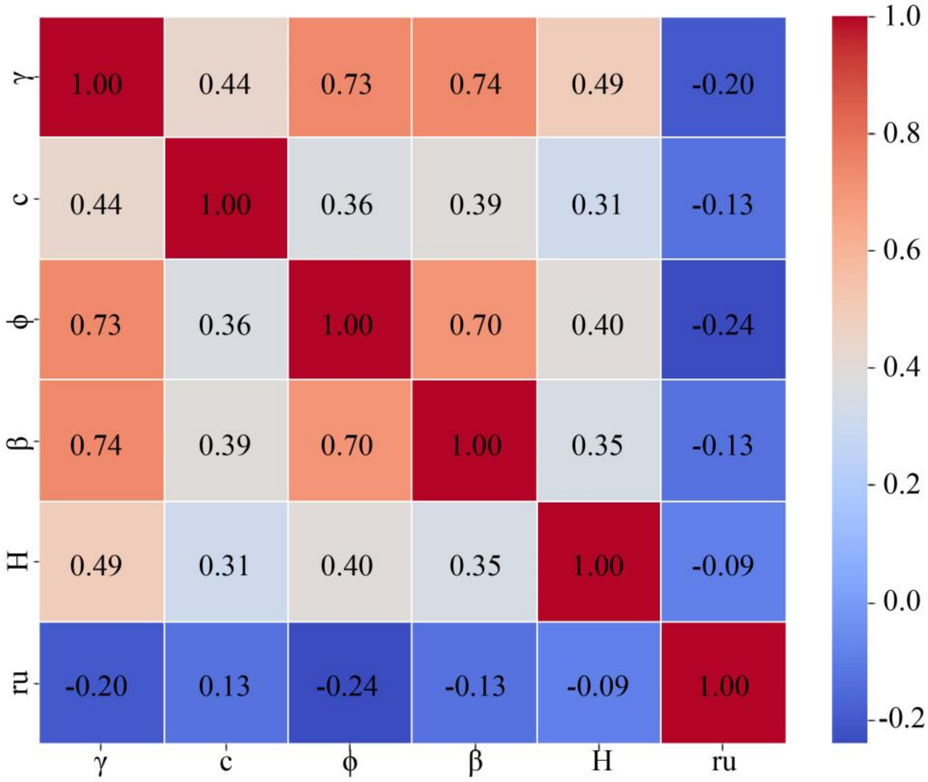

Fig. 5 Correlation matrix of dataset

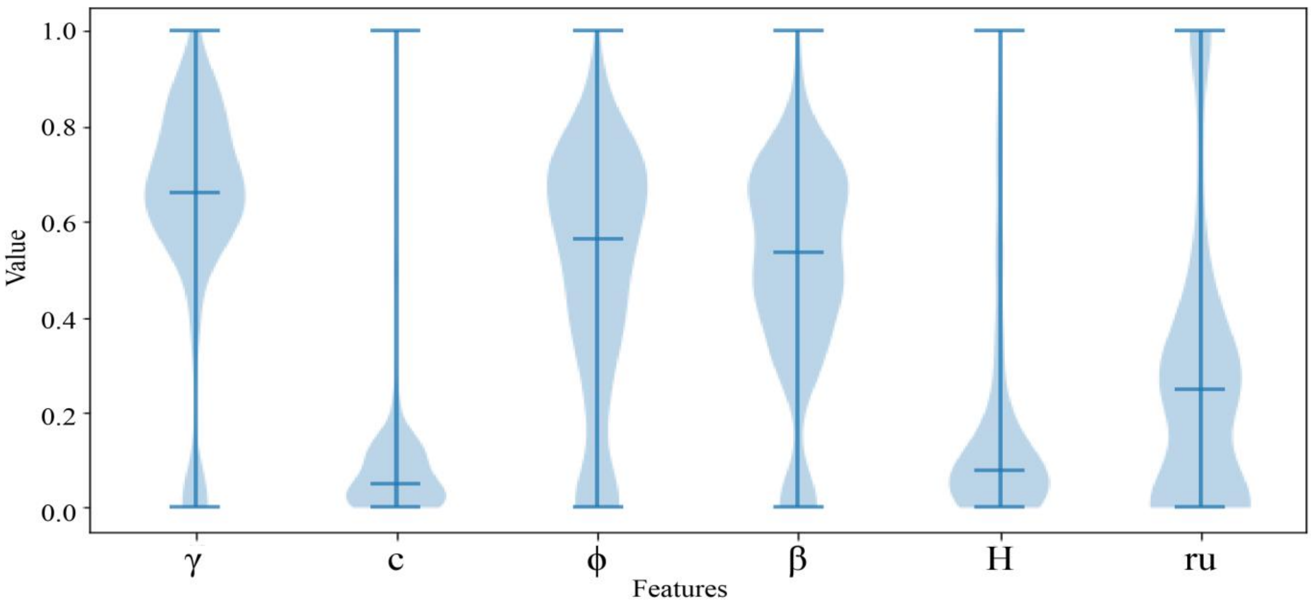

Fig. 6 Violin plots showing the distribution of slope cases

Within this study's framework, around 70% of the original dataset, which equates to 311 cases, is selected as the training set. Approximately 30% of the dataset, which is equivalent to 133 cases, has been selected as the testing set. This split ensures that the model is trained on a sufficiently wide range of datasets, as well as a separate, independent subset for rigorous evaluation of its performance on unseen data. The training process for each ensemble model involves exploring various combinations of hyperparameters, as detailed in Table 1. By systematically evaluating these combinations, the optimal hyperparameters are identified to achieve the best model performance. These optimal hyperparameters are then utilized for making predictions on unseen data, ensuring the model's effectiveness in real-world scenarios. Ensemble learning models are evaluated through a process known as 5-fold cross-validation. This involves randomly dividing the training data into five subsets of equal size, where four subsets are used for training the model, and the remaining subset is used for verifying the model's performance. This process is repeated five times, with each subset being used as the test set in turn. By averaging the results across the five iterations, we can determine the performance of the ensemble learning method on the training data. This technique helps to ensure that the model is robust and can generalize well to new data. The Area Under the Curve (AUC), Accuracy and Sensitivity metrics are used to evaluate the ensemble learning algorithm's overall performance across both the training and testing sets. The AUC metric gives a comprehensive evaluation of the model's predictive capacity, taking into account its ability to discriminate between classes as well as its robustness over varied thresholds. By analyzing the AUC on both training and testing data, the ensemble learning algorithm's ability to capture underlying patterns and generalize to previously encountered instances can be extensively reviewed and validated. Sensitivity provides insight into how well a model can detect positive instances or events. A high sensitivity value indicates that the model has a low rate of false negatives, meaning it is effective at correctly identifying positive instances. The hyperparameter optimization settings for all the models, along with their prediction results, are shown in Table 1.

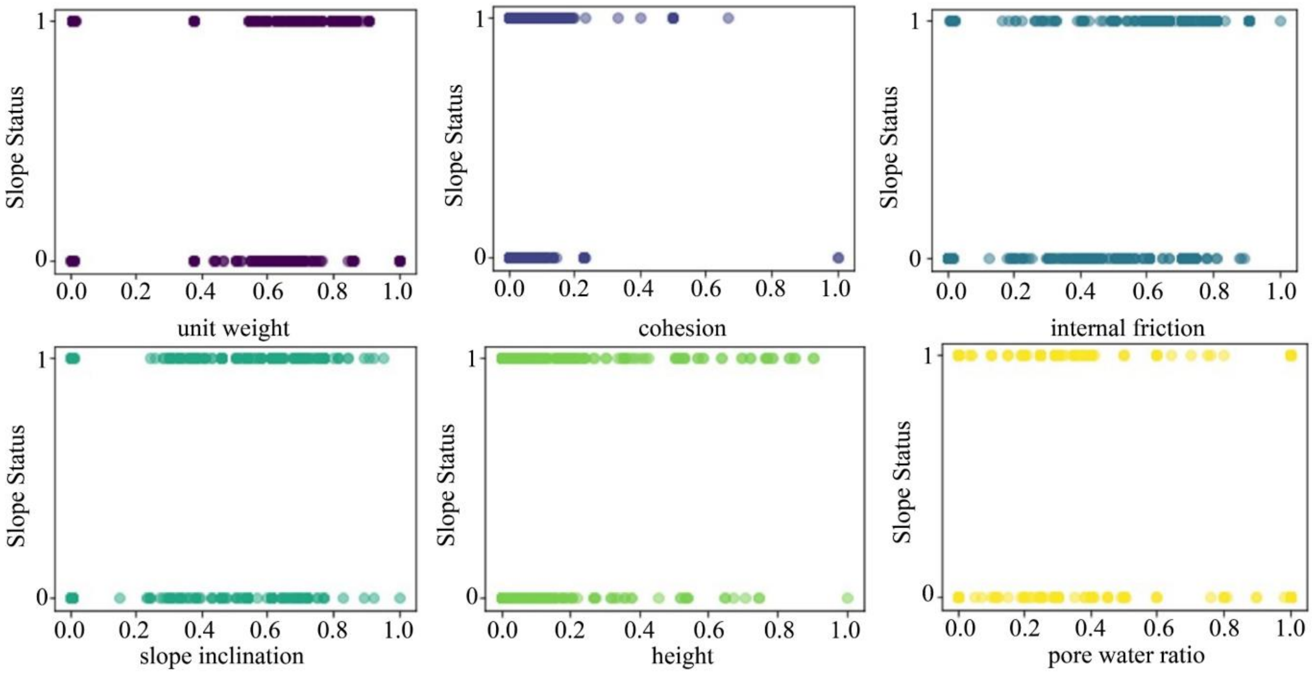

Fig. 7 Slope stability assessment across various parameters

Table 1. Hyperparameters of each model for optimal classification

Model	Hyperparameters	Optimal Hyperparameters	AUC	Accuracy	Sensitivity
RF	n_estimators = [50,100,150,200,250,300,350,400,450,500]	50	0.831	0.867	0.890
CatBoost	learning_rate = [0.1,0.01,0.001]	0.1	0.830	0.854	0.859
	n_estimators = [50,100,200,300,400,500]	400			
SVC	kernel = ['linear', 'poly', 'rbf']	rbf	0.739	0.777	0.750
	C = [1,50,100,150,200,250,300,350,400,450,500]	500			
	Degree = [1,2,3,4,5,6]	3			
KNN	n_neighbor = [10,20,30,40,50,60,70]	30	0.794	0.793	0.859
LR	max_iter = [500]	500	0.615	0.638	0.687
	C = [1,50,100,150,200,250,300,350,400,450,500]	1			
Stacking	Nil	Nil	0.898	0.854	0.859

4. Results and Discussions

This study also employs SVC, KNN, and LR classifiers to assess their performance in slope stability prediction against ensemble models, including RF, CatBoost and Stacking. It is significant to mention that the performance of the classifier is heavily influenced by the AUC, with a value of 1.0 denoting optimal performance. The ROC curves of classification models in Figure 8 show that the AUC of LR is 0.615 SVC is 0.739, KNN is 0.794, CatBoost is 0.830, RF is 0.831, and Stacking is 0.898. The varying AUC values among different classifiers can be attributed to the differences in their underlying algorithms, model complexity, and how well they capture the relationships between features and the target variable. The ROC curves of ensemble learning models such as RF, CatBoost, and Stacking are located slightly higher in the top left corner of the plot in comparison to the other models. These models have AUC values exceeding 0.80, which is somewhat better than that of SVM, LR, and KNN. The results indicate that ensemble classifiers (RF, CatBoost, and Stacking) exhibit superior AUC values, implying better discriminatory ability and overall performance when compared to standalone classifiers (SVC, KNN, and LR). Figure 9 shows the confusion matrix of classification models. It can be seen from the figure that the total misclassifications in RF and CatBoost are 23, Stacking is 27, KNN is 28, SVC is 35, and LR is 52.

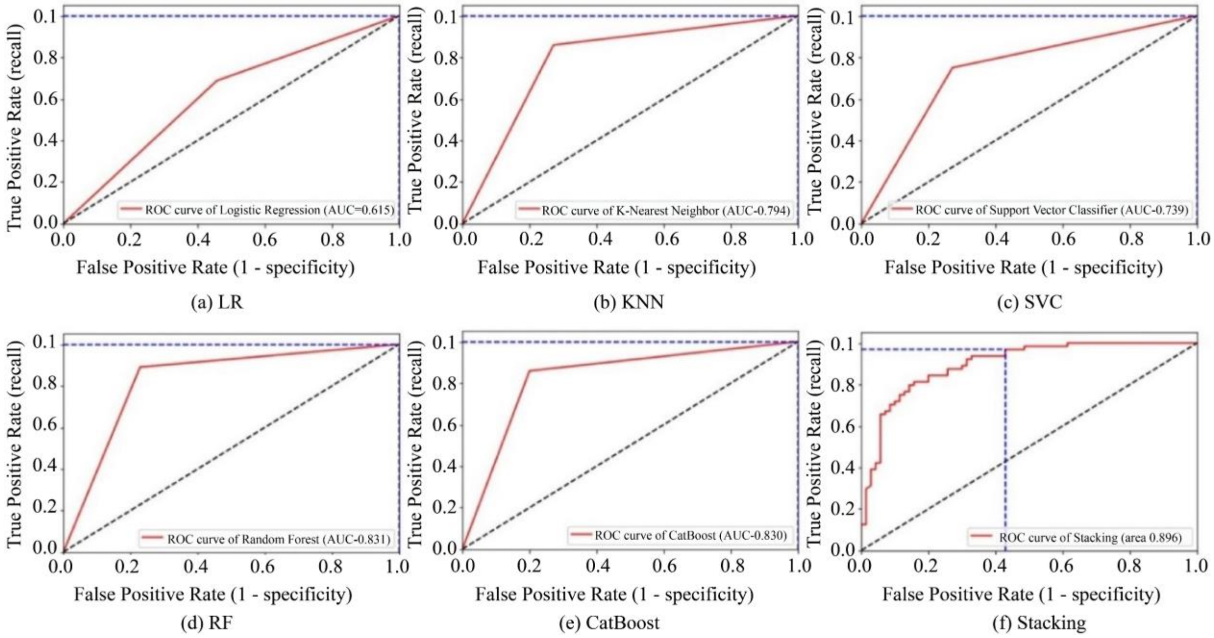

Fig. 8 ROC curves of classification models on the testing dataset

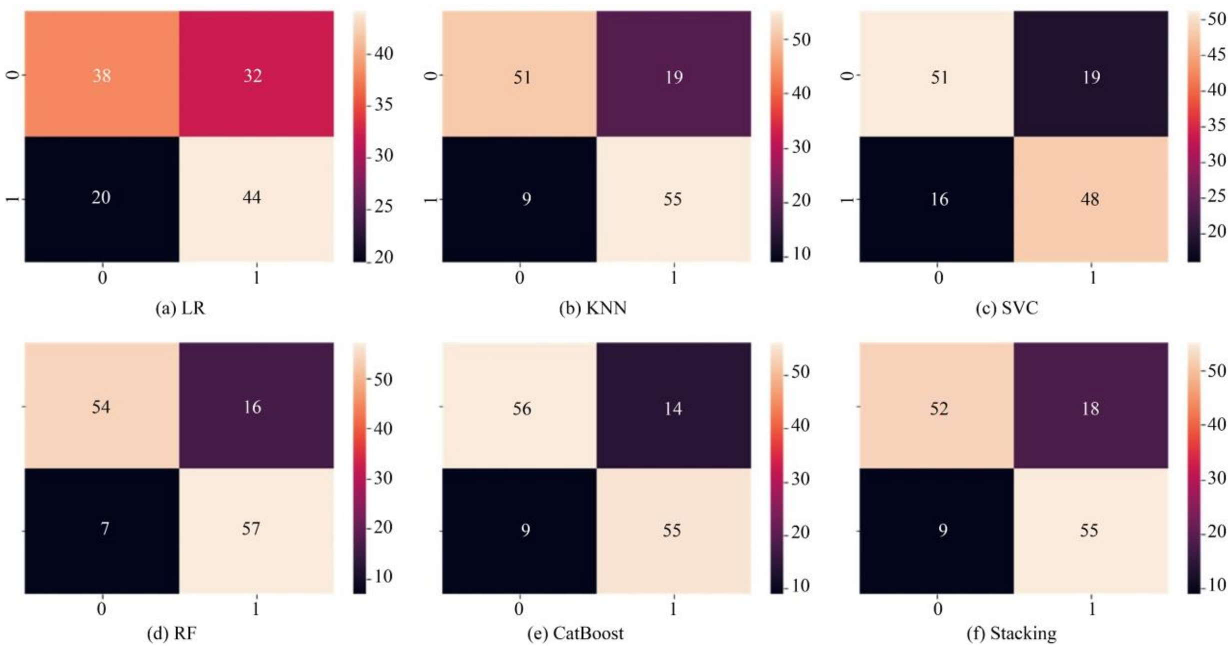

Fig. 9 Confusion matrix of classification models

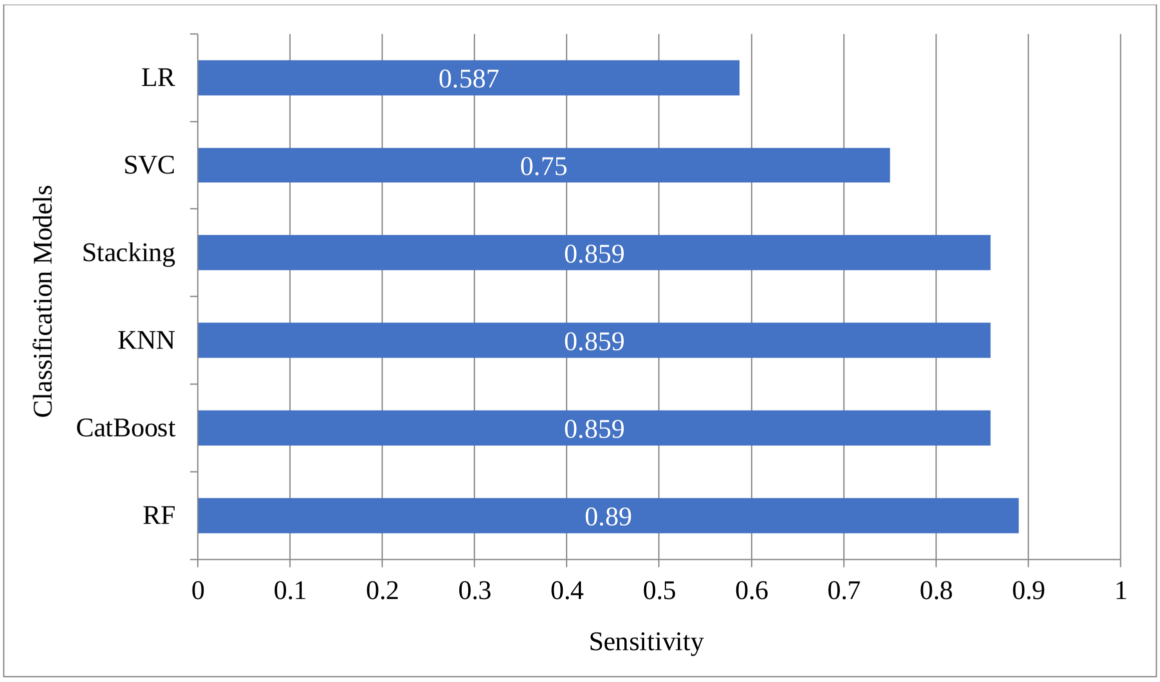

Fig. 10 Sensitivity of classification models

Sensitivity analysis, as illustrated in Figure 10, further emphasized these disparities in performance among classifiers. It can be inferred from Figure 10 that the RF model is highly sensitive, with a sensitivity score of 0.89. Meanwhile, the stacking, KNN, and catboost models have comparable sensitivity scores of 0.859. The LR model, with a sensitivity score of 0.687, is the least sensitive among all the models. These variations underscore the importance of selecting the appropriate classifier based on the specific characteristics of the dataset and the problem at hand to achieve optimal classification results. It is evident from the results that classification models, such as LR, encountered more challenges in prediction compared to Stacking, highlighting the importance of thoughtful model selection in the classification process. However, out of three ensemble models (RF, CatBoost and Stacking), Stacking gave higher results than RF and CatBoost classifier due to its ability to discriminate and better ranking of positive samples despite having lower sensitivity compared to RF.

4.1. Feature Importance Analysis

The concept of feature importance is a crucial element in the development of reliable machine learning models. It involves evaluating the relevance of each input feature in the decision-making process. Essentially, feature importance quantifies the impact of a specific feature on the model's predictions. In Figure 11, the feature importance of all classifiers is illustrated, and the figure indicates that the feature importance for Stacking is significantly higher than that of LR, SVC, RF, KNN, and CatBoost. This suggests that features H, β, c, Υ, and ru exert a substantial influence and are relatively more important for predicting slope stability compared to feature ϕ. Understanding feature importance is crucial for identifying which features have a greater impact on a model's predictive ability. This understanding is essential for interpreting the model's behavior and selecting the most influential features for better performance.

4.2. Sensitivity Analysis

To prevent slope failure, it is critical to assess the sensitivity of features that contribute to triggering such events. Evaluating the sensitivity of these features is crucial for assessing slope stability and building efficient support structures. To quantify the sensitivity of features in each model, their impact on predictive performance is assessed at the optimum hyperparameters for each classification model. This sensitivity study gives vital insights into the influential features triggering slope stability, enabling informed decision-making and preemptive efforts to limit the risk of slope failure. Figure 12 shows the sensitivity of features for each classification model. The results demonstrate that for LR, ru is highly sensitive, whereas ϕ is low sensitive. Similarly, in the case of RF, the sensitivity to H is quite high, while the sensitivity to Υ is relatively low. In the case of KNN, the sensitivity of ru is significantly high, whereas the sensitivity of ϕ is quite modest. In the case of SVC, the variables H and β exhibit a high degree of sensitivity, while Υ and ϕ demonstrate a low level of sensitivity. In CatBoost, the variables H and β exhibit great sensitivity, while Υ has poor sensitivity. When it comes to stacking, H and β exhibit a high level of sensitivity compared to other features. The feature ranking in Figure 13 is determined by averaging the sensitivity of all classification models: H (0.913), ru (0.895), β (0.874), c (0.817), Υ (0.502), and ϕ (0.400). It is evident that H, ru, β, and c are highly sensitive to slope stability. Therefore, ensuring the accurate and reasonable selection of values for H, ru, β, and c in artificial slopes is crucial, taking into account field tests and on-ground conditions. The higher values of H and β indicate that the geometry variables are highly sensitive towards slope stability.

Similarly, c, being the most important soil property is also highly sensitive towards slope stability. The binding property of soil material will change significantly, thereby affecting the integrity of soil structure (internal strength). Optimizing these values is crucial in practical design to ensure slope stability. Additionally, the sensitivity of Ƴ and ϕ is comparatively lower than that of other features.

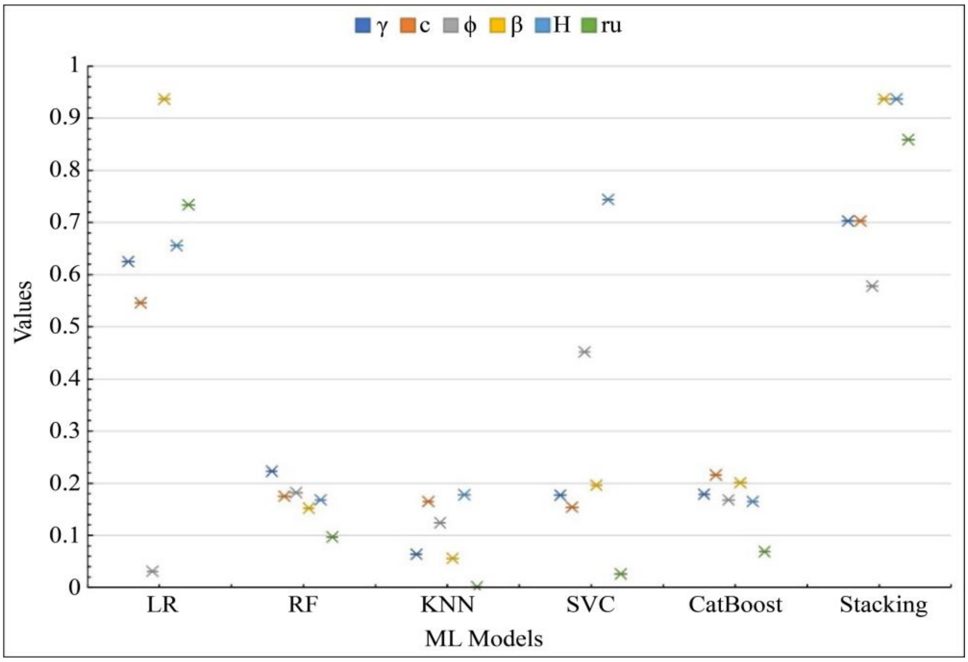

Fig. 11 Representation of feature importance for ML models

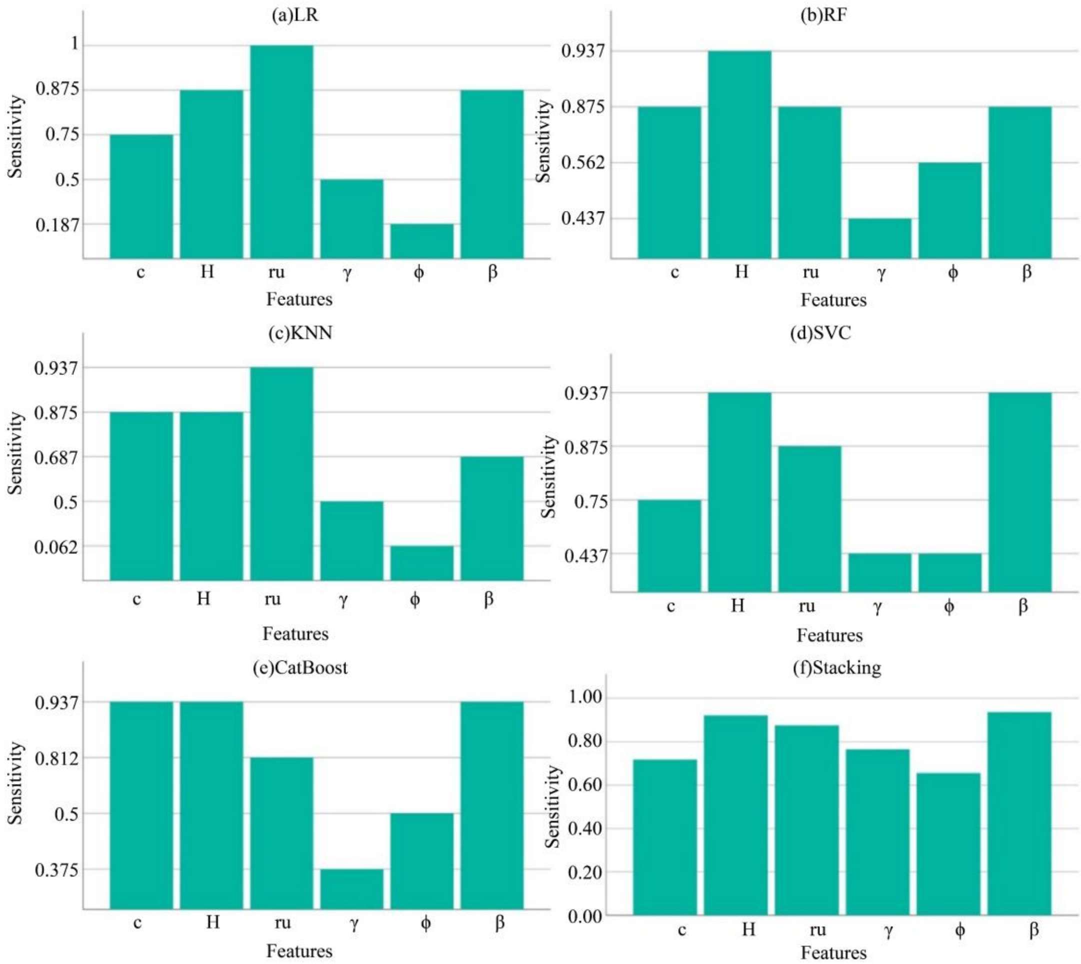

Fig. 12 Sensitivity analysis of features for classification models

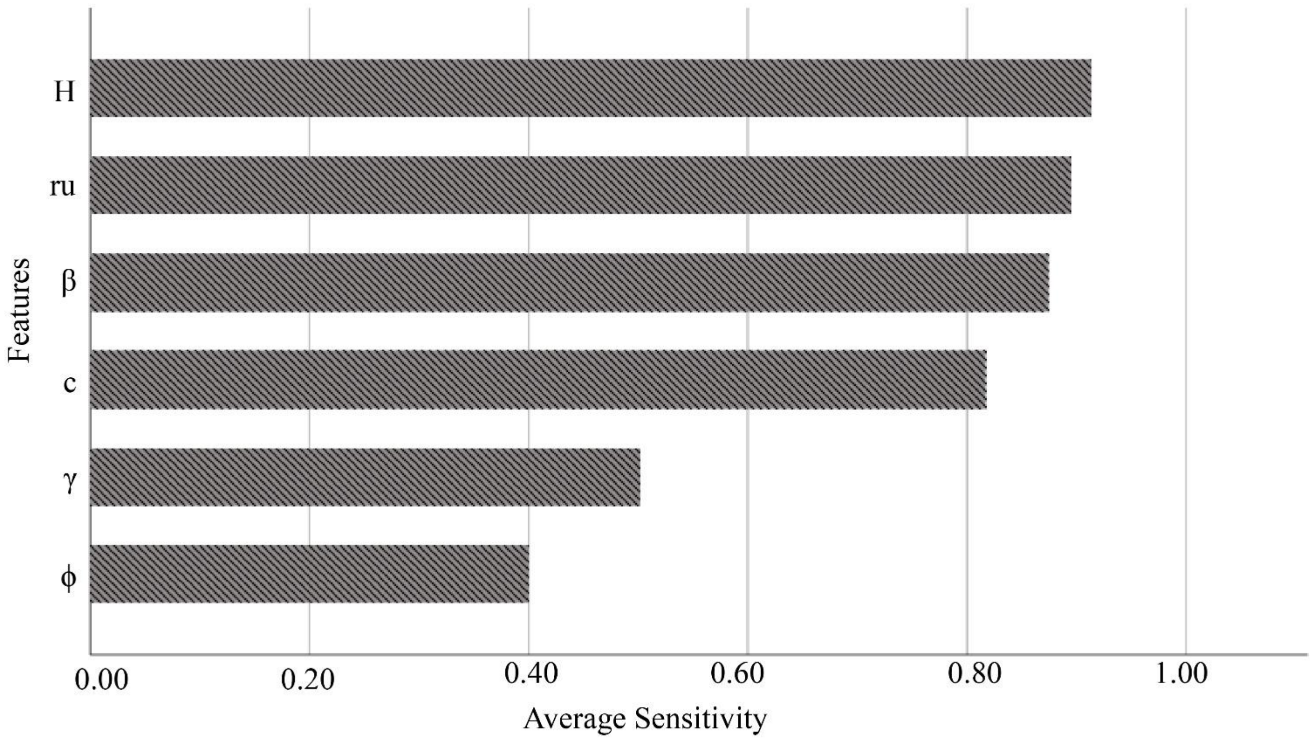

Fig. 13 Ranking of features based on average sensitivity

5. Conclusion

This study proposes a comparative study among three ensemble learning classifiers, including RF, CatBoost and Stacking, in evaluating the stability of 444 slope cases. Six features, including H, ru, β, c, ϒ and ϕ used for the prediction and generalization of classification models. The ensemble models were also compared with LR, SVC and KNN classifiers for slope stability classification. The following conclusions are drawn as per the analysis:

Based on the ROC curves, the Stacking classifier achieved the highest AUC compared to other classifiers. This indicates that Stacking has the best overall performance in discriminating between stable and unstable slopes. Among the ensemble methods (RF, CatBoost, and Stacking), Stacking demonstrated superior performance with the highest AUC. This suggests that an ensemble classifier such as Stacking is a strong alternative to other classifiers for slope stability prediction, particularly when interpretability or handling imbalanced datasets is crucial.

All features taken in the study exhibit sensitivity to slope stability, indicating that relying solely on a single parameter for discriminating slope stability is unreliable. The LR is highly sensitive to ru but less sensitive to ϕ, RF is highly sensitive to H but less sensitive to ϒ, KNN is highly sensitive to ru but less sensitive to ϕ, SVC is highly sensitive to H and β but less sensitive to ϒ and ϕ, CatBoost is highly sensitive to H and β but less sensitive to ϒ, and Stacking is highly sensitive to H and β compared to other features. This in-depth analysis emphasizes the significance of considering geometrical parameters (such as H and β) and soil properties (such as cohesiveness represented by c) while predicting slope stability accurately. It emphasizes the complex nature of evaluating slope stability. It underscores the importance of taking into account a comprehensive range of characteristics instead of relying on individual metrics for precise forecasting.

The complex relationship between slope stability and the various factors that affect it can be difficult to model accurately due to its non-linear and multidimensional nature. However, a study of an engineering case shows that the Stacking classifier is effective in navigating this intricate connection and making precise and reliable predictions. This highlights the importance of supervised learning in assessing slope stability. In the future, it may be possible to improve the effectiveness of ensemble learning algorithms (RF, CatBoost, and Stacking) by incorporating key samples and parameters that influence the dynamics of slope stability. Rainfall patterns, seismic activity, human interventions, and other environmental events are significant factors that affect slope stability outcomes. Integrating these features into the algorithmic structure has the potential to improve prediction accuracy, generalization abilities, and reliability in real-world scenarios. This area of future research aims to expand and improve the effectiveness of ensemble learning methods in addressing complex geotechnical issues.

Author Contribution Statement

SKA: Dataset collection, Conceptualization and Analysis; SKA, DK and SKS: Methodology; SKA: Drafting; DK and SKS: Review and Comments; SKA, DK and SKS: Final drafting and proofreading.

References

[1] James Michael Duncan, "State of the Art: Limit Equilibrium and Finite-Element Analysis of Slopes," *Journal of Geotechnical Engineering,* vol. 122, no. 7, pp. 577-596, 1996. [CrossRef] [Google Scholar] [Publisher Link]

[2] S.Y. Liu, L.T. Shao, and H.J. Li, "Slope Stability Analysis Using the Limit Equilibrium Method and Two Finite Element Methods," *Computers and Geotechnics,* vol. 63, pp. 291-298, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[3] Y.M. Cheng, T. Lansivaara, and W.B. Wei, "Two-Dimensional Slope Stability Analysis by Limit Equilibrium and Strength Reduction Methods," *Computers and Geotechnics,* vol. 34, no. 3, pp. 137-150, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[4] Fei Zheng et al., "Modified Predictor-Corrector Solution Approach for Efficient Discontinuous Deformation Analysis of Jointed Rock Masses," *International Journal for Numerical and Analytical Methods in Geomechanics,* vol. 43, no. 2, pp. 599-624, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[5] Xiaoying Zhuang et al., "A Cover-Based Contact Detection Approach for Irregular Convex Polygons in Discontinuous Deformation Analysis," *International Journal for Numerical and Analytical Methods in Geomechanics,* vol. 45, no. 2, pp. 208-233, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] H. Zheng, D.F. Liu, and C.G. Li, "Slope Stability Analysis Based on Elasto-Plastic Finite Element Method," *International Journal for Numerical Methods in Engineering,* vol. 64, no. 14, pp. 1871-1888, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[7] D.V. Griffiths, and P.A. Lane, "Slope Stability Analysis by Finite Elements," *Geotechnique,* vol. 49, no. 3, pp. 387-403, 1999. [CrossRef] [Google Scholar] [Publisher Link]

[8] Tamotsu Matsui, and Ka-Ching San, "Finite Element Slope Stability Analysis by Shear Strength Reduction Technique," *Soils and Foundations,* vol. 32, no. 1, pp. 59-70, 1992. [CrossRef] [Google Scholar] [Publisher Link]

[9] X. Li, "Finite Element Analysis of Slope Stability Using a Nonlinear Failure Criterion," *Computers and Geotechnics,* vol. 34, no. 3, pp. 127-136, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[10] Shan Lin et al., "Evaluation and Prediction of Slope Stability Using Machine Learning Approaches," *Frontiers of Structural and Civil Engineering,* vol. 15, no. 4, pp. 821-833, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Pijush Samui, "Slope Stability Analysis: A Support Vector Machine Approach," *Environmental Geology,* vol. 56, pp. 255-267, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[12] Min-Yuan Cheng, and Nhat-Duc Hoang, "Slope Collapse Prediction Using Bayesian Framework with K-Nearest Neighbor Density Estimation: Case Study in Taiwan," *Journal of Computing in Civil Engineering,* vol. 30, no. 1, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[13] H. Fattahi, "Prediction of Slope Stability Using Adaptive Neuro-Fuzzy Inference System Based on Clustering Methods," *Journal of Mining and Environment,* vol. 8, no. 2, pp. 163-177, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[14] Nhat-Duc Hoang, and Dieu Tien Bui, "Slope Stability Evaluation Using Radial Basis Function Neural Network, Least Squares Support Vector Machines, and Extreme Learning Machine," *Handbook of Neural Computation,* pp. 333-344, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[15] Sarat Kumar Das et al., "Classification of Slopes and Prediction of Factor of Safety Using Differential Evolution Neural Networks," *Environmental Earth Sciences,* vol. 64, pp. 201-210, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[16] Amin Manouchehrian, Javad Gholamnejad, and Mostafa Sharifzadeh, "Development of a Model for Analysis of Slope Stability for Circular Mode Failure Using Genetic Algorithm," *Environmental Earth Sciences,* vol. 71, pp. 1267-1277, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[17] Yusuf Erzin, and Tulin Cetin, "The Prediction of the Critical Factor of Safety of Homogeneous Finite Slopes Using Neural Networks and Multiple Regressions," *Computers Geosciences,* vol. 51, pp. 305-313, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[18] Chongchong Qi, and Xiaolin Tang, "Slope Stability Prediction Using Integrated Metaheuristic and Machine Learning Approaches: A Comparative Study," *Computers & Industrial Engineering,* vol. 118, pp. 112-122, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[19] Dhruva Karir et al., "Stability Prediction of a Natural and Man-Made Slope Using Various Machine Learning Algorithms," *Transportation Geotechnics,* vol. 34, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[20] Chongchong Qi et al., "Comparative Study of Hybrid Artificial Intelligence Approaches for Predicting Hanging Wall Stability," *Journal of Computing in Civil Engineering,* vol. 32, no. 2, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[21] Min-Yuan Cheng, and Nhat-Duc Hoang, "Typhoon-Induced Slope Collapse Assessment Using a Novel Bee Colony Optimized Support Vector Classifier," *Natural Hazards,* vol. 78, pp. 1961-1978, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[22] Nhat-Duc Hoang, and Anh-Duc Pham, "Hybrid Artificial Intelligence Approach Based on Metaheuristic and Machine Learning for Slope Stability Assessment: A Multinational Data Analysis," *Expert Systems with Applications,* vol. 46, pp. 60-68, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[23] Behrouz Gordan et al., "Prediction of Seismic Slope Stability through Combination of Particle Swarm Optimization and Neural Network," *Engineering with Computers,* vol. 32, pp. 85-97, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[24] Zhenyan Luo et al., "A Novel Artificial Intelligence Technique for Analyzing Slope Stability Using PSO-CA Model," *Engineering with Computers,* vol. 37, pp. 533-544, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[25] Chongchong Qi, and Xiaolin Tang, "A Hybrid Ensemble Method for Improved Prediction of Slope Stability," *International Journal for Numerical and Analytical Methods in Geomechanics,* vol. 42, no. 15, pp. 1823-1839, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[26] Ningthoujam Jibanchand, and Konsam Rambha Devi, "Application of Ensemble Learning in Predicting Shallow Foundation Settlement in Cohesionless Soil," *International Journal of Geotechnical Engineering,* vol. 17, no. 3, pp. 234-245, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[27] Lihui Bai et al., "A Machine Learning Ensemble Model for Predicting Pavement Conditions Using Automatic Laser Crack Measurement Data," *International Journal of Pavement Engineering,* vol. 24, no. 1, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[28] Jingqi Cui et al., "Composite Interpretability Optimization Ensemble Learning Inversion Surrounding Rock Mechanical Parameters and Support Optimization in Soft Rock Tunnels," *Computers and Geotechnics,* vol. 165, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[29] Navid Kardani et al., "Improved Prediction of Slope Stability using a Hybrid Stacking Ensemble Method Based on Finite Element Analysis and Field Data," *Journal of Rock Mechanics and Geotechnical Engineering,* vol. 13, no. 1, pp. 188-201, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[30] Lin Wang et al., "Efficient Reliability Analysis of Earth Dam Slope Stability using Extreme Gradient Boosting Method," *Acta Geotechnica,* vol. 15, pp. 3135-3150, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[31] Wengang Zhang et al., "Assessment of Basal Heave Stability for Braced Excavations in Anisotropic Clay Using Extreme Gradient Boosting and Random Forest Regression," *Underground Space,* vol. 7, no. 2, pp. 233-241, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[32] Zhice Fang et al., "A Comparative Study of Heterogeneous Ensemble-Learning Techniques for Landslide Susceptibility Mapping," *International Journal of Geographical Information Science,* vol. 35, no. 2, pp. 321-347, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[33] Yang Chen et al., "Predicting Uniaxial Tensile Strength of Expansive Soil with Ensemble Learning Methods," *Computers and Geotechnics,* vol. 150, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[34] Gérard Biau, and Erwan Scornet, "A Random Forest Guided Tour," *Test,* vol. 25, pp. 197-227, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[35] Jaime Lynn Speiser et al., "A Comparison of Random Forest Variable Selection Methods for Classification Prediction Modelling," *Expert Systems with Applications,* vol. 134, pp. 93-101, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[36] Mariana Belgiu, and Lucian Drăguţ, "Random Forest in Remote Sensing: A Review of Applications and Future Directions," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 114, pp. 24-31, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[37] Husein Ali Zeini et al., "Random Forest Algorithm for the Strength Prediction of Geopolymer Stabilized Clayey Soil," *Sustainability,* vol. 15, no. 2, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[38] Binh Thai Pham et al., "A Novel Hybrid Soft Computing Model using Random Forest and Particle Swarm Optimization for Estimation of Undrained Shear Strength of Soil," *Sustainability,* vol. 12, no. 6, pp. 1-16, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[39] Nitish Puri, Harsh Deep Prasad, and Ashwani Jain, "Prediction of Geotechnical Parameters Using Machine Learning Techniques," *Procedia Computer Science,* vol. 125, pp. 509-517, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[40] Vinicius Luiz Pacheco et al., "Cone Penetration Test Prediction Based on Random Forest Models and Deep Neural Networks," *Geotechnical and Geological Engineering,* vol. 41, no. 8, pp. 4595-4628, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[41] Anna Veronika Dorogush, Vasily Ershov, and Andrey Gulin, "Catboost: Gradient Boosting with Categorical Features Support," *arXiv preprint*, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[42] Liudmila Prokhorenkova et al., "Catboost: Unbiased Boosting with Categorical Features," *Advances in Neural Information Processing Systems,* vol. 31, 2018. [Google Scholar] [Publisher Link]

[43] Mário Papík et al., "CatBoost: The Case of Bankruptcy Prediction," *International Conference on Business and Technology,* pp. 3-17, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[44] B. Dhananjay, and J. Sivaraman, "Analysis and Classification of Heart Rate Using Catboost Feature Ranking Model," *Biomedical Signal Processing and Control,* vol. 68, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[45] Abhisht Joshi et al., "CatBoost-An Ensemble Machine Learning Model for Prediction and Classification of Student Academic Performance," *Advances in Data Science and Adaptive Analysis,* vol. 13, no. 3, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[46] Li Diao et al., "Short-Term Weather Forecast Based on Wavelet Denoising and Catboost," *Chinese Control Conference,* pp. 3760-3764, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[47] Guomin Huang et al., "Evaluation of Catboost Method for Prediction of Reference Evapotranspiration in Humid Regions," *Journal of Hydrology,* vol. 574, pp. 1029-1041, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[48] David H. Wolpert, "Stacked Generalization," *Neural Networks,* vol. 5, no. 2, pp. 241-259, 1992. [CrossRef] [Google Scholar] [Publisher Link]

[49] Shan Lin et al., "Comparative Performance of Eight Ensemble Learning Approaches for the Development of Models of Slope Stability Prediction," *Acta Geotechnica,* vol. 17, no. 4, pp. 1477-1502, 2022. [CrossRef] [Google Scholar] [Publisher Link]