Analysis Grid Search Optimization of Machine Learning Models for Slope Stability Prediction Supports the Design Construction of Geotechnical Structures and Environmental Development

Saurabh Kumar Anuragi[1], D. Kishan[2]

[1]*Ph.D Scholar, Department of Civil Engineering, Maulana Azad National Institute of Technology, India*
[2]*Associate Professor, Department of Civil Engineering, Maulana Azad National Institute of Technology, India*

Abstract: The accurate prediction of slope stability is crucial for the design and construction of geotechnical structures, as well as for environmental development and risk mitigation. This study explores the application of machine learning (ML) models optimized using the Grid Search method to enhance slope stability predictions. In this study, an in-depth analysis of seven prediction models Logistic Regression (LR), K-Nearest Neighbor (KNN), Support Vector Classifier (SVC), CatBoost, and RUSBoost is provided. These models were evaluated using a grid search approach to find the optimal hyperparameters. The model takes several input features, including pore water ratio (ru), height (H), unit weight (Y), cohesion (c), slope angle (β), and angle of internal friction (ϕ). The output is the slope status, either stable (1) or unstable (0). The generalization ability of classification models is improved by using a 5-fold cross-validation (CV). Evaluation indicators such as AUC, accuracy, and kappa were analyzed, and CatBoost outperformed other machine learning models with the highest AUC of 0.823, accuracy of 0.874, and kappa of 0.642. The results indicate that CatBoost is a highly effective tool for predicting slope stability, surpassing other models in classification accuracy. The capacity and efficiency of RF in deformation prediction models make it the most accurate tool available for forecasting slope stability. Additionally, a comprehensive analysis of feature sensitivity was conducted to determine the most significant characteristics for predicting slope stability. These findings not only enhance geotechnical safety but also contribute to sustainable environmental development by preventing landslides, reducing soil erosion, and supporting responsible land-use planning. The integration of machine learning into slope stability analysis promotes ecological preservation and long-term resilience in infrastructure projects.

Keywords: Logistic Regression, Slope Stability, Random Forest, Hyperparameters, RUSBoost, CatBoost, Grid Search, LightGBM. Geotechnical Engineering, Environmental Development

1. Introduction

Slopes are a common feature in geology and geography that are often encountered in various human activities such as engineering and economics. However, slope instability, including landslides, mudslides, and collapses, has caused significant human casualties, property damage, and economic implications in recent times. Therefore, it is crucial to accurately evaluate the stability of slopes to ensure the success of slope engineering. The accurate prediction of slope stability holds immense engineering importance. Slope stability estimation is a challenging task due to the complex physical state of the soil. The increasing number of slope failures causing significant economic and social losses has raised concerns among researchers and engineers. To avoid or reduce such damages, it is essential to conduct slope stability analysis and implement appropriate stabilization measures. A comprehensive understanding of the mechanisms that cause slope failure is crucial to effectively address these incidents. Therefore, researchers make efforts to minimize losses through numerical and analytical modelling, enabling accurate predictions and informed actions. In general, the methods used for slope stability analysis are the finite element method and the limit equilibrium method [1, 2]. The Finite Element Method (FEM) is a technique that can be used to analyze the limit equilibrium status of a block and determine its safety factor through straightforward calculations and clear concepts. However, it has a limitation in that it assumes that all sliding surfaces reach an ultimate state of failure simultaneously, which does not accurately reflect the actual stress status of slip surfaces. Therefore, it cannot account for nonlinear structural deformation [3–7]. Although the FEM can be used to determine the stress field and displacement field of a slope [8], it cannot produce an accurate value for the slope stability safety factor. Some researchers have calculated slope stability safety factors by combining finite element analysis and strength reduction, this method involves the use of specific failure criteria in order to determine whether a system enters a limit equilibrium state [9–14]. As a result, estimating slope stability using traditional methods is a difficult process that necessitates the use of complex modelling approaches, large experimental data, in-depth engineering knowledge and expertise.

In recent years, researchers have been using machine learning methods to analyze slope stability due to advancements in computational techniques. These methods evaluate the stability of slopes by taking into account factors such as the slope's geometry and material properties, which produce significant results. Li et al. [15] carried out a comparative study among 11 ML models for the prediction of the factor of safety considering six slope factors. Zhou et al. [16] utilized the gradient boosting machine method for slope stability prediction as a classification model. Qi et al. [17] carried out a study in which they compared six different artificial intelligence (AI) techniques, namely random forest (RF), support vector machine (SVM), gradient boosting machine (GBM), logistic regression (LR), decision tree (DT), and multi-layer perceptron neural network (MLPNN), to predict slope stability. In addition, they used the firefly algorithm (FA) to optimize the hyper-parameters of these models. A comparative study conducted by Das et al. [18] between artificial neural network (ANN), SVM and genetic programming to predict factor of safety and classify the slope as stable or unstable. Manouchehrian et al. [19] implemented genetic algorithm (GA) for slope stability analysis specifically for circular mode failure. Sakellariou et al. [20] applied ANN using back-propagation technique to estimate slope status. Kardani et al. [21] developed a hybrid stacking ensemble model on a database generated using finite element analysis to enhance slope stability prediction. Zhang et al. [22] compared four ML techniques namely, random forest (RF), support vector machine (SVM), logistic regression (LR) and extreme gradient boosting (XGBoost) for slope stability prediction. Gordan et al. [23] developed ANN and particle swarm optimization (PSO)-ANN models on 699 samples generated under different conditions using GeoStudio software for slope stability prediction. Pham et al. [24] developed ensemble models using sequential and parallel techniques for slope stability estimation. Apart from slope stability, machine learning techniques have been widely used in various fields of geotechnical engineering such as landslide susceptibility assessment [25–28] , settlement prediction of shallow foundations [29, 30], predicting bearing capacity of pile foundations [31–33] , predicting physical properties of soil [34, 35], soil quality assessment [36] and liquefaction assessment [37]. Meanwhile, remarkable results were obtained with increasing available slope features for slope stability prediction using supervised learning methods. Machine learning (ML) has become an important tool for predicting slope deformation. With the help of various intelligent algorithms and technical methods, the slope deformation prediction model has significantly improved, leading to a new era of slope deformation prediction. However, it is still essential to find an appropriate technique for predicting slope deformation, as each of the mentioned strategies has its own drawbacks.

Therefore, the objective of this study is to conduct a comparative analysis among different machine learning classifiers with the specific goal of predicting slope stability modelled as a classification problem. The study will explore and evaluate the performance of Logistic Regression (LR), K-Nearest Neighbor (KNN), Support Vector Classifier (SVC), CatBoost and RUSBoost classifiers. These techniques have been chosen due to their popularity in various projects, but rarely compared with each other. This study is outlined as follows: Section 2 provides a concise introduction to the machine 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 conclusion is given in Section 5.

1.1 Benefits of Slope Stability Prediction for Geotechnical Structures and Environmental Development

Accurate slope stability prediction provides numerous benefits in both geotechnical engineering and environmental development:

- Enhanced Infrastructure Safety – By accurately predicting slope failures, engineers can design safer roads, embankments, and retaining walls, reducing the risk of structural collapse and minimizing maintenance costs.
- Optimized Construction Planning – Reliable slope stability analysis enables efficient land-use planning, reducing excessive excavation and ensuring that construction activities are carried out in stable zones.
- Landslide and Disaster Prevention – Early detection of unstable slopes helps mitigate landslide risks, protecting human settlements, transportation networks, and industrial sites from natural disasters.
- Environmental Conservation – Slope instability often leads to deforestation, soil erosion, and sedimentation in water bodies, negatively impacting biodiversity and ecosystem balance. Predictive models aid in sustainable land management, reducing environmental degradation.
- Climate Resilience and Sustainable Development – As climate change increases the frequency of extreme weather events, robust slope stability analysis helps in adapting infrastructure to withstand heavy rainfall, earthquakes, and other environmental stressors.
- Improved Resource Management – Machine learning-driven slope stability analysis enables better decision-making in resource allocation, minimizing financial losses associated with slope failures and infrastructure damage.

By integrating machine learning models into geotechnical engineering, this study contributes to the development of resilient infrastructure while promoting ecological balance and sustainable environmental practices.

2. Computational Techniques

2.1 Logistic Regression

Logistic regression (LR) is a statistical method used to model the relationship between a binary dependent variable and one or more independent variables. It allows for the estimation of probabilities and odds ratios, as well as the making of predictions for new observations [38]. The dependent variable in logistic regression is a binary variable, which means it takes one of two possible values. For example, it could represent the presence or absence of a disease, or whether a customer purchased a product or not. The independent variables can be continuous, categorical, or a combination of both. The logistic regression model is used to estimate the probability of the dependent variable taking on the value of 1 (or 0, depending on how the variable is coded) based on the values of the independent variables. This probability is modelled as a function of the linear combination of the independent variables, which is then transformed by the logistic function (as shown in Equation 1):

$$P(Y=1|X)=\frac{1}{(1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_p X_p)})} \qquad (1)$$

where P(Y=1|X) is the probability of the dependent variable taking the value 1 given the values of the independent variables X_1, X_2, ..., X_p, and β_0, β_1, β_2, ..., β_p are the coefficients of the model.

The logistic function is a curve that takes a linear combination of independent variables and maps it onto a probability value between 0 and 1. It has an S-shape, with a steeper curve around the middle (where the probability is 0.5) and a flatter curve towards the extremes (where the probability is close to 0 or 1). Logistic regression is a statistical model that uses maximum likelihood estimation to determine the values of the coefficients that best fit the data. These coefficients can be interpreted as the change in the log odds of the dependent variable for an increase of one unit in the associated independent variable. Additionally, you can use the exponentiated value of the coefficient to calculate the odds ratio.

2.2 K-Nearest Neighbor

K-Nearest Neighbors (KNN) is a non-parametric technique that is extensively used in supervised machine learning for classification and regression tasks. The fundamental principle behind the KNN algorithm is to categorize a new data point by comparing it to the existing data points. The algorithm saves all the existing data points and identifies the K-nearest neighbors using a distance metric like euclidean distance. The algorithm then assigns the new data point to the category that is most common among its K-nearest neighbors. While KNN can be applied to both classification and regression tasks, it is mainly used for classification [39, 40].

2.3 Support Vector Classifier

Support Vector Classifier (SVC) is a supervised learning model based on the statistical learning theory used to perform classification analysis. The theory of SVC focuses on linear separability, decision limits, and margin maximization. In 1963, Vapnik and Chervonenkis [41] established the Vapnik-Chervonenkis (VC) dimension, which helped explain the generalization performance of SVC. The hypothesis space, which is the set of decision boundaries that a learning algorithm may learn, is measured by its capacity or complexity through the VC dimension. The VC dimension of SVC represents the greatest number of points that the algorithm's decision border can shatter or perfectly separate. Understanding the VC dimension is key to understanding SVC generalization performance since it provides a theoretical framework for analyzing the model's complexity versus generalization to unknown data. In 1992, Boser et al. [42] introduced kernel functions to SVC to handle non-linearly separable data. In 1995, Cortes and Vapnik [43] introduced the SVC formulation, which penalty factor C to discover optimal separating hyperplanes with the largest margins. The penalty factor C determines the trade-off between misclassifications in training data.

Numerous studies have been conducted to enhance the training and performance of SVC. Platt (1999) [44] introduced the Sequential Minimum Optimization approach to expedite the training of large datasets. Joachims (2006) [45] proposed the Budgeted Support Vector Machine that selects a subset of support vectors to accelerate the training process. Additionally, parallel computing, distributed learning, and active learning have been explored to enhance the efficiency and scalability of SVC. Despite some challenges with parameter tuning and scalability, SVC has demonstrated remarkable performance and versatility [46–51]. Researchers are continuously investigating ways to improve the algorithm's performance. It is essential to comprehend the progress made by SVC and future directions to utilize it effectively in solving complicated categorization problems.

2.4 CatBoost

CatBoost is a gradient boosting decision tree algorithm designed for handling categorical variables in datasets [52, 53]. It has built-in support for categorical variables, eliminating the need for preprocessing like one-hot encoding. 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. Additionally, 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. CatBoost utilizes binary decision trees as its fundamental predictor, which greatly enhances its strong performance across various domains and datasets [54].

2.5 RUSBoost

Data sampling is a technique used in machine learning to address class imbalance. Class imbalance occurs when one class has significantly more data points than the other class(es) in a dataset. This can result in a biased model that performs poorly on the minority class. To address this issue, data sampling methods can be used to modify the distribution of the classes in the training data set. One common method of data sampling is oversampling. Oversampling involves adding more instances of the minority class to the training data set. This can be done using techniques such as SMOTE (Synthetic Minority Over-sampling Technique), which creates synthetic data points in the minority class. Another method is undersampling, which involves removing instances from the majority class to achieve a more balanced distribution. The RUS (Random Under-Sampling) algorithm is an example of an undersampling method. Unlike other undersampling methods, RUS does not selectively remove instances from the majority class. Instead, it randomly selects instances from the majority class to remove until a desired class distribution is achieved. The detailed literature may be referred to at [55].

3. Model Development and Optimization

3.1 Data Preprocessing and Visualization

When developing a classification model for analyzing slope stability, it is crucial to carefully select features that significantly impact slope stability. This involves adhering to specific feature selection principles, such as selecting essential features from the available feature set and eliminating irrelevant features that do not contribute much to the succeeding learning stages. These principles help to train the model efficiently and prevent problems associated with high-dimensional data. It also ensures that the model focuses on the most important features of slope stability, which enhances the model's overall prediction performance. Applying these principles in feature selection is crucial for developing a robust classification model designed to handle the complexities of slope stability evaluation. Currently, factors such as the magnitude of pore water ratio (ru), height (H), unit weight (Ɣ), cohesion (c), slope angle (β), and angle of internal friction (ϕ) are extensively used in slope stability prediction. This study uses 444 slope stability cases and modelled them as a classification problem (Figure 1) to predict slope status [stable (1) or unstable (0)] collected from [56]. Before conducting the analysis, the dataset is normalized using Equation 2 to remove the effects of different scales, units, and distributions that can result in biased model training and reduced accuracy. By scaling the data to a common range, we enhanced the model's ability to generalize and make accurate predictions on unseen data.

$$y_{normalization} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

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 important factors that need to be analyzed in order to gain a better understanding of the dataset. Figure 2 shows the distribution and variability of each input variable. The violin plots of the dataset are presented in Figure 3. The violin plot is a useful visualization tool that displays the distribution and density of a dataset across different categories or groups. The width of the violin at any point represents the density of data at that particular point, with the thicker parts of the violin indicating regions of high density and the thinner parts indicating regions of low density. The horizontal line inside the violin represents the median value of the data. By examining the violin plots, we can clearly observe that the variables Ɣ, 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 and have a large range of values. 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. This indicates that the data points for these variables are closely grouped and have a smaller range of values. Such insights can be useful in identifying trends and patterns and can help in formulating strategies to manage slope stability. The distribution of the variables and the relationship between slope status and other input variables are shown in the pair plots (Figure 4). The figure provides a graphical representation of the pairwise relationships between the different features, including the frequency distribution of each parameter. The diagonal histogram displays the numerical distribution of each attribute, which helps in understanding the data distribution. The upper and lower triangles represent scatter plots, which depict the overall correlation between the variables plotted on the x and y axes.

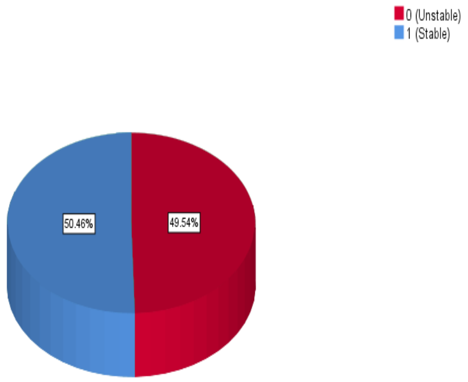

Figure 1: Dataset Pie Chart

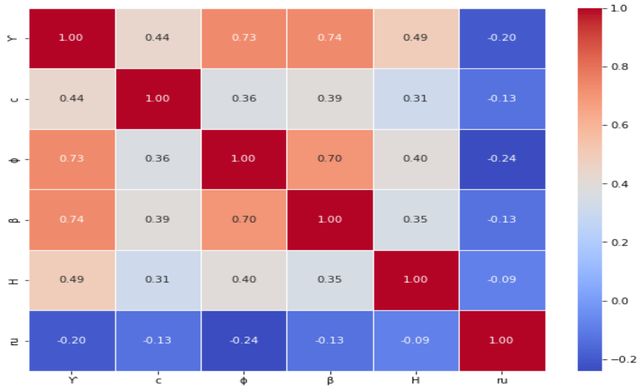

Figure 2: Correlation Matrix of Dataset

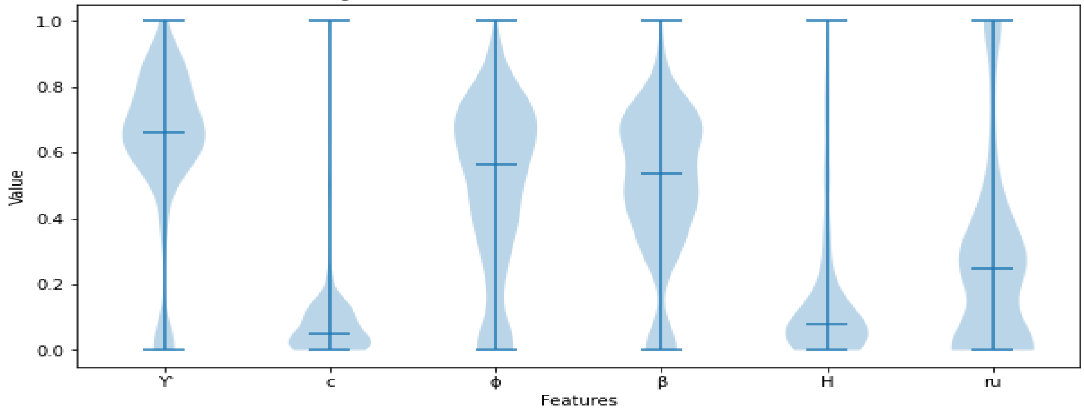

Figure 3: Violin Plots showing the distribution of slope cases

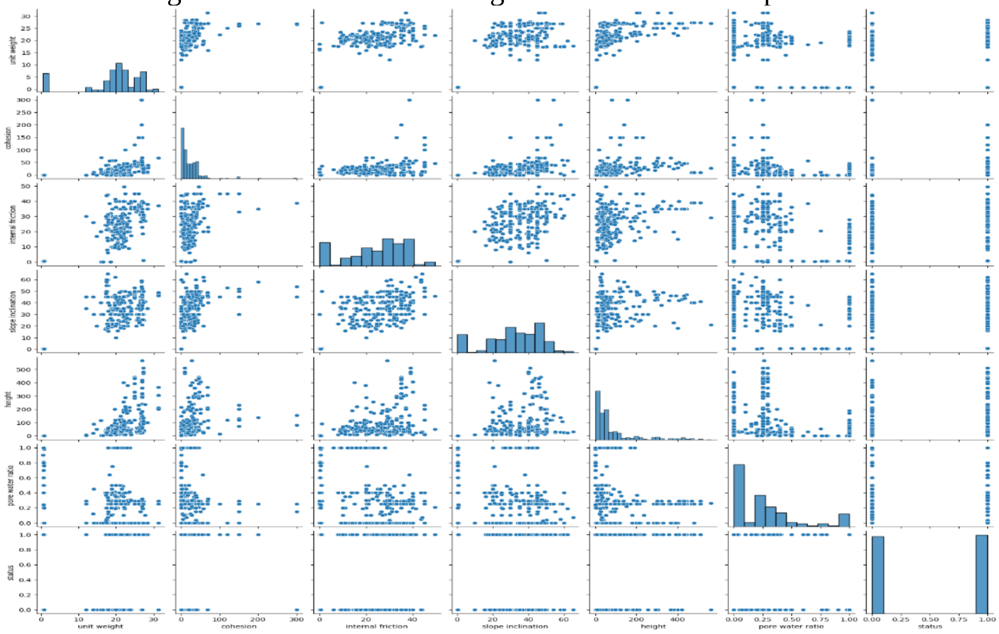

Figure 4: Pairplots of variables

In order to achieve the most accurate classification, it is best if each value of every feature on the diagram is linked exclusively to one class label, either stable or unstable. Figure 5 illustrates the categorization of slope stability using different indicators. The graphic shows instances where one indicator value corresponds to multiple slope classifications. A possible explanation for this behavior is that the data lacks linear separability, which makes it difficult to establish clear boundaries for the feature value

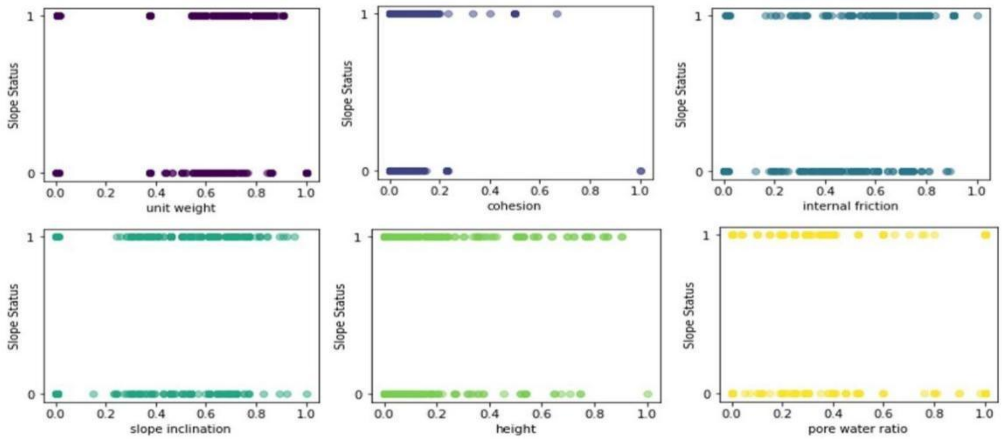

Figure 5: Slope Stability assessment across various features

3.2 Model Development and Optimization

This study investigates the use of five machine learning algorithms, namely logistic regression (LR), k-nearest neighbor (KNN), support vector classifier (SVC), catboost and rusboost in slope stability classification. When dealing with supervised classification problems, it is crucial to evaluate the performance of classification models on new data to determine their ability to generalize. To achieve this, the dataset is typically divided into two subsets - the training set and the testing set (Figure 6). The training set, which contains most of the data, is used to train the model and optimize hyperparameters. On the other hand, the smaller testing set, consisting of approximately 30% of the dataset or 133 cases, is exclusively used to assess the models ability to generalize to new, unseen instances. Around 70% of the original dataset, which equates to 311 cases, is selected as the training set. This 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 model involves exploring various combinations of hyperparameters using grid search, 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 models effectiveness in real-world scenarios. Machine learning models are assessed using a technique called 5-fold cross-validation (CV). This process involves dividing the training data into five random folds. Four of these folds are used to train the machine learning model, while the remaining fold is used as a verification set to assess the models performance. The training and testing process is repeated five times, using a different subset each time as the testing set. This allows us to calculate the performance of the machine learning model on the training set by averaging the results of the five training iterations. The Area Under the Curve (AUC), Accuracy, and Sensitivity metrics are used to evaluate the machine learning algorithms overall performance across both the training and testing sets. The AUC metric provides 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 machine 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.

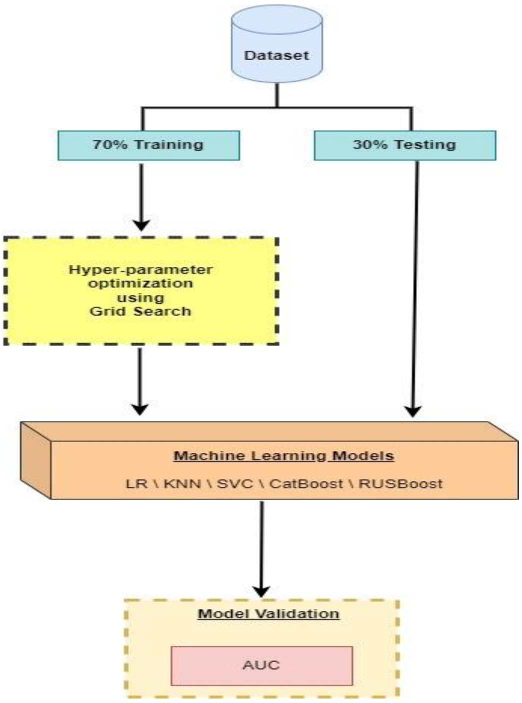

Figure 6: Machine Learning Model for the study

Table 1: Obtained optimal hyperparameters of each model by grid search

Model	Hyperparameters	Step Size	Optimal Hyperparameters	AUC	Accuracy	Sensitivity
LR	C = [1 – 500]	1	2	0.615	0.638	0.687
KNN	n_neighbors = [10 – 300]	1	32	0.779	0.796	0.843
SVC	kernel = ['linear', 'poly', 'rbf']		rbf	0.815	0.809	0.859
	C = [1-500]	1	352			
	Gamma = [0.001, 0.01, 0.1]	nil	0.01			
CatBoost	learning_rate = [0.1,0.01,0.001]	Nil	0.1	0.823	0.874	0.859
	n_estimators = [50 – 500]	1	295			
RUSBoost	learning_rate = [0.1,0.01,0.001]	Nil	0.1	0.771	0.819	0.828
	n_estimators = [50 – 500]	1	317			

4. Results and Discussion

This study utilized several classifiers such as LR, KNN, SVC, CatBoost and RUSBoost for slope stability prediction. The performance of each classifier was measured using the AUC metric, with 1.0 indicating optimal performance. The ROC curves of these classifiers, depicted in Figure 7, showed that LR had an AUC of 0.615, KNN had an AUC of 0.779, SVC had an AUC of 0.815, CatBoost had an AUC of 0.823 and RUSBoost had an AUC of 0.771. The various AUC values for different classifiers can be attributed to differences in their underlying algorithms, model complexity, and how well they capture relationships between features and the target variable. The ROC curves of machine learning models such as SVC and CatBoost were located slightly higher in the top left corner of the plot compared to the other models, indicating superior performance. These models had AUC values exceeding 0.80, which is somewhat better than that of LR, KNN and RUSBoost. Figure 8, which shows the confusion matrix, further highlights the differences in performance between classifiers. From Figure 8, it can be inferred that LR had 52 misclassifications, RUSBoost had 31, KNN had 30, SVC had 25 and CatBoost had 24. Based on these results, classifiers such as SVC and CatBoost had superior performance in terms of AUC values and had fewer misclassifications, implying better discriminatory ability and overall performance when compared to classifiers such as LR, KNN, and RUSBoost.

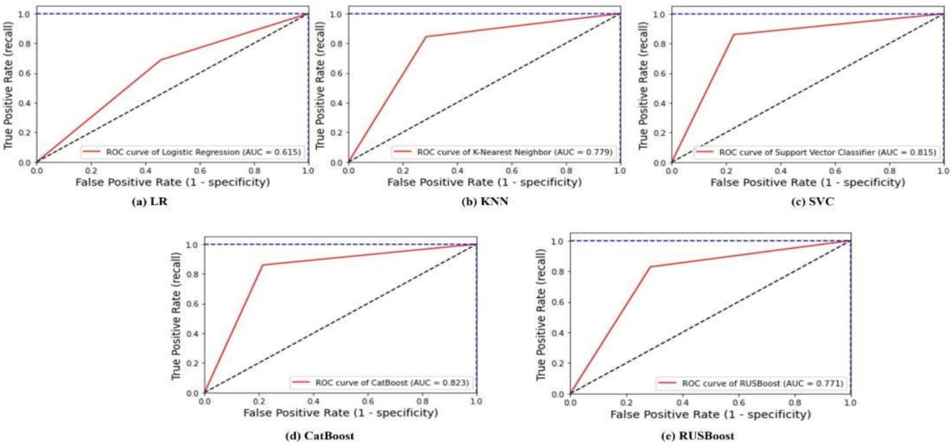

Figure 7: ROC curves of classification models on testing dataset

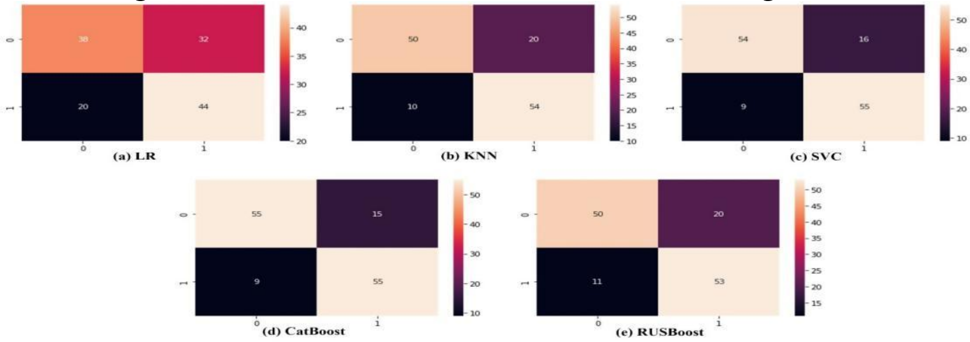

Figure 8: ROC curves of classification models on testing dataset

To validate, the performance of different classifiers was evaluated using kappa values, which range from -1 to 1. The closer the value is to 1, the more reliable the classification is. Among the classifiers evaluated, CatBoost had a kappa value of 0.642, followed closely by SVC at 0.627, while KNN and RUSBoost had values of 0.554 and 0.539, respectively. LR had the lowest kappa value at 0.228. Based on the kappa values, CatBoost outperformed other models. The ranking of models based on the kappa value is shown in Figure 9.

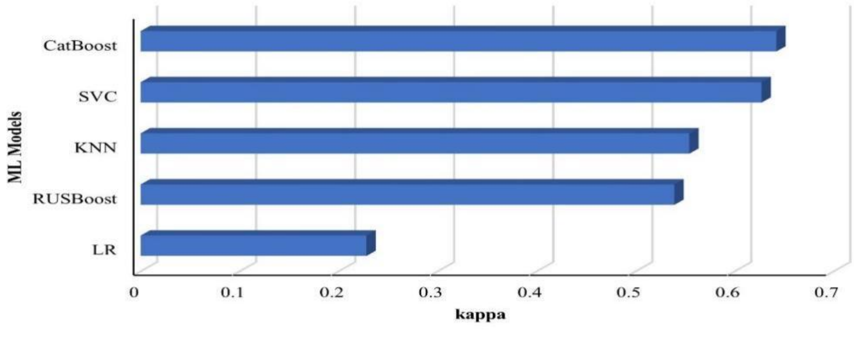

Figure 9: Performance ranking of machine learning models based on kappa

4.1 Sensitivity Analysis

To prevent slope failures, it is crucial to assess the sensitivity of features that contribute to triggering such events. Evaluating the sensitivity of these features is essential to assess slope stability and build effective support structures. In order to quantify the sensitivity of features for CatBoost model, their impact on predictive performance is assessed at the optimum hyperparameters. This sensitivity study provides valuable insights into the influential features that trigger slope stability, enabling informed decision-making and preemptive efforts to limit the risk of slope failure. Figure 10 shows the sensitivity of features for CatBoost classification model. The results reveal that the sensitivity of β is 0.921, H is 0.921, ru is 0.859, c is 0.687, ϕ is 0.515 and Υ is 0.468. It can be observed that β, H, ru, and c are highly sensitive to slope stability. Therefore, the values of β, H, ru, and c must be selected accurately and reasonably in artificial slopes, considering field tests and on-ground conditions. The values of H and β that are higher indicate that the geometry variables are highly sensitive to slope stability. Similarly, the soil property c, which is the most important, is also highly sensitive to slope stability. This means that the binding property of soil material will change significantly, thereby affecting the integrity of soil structure (internal strength). Additionally, ru also has high sensitivity, and it directly influences soil strength and stability by the ability to retain or drain water. It is crucial to optimize these values in practical design to ensure slope stability. Finally, it is worth noting that the sensitivity of ϕ and ϒ is comparatively lower than that of other features.

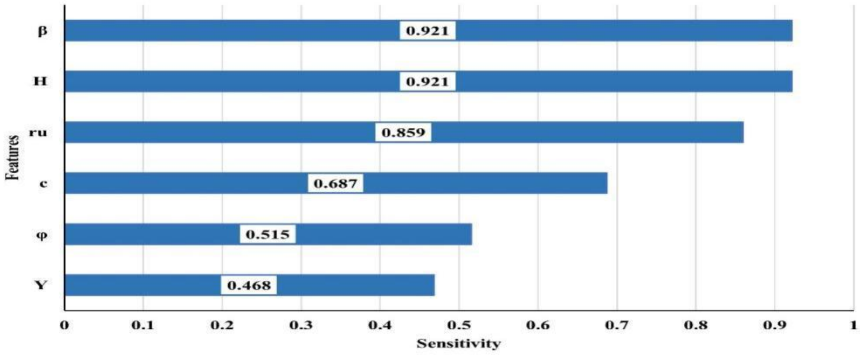

Figure 10: Representation of feature sensitivity for CatBoost model

4.2 Feature Importance Analysis

In slope stability analysis, it is crucial to identify the most influential parameter for making accurate predictions. This section assesses the feature importance using the CatBoost model, which represents the contribution of each feature in a model (as illustrated in Figure 11). The most important feature for predicting slope stability was the parameter c (0.216), followed by β (0.198), ϒ (0.185), H (0.166), ϕ (0.17), and ru (0.063). The pore water pressure (ru) had the smallest impact on the slope, with a value of 0.063. This observation is consistent with previous research presented by [21], indicating that the impact of pore water pressure on slope stability is almost negligible. Most other influential variables have significant influences that cannot be ignored. Therefore, these factors are also an indispensable part of slope stability analysis. It is important to note that different influence factor rankings may be obtained when different data sets and classification models are used. More representative results may be obtained with the emergence of more effective slope cases and algorithms in the future.

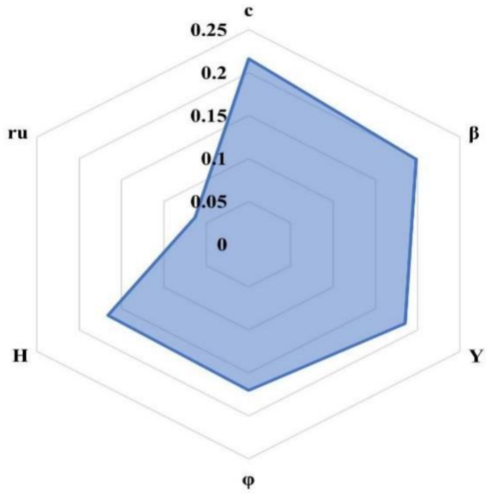

Figure 11: Representation of feature importance for CatBoost model

5. Conclusion

This study aims to compare the performance of five different machine learning classifiers, namely LR, SVC, KNN, CatBoost, and RUSBoost, in predicting the stability of 444 slope cases. Six features, including H, ru, β, c, ϒ and ϕ, were utilized to develop and evaluate the classification models. To optimize the performance of the models, grid search was used to determine the optimal hyper-parameters. Based on the analysis, the conclusion of the study are as follows:

1.Based on the evaluation metrics, it can be determined that the CatBoost model outperformed other classifiers in terms of accuracy and generalization. To validate the classification performance of the models, confusion matrix and kappa value were used in the analysis. The confusion matrix provides a comprehensive summary of the models classification results, including true positives, true negatives, false positives, and false negatives. On the other hand, kappa is a reliability measure that ranges from -1 to 1, where values closer to 1 indicate better classification performance. After analyzing the results, it can be concluded CatBoost has the best overall performance in discriminating between stable and unstable slopes. Among the machine learning methods (LR, SVC, KNN, CatBoost, and RUSBoost), CatBoost demonstrated superior performance with the highest AUC (= 0.823), least misclassification of 24 and highest kappa of 0.642. This suggests that machine learning classifier such as CatBoost is a strong alternative to other classifiers for slope stability prediction, particularly when interpretability or handling imbalanced datasets is crucial.

2.The study clearly establishes that slope stability is a complex phenomenon that cannot be accurately predicted by relying solely on one parameter. The results obtained from the CatBoost model indicate that several features exhibit high sensitivity to slope stability. Specifically, the values of β (= 0.921), H (= 0.921), ru (= 0.859) and c (= 0.687) are highly sensitive to slope stability, which means that accurate selection of these features is crucial for accurate prediction of slope stability. The importance of considering both geometrical features (such as H and β) and soil properties (such as c, and ϕ) while predicting slope stability cannot be overstated. The feature importance scores were also evaluated to measure the contribution of each feature in improving the predictive power of the model. The results from the CatBoost model indicate that the model relies most on parameter c, followed by β, ϒ, H, ϕ, and ru for making predictions. The feature importance score of ru is mainly concentrated near zero, implying that this feature has little impact on the model's predictions. This in-depth analysis highlights the intricate nature of evaluating slope stability and emphasizes the need for taking into account a comprehensive range of characteristics instead of relying on individual metrics for precise forecasting. It is important to note that accurate prediction of slope stability is essential for ensuring the safety of geotechnical structures such as roads, bridges, and buildings. The findings of this study provide valuable insights into the features that affect slope stability and emphasize the need for taking a holistic approach to slope stability analysis. By considering a range of features and characteristics, geotechnical engineers can make more informed decisions about the design and construction of geotechnical structures, thereby enhancing their safety and reliability.

3.The complex relationship between slope stability and the various features that affect it can be difficult to model accurately due to its nonlinear and multidimensional nature. However, this study case shows that the CatBoost classifier is effective in navigating this intricate connection, making precise and reliable predictions. This highlights the importance of supervised learning in assessing slope stability. To improve and expand upon the effectiveness of machine learning algorithms in assessing slope stability, future research should focus on incorporating key samples and features that influence the dynamics of slope stability. This could include integrating rainfall patterns, seismic activity, human interventions, and other environmental events into the algorithmic structure to improve prediction accuracy, generalization abilities, and reliability in real-world scenarios. By doing so, we can further expand and improve the effectiveness of machine learning methods in addressing complex geotechnical issues, ultimately helping to mitigate risks and ensure the safety of infrastructure and communities.

4.Accurate slope stability predictions enable proactive risk assessment, reducing the likelihood of landslides and structural failures in geotechnical engineering projects.

5. Machine learning-based slope stability analysis supports sustainable land-use planning, minimizes environmental degradation, and prevents soil erosion and sedimentation in water bodies.

6. The findings contribute to the development of safer and more resilient geotechnical structures, reducing maintenance costs and enhancing long-term infrastructure stability.

7. Further studies can explore deep learning techniques, hybrid ML models, and real-time monitoring systems to enhance prediction accuracy and adaptability in varying geotechnical conditions.

8. By integrating machine learning into geotechnical engineering, this research contributes to both infrastructure safety and environmental conservation, promoting a balanced approach to development and sustainability.

References

1. State of the Art: Limit Equilibrium and Finite-Element Analysis of Slopes | Journal of Geotechnical Engineering | Vol 122, No 7. https://ascelibrary.org/doi/abs/10.1061/(ASCE)0733-9410(1996)122:7(577). Accessed 1 May 2023

2. Duncan JM, Wright SG (1980) The accuracy of equilibrium methods of slope stability analysis. Eng Geol 16:5–17. https://doi.org/10.1016/0013-7952(80)90003-4
3. Slope stability analysis based on a multigrid method using a nonlinear 3D finite element model | SpringerLink. https://link.springer.com/article/10.1007/s11709-013-0190-1. Accessed 1 May 2023
4. Search for Critical Slip Surface in Slope Stability Analysis by Spline-Based GA Method | Journal of Geotechnical and Geoenvironmental Engineering | Vol 134, No 2. https://ascelibrary.org/doi/full/10.1061/%28ASCE%291090-0241%282008%29134%3A2%28252%29. Accessed 1 May 2023
5. Zhang W, Chen J, Zhang W, et al (2013) Determination of critical slip surface of fractured rock slopes based on fracture orientation data. Sci China Technol Sci 56:1248–1256. https://doi.org/10.1007/s11431-012-5129-6
6. Uncertainty analysis of correlated non-normal geotechnical parameters using Gaussian copula | SpringerLink. https://link.springer.com/article/10.1007/s11431-012-4937-z. Accessed 1 May 2023
7. Rong G, Huang K, Zhou C, et al (2012) A new constitutive law for the nonlinear normal deformation of rock joints under normal load. Sci China Technol Sci 55:555–567. https://doi.org/10.1007/s11431-011-4654-z
8. Dawson EM, Roth WH, Drescher A (1999) Slope stability analysis by strength reduction. Géotechnique 49:835–840. https://doi.org/10.1680/geot.1999.49.6.835
9. Liu SY, Shao LT, Li HJ (2015) Slope stability analysis using the limit equilibrium method and two finite element methods. Comput Geotech 63:291–298. https://doi.org/10.1016/j.compgeo.2014.10.008
10. Matsui T, San K-C (1992) Finite Element Slope Stability Analysis by Shear Strength Reduction Technique. Soils Found 32:59–70. https://doi.org/10.3208/sandf1972.32.59
11. Cheng YM, Lansivaara T, Wei WB (2007) Two-dimensional slope stability analysis by limit equilibrium and strength reduction methods. Comput Geotech 34:137–150. https://doi.org/10.1016/j.compgeo.2006.10.011
12. You G, Mandalawi MA, Soliman A, et al (2018) Finite Element Analysis of Rock Slope Stability Using Shear Strength Reduction Method. In: Frikha W, Varaksin S, Viana da Fonseca A (eds) Soil Testing, Soil Stability and Ground Improvement. Springer International Publishing, Cham, pp 227–235
13. Jiang Q, Qi Z, Wei W, Zhou C (2015) Stability assessment of a high rock slope by strength reduction finite element method. Bull Eng Geol Environ 74:1153–1162. https://doi.org/10.1007/s10064-014-0698-1
14. Wei WB, Cheng YM, Li L (2009) Three-dimensional slope failure analysis by the strength reduction and limit equilibrium methods. Comput Geotech 36:70–80. https://doi.org/10.1016/j.compgeo.2008.03.003
15. Lin S, Zheng H, Han C, et al (2021) Evaluation and prediction of slope stability using machine learning approaches. Front Struct Civ Eng 15:821–833. https://doi.org/10.1007/s11709-021-0742-8
16. Zhou J, Li E, Yang S, et al (2019) Slope stability prediction for circular mode failure using gradient boosting machine approach based on an updated database of case histories. Saf Sci 118:505–518. https://doi.org/10.1016/j.ssci.2019.05.046
17. Qi C, Tang X (2018) Slope stability prediction using integrated metaheuristic and machine learning approaches: A comparative study. Comput Ind Eng 118:112–122. https://doi.org/10.1016/j.cie.2018.02.028
18. Das SK, Biswal RK, Sivakugan N, Das B (2011) Classification of slopes and prediction of factor of safety using differential evolution neural networks. Environ Earth Sci 64:201–210. https://doi.org/10.1007/s12665-010-0839-1
19. Manouchehrian A, Gholamnejad J, Sharifzadeh M (2014) Development of a model for analysis of slope stability for circular mode failure using genetic algorithm. Environ Earth Sci 71:1267–1277. https://doi.org/10.1007/s12665-013-2531-8
20. Sakellariou MG, Ferentinou MD (2005) A study of slope stability prediction using neural networks. Geotech Geol Eng 23:419–445. https://doi.org/10.1007/s10706-004-8680-5
21. Kardani N, Zhou A, Nazem M, Shen S-L (2021) Improved prediction of slope stability using a hybrid stacking ensemble method based on finite element analysis and field data. J Rock Mech Geotech Eng 13:188–201. https://doi.org/10.1016/j.jrmge.2020.05.011
22. Zhang W, Li H, Han L, et al (2022) Slope stability prediction using ensemble learning techniques: A case study in Yunyang County, Chongqing, China. J Rock Mech Geotech Eng 14:1089–1099. https://doi.org/10.1016/j.jrmge.2021.12.011
23. Gordan B, Jahed Armaghani D, Hajihassani M, Monjezi M (2016) Prediction of seismic slope stability through combination of particle swarm optimization and neural network. Eng Comput 32:85–97. https://doi.org/10.1007/s00366-015-0400-7
24. Pham K, Kim D, Park S, Choi H (2021) Ensemble learning-based classification models for slope stability analysis. CATENA 196:104886. https://doi.org/10.1016/j.catena.2020.104886
25. Feizizadeh B, Roodposhti MS, Blaschke T, Aryal J (2017) Comparing GIS-based support vector machine kernel functions for landslide susceptibility mapping. Arab J Geosci 10:122. https://doi.org/10.1007/s12517-017-2918-z

26. Full article: Comparison of four kernel functions used in support vector machines for landslide susceptibility mapping: a case study at Suichuan area (China). https://www.tandfonline.com/doi/full/10.1080/19475705.2016.1250112. Accessed 1 May 2023
27. A Comparative Study of Landslide Susceptibility Mapping Using SVM and PSO-SVM Models Based on Grid and Slope Units. https://www.hindawi.com/journals/mpe/2021/8854606/. Accessed 1 May 2023
28. Ballabio C, Sterlacchini S (2012) Support Vector Machines for Landslide Susceptibility Mapping: The Staffora River Basin Case Study, Italy. Math Geosci 44:47–70. https://doi.org/10.1007/s11004-011-9379-9
29. Shahin MA, Maier HR, Jaksa MB (2002) Predicting Settlement of Shallow Foundations using Neural Networks. J Geotech Geoenvironmental Eng 128:785–793. https://doi.org/10.1061/(ASCE)1090-0241(2002)128:9(785)
30. Settlement prediction of shallow foundations on granular soils using B-spline neurofuzzy models - ScienceDirect. https://www.sciencedirect.com/science/article/pii/S0266352X03000946. Accessed 1 May 2023
31. Prediction of pile bearing capacity using support vector machine: International Journal of Geotechnical Engineering: Vol 5, No 1. https://www.tandfonline.com/doi/abs/10.3328/IJGE.2011.05.01.95-102. Accessed 1 May 2023
32. Lee I-M, Lee J-H (1996) Prediction of pile bearing capacity using artificial neural networks. Comput Geotech 18:189–200. https://doi.org/10.1016/0266-352X(95)00027-8
33. Modeling Pile Capacity Using Support Vector Machines and Generalized Regression Neural Network | Journal of Geotechnical and Geoenvironmental Engineering | Vol 134, No 7. https://ascelibrary.org/doi/full/10.1061/%28ASCE%291090-0241%282008%29134%3A7%281021%29. Accessed 1 May 2023
34. Besalatpour A, Hajabbasi M, Ayoubi S, et al (2012) Prediction of soil physical properties by optimized support vector machines. Int Agrophysics 26:109–115. https://doi.org/10.2478/v10247-012-0017-7
35. Prediction of Maximum Dry Density and Unconfined Compressive Strength of Cement Stabilised Soil Using Artificial Intelligence Techniques | SpringerLink. https://link.springer.com/article/10.1007/s40891-016-0051-9. Accessed 1 May 2023
36. A comprehensive support vector machine-based classification model for soil quality assessment - ScienceDirect. https://www.sciencedirect.com/science/article/pii/S0167198715001543. Accessed 1 May 2023
37. "APPLICATION OF A SUPPORT VECTOR MACHINE FOR LIQUEFACTION ASSESSMENT" by Ching-Yinn Lee and Shuh-Gi Chern. https://jmstt.ntou.edu.tw/journal/vol21/iss3/10/. Accessed 1 May 2023
38. DeMaris A (1995) A Tutorial in Logistic Regression. J Marriage Fam 57:956. https://doi.org/10.2307/353415
39. Guo G, Wang H, Bell D, et al (2003) KNN Model-Based Approach in Classification. In: Meersman R, Tari Z, Schmidt DC (eds) On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE. Springer, Berlin, Heidelberg, pp 986–996
40. Zhang M-L, Zhou Z-H (2007) ML-KNN: A lazy learning approach to multi-label learning. Pattern Recognit 40:2038–2048. https://doi.org/10.1016/j.patcog.2006.12.019
41. V. N. Vapnik, A. Ya. Chervonenkis, "On a perceptron class", Avtomat. i Telemekh., 25:1 (1964), 112–120. https://www.mathnet.ru/php/archive.phtml?wshow=paper&jrnid=at&paperid=11558&option_lang=eng. Accessed 7 Jun 2023
42. Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. In: Proceedings of the fifth annual workshop on Computational learning theory. Association for Computing Machinery, New York, NY, USA, pp 144–152
43. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20:273–297. https://doi.org/10.1007/BF00994018
44. Platt J (2000) Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. Adv Large Margin Classif 10:
45. Joachims T (2006) Training linear SVMs in linear time. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. Association for Computing Machinery, New York, NY, USA, pp 217–226
46. Mashford J, Marlow D, Tran D, May R (2011) Prediction of Sewer Condition Grade Using Support Vector Machines. J Comput Civ Eng 25:283–290. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000089
47. Gui G, Pan H, Lin Z, et al (2017) Data-driven support vector machine with optimization techniques for structural health monitoring and damage detection. KSCE J Civ Eng 21:523–534. https://doi.org/10.1007/s12205-017-1518-5
48. Zhou Y, Su W, Ding L, et al (2017) Predicting Safety Risks in Deep Foundation Pits in Subway Infrastructure Projects: Support Vector Machine Approach. J Comput Civ Eng 31:04017052. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000700

49. An S-H, Park U-Y, Kang K-I, et al (2007) Application of Support Vector Machines in Assessing Conceptual Cost Estimates. J Comput Civ Eng 21:259–264. https://doi.org/10.1061/(ASCE)0887-3801(2007)21:4(259)
50. Mahfouz T (2012) Unstructured Construction Document Classification Model through Support Vector Machine (SVM). 126–133. https://doi.org/10.1061/41182(416)16
51. COVID-19 Infection Detection from Chest X-Ray Images Using Hybrid Social Group Optimization and Support Vector Classifier - PubMed. https://pubmed.ncbi.nlm.nih.gov/33688379/. Accessed 5 May 2024
52. Dorogush AV, Ershov V, Gulin A (2018) CatBoost: gradient boosting with categorical features support
53. Prokhorenkova L, Gusev G, Vorobev A, et al (2018) CatBoost: unbiased boosting with categorical features. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.
54. Huang G, Wu L, Ma X, et al (2019) Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. J Hydrol 574:1029–1041. https://doi.org/10.1016/j.jhydrol.2019.04.085
55. Seiffert C, Khoshgoftaar TM, Van Hulse J, Napolitano A (2008) RUSBoost: Improving classification performance when training data is skewed. In: 2008 19th International Conference on Pattern Recognition. IEEE, Tampa, FL, USA, pp 1–4
56. Comparative performance of eight ensemble learning approaches for the development of models of slope stability prediction | Acta Geotechnica. https://link.springer.com/article/10.1007/s11440-021-01440-1. Accessed 11 Apr 2024