Efficient Seismic Stability Analysis of Embankment Slopes Subjected to Water Level Changes Using Gradient Boosting Algorithms

Luqi Wang [1,2], Jiahao Wu [1], Wengang Zhang [1,2], Lin Wang [1,2]* and Wei Cui [3]

[1]School of Civil Engineering, Chongqing University, Chongqing, China, [2]Key Laboratory of New Technology for Construction of Cities in Mountain Area, Chongqing University, Chongqing, China, [3]CCCC Infrastructure Maintenance Group Co., LTD., Beijing, China

Embankments are widespread throughout the world and their safety under seismic conditions is a primary concern in the geotechnical engineering community since the failure events may lead to disastrous consequences. This study proposes an efficient seismic slope stability analysis approach by introducing advanced gradient boosting algorithms, namely Categorical Boosting (CatBoost), Light Gradient Boosting Machine (LightGBM), and Extreme Gradient Boosting (XGBoost). A database consisting of 600 datasets is prepared for model calibration and evaluation, where the factor of safety (FS) is regarded as the output and four influential factors are selected as the inputs. For each dataset, the FS corresponding to the four inputs is evaluated using the commercial geotechnical software of Slide2. As an illustration, the proposed approach is applied to the seismic stability analysis of a hypothetical embankment example subjected to water level changes. For comparison, the predictive performance of CatBoost, LightGBM, and XGBoost is investigated. Moreover, the Shapley additive explanations (SHAP) method is used in this study to explore the relative importance of the four features. Results show that all the three gradient boosting algorithms (i.e., CatBoost, LightGBM, and XGBoost) perform well in the prediction of FS for both the training dataset and testing dataset. Among the four influencing factors, the friction angle φ is the most important feature variable, followed by horizontal seismic coefficient K_h, cohesion c, and saturated permeability k_s.

Keywords: machine learning, seismic slope stability, embankment, CatBoost, LightGBM, XGBoost

INTRODUCTION

The embankment is one of the most important infrastructures distributed around the world and has gained increasing attention in geotechnical and hydrogeological communities because its failure may induce disastrous consequences (e.g., Hicks and Li, 2018; Wang et al., 2018; Gordan et al., 2021). Rational stability assessment of embankments is a prerequisite for disaster prevention and reduction, and the index of the factor of safety (FS) obtained from deterministic slope stability analysis methods (e.g., limit equilibrium method and finite element method) is frequently applied to measure the slope stability due to its conceptual simplicity. It is well recognized that the embankment slope stability is significantly affected by the combined effects of several internal factors (e.g., shear strength

parameters and hydraulic parameters) and external factors (e.g., earthquakes, water level fluctuations, and rainfall). Under such circumstances, slope stability prediction can offer a fast estimation of the stability status and further provide a scientific basis for decision-making in disaster mitigation (Qi and Tang, 2018).

In the past few decades, many researchers have contributed to slope stability prediction and significant progress has been achieved in landslide disaster prevention (e.g., Sakellariou and Ferentinou, 2005; Gordan et al., 2016; Mahdiyar et al., 2017; Mojtahedi et al., 2019; Bui et al., 2020; Wang et al., 2020c, Wang L. et al., 2021; Zeng et al., 2021; Zhuang and Xing, 2021). For example, Sakellariou and Ferentinou (2005) introduced neural networks to predict slope stability. The geotechnical and geometrical parameters were taken as inputs and the FS or stability status was considered as output in their study. Gordan et al. (2016) developed a hybrid prediction model for predicting the FS of homogeneous slopes through combining the particle swarm optimization (PSO) algorithm and artificial neural network (ANN). They found that the proposed PSO-ANN method performs better than the ANN model in the prediction of FS. Mahdiyar et al. (2017) employed Monte Carlo (MC) technique to predict the FS of slopes under seismic conditions based on the five important input parameters, including slope height, slope angle, cohesion, angle of internal friction, and peak ground acceleration. Results showed that the MC-based approach is able to predict the FS appropriately. Qi and Tang (2018) compared the predictive performance of six machine learning algorithms (i.e., logistic regression, decision tree, random forest, gradient boosting machine, support vector machine, and multilayer perceptron neural network), and concluded that integrated artificial intelligence techniques had great potential in the prediction of slope stability.

Recently, Koopialipoor et al. (2019) compared the performance of four hybrid intelligent models in the stability prediction of slopes under static and dynamic conditions, namely imperialist competitive algorithm (ICA)-ANN, genetic algorithm (GA)-ANN, particle swarm optimization (PSO)-ANN, and artificial bee colony (ABC)-ANN. It was observed that the PSO-ANN model was superior to the remaining three hybrid intelligent models in predicting the FS of slopes. Mojtahedi et al. (2019) proposed an MC-based probabilistic approach for forecasting the FS of slopes and found that the internal friction angle was the most influential factor among the four inputs through conducting sensitivity analysis. Zhou et al. (2019) applied a gradient boosting machine (GBM) approach to predict the stability status of slopes based on an updated database that records a total of 221 historical cases gathered from the literature. They found that the proposed GBM classifier can accurately capture the nonlinear relationship between slope stability status and the six influential factors. Bui et al. (2020) presented an optimized ANN model for predicting the FS of slopes by introducing the Levenberg–Marquardt backpropagation technique. Luo et al. (2021) proposed a new hybrid intelligent model to analyze the slope stability in open-pit mines by combining the PSO and cubist algorithm (CA), and results

indicated that the proposed PSO-CA model was able to provide satisfactory performance in the prediction of FS. Zeng et al. (2021) investigated the predictive performance of three hybrid least squares support vector machine (LSSVM) models and found that both the gravitational search algorithm (GSA) and whale optimization algorithm (WOA) could improve the predictive accuracy.

It can be observed that previous research focused more on geometric parameters, shear strength parameters, and seismic coefficients. In contrast, hydraulic parameters (e.g., saturated permeability) are rarely considered in slope stability prediction. In engineering practice, embankments are usually subjected to water level changes, which may pose potentially destabilizing effects on the embankment slope stability. Generally, the hydraulic parameters play an indispensable role in the seepage analysis and slope stability analysis, and thus it is necessary to take the hydraulic parameters into account in the slope stability prediction of embankments. Benefited from the rapid development of artificial intelligence, many machine learning algorithms have been proposed, and they are served as a promising tool for tackling geotechnical-related topics, such as tunnels (Zheng et al., 2019; Zhang et al., 2020; Zhu et al., 2021), embankments (Wang et al., 2020a,b), landslides (Huang et al., 2020; Wang H. et al., 2021; Liu et al., 2021; Xiao et al., 2021), and other issues (Atangana Njock et al., 2021; Jamei et al., 2021; Shen et al., 2021).

This study aims to develop an efficient seismic slope stability analysis approach by introducing three advanced machine learning algorithms, namely Categorical Boosting (CatBoost), Light Gradient Boosting Machine (LightGBM), and Extreme Gradient Boosting (XGBoost). The four influential factors (i.e., cohesion, friction angle, horizontal seismic coefficient, and saturated permeability) are selected as the inputs and the FS is regarded as the output. The remainder of this paper starts with the introduction of CatBoost, LightGBM, and XGBoost, followed by a description of the associated implementation procedures. Then, the proposed approach is applied to the seismic stability analysis of a hypothetical embankment example subjected to water level changes. A database consisting of 600 datasets is compiled for model calibration and evaluation, where the four influential factors are selected as the inputs and the factor of safety (FS) is regarded as the output. Finally, the performance of CatBoost, LightGBM, and XGBoost in the prediction of FS is investigated, and the relative importance of features is ranked using the Shapley additive explanations (SHAP) method.

METHODOLOGIES

Categorical Boosting

CatBoost is a new open-source library shared by the Yandex company, which aims to handle the categorical features and prediction shift problems in machine learning (Dorogush et al., 2018; Prokhorenkova et al., 2018). Besides numerical features, categorical features are also frequently encountered in the application of machine learning, which contains a discrete set

of values that are not necessarily comparable with each other. It is evident that such categorical features can not be identified in the binary decision trees and requires to be converted to numerical features through encoding techniques. As a widely used encoding technique, the one-hot encoding may cause the curse of dimensionality in tackling the high cardinality features and tends to be more efficient in handling the low-cardinality features. To address this issue, CatBoost uses the target statistics (TS) as new numerical features to deal with the categorical features, which has been proved to be the most efficient method with minimum information loss (Prokhorenkova et al., 2018). It generates a random permutation of the dataset and then calculates the average label value of the training examples with the same category in the permutation. Following Prokhorenkova et al. (2018), if $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ is a permutation, the category $x_{\sigma_p,k}$ can be substituted with the average label value $\hat{x}_{\sigma_p,k}$:

$$\hat{x}_{\sigma_p,k} = \frac{\sum_{j=1}^{p-1}\left[x_{\sigma_j,k} = x_{\sigma_p,k}\right]Y_{\sigma_j} + a \cdot P}{\sum_{j=1}^{p-1}\left[x_{\sigma_j,k} = x_{\sigma_p,k}\right] + a} \quad (1)$$

where P is a prior value; a is the weight of the prior; Y_{σ_j} is a label value; $[\cdot]$ denotes the Iverson bracket, namely $[x_{\sigma_j,k} = x_{\sigma_p,k}]$ equals 1 if $x_{\sigma_j,k} = x_{\sigma_p,k}$, and otherwise, it is equal to 0.

Traditional gradient boosting decision tree algorithms generally suffer from an inevitable problem of gradient bias, which will eventually lead to prediction shift. Although the ordered boosting algorithm can avoid the prediction shift, it may be infeasible in practical applications due to the computational complexity and memory requirements in the process of training a larger number of supporting models. In such a case, CatBoost uses a modification of the ordered boosting algorithm in which the gradient boosting algorithm with decision trees are taken as base predictors. Furthermore, CatBoost also has superiority in the aspects of fast scorer and fast training on GPU. Interested readers are referred to Prokhorenkova et al. (2018) and Dorogush et al. (2018) for more details about the CatBoost.

Light Gradient Boosting Machine

LightGBM is a novel member of the histogram-based gradient boosting decision tree (GBDT) developed by Microsoft in 2017 for tackling the problems with big data and a large number of features (Ke et al., 2017). Conventional GBDT models require scanning all the data to evaluate the information gain of all the possible split points for each feature, indicating that the computational efforts may become prohibitively expensive when the data size is large and the feature dimension is high. To address this issue, LightGBM introduces two advanced techniques called Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) to reduce the number of the data instances and features in a rational manner.

The gradient of data instance generally poses a significant effect on the evaluation of information gain. Compared with the data instances with larger gradients, the data instances with small gradients contribute less to the estimation of information gain. In other words, more attention should be paid to the data instances with larger gradients. Inspires by this thought, GOSS reduces the

number of data instances by excluding the data instances with small gradients and simply using the rest to calculate the information gain. Moreover, many features may be mutually exclusive in a sparse feature space, and these mutually exclusive features are unable to take nonzero values simultaneously. The basic idea of EFB is to reduce the number of features by bundling mutually exclusive features. These two novel techniques (i.e., GOSS and EFB) enable the LightGBM to achieve excellent performance in terms of computational efficiency and memory consumption. More detailed explanations of the LightGBM can refer to Ke et al. (2017).

Extreme Gradient Boosting

XGBoost is a scalable end-to-end tree boosting method developed by Chen and Guestrin (2016), which has gained increasing attention in the famous Kaggle machine learning competitions due to its advantages of high efficiency and sufficient flexibility. The main idea of XGBoost is to build classification or regression trees one by one in an additive manner, and each tree learns from its predecessors and updates the residual errors in the estimated values (Zhang W. et al., 2021). Specifically, the prediction result of the gradient boosting tree model can be evaluated by integrating the values calculated from all the previously trained trees. The depth and number of trees play a significant role in the XGBoost model construction, which affect the predictive accuracy directly and can be determined by optimizing the objective function. Inspired by Chen and Guestrin (2016), the objective function $Obj^{(t)}$ is expressed as:

$$Obj^{(t)} = \sum_{i=1}^{t} l\left(Y_i, y_i^{(t)}\right) + \sum_{i=1}^{t} \Omega(f_i) \quad (2)$$

where Y_i is the actual value; $y_i^{(t)}$ is the predicted value; $l\left(Y_i, y_i^{(t)}\right)$ is the loss function describing that how well the model fits training data; $\Omega(f_i)$ is a regularization term to penalize model complexity and avoid potential over-fitting problems. For more detailed information about the XGBoost algorithm, interested readers can refer to Chen and Guestrin (2016).

IMPLEMENTATION PROCEDURE

Figure 1 shows the implementation procedures of seismic stability analysis of embankment slopes using gradient boosting algorithms. Firstly, the database used for model calibration should be prepared, which contains the necessary information about the input parameters (e.g., shear strength parameters), and output quantity of interest (e.g., FS). Then, divide the database into the training dataset and testing dataset according to a rational ratio. Thereafter, the three variants of the gradient boosting algorithms, namely CatBoost, LightGBM, and XGBoost, are used to construct the machine learning models, where the associated hyper-parameters can be determined by optimization techniques (e.g., Bayesian optimization). Finally, the predictive performance of these constructed machine learning models can be quantitatively measured using statistical indicators (e.g., the coefficient of determination R^2). For illustration, the
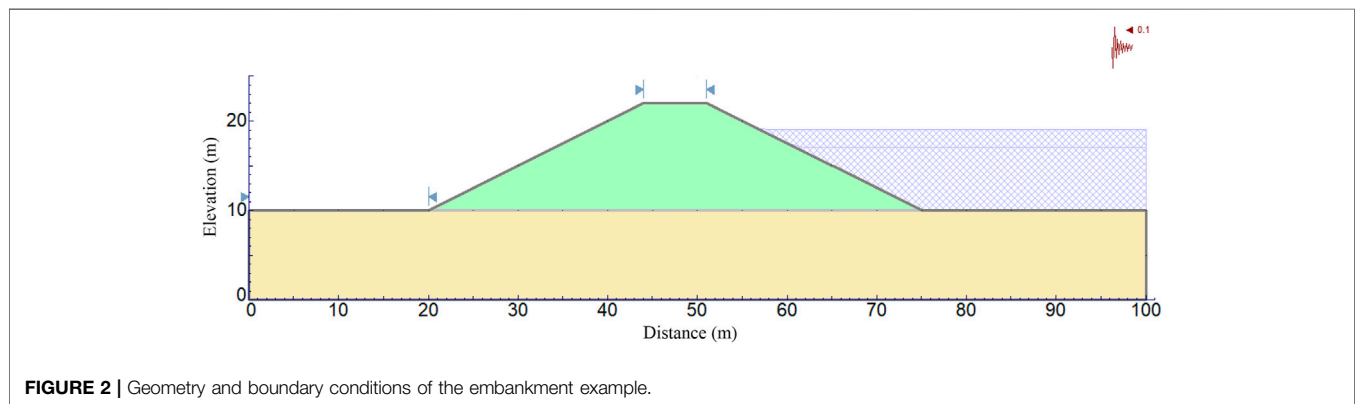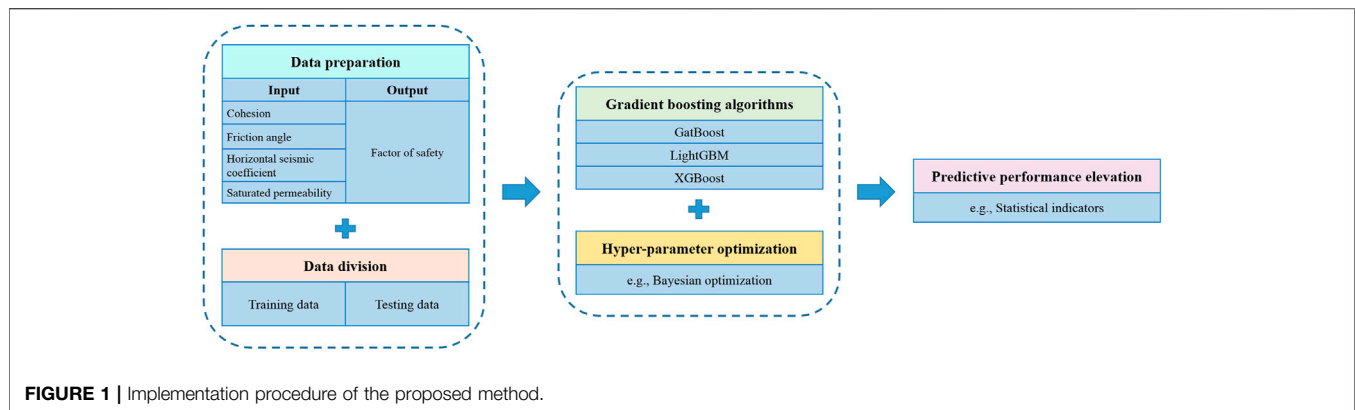
FIGURE 1 | Implementation procedure of the proposed method.

FIGURE 2 | Geometry and boundary conditions of the embankment example.

TABLE 1 | Statistical properties of parameters used in this example.

Parameters	Mean value	COV	Distribution
Cohesion c (kPa)	8	0.25	Lognormal
Friction angle φ (°)	28	0.15	Lognormal
Horizontal seismic coefficient K_h	0.1	0.3	Lognormal
Saturated permeability k_s (m/s)	1.0×10^{-6}	0.5	Lognormal

proposed approach is applied to the seismic stability analysis of a hypothetical embankment case in the next section.
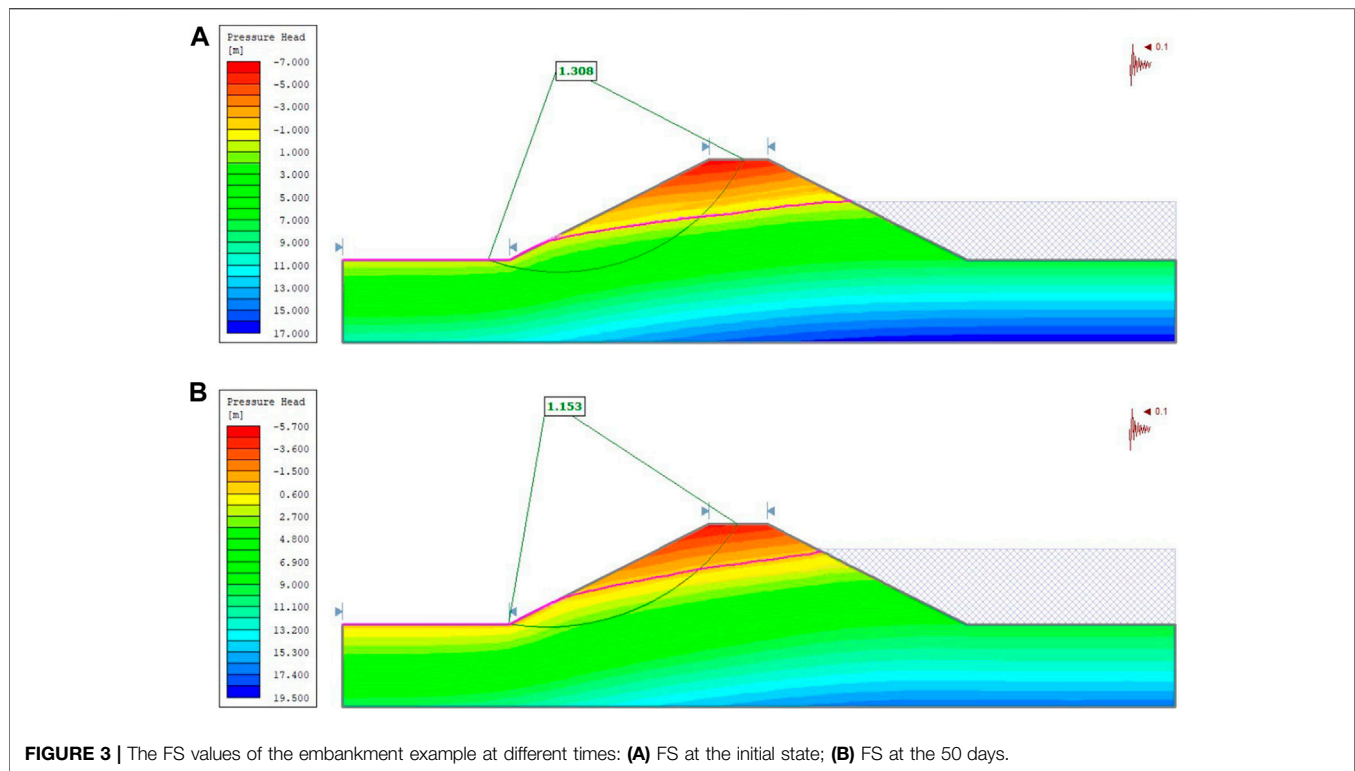
ILLUSTRATIVE EXAMPLE

For illustration, a hypothetical embankment example with a height of 12 m and a slope of 27° is used in this study for illustration, as shown in **Figure 2**. It is situated on a foundation of 100 m. Due to the fact that the embankments suffer from water level changes frequently, and thus a constant total head equal to the upstream water level is applied to the embankment below the water level. For the foundation, a zero flux boundary is assigned to both sides and the bottom. In this example, the 2D limit equilibrium slope stability software Slide2 (Rocscience Inc., 2018) is applied to perform seepage and slope stability analysis of the embankment example under combined

effects of seismic loading and water level changes. The water level is assumed to rise uniformly from the initial water level (i.e., 17 m) to the highest water level (i.e., 19 m) after 8 days. **Table 1** tabulates the mean values of the four main influential factors the govern the stability of embankment slopes, including the cohesion c, friction angle φ, horizontal seismic coefficient K_h, and saturated permeability k_s. Based on these mean values, the simplified Bishop method embedded in the Slide2 software can be applied to calculate the FS of the downstream slope. **Figures 3A,B** plot the FS values of embankment slope example at the initial state and 50 days, respectively. The FS at 50 days reaches a steady-state, and it is used as a baseline in the following database preparation.

Database Preparation for Model Calibration

A database containing the four input parameters (i.e., c, φ, K_h, and k_s) and the corresponding output of FS should be prepared for calibrating the machine learning models. Inspired by previous research (e.g., Cho, 2012; Li et al., 2015; Zhang W. G. et al., 2021), the four input parameters are assumed to follow lognormal distributions, so as to avoid possible negative values that may be physically meaningless. Based on the mean values, coefficients of variation (COVs), and probability distributions tabulated in **Table 1**, a total of 600 groups of data are generated using the Latin hypercube sampling method. **Figure 4A–D** plot the histogram of the cohesion c, friction angle φ, horizontal seismic coefficient K_h,

FIGURE 3 | The FS values of the embankment example at different times: **(A)** FS at the initial state; **(B)** FS at the 50 days.

and saturated permeability k_s, respectively. The possible ranges for c, φ, K_h, and k_s are [3.74 kPa, 17.66 kPa] [17.29°, 45.08°] [0.04, 0.25], and [1.73×10^{-7}, 5.69×10^{-6}], respectively. Each data group containing the c, φ, K_h, and k_s is used as input in the Slide2 software for calculating the FS of the embankment slope example. With the aid of Slide2 software, all the FS values corresponding to the 600 groups of data can be evaluated. As plotted in **Figure 4E**, the FS values range from 0.747 to 1.507. These input parameters and output consequences constitute a database with a total of 600 datasets and each dataset consists of four input parameters (i.e., c, φ, K_h, and k_s) associated with the corresponding FS value. Although the Slide2 software is used in this study to perform seismic stability analyses of the 600 groups of data, other geotechnical commercial software of interest can also be applied.

The compiled database can be divided into training dataset and testing dataset for model construction and evaluation. In this study, 400 groups of data are used as the training dataset and 200 groups of data are regarded as the testing dataset. Then, the three gradient boosting algorithms (i.e., CatBoost, LightGBM, and XGBoost) are used to construct the machine learning models. The performance of different machine models in the prediction of FS can be evaluated using statistical indicators.
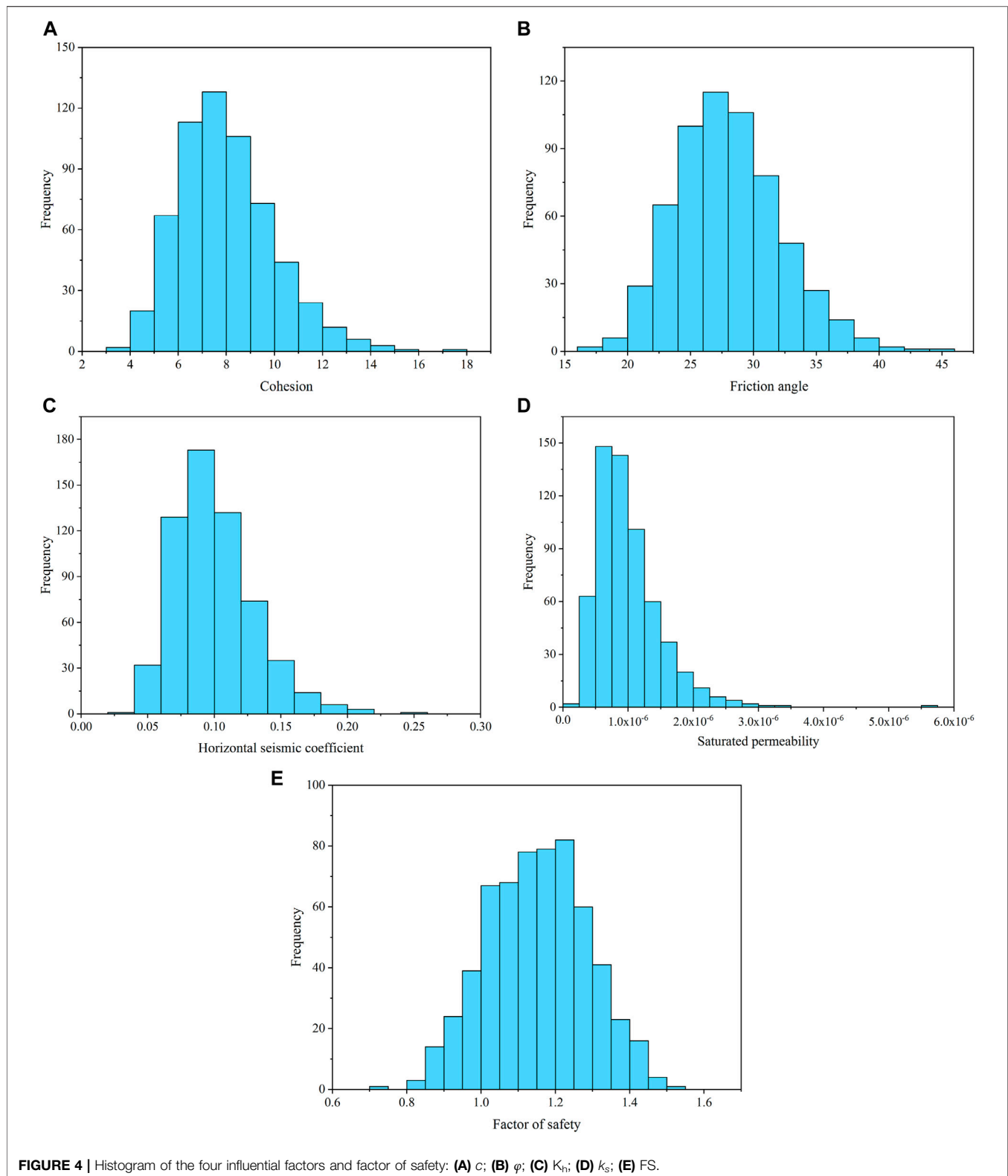
Predictive Performance of Different Models

Figure 5A compares the FS values obtained from the established CatBoost model and actual values calculated from the Slide2 software for all the 600 groups of data. It can be observed that the predicted FS values obtained from the established CatBoost model agree well with those calculated from the Slide2 software for both the training dataset (i.e., 400 groups of data)

and testing dataset (i.e., 200 groups of data). To quantitatively evaluate the model performance, the frequently used index of the coefficient of determination (R^2) is used in this study. As shown in **Figure 5A**, the R^2 values of training dataset and testing dataset are larger than 0.90, indicating that the established CatBoost model is able to predict the FS of the embankment slope example with satisfactory accuracy. Likewise, **Figure 5B** compares the FS values predicted from the constructed LightGBM model and actual values calculated from the Slide2 software. Both the training dataset and testing dataset can achieve a relatively high R^2 value, illustrating the excellent capability of LightGBM model in predicting the FS. Furthermore, **Figure 5C** compares the prediction results of XGBoost model and actual values calculated from the Slide2 software. It is shown that most of the points gather around the reference line (i.e., 1:1 line), and the corresponding R^2 values of training dataset and testing dataset are also relatively high. This implies that the XGBoost model performs well in the prediction of FS. In general, it can be concluded that all the three machine learning models (i.e., CatBoost, LightGBM, and XGBoost model) are able to provide satisfactory performance in the prediction of FS for the embankment slope example, which offers a promising approach for seismic stability analysis by introducing advanced gradient boosting algorithms.

Feature Importance Analysis

To investigate the relative importance of features on the predictive performance of machine learning models, the Shapley additive explanations (SHAP) method is used in this study due to its fast implementation for tree-based models. It uses

FIGURE 4 | Histogram of the four influential factors and factor of safety: **(A)** c; **(B)** φ; **(C)** K_h; **(D)** k_s; **(E)** FS.

the Shapley values to quantify the contribution of each feature to the prediction based on the coalitional game theory (Lundberg and Lee, 2017; Guo et al., 2021). Generally, the features with

higher positive SHAP values tend to pose a more significant influence on the final prediction. **Figure 6** plots the SHAP values of the four features calculated from the CatBoost model. Each

FIGURE 5 | Predictive performance of the three gradient boosting algorithms: **(A)** CatBoost model; **(B)** LightGBM model; **(C)** XGBoost model.

FIGURE 6 | SHAP values of features calculated from the CatBoost model.

FIGURE 7 | Relative importance of features calculated from the CatBoost model.
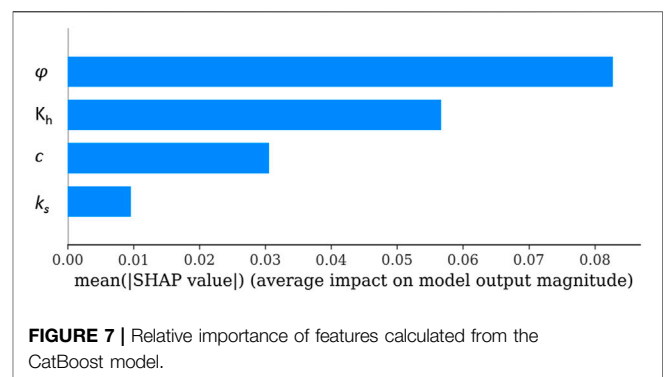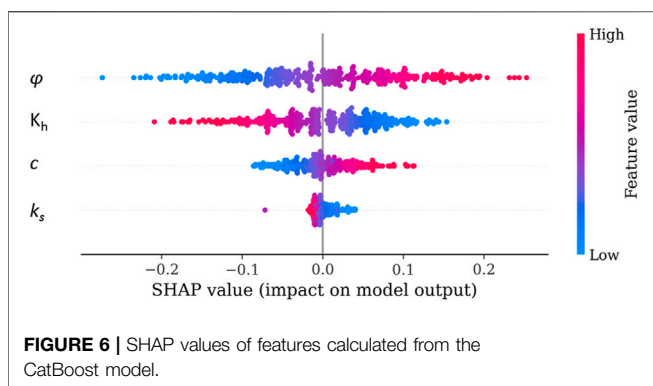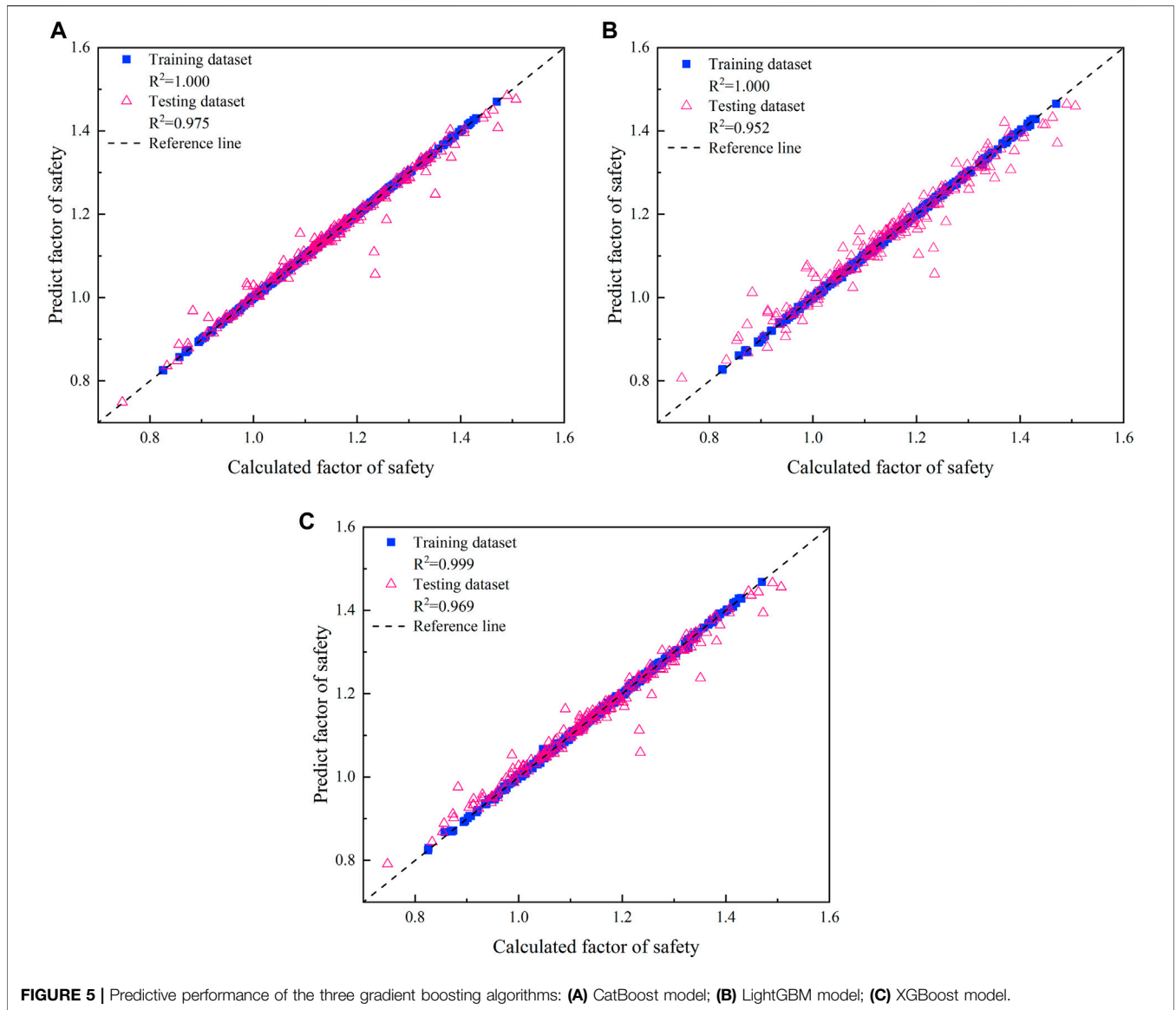
scattered point on the figure represents one sample, and the points with red colors indicate that the associated feature values are high. On the other hand, the blue colors imply that the feature

values are low. For the friction angle φ, it can be observed that many sample points with red colors gather around the zone with positive SHAP values, indicating that the friction angle affects the

FS of the embankment slopes significantly, and the larger value of friction angle will enhance the embankment slope stability. In contrast, for the horizontal seismic coefficient K_h, a large number of sample points with red colors locate in the zone with negative SHAP values. This means that the horizontal seismic coefficient will weaken the embankment slope stability.

In general, the friction angle φ has the most significant influence on the prediction of FS, followed by horizontal seismic coefficient K_h, cohesion c, and saturated permeability k_s. Among the four features, the shear strength parameters (i.e., φ and c) have positive influences on the embankment slope stability, while the increasing K_h and k_s will destabilize the embankment slope stability. Furthermore, **Figure 7** ranks feature importance of the four features. The arrangement of these four features from bottom to top is based on their relative importance. Similarly, it can be found that the friction angle φ has the most significant influence on the prediction of FS, followed by K_h, c, k_s. This finding is consistent with that observed in **Figure 6**, further validating the significance of shear strength parameters (i.e., φ and c) and seismic coefficient (i.e., K_h) in the seismic stability evaluation of embankment slopes.

SUMMARY AND CONCLUSION

This paper developed a gradient boosting algorithm-based approach for seismic stability analysis of embankment slopes. Three advanced gradient boosting algorithms, namely Categorical Boosting (CatBoost), Light Gradient Boosting Machine (LightGBM), and Extreme Gradient Boosting (XGBoost), were calibrated and evaluated in this study using a well-established database that contains a total of 600 datasets. Each dataset records the four features (i.e., the cohesion, friction angle, horizontal seismic coefficient, and saturated permeability) associated with the factor of safety (FS). For illustration, the proposed approach was applied to the seismic stability analysis of a hypothetical embankment example subjected to water level changes. The predictive performance of CatBoost, LightGBM, and XGBoost were compared, and the relative importance of features on the prediction was also quantified by the Shapley additive explanations (SHAP) method.

Results showed that all the coefficient of determination (R^2) values of the three gradient boosting algorithms (i.e., CatBoost, LightGBM, and XGBoost) were larger than 0.90 for both the training dataset and testing dataset, indicating that the proposed approach is able to predict the FS of embankment slopes with satisfactory accuracy. Among the four influencing factors, the friction angle φ had the most significant influence on the prediction of FS, followed by horizontal seismic coefficient K_h, cohesion c, and saturated permeability k_s. Different from the shear strength parameters (i.e., φ and c) that had positive influences on the embankment slope stability, the increasing K_h and k_s tended to destabilize the embankment slope stability. The proposed approach making the best use of advanced gradient boosting algorithms can serve as a useful tool for geotechnical practitioners to grasp the stability status of slopes accurately and fastly, and further provides a scientific basis for decision making in disaster prevention and mitigation. Besides the above four influential factors, other geometric and geotechnical parameters of interest can also be considered in future studies. This study provided a preliminary exploration of the machine learning-aided seismic stability analysis of embankment slopes subjected to water level changes, and a practical engineering case considering more influential factors warrants further research.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

LQW contributed to conceptualization, original manuscript preparation and funding acquisition. JW contributed to formal analysis, methodology and software. WZ contributed to supervision and polished the manuscript. LW contributed to original manuscript preparation and revised the manuscript. WC contributed to data curation and investigation.

ACKNOWLEDGMENTS

REFERENCES

Atangana Njock, P. G., Shen, S.-L., Zhou, A., and Modoni, G. (2021). Artificial Neural Network Optimized by Differential Evolution for Predicting Diameters of Jet Grouted Columns. *J. Rock Mech. Geotechnical Eng.* doi:10.1016/j.jrmge.2021.05.009

Bui, X.-N., Muazu, M. A., and Nguyen, H. (2020). Optimizing Levenberg-Marquardt Backpropagation Technique in Predicting Factor of Safety of Slopes after Two-Dimensional OptumG2 Analysis. *Eng. Comput.* 36 (3), 941–952. doi:10.1007/s00366-019-00741-0

Chen, T., and Guestrin, C. (2016). "Xgboost: A Scalable Tree Boosting System," in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 785–794.

Cho, S. E. (2012). Probabilistic Analysis of Seepage that Considers the Spatial Variability of Permeability for an Embankment on Soil Foundation. *Eng. Geology.* 133-134, 30–39. doi:10.1016/j.enggeo.2012.02.013

Dorogush, A. V., Ershov, V., and Gulin, A. (2018). *CatBoost: Gradient Boosting with Categorical Features Support.* arXiv preprint, 181011363.

Gordan, B., Jahed Armaghani, D., Hajihassani, M., and Monjezi, M. (2016). Prediction of Seismic Slope Stability through Combination of Particle Swarm Optimization and Neural Network. *Eng. Comput.* 32 (1), 85–97. doi:10.1007/s00366-015-0400-7

Gordan, B., Raja, M. A., Armaghani, D. J., and Adnan, A. (2021). Review on Dynamic Behaviour of Earth Dam and Embankment during an Earthquake. *Geotech Geol. Eng.* 4. doi:10.1007/s10706-021-01919-4

Guo, D., Chen, H., Tang, L., Chen, Z., and Samui, P. (2021). Assessment of Rockburst Risk Using Multivariate Adaptive Regression Splines and Deep forest Model. *Acta Geotech.* 2. doi:10.1007/s11440-021-01299-2

Hicks, M. A., and Li, Y. (2018). Influence of Length Effect on Embankment Slope Reliability in 3D. *Int. J. Numer. Anal. Methods Geomech* 42 (7), 891–915. doi:10.1002/nag.2766

Huang, F., Cao, Z., Jiang, S.-H., Zhou, C., Huang, J., and Guo, Z. (2020). Landslide Susceptibility Prediction Based on a Semi-supervised Multiple-Layer Perceptron Model. *Landslides* 17 (12), 2919–2930. doi:10.1007/s10346-020-01473-9

Jamei, M., Hasanipanah, M., Karbasi, M., Ahmadianfar, I., and Taherifar, S. (2021). Prediction of Flyrock Induced by Mine Blasting Using a Novel Kernel-Based Extreme Learning Machine. *J. Rock Mech. Geotechnical Eng.* doi:10.1016/j.jrmge.2021.07.007

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al. (2017). "LightGBM: a Highly Efficient Gradient Boosting Decision Tree," in *Advances in Neural Information Processing Systems* (San Mateo, CA, USA: Morgan Kaufmann Publishers), 3146–3154.

Koopialipoor, M., Jahed Armaghani, D., Hedayat, A., Marto, A., and Gordan, B. (2019). Applying Various Hybrid Intelligent Systems to Evaluate and Predict Slope Stability under Static and Dynamic Conditions. *Soft Comput.* 23 (14), 5913–5929. doi:10.1007/s00500-018-3253-3

Li, D.-Q., Jiang, S.-H., Cao, Z.-J., Zhou, W., Zhou, C.-B., and Zhang, L.-M. (2015). A Multiple Response-Surface Method for Slope Reliability Analysis Considering Spatial Variability of Soil Properties. *Eng. Geology.* 187, 60–72. doi:10.1016/j.enggeo.2014.12.003

Liu, Z., Gilbert, G., Cepeda, J. M., Lysdahl, A. O. K., Piciullo, L., Hefre, H., et al. (2021). Modelling of Shallow Landslides with Machine Learning Algorithms. *Geosci. Front.* 12 (1), 385–393. doi:10.1016/j.gsf.2020.04.014

Lundberg, S. M., and Lee, S. I. (2017). "A Unified Approach to Interpreting Model Predictions," in Proceedings of the 31st Conference on Neural Information Processing Systems, 4768–4777.

Luo, Z., Bui, X.-N., Nguyen, H., and Moayedi, H. (2021). A Novel Artificial Intelligence Technique for Analyzing Slope Stability Using PSO-CA Model. *Eng. Comput.* 37 (1), 533–544. doi:10.1007/s00366-019-00839-5

Mahdiyar, A., Hasanipanah, M., Armaghani, D. J., Gordan, B., Abdullah, A., Arab, H., et al. (2017). A Monte Carlo Technique in Safety Assessment of Slope under Seismic Condition. *Eng. Comput.* 33 (4), 807–817. doi:10.1007/s00366-016-0499-1

Mojtahedi, S. F. F., Tabatabaee, S., Ghoroqi, M., Soltani Tehrani, M., Gordan, B., and Ghoroqi, M. (2019). A Novel Probabilistic Simulation Approach for Forecasting the Safety Factor of Slopes: a Case Study. *Eng. Comput.* 35 (2), 637–646. doi:10.1007/s00366-018-0623-5

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A. (2018). CatBoost: Unbiased Boosting with Categorical Features. *Adv. Neural Inf. Process. Syst.*, 6638–6648.

Qi, C., and Tang, X. (2018). Slope Stability Prediction Using Integrated Metaheuristic and Machine Learning Approaches: A Comparative Study. *Comput. Ind. Eng.* 118, 112–122. doi:10.1016/j.cie.2018.02.028

Rocscience Inc (2018). *Slide2 Version 2018 – 2D Limit Equilibrium Slope Stability Analysis.* Toronto, Ontario: Canada. www.rocscience.com.

Sakellariou, M. G., and Ferentinou, M. D. (2005). A Study of Slope Stability Prediction Using Neural Networks. *Geotech Geol. Eng.* 23 (4), 419–445. doi:10.1007/s10706-004-8680-5

Shen, H., Li, J., Wang, S., and Xie, Z. (2021). Prediction of Load-Displacement Performance of Grouted Anchors in Weathered Granites Using FastICA-MARS as a Novel Model. *Geosci. Front.* 12 (1), 415–423. doi:10.1016/j.gsf.2020.05.004

Wang, B., Chen, Y., Wu, C., Peng, Y., Song, J., Liu, W., et al. (2018). Empirical and Semi-analytical Models for Predicting Peak Outflows Caused by Embankment Dam Failures. *J. Hydrol.* 562, 692–702. doi:10.1016/j.jhydrol.2018.05.049

Wang, H., Zhang, L., Yin, K., Luo, H., and Li, J. (2021). Landslide Identification Using Machine Learning. *Geosci. Front.* 12 (1), 351–364. doi:10.1016/j.gsf.2020.02.012

Wang, L., Wu, C., Gu, X., Liu, H., Mei, G., and Zhang, W. (2020a). Probabilistic Stability Analysis of Earth Dam Slope under Transient Seepage Using Multivariate Adaptive Regression Splines. *Bull. Eng. Geol. Environ.* 79 (6), 2763–2775. doi:10.1007/s10064-020-01730-0

Wang, L., Wu, C., Tang, L., Zhang, W., Lacasse, S., Liu, H., et al. (2020b). Efficient Reliability Analysis of Earth Dam Slope Stability Using Extreme Gradient Boosting Method. *Acta Geotech.* 15 (11), 3135–3150. doi:10.1007/s11440-020-00962-4

Wang, L., Yin, Y., Huang, B., and Dai, Z. (2020c). Damage Evolution and Stability Analysis of the Jianchuandong Dangerous Rock Mass in the Three Gorges Reservoir Area. *Eng. Geology.* 265, 105439. doi:10.1016/j.enggeo.2019.105439

Wang, L., Zhang, Z., Huang, B., Hu, M., and Zhang, C. (2021). Triggering Mechanism and Possible Evolution Process of the Ancient Qingshi Landslide in the Three Gorges Reservoir. *Geomatics, Nat. Hazards Risk* 12, 3160–3174. doi:10.1080/19475705.2021.1998230

Xiao, T., Yu, L. B., Tian, W. M., Zhou, C., and Wang, L. Q. (2021). Reducing Local Correlations Among Causal Factor Classifications as a Strategy to Improve Landslide Susceptibility Mapping. *Front. Earth Sci.* doi:10.3389/feart.2021.781674

Zeng, F., Nait Amar, M., Mohammed, A. S., Motahari, M. R., and Hasanipanah, M. (2021). Improving the Performance of LSSVM Model in Predicting the Safety Factor for Circular Failure Slope through Optimization Algorithms. *Eng. Comput.* doi:10.1007/s00366-021-01374-y

Zhang, W. G., Meng, F. S., Chen, F. Y., and Liu, H. L. (2021). Effects of Spatial Variability of Weak Layer and Seismic Randomness on Rock Slope Stability and Reliability Analysis. *Soil Dyn. Earthquake Eng.* 146, 106735. doi:10.1016/j.soildyn.2021.106735

Zhang, W., Han, L., Gu, X., Wang, L., Chen, F., and Liu, H. (2020). Tunneling and Deep Excavations in Spatially Variable Soil and Rock Masses: A Short Review. *Underground Space.* doi:10.1016/j.undsp.2020.03.003

Zhang, W., Wu, C., Zhong, H., Li, Y., and Wang, L. (2021). Prediction of Undrained Shear Strength Using Extreme Gradient Boosting and Random forest Based on Bayesian Optimization. *Geosci. Front.* 12 (1), 469–477. doi:10.1016/j.gsf.2020.03.007

Zheng, G., Yang, P., Zhou, H., Zeng, C., Yang, X., He, X., et al. (2019). Evaluation of the Earthquake Induced Uplift Displacement of Tunnels Using Multivariate Adaptive Regression Splines. *Comput. Geotechnics* 113, 103099. doi:10.1016/j.compgeo.2019.103099

Zhou, J., Li, E., Yang, S., Wang, M., Shi, X., Yao, 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. doi:10.1016/j.ssci.2019.05.046

Zhu, X., Chu, J., Wang, K., Wu, S., Yan, W., and Chiam, K. (2021). Prediction of Rockhead Using a Hybrid N-XGBoost Machine Learning Framework. *J. Rock Mech. Geotechnical Eng.* doi:10.1016/j.jrmge.2021.06.012

Zhuang, Y., Xing, A., Leng, Y., Bilal, M., Zhang, Y., Jin, K., et al. (2021). Investigation of Characteristics of Long Runout Landslides Based on the Multi-Source Data Collaboration: A Case Study of the Shuicheng Basalt Landslide in Guizhou, China. *Rock Mech. Rock Eng.* 54, 3783–3798. doi:10.1007/s00603-021-02493-0