Full Length Article

Slope stability prediction using ensemble learning techniques: A case study in Yunyang County, Chongqing, China

Wengang Zhang [a,b,c], Hongrui Li [a], Liang Han [a], Longlong Chen [a], Lin Wang [a,*]

[a] School of Civil Engineering, Chongqing University, Chongqing, 400045, China
[b] Key Laboratory of New Technology for Construction of Cities in Mountain Area, Chongqing University, Chongqing, 400045, China
[c] National Joint Engineering Research Center of Geohazards Prevention in the Reservoir Areas, Chongqing University, Chongqing, 400045, China

ABSTRACT

Slope stability prediction plays a significant role in landslide disaster prevention and mitigation. This study develops an ensemble learning-based method to predict the slope stability by introducing the random forest (RF) and extreme gradient boosting (XGBoost). As an illustration, the proposed approach is applied to the stability prediction of 786 landslide cases in Yunyang County, Chongqing, China. For comparison, the predictive performance of RF, XGBoost, support vector machine (SVM), and logistic regression (LR) is systematically investigated based on the well-established confusion matrix, which contains the known indices of recall rate, precision, and accuracy. Furthermore, the feature importance of the 12 influencing variables is also explored. Results show that the accuracy of the XGBoost and RF for both the training and testing data is superior to that of SVM and LR, revealing the superiority of the ensemble learning models (i.e. XGBoost and RF) in the slope stability prediction of Yunyang County. Among the 12 influencing factors, the profile shape is the most important one. The proposed ensemble learning-based method offers a promising way to rationally capture the slope status. It can be extended to the prediction of slope stability of other landslide-prone areas of interest.

© 2022 Institute of Rock and Soil Mechanics, Chinese Academy of Sciences. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Landslide is one of the most severe natural hazards occurring in mountainous areas, which has attracted increasing concern in geotechnical and geological engineering researches, because it may induce considerably detrimental social and economic impacts (e.g. Huang et al., 2018, 2020a; Tang et al., 2019). China is one of the countries with the most severe landslides in the world. For example, there have been more than 5000 landslides or potential landslides distributed in the Three Gorges Reservoir area since the first impoundment in 2003 (Gu et al., 2017). Thus, it is of great significance to evaluate the slope stability for designing remedial and mitigation measures. Generally, for a specific slope, its stability can be rationally and explicitly quantified by performing slope stability analysis with existing commercial geotechnical software. However, when facing numerous landslides or potential landslides,

it is unrealistic to conduct slope stability analysis for all of them in practice. In such case, slope stability prediction has gained popularity in geotechnical engineering and geological engineering.

In the past few decades, many researchers have contributed to slope stability prediction (e.g. Qi and Tang, 2018; Zhou et al., 2019; Kardani et al., 2021; Pham et al., 2021; Zeng et al., 2021). For example, Qi and Tang (2018) compared the predictive performance of six integrated artificial intelligence approaches in slope stability prediction based on the 168 slope cases in the literature. Zhou et al. (2019) introduced the gradient boosting machine method to slope stability prediction using a database that contains 221 actual slope cases with circular mode failure. Kardani et al. (2021) developed a hybrid stacking ensemble approach for improving the slope stability prediction based on synthetic and field data. Besides, it can be observed that the previous studies have paid more attention to the mechanical parameters and geometric variables. For example, Mojtahedi et al. (2019) considered the four input parameters (i.e. slope height, slope angle, cohesion, and friction angle). Gordan et al. (2016), Mahdiyar et al. (2017), Koopialipoor et al. (2019), Luo et al. (2021), and Pham et al. (2021) used the five inputs, i.e. slope height, angle, cohesion, angle of internal friction, and unit weight (or peak

ground acceleration). Qi and Tang (2018), Zhou et al. (2019), Kardani et al. (2021), and Zeng et al. (2021) selected six influencing factors (i.e. unit weight, cohesion, angle of internal friction, slope angle, slope height, and coefficient of pore water pressure). It is well recognized that slope stability is influenced by many factors, such as mechanical parameters, geometric variables, topographic features, and geological conditions. However, the latter two factors are rarely considered. With the rapid advancement of remote sensing technology and geological exploration technology, the topographic features and geological conditions of a specified region can be acquired rationally (Huang et al., 2020b; Ji et al., 2020). Benefited from the fast development of computer technology, many machine learning algorithms have been successfully applied to addressing geotechnical-related problems and significant progress has been achieved (e.g. Kamrava et al., 2020; Liu et al., 2020; Zhang et al., 2020a, b; Kardani et al., 2021; Li et al., 2021; Wang et al., 2021). With the application of machine learning, it is possible to reveal the relative importance of topographic features and geological conditions in slope stability prediction.

As a branch of machine learning, ensemble learning techniques make full use of multiple predictors to form a superior one for improving the performance of machine learning, which has attracted increasing attention in geotechnical engineering (e.g. Wang et al., 2020a, b; Liu et al., 2021; Zhang et al., 2021a, b; Zhou et al., 2021; Zhu et al., 2021). For example, Wang et al. (2020a) proposed an extreme gradient boosting (XGBoost)-based reliability analysis approach for calculating the failure probability of earth dam slope. Zhang et al. (2021c) applied the random forest (RF) and XGBoost to predicting the undrained shear strength of soft sensitive clays. Zhou et al. (2021) compared the performance of six hybrid XGBoost models in the prediction of tunnel boring machine's penetration rate. Generally, ensemble learning techniques can be broadly categorized into two groups according to their structures, including bagging (parallel) and boosting (sequential) (Zhang et al., 2021c). The RF mentioned above belongs to the bagging method (Breiman, 2001; Xia et al., 2017) and the XGBoost is developed within the boosting framework (Chen and Guestrin, 2016). Inspired by the contribution of previous researches, this study tries to investigate the performance of these two well-known ensemble learning techniques (i.e. RF and XGBoost) in the prediction of slope stability.

This study aims to develop an approach to predict the slope stability using the RF and XGBoost. As an illustration, the proposed approach is applied to stability data of 786 landslide cases in Yunyang County. The remainder of this paper starts with the introduction of this area, followed by a brief description of the RF and XGBoost. Finally, the performance of RF, XGBoost, support vector machine (SVM), and logistic regression (LR) in the slope stability prediction of Yunyang County is systematically explored, and the feature importance of the 12 influencing factors is ranked.

2. Investigated area

The Yunyang County covers an area of approximately 3649 km^2, located in Chongqing, China. The shoreline of the Three Gorges Reservoir in this area is approximately 707.8 km, accounting for about 11% of the overall length (i.e. 6300 km).

2.1. Topographic conditions

Elevation accurately reflects the altitude of a specific geographical location, and the visualized elevation map of Yunyang County can be obtained from the ArcMap 10.2 software, as shown in Fig. 1. According to the investigation report of geological disasters in Yunyang County, this area contains numerous landslides, and

more landslides occurred in low altitude regions than that in high altitude regions. Fig. 2a and b plots the distribution of the front edge and back edge elevations, respectively. It can be seen that the heights of landslide cases in Yunyang County are mostly between 200 m and 700 m, and the cases with an elevation of about 400 m are the most.

The height describing the relative relief between the front edge and back edge elevations is significantly related to the inner stress of the slope, which plays a controlling role in slope stability. Generally, higher slopes are accompanied by stronger disturbance sensitivity. The combined effect of the slope height and angle is crucial in slope stability. According to the distribution of the heights shown in Fig. 2c, heights of most slopes are between 40 m and 120 m, while those higher than 180 m account for a small proportion.

The slope directly affects the amount of geomaterial deposited on a slope and further affects the stability of the slope. The map of slope angle can be obtained by ArcGIS software using grid analysis to transform the coordinate system and extract the slope value in the elevation figure (Fig. 3). Besides, Fig. 2d shows the distribution of slope angle. It can be observed that the slope angles are mainly between 10° and 30°, slopes with an angle more than 30° take up a few parts and only one slope with angle more than 50°. Besides, Fig. 4 also portrays the aspect map of the investigated area.

2.2. Geological conditions

Due to differences in tectonic movement, different rock masses have different cohesions and shear strengths, which thus lead to
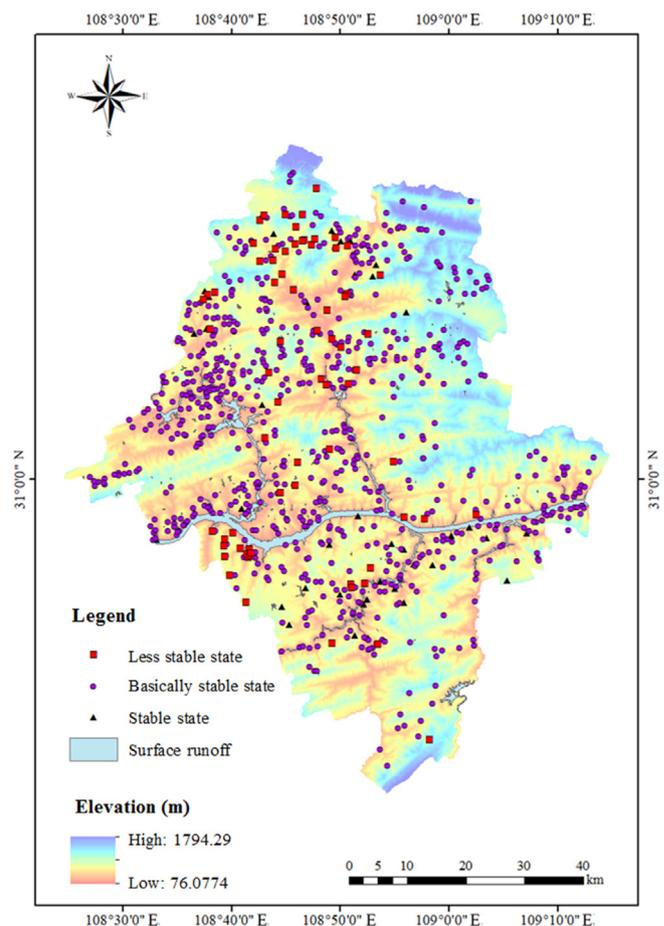

Fig. 1. Elevation map of the investigated area.
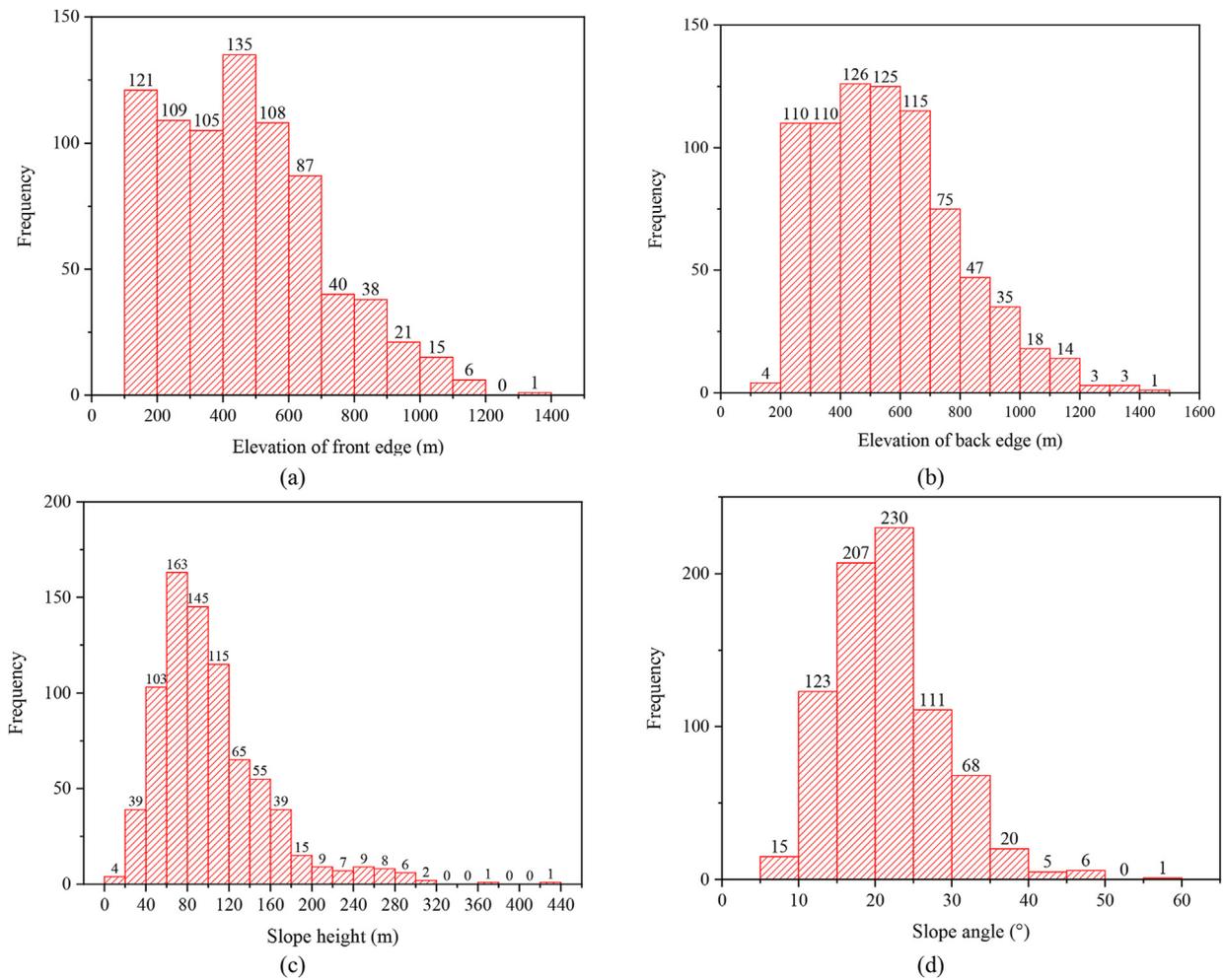
Fig. 2. Distributions of topographic features: (a) Elevation of front edge, (b) Elevation of back edge, (c) Slope height, and (d) Slope angle.

various evolution trends of slopes subject to the identical external conditions. The statistics show that rock masses in the investigated area are mainly mudstone, argillaceous limestone, sandstone, sandy mudstone, and shale. According to the classification based on rock hardness, slope rock masses in Yunyang County are divided into extremely soft rock, soft rock, moderately soft rock, and moderately hard rock. Fig. 5a plots the distribution of the four types. It shows that moderately soft rock accounts for the highest proportion while extremely soft rock is the lowest one.

Rock mass with larger inclination angle is more likely to suffer weathering, denudation, and deformation, which thus result in changes in the slope morphology. Fig. 5b shows the distribution of inclination angle. It can be observed that the inclination angles are mainly 5°–25° and the number of slopes decreases with the increase in inclination angle when larger than 10°.

The dip direction is one of the three major factors that affect the attitude of rock formations, as it determines the spatial orientation of a layer. There are many dip directions of rock layers in this area (Fig. 5c). The statistics show that slopes with north and south dip directions are most common, which is mainly related to the geographical location of Yunyang County.

Slope structures reflect the relative positions of the rock layers in the slope. According to the intersection angle between the dip and slope angles, the slopes can be categorized into five structures: dip, anti-dip, oblique-dip, cross-dip, and horizontally layered slopes (Fig. 5d). It is shown that the dip, anti-dip, and oblique-dip slopes

are the most common ones, followed by the cross-dip slope, while the horizontally layered slope is the least one.

2.3. The features of landslide cases

The plane morphology and profile shape reflect the deformation area and overall shape of the landslide, which can be obtained by the projection method in geotechnical and geological engineering researches. The plane morphology reflects the projection of the landslide in the horizontal direction (i.e. x-y plane), while the profile shape portrays the projection of the landslide in the vertical direction (i.e. y-z plane). The statistics of different plane morphologies and profile shapes in the investigated area are illustrated in Fig. 6a and b. Among the six types of plane morphology, the tongue and irregular shapes are the dominant, followed by rectangular, semicircular, laterally long, and dustpan shapes (Fig. 6a). As shown in Fig. 6b, concave, convex, and flat shapes are the majority among the five types of profile shapes.

Through the statistics of the landslide volume in this area, it can be seen that the volumes of most slopes are in the range of 0–500,000 m³, accounting for more than half of the total number (Fig. 6c). The number of landslides gradually decreases with the increase of landslide volume. The engineering construction activities may exert adverse influence on slope stability. In the investigated area, the main human activities involve underground excavation, post-slope loading, vegetation destruction, slope
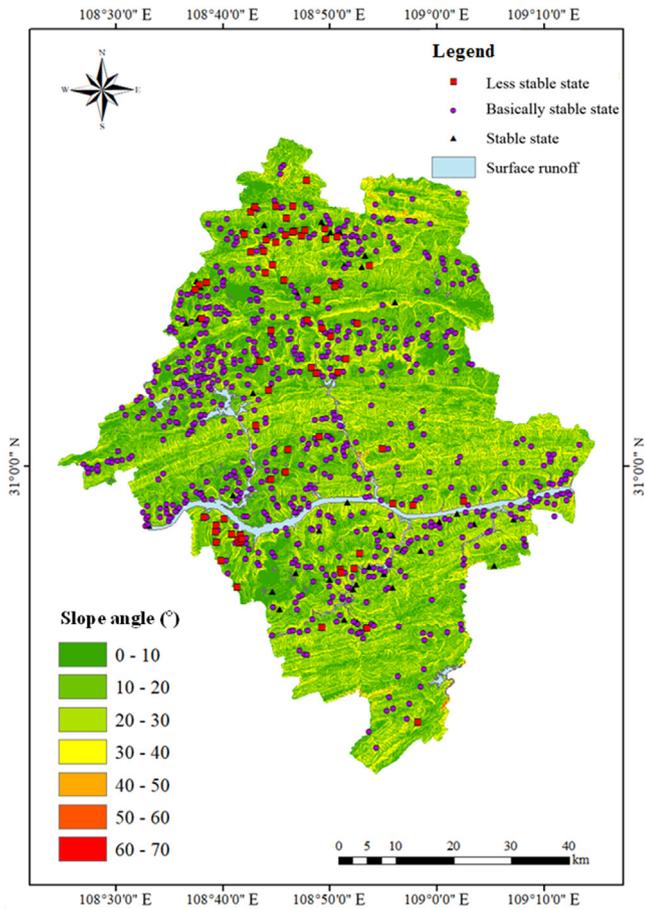
**Fig. 3.** Map of slope angle of the investigated area.
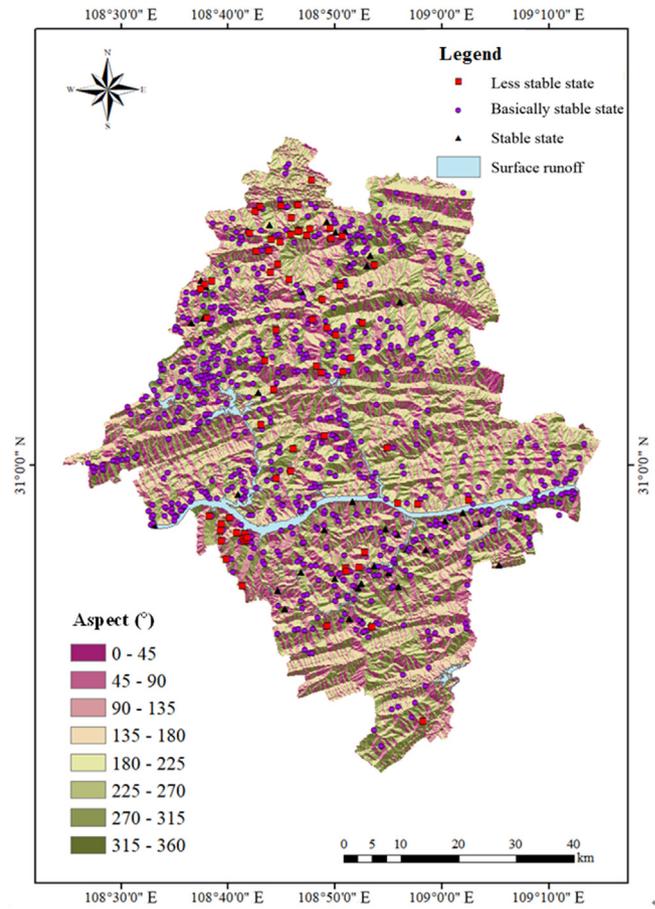
Fig. 4. Aspect map of the investigated area.

cutting, and blasting vibration. The influence degree of each human activity is defined as 1, and the influence degree will be superimposed when subjected to multiple activities. As plotted in Fig. 6d, the undisturbed slopes and the slopes subjected to a single factor account for a relatively large proportion, and a few landslides are affected by two or three factors.

3. Methodology

3.1. XGBoost

The XGBoost uses a gradient boosting framework and is also a decision-tree-based ensemble method (Chen and Guestrin, 2016), which has been widely used in the renowned Kaggle competitions due to its high efficiency and sufficient flexibility. The core principle of this method is that it builds classification or regression trees (RTs) one by one, and then the residuals of the previous trees are used to train the subsequent model. It integrates the values calculated from the previously trained trees to achieve a better outcome in the training process. To avoid overfitting, the pruning procedure is necessary, which reduces the size of a decision tree by removing nodes that contribute little to target values. The prediction is calculated as follows (Chen and Guestrin, 2016):

$$\widehat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = \widehat{y}_i^{(t-1)} + f_t(x_i) \tag{1}$$

where $\widehat{y}_i^{(t)}$ denotes the final tree model, $\widehat{y}_i^{(t-1)}$ is the previous tree model, x_i represents the features corresponding to the sample i, $f_t(x_i)$ is the newly generated tree model, and t is the total number of base tree models.

It is important to select appropriate values of depth and number of trees for achieving optimal performance. Accordingly, the objective function $Obj^{(t)}$ can be given by

$$Obj^{(t)} = \sum_{i=1}^{t} l\left(y_i, \widehat{y}_i^{(t)}\right) + \sum_{i=1}^{t} \Omega(f_i) \tag{2}$$

where y_i is the actual value, $l(y_i, \widehat{y}_i^{(t)})$ is the loss function describing how well the model fit with training data, and $\Omega(f_i)$ is the penalty term for regularization to avoid overfitting.

Following Chen and Guestrin (2016), the objective function can be transformed into

$$Obj^{(t)} = \sum_{i=1}^{t} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{3}$$

where $g_i = \partial \widehat{y}_i^{(t-1)} l(y_i, \widehat{y}_i^{(t-1)})$ and $h_i = \partial^2 \widehat{y}_i^{(t-1)} l(y_i, \widehat{y}_i^{(t-1)})$ are the first and second-order partial derivatives of the loss function, respectively. The penalty term $\Omega(f_t)$ is used to reduce complexity of model, avoid overfitting, and further enhance the generalization capacity. It is evaluated by
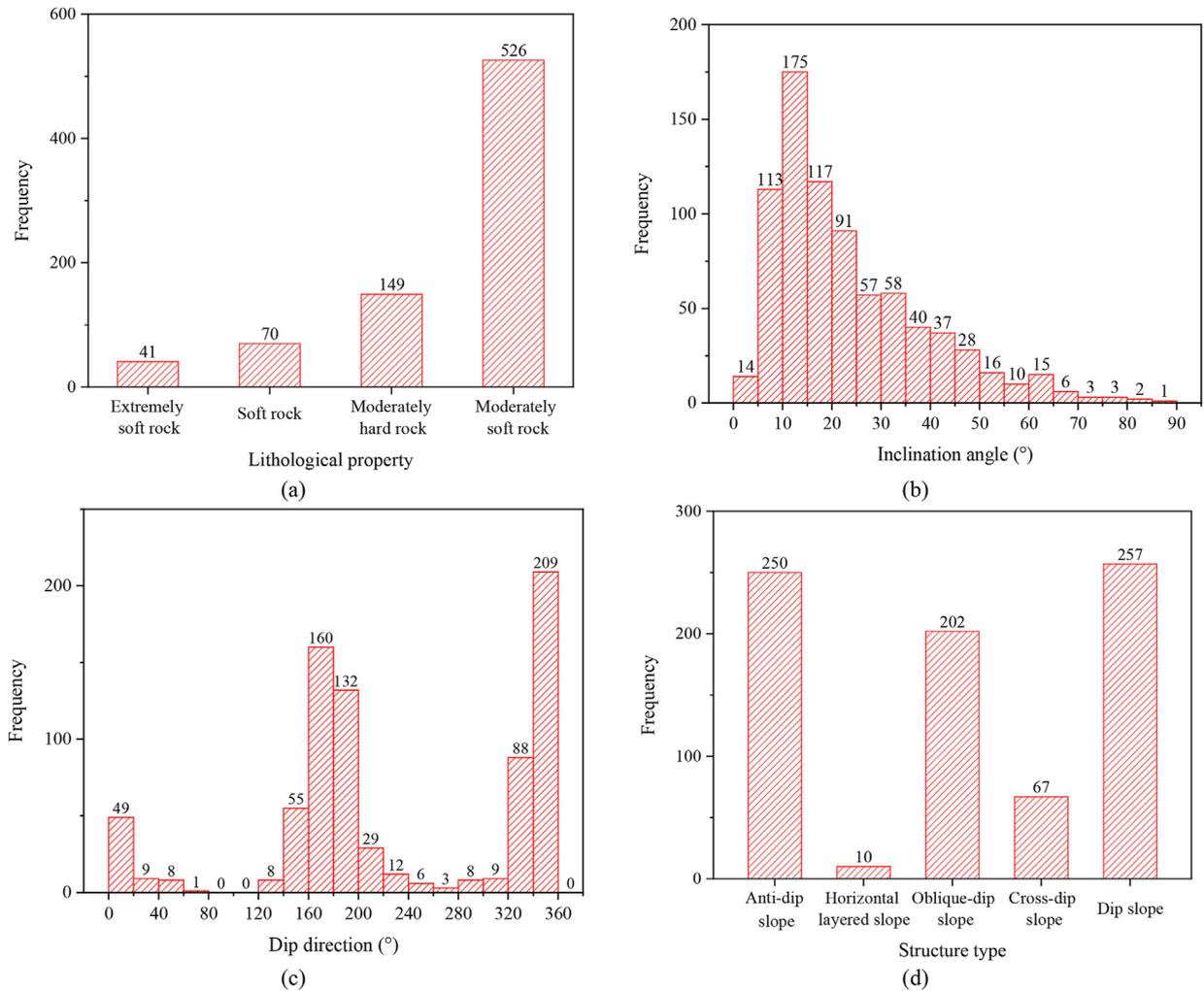
Fig. 5. Distributions of geological conditions: (a) Lithological property, (b) Inclination angle, (c) Dip direction, and (d) Structure type.

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda\|w\|^2 \qquad (4)$$

where T is the number of leaves; w is the corresponding weight of the leaf; and λ and γ are the coefficients, whose default values are 1 and 0, respectively.

3.2. RF

RF is a tree-based ensemble learning algorithm based on evaluations of several decision trees (Ho, 1995; Breiman, 2001). The main function of RF is to combine the predicted results of many decision trees to provide an ensemble result. In other words, the final prediction can be obtained by averaging the results from all decision trees, which may be more reliable and convincing in many applications compared with a single decision tree. It uses random sampling and construction of tree nodes to improve generalization capacity and avoid overfitting (Guo et al., 2021). In this study, the base evaluator of several RTs can be further specified. For every branch of an RT, the mean of the data from the leaf nodes will be calculated. The RTs will continue growing until the mean square error (MSE) between each sample reaches the minimum or no more features are available. To obtain an ideal RF model, two key parameters should be optimized, i.e. the number of RTs

(n_estimators) and the maximum depth of the RT (max_depth). The RF can be applied to tackling classification and regression problems, which has been widely used in geotechnical engineering with satisfactory performance (Guo et al., 2021; Zhang et al., 2021c, d).

3.3. Data preprocessing and performance measurement

Data preprocessing is an essential procedure for building a model, because the original database may have missing values, duplicate values, and outliers. In this study, the main steps for data preprocessing are summarized as follows:

(1) Determine the factors influencing the slope stability, in which the continuous and categorical variables should be distinguished and the text-descriptive variables need to be coded. Among the influencing factors selected in this study, there are seven continuous variables, i.e. front edge and back edge elevations, slope height, slope angle, inclination angle, dip direction, and landslide volume. In addition, there are five categorical variables, i.e. lithological property, structure type, plane morphology, profile shape, and influence degree of human activities. For the continuous variables, they are not further coded as the established model can distinguish their magnitude. For the categorical variables, since the
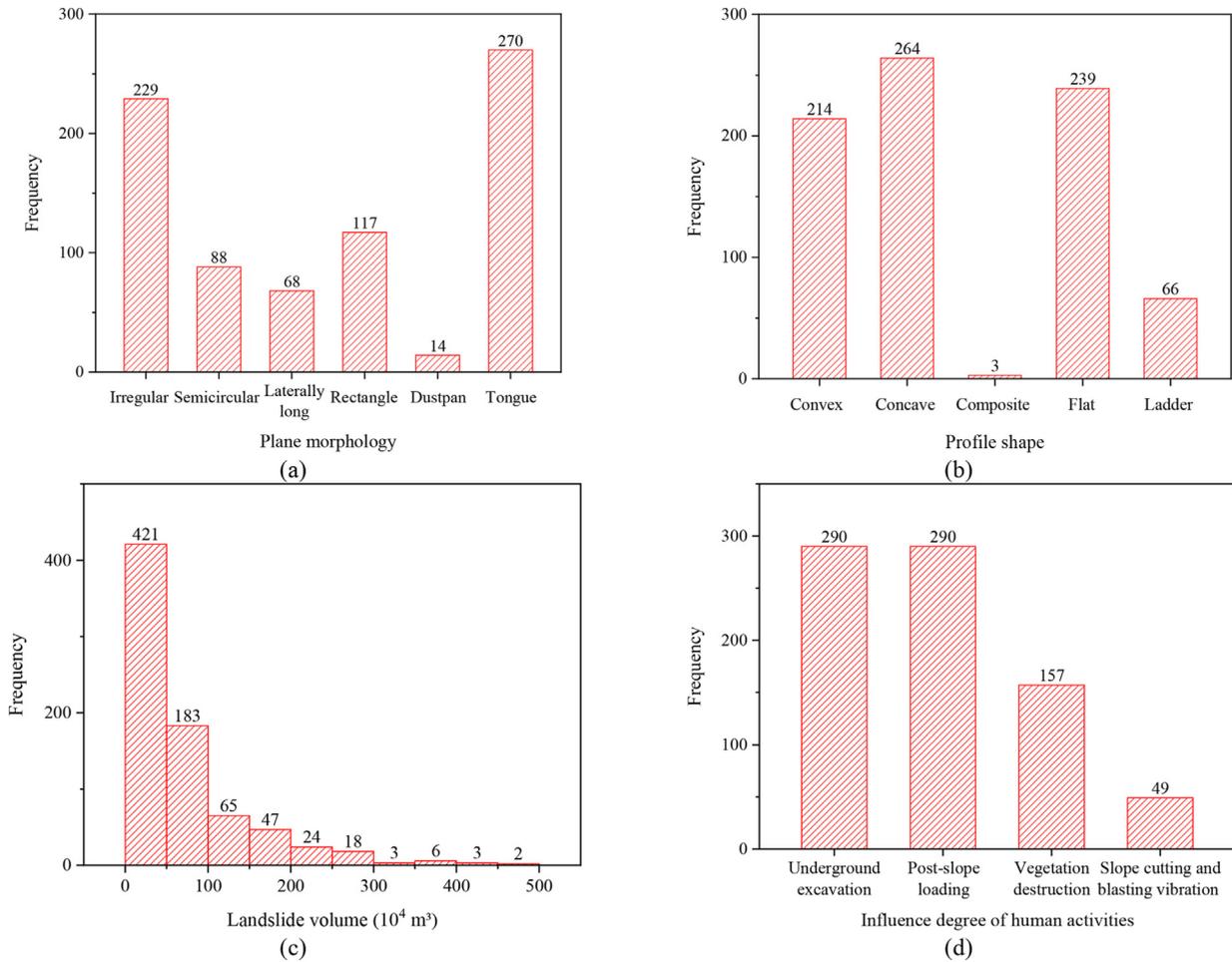
Fig. 6. Distributions of slope features: (a) Plane morphology, (b) Profile shape, (c) Landslide volume, and (d) Influence degree of human activities.

model is not distinguishable for the text-descriptive features, each categorical variable will be numbered, i.e. each categorical variable is represented by a number.

(2) Filter the original data. In this study, the missing values in the database mainly exist in the feature of profile shape. Considering that the strategy of filling missing values does not necessarily conform to reality and has an impact on the true statistical characteristics of the sample, we directly deleted them. There are no repeated data in this new database. At last, there are 786 sample data.

For the categorical variables, the numbers are simply used as a label for denoting different categories. Accordingly, it may be unnecessary to normalize these categorical variables, and thus normalization procedure is not performed in this study. When the number of features is relatively large, the feature selection generally needs to be taken into consideration. In this study, through analyzing the influencing factors and data availability, 12 influencing factors are used for establishing models and further assessing the slope stability. To facilitate the coding and analysis, these factors are numbered as listed in Table 1.

The database consists of 786 landslide sample data in Yunyang County, which are randomly divided into two groups of data with the ratio of 8:2, and accordingly, 628 of them are grouped into training data and 158 into the testing data (Table 2). The training and testing data are distinguished using different colors, which are visualized in the regional map, as plotted in Fig. 7. It can be

Table 1
The number of influencing factors.

Influencing factor	Number
Elevation of front edge	F1
Elevation of back edge	F2
Slope height	F3
Slope angle	F4
Lithological property	F5
Inclination angle	F6
Dip direction	F7
Structure type	F8
Plane morphology	F9
Profile shape	F10
Landslide volume	F11
Influence degree of human activities	F12

Table 2
Landslide dataset partition.

State	Training data	Testing data	Total
Stable	34	5	39
Basically stable	543	140	683
Less stable	51	13	64
Sum	628	158	786

observed that these two groups of data are successfully randomly divided, and in other words, these two groups of data meet the requirement of uniform distribution.
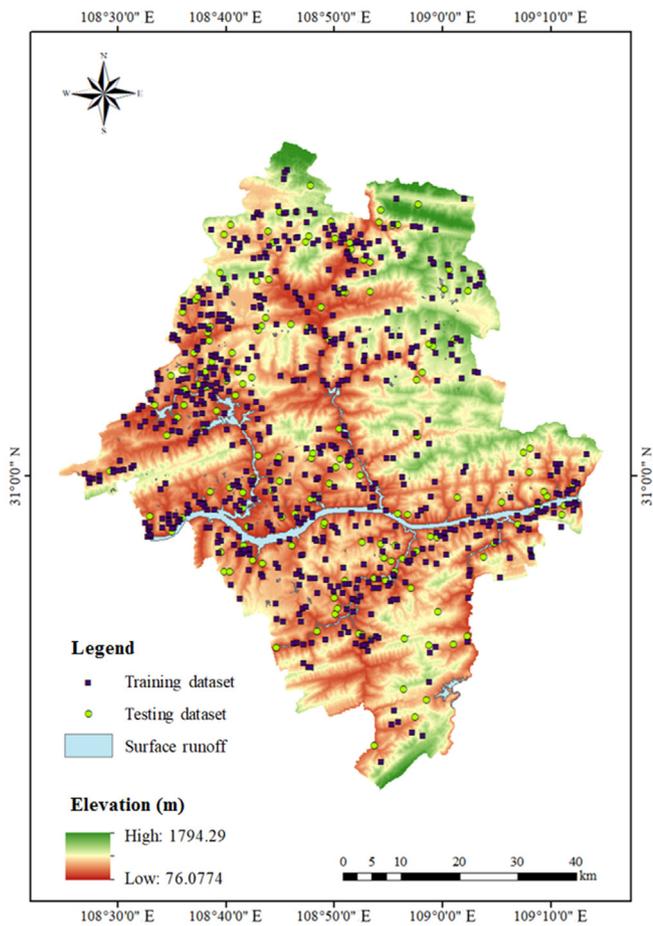
Fig. 7. Distribution map of two groups of data.

In the evaluation of model accuracy, the correct rate and error rate are generally considered to assess the model performance. However, these two indicators are not enough to comprehensively describe the model accuracy. Thus, the confusion matrix is necessary to be adopted to judge the correct rate and error rate of models for each situation. The confusion matrix is a table used for evaluating the good or poor performance of a classification model, where its columns are the actual conditions of classification, and its rows are the corresponding predicted conditions of classification. Based on the confusion matrix, the indices of recall rate, precision, and accuracy can be conveniently calculated for evaluating the performance of the established models. Accuracy is the ratio of correct

predictions to the total number of dataset. Precision describes the ratio of correctly predicted positive data to all the predicted positive ones. Recall rate represents the proportion of correctly predicted positive data to the total number of positive ones. These three indices are widely used in geotechnical engineering (e.g. Hu and Liu, 2019; Chen et al., 2021).

In this study, a comparative study is conducted to compare the predictive performance of RF, XGBoost, SVM, and LR in the slope stability of Yunyang County. For more detailed information about the SVM and LR algorithms, interested readers can refer to Goh et al. (2017) and Zhu et al. (2021). During the establishment of the model, 12 influencing factors are regarded as the input, and the target output is the slope stability state which can be categorized into three groups, i.e. stable, basically stable, and less stable. Based on the constructed models, the predictive performance of the four machine learning methods is systematically investigated and compared.

4. Results and discussion

4.1. Predictive performance of different machine learning methods

The Gaussian radial basis kernel function is adopted in the construction of SVM model, and the weight parameter, i.e. class_weight = 1:10, was used. Fig. 8 summarized the confusion matrix calculated from the SVM model. Results show that the recall rates of the basically stable state in the training and testing data are the same (i.e. 1), and the overall accuracies of them are 0.865 and 0.886, respectively. However, for the stable and less stable states, both of them are predicted to be basically stable when using the SVM model. Accordingly, the recall rates of these two cases are 0. This means that although the SVM model can classify the majority of data correctly, there are shortcomings in classify the minority, which may be attributed to the deficiency in tackling the imbalanced data problem.

For the RF model, N_estimators and Max_depth are the two important parameters. In this study, the grid search method was used to determine the optimal values of these parameters, i.e. N_estimators = 16 and Max_depth = 15. Fig. 9 lists the confusion matrix evaluated from the RF model. The overall accuracy of the model reaches 0.99 in the training data, and the recall rates for the stable, basically stable, and less stable states are 0.853, 1, and 0.98, respectively. In the testing data, the overall accuracy of is 0.911. In terms of the recall rate, that of the basically stable state is 1. For stable state, all of them are incorrectly predicted to be basically stable. For less stable state, nine of them are mistakenly identified to be basically stable. It can be observed that the prediction results of the RF model on the training data are relatively satisfactory. The

		Predicted class			
		Stable	Basically stable	Less stable	Recall
True class	Stable	0	34	0	0.000
	Basically stable	0	543	0	1.000
	Less stable	0	51	0	0.000
	Precision	0.000	0.865	0.000	0.865

(a)

		Predicted class			
		Stable	Basically stable	Less stable	Recall
True class	Stable	0	5	0	0.000
	Basically stable	0	140	0	1.000
	Less stable	0	13	0	0.000
	Precision	0.000	0.886	0.000	0.886

(b)

Fig. 8. Confusion matrix of SVM algorithm prediction value: (a) Training and (b) Testing data.
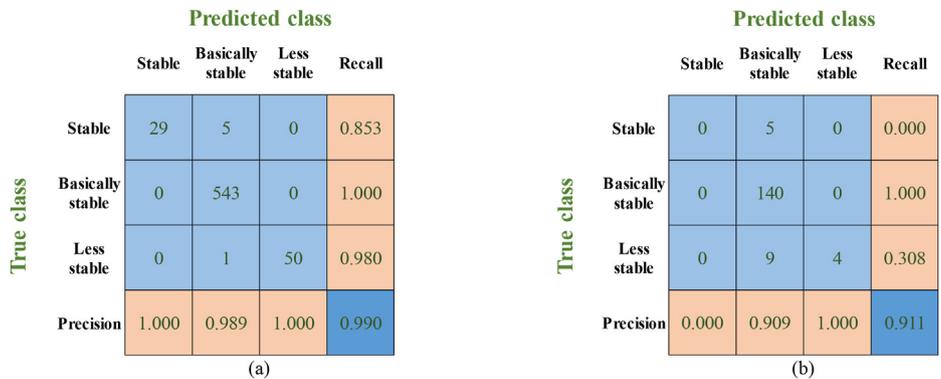
Fig. 9. Confusion matrix of RF algorithm prediction value: (a) Training and (b) Testing data.
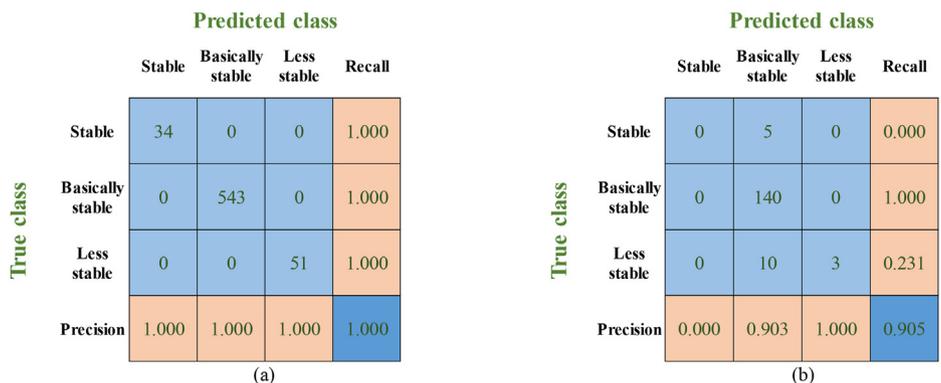
Fig. 10. Confusion matrix of XGBoost algorithm prediction value: (a) Training and (b) Testing data.
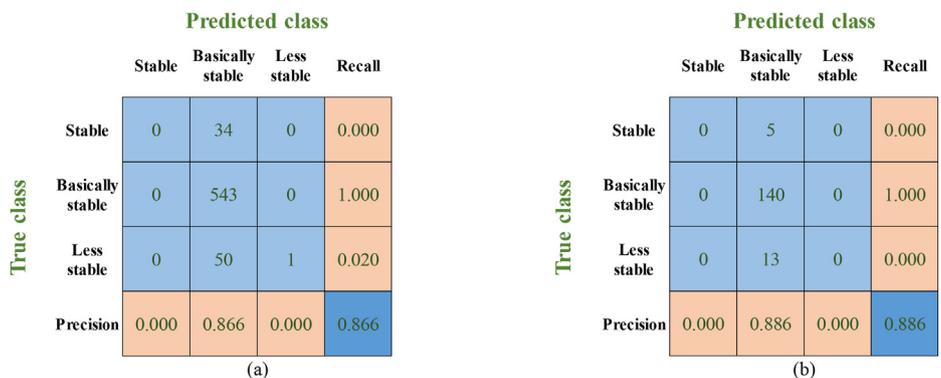
Fig. 11. Confusion matrix of LR algorithm prediction value: (a) Training and (b) Testing data.

predictive performance of the RF model is better than that of the SVM model.

Fig. 10 shows the confusion matrix calculated from the XGBoost model. It is shown that the overall accuracies and recall rates of the three cases of stable state, basically stable state, and less stable state are all 1 in the training data. In the testing data, the overall accuracy of the XGBoost model is 0.905, and the recall rate of the basically stable state is 1. All the stable state cases are mistakenly judged to be basically stable, and the proportion of correct judgment for the less stable state case is 0.231. In general, both ensemble learning models (i.e. XGBoost and RF) outperform the SVM model.

Furthermore, Fig. 11 summarizes the confusion matrix obtained from the LR model. The overall accuracy of the LR model is 0.866 in the training data, and the recall rate for the basically stable state

cases is 1. However, all the stable state cases are mistakenly judged to be basically stable, and only one less stable case is correctly identified. In the testing data, the overall accuracy of the LR model is 0.886, and the recall rate of the basically stable state is 1. In contrast, both the recall rates of the stable and less stable states are 0. In brief, the performance of LR model is similar with that of SVM model in the slope stability prediction. Both of them lag behind the two ensemble learning models (i.e. XGBoost and RF).

Table 3 summarizes the predictive performance of the four models. In general, the XGBoost and RF perform better than SVM and LR in the slope stability prediction of Yunyang County. The main advantage of ensemble learning models over the SVM and LR models is their capacity to make full use of multiple predictors to form a better one with high accuracy. Overfitting is a tricky issue frequently

Table 3
Predictive performance of the four models.

Model	Accuracy		$|Atrain - Atest|$
	Training data ($Atrain$)	Testing data ($Atest$)	
SVM	0.865	0.886	0.021
RF	0.99	0.911	0.079
XGBoost	1	0.905	0.095
LR	0.866	0.886	0.02

encountered in the process of model training, which means the predictive performance on the training data is much better than that on the testing data and will weaken the generalization capacity of models (Cawley and Talbot, 2010). In other words, overfitting may occur if there is a significant difference between the prediction accuracies of the training and testing data. As tabulated in the last column of Table 3, the absolute differences between the prediction accuracies of the training ($Atrain$) and testing data ($Atest$) for the four models are less than 0.1. Although there is still no explicit rule about how much difference between the prediction accuracies of the training and testing data belongs to overfitting in engineering practice (Hou et al., 2021), the values of $|Atrain - Atest|$ for the four models are relatively small (i.e. less than 0.1) in this study.

4.2. Feature importance analysis

The feature importance is an important aspect that should be considered in feature selection and model interpretability in machine learning. The feature importance of the 12 influencing factors for the XGBoost model is ranked in Fig. 12. It can be seen that the profile shape is the most important variable, accounting for more than 0.1. This implies that the profile shape plays a vital role in the slope stability prediction of Yunyang County.

5. Discussion

As one of the popular and powerful machine learning algorithms, ensemble learning techniques make full use of multiple predictors to form a superior one for improving the predictive performance and generalization capacity. It converts the composition models into a stronger learner in bagging, boosting, or stacking manner, allowing it to achieve a better predictive performance compared to a single model (Pan and Zhang, 2020; Hou et al., 2021). This study developed the RF- and XGBoost-based method to predict the slope stability, and the obtained results revealed their superiority over the conventional method (i.e. SVM and LR) in the slope stability prediction of Yunyang County. This may be attributed to the powerful learning capacity that inherits from multiple individual learners, allowing ensemble learning models to approximate the implicit high-dimensional relationship between the various influencing factors and slope stability status in this study. It provides an effective way to rationally predict the stability status of slopes in geotechnical and geological engineering practice.

Besides the 12 influencing factors considered in this study, the slope stability may also be affected by water level fluctuation and rainfall (e.g. Tang et al., 2019; Wang et al., 2020c; Xing et al., 2021). Since the information about the groundwater table and rainfall is missed, thus these two factors were not considered in this study. This may be the weakness of this study. It is worth noting that although 12 influencing factors are considered in this study, the proposed approach is also applicable to taking into account more influencing factors, including groundwater table and rainfall, provided that the relevant data are available. If more influencing factors are considered, geotechnical practitioners simply need to change the number of input features and recalibrate the ensemble learning models. Increasing the number of inputs does not change the methodology and implementation framework of the proposed approach.

Furthermore, imbalanced data problem is frequently encountered in geotechnical engineering practice, which can be addressed by ensemble learning techniques and specially designed sampling algorithms (Hou et al., 2021). Previous studies have confirmed that ensemble learning algorithms performed better in tackling imbalanced data problems due to their powerful generalization capacity (e.g. Feng et al., 2020). Concerning the specially designed sampling algorithms, it contains the oversampling and undersampling algorithms (Elrahman and Abraham, 2013). It aims to finely tune the number of data with different classes for achieving a rational proportion. In other words, the oversampling algorithm increases the number of minority class data, while the undersampling algorithm is to decrease the number of majority class data. In this study, random sampling and construction of tree nodes are used in RF to
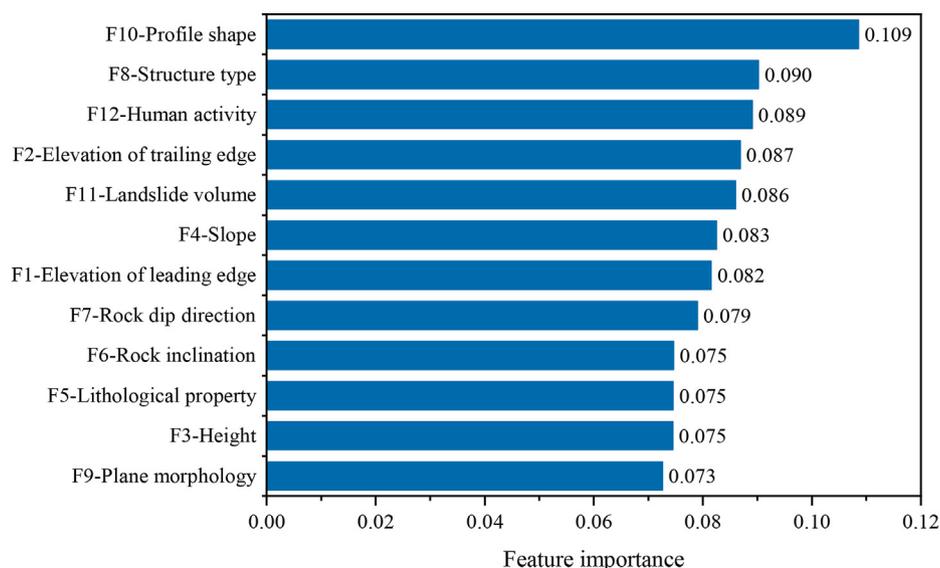
Fig. 12. Feature importance of influencing factors.

improve generalization capacity and avoid overfitting (Guo et al., 2021). For the XGBoost, a penalty term $\Omega(f_i)$ was added to penalize complicated models and avoid overfitting, as presented in Eq. (2) (Chen and Guestrin, 2016). The combination of ensemble learning techniques and oversampling algorithm is worthy of further research.

6. Conclusions

This study combined two promising ensemble learning techniques called RF and XGBoost to predict the slope stability, which produces the RF-based method and XGBoost-based method. For illustration, the proposed approach was applied to the stability prediction of 786 landslide cases in Yunyang County. A comparative study was conducted to compare the predictive performance of RF, XGBoost, SVM, and LR. Furthermore, the feature importance of the 12 influencing variables (i.e. elevations of front edge and back edge, slope height, slope angle, lithological property, inclination angle, dip direction, structure type, plane morphology, profile shape, landslide volume, and influence degree of human activities) was also explored using the results obtained from the XGBoost model.

Results showed that the accuracy of the XGBoost, RF, LR, and SVM on the training data was 1, 0.99, 0.866, and 0.865, respectively, and the accuracy of those on the testing data was 0.905, 0.911, 0.886, and 0.886, respectively. The obtained results revealed the superiority of ensemble learning models (i.e. the XGBoost and RF) over the conventional SVM and LR models in the slope stability prediction of Yunyang County. The main advantage of ensemble learning models over the SVM and LR models is their capacity to make full use of several predictors to form a better one with high accuracy. It provides a possibility of integrating ensemble learning techniques into slope stability prediction for rationally predict the slope stability in geotechnical and geological engineering practice. Among the 12 influencing factors, the profile shape was the most influencing one in the slope stability prediction. Last but not least, it is worth pointing out that Yunyang County was used in this study for illustration. The proposed ensemble learning-based method may also be applied to other landslide-prone areas provided that the database for model calibration is available.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Cawley, G.C., Talbot, N.L.C., 2010. On over-fitting in model selection and subsequent selection bias in performance evaluation. J. Mach. Learn. Res. 11, 2079–2107.

Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining. ACM, pp. 785–794.

Chen, J., Yang, T., Zhang, D., Huang, H., Tian, Y., 2021. Deep learning based classification of rock structure of tunnel face. Geosci. Front. 12 (1), 395–404.

Elrahman, S.M.A., Abraham, A., 2013. A review of class imbalance problem. J. Netw. Innovat. Comput. 1, 332–340.

Feng, D.C., Liu, Z.T., Wang, X.D., Jiang, Z.M., Liang, S.X., 2020. Failure mode classification and bearing capacity prediction for reinforced concrete columns based on ensemble machine learning algorithm. Adv. Eng. Inf. 45, 101126.

Goh, A.T.C., Zhang, Y., Zhang, R., Zhang, W., Xiao, Y., 2017. Evaluating stability of underground entry-type excavations using multivariate adaptive regression splines and logistic regression. Tunn. Undergr. Space Technol. 70, 148–154.

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 (1), 85–97.

Gu, D.M., Huang, D., Yang, W.D., Zhu, J.L., Fu, G.Y., 2017. Understanding the triggering mechanism and possible kinematic evolution of a reactivated landslide in the Three Gorges Reservoir. Landslides 14, 2073–2087.

Guo, D., Chen, H., Tang, L., Chen, Z., Samui, P., 2021. Assessment of rockburst risk using multivariate adaptive regression splines and deep forest model. Acta Geotech 2.

Ho, T.K., 1995. Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition, vol. 1. IEEE, pp. 278–282.

Hou, S., Liu, Y., Yang, Q., 2021. Real-time prediction of rock mass classification based on TBM operation big data and stacking technique of ensemble learning. J. Rock Mech. Geotech. Eng.

Hu, J.L., Liu, H.B., 2019. Bayesian network models for probabilistic evaluation of earthquake-induced liquefaction based on CPT and vs databases. Eng. Geol. 254, 76–88.

Huang, D., Gu, D.M., Song, Y.X., Cen, D.F., Zeng, B., 2018. Towards a complete understanding of the triggering mechanism of a large reactivated landslide in the Three Gorges Reservoir. Eng. Geol. 238, 36–51.

Huang, D., Luo, S., Zhong, Z., 2020a. Analysis and modeling of the combined effects of hydrological factors on a reservoir bank slope in the Three Gorges Reservoir area. China. Eng. Geol. 279, 105858.

Huang, F., Cao, Z., Guo, J., Jiang, S.H., Li, S., Guo, Z., 2020b. Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. Catena 191, 104580.

Ji, S., Yu, D., Shen, C., Li, W., Xu, Q., 2020. Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks. Landslides 17 (6), 1337–1352.

Kamrava, S., Tahmasebi, P., Sahimi, M., 2020. Linking morphology of porous media to their macroscopic permeability by deep learning. Transp. Porous Media 131 (2), 427–448.

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 (1), 188–201.

Koopialipoor, M., Jahed Armaghani, D., Hedayat, A., Marto, A., 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.

Li, J.H., Li, P.X., Guo, D., Li, X., Chen, Z.Y., 2021. Advanced prediction of tunnel boring machine performance based on big data. Geosci. Front. 12 (1), 331–338.

Liu, Z.Q., Guo, D., Lacasse, S., Li, J.H., Yang, B.B., Choi, J.C., 2020. Algorithms for intelligent prediction of landslide displacements. J. Zhejiang Univ. - Sci. 21 (6), 412–429.

Liu, Z., Gilbert, G., Cepeda, J.M., Lysdahl, A.O.K., Piciullo, L., Hefre, H., Lacasse, S., 2021. Modelling of shallow landslides with machine learning algorithms. Geosci. Front. 12 (1), 385–393.

Luo, Z., Bui, X.N., Nguyen, H., Moayedi, H., 2021. A novel artificial intelligence technique for analyzing slope stability using PSO-CA model. Eng. Comput. 37 (1), 533–544.

Mahdiyar, A., Hasanipanah, M., Armaghani, D.J., Gordan, B., Abdullah, A., Arab, H., Majid, M.Z.A., 2017. A Monte Carlo technique in safety assessment of slope under seismic condition. Eng. Comput. 33 (4), 807–817.

Mojtahedi, S.F.F., Tabatabaee, S., Ghoroqi, M., Soltani Tehrani, M., Gordan, B., Ghoroqi, M., 2019. A novel probabilistic simulation approach for forecasting the safety factor of slopes: a case study. Eng. Comput. 35 (2), 637–646.

Pan, Y., Zhang, L., 2020. Data-driven estimation of building energy consumption with multi-source heterogeneous data. Appl. Energy 268, 114965.

Pham, K., Kim, D., Park, S., Choi, H., 2021. Ensemble learning-based classification models for slope stability analysis. Catena 196, 104886.

Qi, C., Tang, X., 2018. Slope stability prediction using integrated metaheuristic and machine learning approaches: a comparative study. Comput. Ind. Eng. 118, 112–122.

Tang, H., Wasowski, J., Juang, C.H., 2019. Geohazards in the three Gorges reservoir area, China – lessons learned from decades of research. Eng. Geol. 261, 105267.

Wang, L., Wu, C., Tang, L., Zhang, W., Lacasse, S., Liu, H., Gao, L., 2020a. Efficient reliability analysis of earth dam slope stability using extreme gradient boosting method. Acta Geotech 15 (11), 3135–3150.

Wang, L., Wu, C., Gu, X., Liu, H., Mei, G., Zhang, W., 2020b. Probabilistic stability analysis of earth dam slope under transient seepage using multivariate adaptive regression splines. Bull. Eng. Geol. Environ. 79 (6), 2763–2775.

Wang, L.Q., Yin, Y.P., Huang, B.L., Dai, Z.W., 2020c. Damage evolution and stability analysis of the jianchuandong dangerous rock mass in the three Gorges reservoir area. Eng. Geol. 265, 105439.

Wang, L.Q., Wu, J.H., Zhang, W.G., Wang, L., Cui, W., 2021. Efficient seismic stability analysis of embankment slopes subjected to water level changes using gradient boosting algorithms. Front. Earth Sci. 9, 1–9.

Xia, Y., Liu, C., Li, Y.Y., Liu, N., 2017. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. Expert Syst. Appl. 78, 225–241.

Xing, X., Wu, C., Li, J., Li, X., Zhang, L., 2021. Susceptibility assessment for rainfall-induced landslides using a revised logistic regression method. Nat. Hazards 106 (1), 97–117.

Zeng, F., Nait Amar, M., Mohammed, A.S., Motahari, M.R., Hasanipanah, M., 2021. Improving the performance of LSSVM model in predicting the safety factor for circular failure slope through optimization algorithms. Eng. Comput.

Zhang, W., Han, L., Gu, X., Wang, L., Chen, F., Liu, H., 2020a. Tunneling and deep excavations in spatially variable soil and rock masses: a short review. Undergr. Space.

Zhang, W., Zhang, R., Wu, C., Goh, A.T.C., Lacasse, S., Liu, Z., Liu, H., 2020b. State-of-the-art review of soft computing applications in underground excavations. Geosci. Front. 11 (4), 1095–1106.

Zhang, W., Wu, C., Li, Y., Wang, L., Samui, P., 2021a. Assessment of pile drivability using random forest regression and multivariate adaptive regression splines. Georisk 15 (1), 27–40.

Zhang, J.Z., Phoon, K.K., Zhang, D.M., Huang, H.W., Tang, C., 2021b. Novel approach to estimate vertical scale of fluctuation based on CPT data using convolutional neural networks. Eng. Geol. 294, 106342.

Zhang, W.G., Wu, C.Z., Zhong, H.Y., Li, Y.Q., Wang, L., 2021c. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. Geosci. Front. 12 (1), 469–477.

Zhang, W., Li, H., Li, Y., Liu, H., Chen, Y., Ding, X., 2021d. Application of deep learning algorithms in geotechnical engineering: a short critical review. Artif. Intell. Rev.

Zhou, J., Li, E., Yang, S., Wang, M., Shi, X., Yao, S., Mitri, H.S., 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.

Zhou, J., Qiu, Y., Armaghani, D.J., Zhang, W., Li, C., Zhu, S., Tarinejad, R., 2021. Predicting TBM penetration rate in hard rock condition: a comparative study among six XGB-based metaheuristic techniques. Geosci. Front. 12 (3), 101091.

Zhu, X., Chu, J., Wang, K., Wu, S., Yan, W., Chiam, K., 2021. Prediction of rockhead using a hybrid N-XGBoost machine learning framework. J. Rock Mech. Geotech. Eng.

Wengang Zhang is currently full professor in School of Civil Engineering, Chongqing University, China. He obtained his BSc and MSc degrees from the Hohai University in China and his PhD from Nanyang Technological University, Singapore. He is now the member of the International Society for society Mechanics and Geotechnical Engineering TC304 (Reliability), TC309 (Machine Learning), TC219 (System Performance of Geotechnical Structures) and TC222 (Building Information Modelling and Digital Twins). Dr. Zhang has been selected as the World's Top 2% Scientists in 2020 and 2021. He received 2021 *Underground Space* Outstanding Paper Award, *Computers and Geotechnics* Sloan Outstanding Paper Award, Tunnelling and Underground Construction Society (Singapore) *Hulme* Best Paper Award and the Best Referees 2020 of *Geoscience Frontiers*. He served as Associate Editor for *Geoscience Frontiers*, editorial board member of *Journal of Rock Mechanics and Geotechnical Engineering* and *Georisk*.