

Research Article

Improving the Landslide Susceptibility Prediction Accuracy by Using Genetic Algorithm Optimized Machine Learning Approach

Binbin Zheng ^{1,2}, Jiahe Wang ¹, Tingting Feng ¹, Wensong Wang ³, Yufei Wang ¹,
Yonghao Yang ⁴, and Guangjin Wang ⁵

¹School of Management Science and Engineering, Shandong Technology and Business University, Yantai 264005, China

²Shandong Emergency Management Institute, Shandong Technology and Business University, Yantai 264005, China

³State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu, China

⁴School of Civil Engineering, Chongqing Jiaotong University, Chongqing 400074, China

⁵Faculty of Land Resources Engineering, Kunming University of Science and Technology, Kunming 650093, China

Correspondence should be addressed to Wensong Wang; wws@cqu.edu.cn

Received 12 April 2023; Revised 2 August 2023; Accepted 25 October 2023; Published 9 November 2023

Academic Editor: Mohammad R. Khosravi

Copyright © 2023 Binbin Zheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Landslide susceptibility prediction is critical in open pit mines and geotechnical fields. Prediction accuracy is very essential to reduce the risk of slope instability. Traditional statistical learning methods have been widely used in early warning systems, but they cannot thoroughly explore the coupling effect among related factors, which often results in low prediction accuracy. This paper establishes an ensemble learning prediction model optimized by a genetic algorithm (GA) to determine landslide susceptibility more quickly and efficiently. The model is based on 290 sets of slope cases containing height (H), slope angle (α), unit weight (γ), cohesion (c), friction angle (φ), and pore water pressure (ru). Two common algorithms are incorporated into the ensemble learning model: Xgboost and gradient boosting decision tree (GBDT). The area under the curve (AUC) of GA-GBDT and GA-Xgboost were found to be 0.928 and 0.933, respectively, both of which could predict landslide susceptibility better. Compared with multiple logistic regression and other machine learning algorithms, both GA-GBDT and GA-Xgboost models perform better in terms of accuracy and applicability. The study results demonstrate that the developed optimized machine learning model can accurately predict landslide susceptibility and that the parameters should be optimized on a case-by-case basis to achieve more accurate results after building a suitable machine model. The optimization model proposed in this paper can be an effective new method for the intelligent prediction of landslide susceptibility.

1. Introduction

Instability can form on slopes during the construction of mines, water conservancy actions, and construction projects, which can have substantial catastrophic consequences [1, 2]. However, there are few effective and accurate tools for prediction and prevention [3, 4]. Accurate prediction of landslide susceptibility provides a basis for the safe operation of projects and helps to determine the stability state of slopes, prevent risk hazards and provide safety management solutions. Therefore, constructing a reliable and efficient landslide susceptibility prediction model is of great theoretical and practical significance.

The current research broadly classifies landslide susceptibility prediction models into two categories: traditional and machine learning. Traditional models mainly include analytical methods [5, 6], numerical simulations [7, 8], and others. These evaluate landslide susceptibility based on the corresponding mechanical theories (ex., elastoplasticity theory and viscoelastic theory). However, the above methods have several limitations. The limit equilibrium method is convenient and straightforward. However, the theoretical basis is imperfect and requires the assumption of a slip surface, which does not reflect the actual stress conditions of the slip surface, resulting in reduced accuracy because of the simplified assumptions [9]. Numerical

methods are usually time-consuming, and their accuracy depends heavily on assessing the geotechnical and physical parameters [10, 11]. As a nonlinear, dynamic, and complex open-loop system, slopes have many of risk factors that are difficult to assess. This leads to applying traditional models with significant limitations to achieve the desired prediction accuracy and effectiveness.

Conventional landslide susceptibility analysis methods are complex, iterative, and overload the computational system [12]. This has inspired researchers to look for alternative methods for calculating landslide susceptibility. Soft computing techniques can solve highly complex, nonlinear, multivariate problems [13]. Machine learning algorithms such as neural networks and support vector machines have also been used to solve landslide susceptibility problems [14, 15]. The increased use of machine learning algorithms has promoted the crossover, integration, and development of such tools with slope engineering problems, providing new ideas and methods for landslide susceptibility prediction problems [16]. Qi and Tang [17] proposed and compared six integrated artificial intelligence (AI) methods for landslide susceptibility prediction based on metaheuristics and machine learning algorithms and demonstrated that integrated AI methods have great potential in predicting landslide susceptibility. Tien Bui et al. [18] employed machine learning-based techniques to predict the safety factor of slope failures, showing that the multilayer perceptron (MLP) outperformed other machine learning-based models. Chang et al. [19] investigated the performance of eight commonly used machine learning models for predicting slope safety coefficients. Parameter optimization and cross-validation combined historical slope data to establish a machine learning-based slope safety coefficient prediction system. Lin et al. [20] developed a machine learning (ML) model for landslide susceptibility evaluation and found that the performance and reliability of nonlinear regression methods were slightly better than linear regression methods. Huang et al. [21] used 369 recorded landslides and 13 associated conditional factors to study landslide areas. They compared analytic hierarchy process (AHP), general linear model (GLM), information value (IV), binary logistic regression (BLR), multilayer perceptron (MLP), BPNN, support vector machine (SVM), and C5.0 decision tree (C5.0DT) models for prediction. It is found that machine learning models are more suitable for landslide susceptibility prediction than the other two types of heuristic and general statistical models. Machine learning (ML) models based on remote sensing (RS) imagery and geographic information systems (GIS) have been widely and effectively implemented for landslide susceptibility prediction. Chang et al. [22] compare the landslide susceptibility prediction performance of these supervised machine learning (SML) and unsupervised machine learning (USML) models to further explore their strengths and weaknesses. They were able to achieve more accurate and reliable prediction results. In summary, machine learning algorithms have become a hot research topic in data mining and classification prediction in landslide susceptibility, but different prediction algorithms have their limitations [23, 24].

Landslide susceptibility evaluation research focuses on high prediction accuracy, which requires algorithms to continuously find newer and more robust algorithms to build landslide susceptibility evaluation models for better prediction results [25]. Therefore, it is necessary to find intelligent algorithms with high accuracy and better applicability. Integrated learning can train multiple algorithms, resulting in complementary advantages and better landslide susceptibility prediction results than a single algorithm. Both gradient boosting decision tree (GBDT) and eXtreme gradient boosting (Xgboost) are classified as ensemble learning, which is an excellent engineering implementation and optimization improvement of random forest (RF). The purpose of integrated learning is to improve a single learner's generalization ability and robustness by combining the prediction results of multiple base learners. Achour and Pourghasemi [26] found that the RF model achieved the highest prediction accuracy by comparing RF, support vector machine (SVM), and boosted regression tree (BRT) to assess the susceptibility of landslides near roads. Pham et al. [27] used 16 landslide condition factors to predict landslide susceptibility, which was found to be more accurate with RF prediction capability after comparing traditional models with machine learning. Achour et al. [28] prepared an inventory map with 12 variables (including geomorphological, geological, hydrological, and environmental factors) to predict landslide susceptibility. The RF and Xgboost models had the same prediction accuracy (AUC) and better prediction performance. Drid et al. [29] selected and evaluated eleven gully erosion condition factors to identify the areas most vulnerable to this hazard, and the results showed that the Xgboost model had the best predictive performance. Xgboost and GBDT have been widely used in various scenarios and achieved good results, but the single integrated learning model is affected by its parameters. In this paper, we explore the problem of algorithm accuracy based on the prediction of landslide susceptibility optimized by heuristic algorithms.

This study investigates the feasibility of GA-GBDT and GA-Xgboost algorithms with numerous machine-learning algorithms for landslide susceptibility prediction. Firstly, the collected data are described and analyzed. Then the principle process of GA-optimized GBDT and Xgboost algorithms and the accuracy evaluation criteria of prediction models are introduced. The performance of different prediction models under the same data is compared and analyzed by calculating various prediction model evaluation indexes and receiver operating characteristic curve (ROC) curve quantitative tests to explore the feasibility of the method in this paper.

2. Data Collection and Description

2.1. Dataset and Predictor Variables. The database includes 290 cases (156 and 134 for stable slopes and failed slopes, respectively) derived from the information in Feng et al. [30] and Zhou et al. [31]. The database contains the basic geometric slope design parameters, such as slope height (H), slope angle (β), unit weight (γ), cohesion (c), friction angle (φ), and pore water pressure coefficient (ru). The external

trigger considered in this study is the pore water pressure (ru), defined as the pore water pressure to overburden pressure (Michalowski, 1995; Kim et al., 1999) [6]. The six parameters chosen are strongly correlated with the geometry and geotechnical properties of the slope and have different degrees of influence on slope stability.

Slope instability is recorded as 0, and stability is recorded as 1. Figure 1 shows the violin plots of the six influencing factors for each of the 290 historical slope cases under the failure and stability scenarios. The advantage of the violin diagram is that it can be more intuitive to visualize the distribution of different influencing factors when the slope is stabilized or damaged. The violin plot combines a box-line plot and a density plot to show the data dispersion statistics for each group and to provide the density of data distribution. Wider violin curves correspond to higher densities and represent areas of concentrated distribution. Here, the data are widely distributed. However, the distribution of variables is asymmetric, and the data distribution of different influencing factors under stable and unstable conditions is broad. From basic plots such as Figure 1, it is impossible to visually distinguish the essential parameters affecting landslide susceptibility.

2.2. Principal Component Analysis. Principal component analysis (PCA) was used to further determine the influence of the six basic characteristics study on landslide susceptibility. First, PCA investigated the relevant contribution of each factor to landslide susceptibility, summarizing and visualizing the collected landslide susceptibility data for interpreting its variance-covariance structure. PCA also assessed the database to ensure a representative dataset.

As shown in Figure 2, slope height has the highest contribution to PC1 at 41.983%, and pore water pressure has the lowest contribution to PC1 at 5.992%. PCA allows for the visualization of the classification function of the slope dataset in a two-dimensional plane. There are overlapping domains between the two types of landslide susceptibility from the two-dimensional space. In addition, some indicators with significant skewness can impact the prediction model. According to the PCA results (shown in Figure 2(a)), the components of the first and second dimensions are visualized (Figure 2(b)). The data distribution areas for the two types of slope states on the first two components are relatively close with overlapping areas.

A comprehensive analysis of the statistical characteristics of the six influencing factors of the slopes shows that each characteristic presents a different distribution, and the span and density of the values are relatively large. This indicates that the database contains slopes of different heights, slope angles, lithologies, and types. Table 1 shows the variability of the effect of different influencing factors on slope stability. By using completely different types of slopes and using their common characteristics for slope instability prediction, the powerful ability of machine learning algorithms to handle nonlinear data can be better reflected.

3. Methodology

3.1. Ensemble Learning Algorithm. The ensemble learning algorithm, commonly known as GBDT or Xgboost, is implemented by changing the data distribution to determine the weights of each sample based on whether a sample is correctly classified in each training set and the accuracy of the last overall classification. The new dataset with modified weights is sent to the lower classifier for training. The classifiers obtained from each training are finally combined as the final decision classifier.

3.2. GBDT Model. GBDT is one of the boosting integrated learning algorithms, often used for classification and regression problems. Rather than simply adjusting the weights of weak learners, GBDT reduces the residuals after each computation by building a new model in the direction of gradient descent of the residuals [32]. The GBDT model inherits the advantages of statistical models and artificial intelligence methods, using the advantage of calculating the relative importance between variables while identifying complex nonlinear relationships [33]. Because of the complex nonlinear relationship of a slope system, this paper selected GBDT as the core model to study the landslide susceptibility judgment problem. The steps of the GBDT training model are as follows:

Step 1: Initialize the model function with the side slope training set and loss function:

Each i^{th} record of the training slope dataset is of the form $\{y_i, \vec{x}_i\}$, where (y, \vec{x}) are known and $\vec{x} = \{x_1, x_2, x_3, \dots, x_n\}$ refer to the slope features (e.g., unite weight, slope height, etc.). The aim is to predict the value of y (landslide susceptibility) based on \vec{x} . Hence, a mapping $F^*(\vec{x})$: \vec{x} needs to be identified such that the expected value of the loss function $\vartheta(y, F(\vec{x}))$ is minimized as given in the following equation:

$$F^*(\vec{x}) = \arg \min_{F(\vec{x})} E_{y, \vec{x}} \vartheta(y, F(\vec{x})). \quad (1)$$

The $F(\vec{x})$ is expanded as $F(\vec{x}) = \sum_{m=0}^M \beta_m \tau(\vec{x}; \vec{a}_m)$. $\tau(\vec{x}, \vec{a})$ is a base learner with its parameters given by $\vec{a} = \{a_1, a_2, \dots\}$. Every iteration attempts to find a better fit for the expansion coefficients $\{\beta_m\}_0^M$ and the parameters of the base function $\{a_m\}_0^M$ to achieve a better prediction. In the beginning, the training is initialized by guessing $F_0(\vec{x})$.

Step 2: For $m = 1, 2, \dots, M$, M regression trees are generated iteratively:

$$\begin{aligned} (\beta_m, \vec{a}_m) &= \arg \min_{\beta, \vec{a}} \sum_{i=1}^N \vartheta(y_i, F_{m-1}(\vec{x}_i) + \beta \tau(\vec{x}_i; \vec{a})), \\ F_m(\vec{x}) &= F_{m-1}(\vec{x}) + \beta_m \tau(\vec{x}, \vec{a}_m). \end{aligned} \quad (2)$$

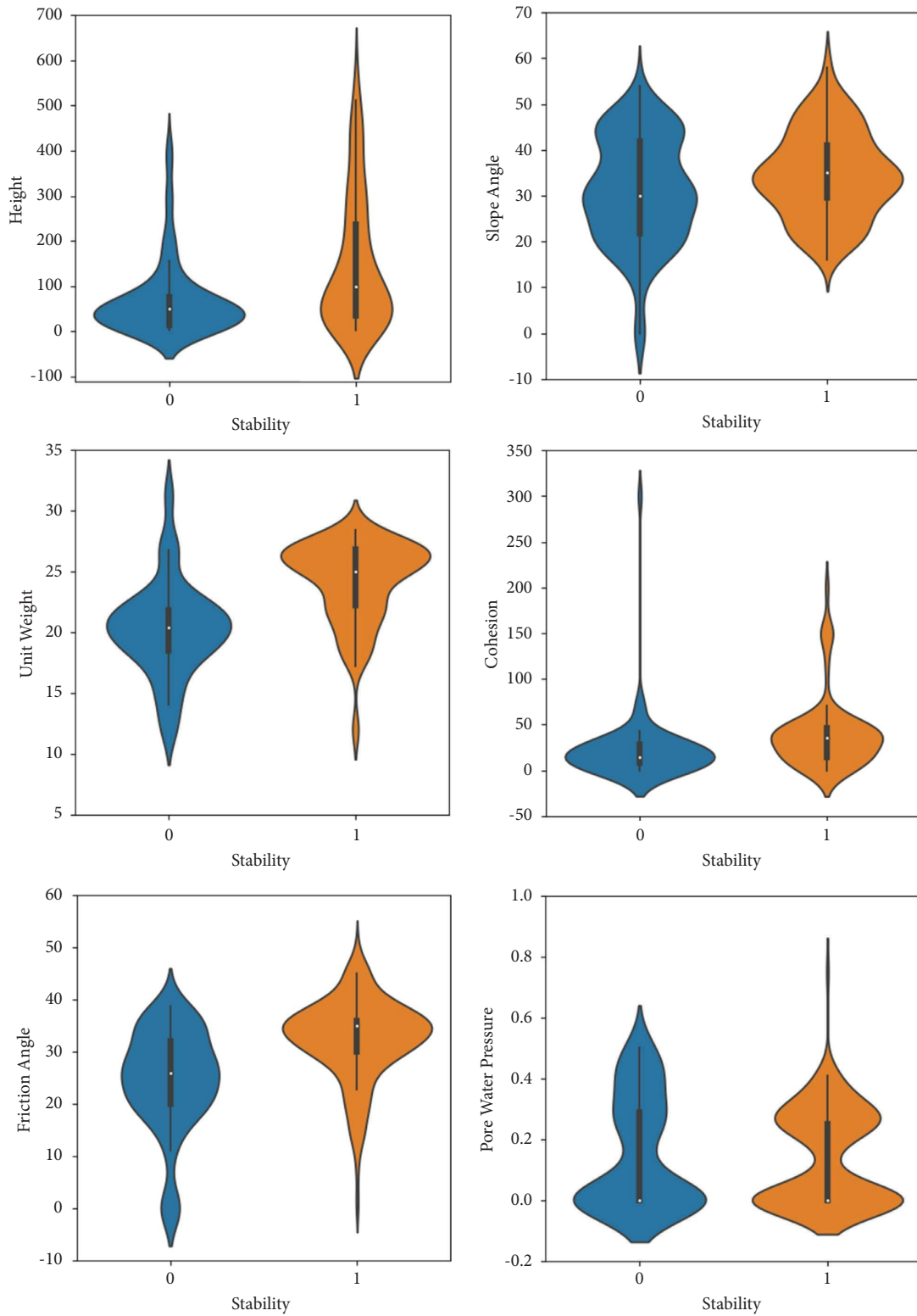


FIGURE 1: Violin diagram of data distribution of influencing factors.

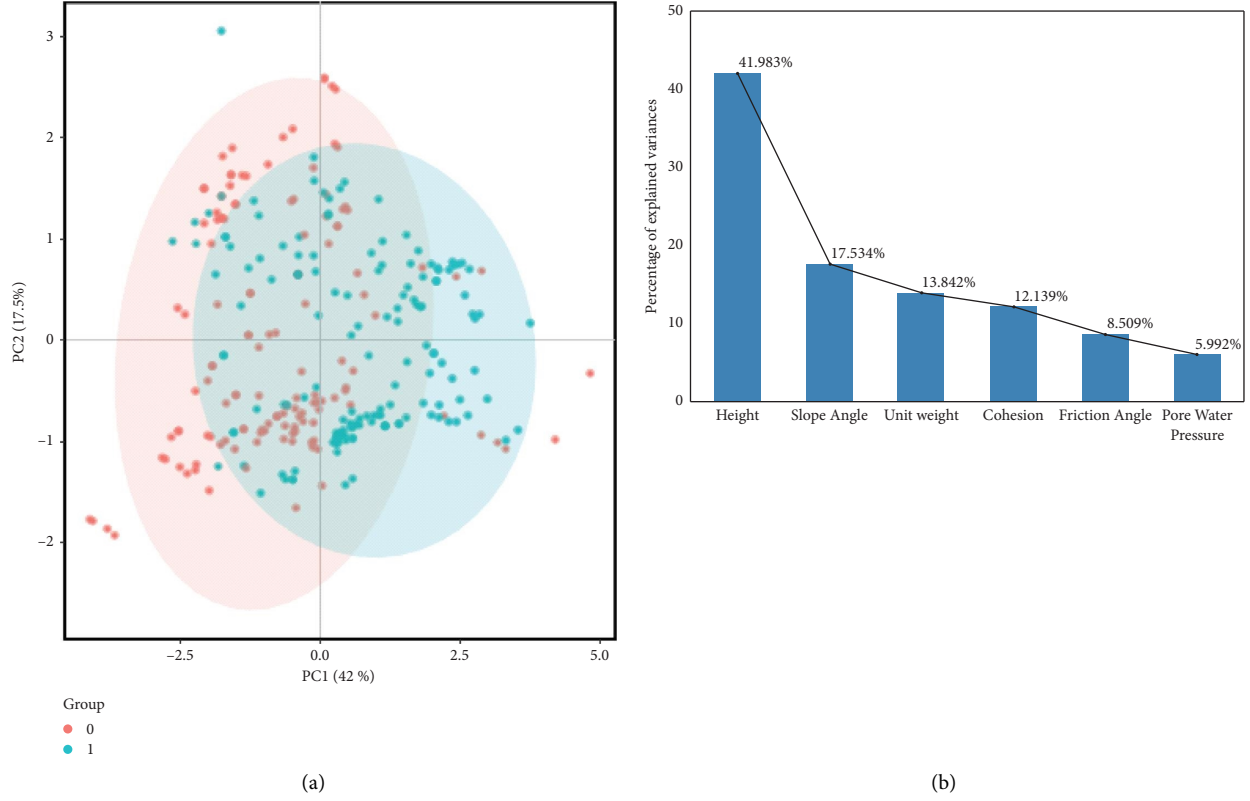


FIGURE 2: PCA distribution of landslide susceptibility data. (a) PCA distribution chart. (b) Factor contribution chart.

TABLE 1: Variance explanation rate table.

No	Feature root	Explanation of variance (%)	Accumulation (%)	Feature root and load factor		
				Name	Principal components 1	Principal components 2
1	2.519	41.983	41.983	Slope height	0.590	-0.093
2	1.052	17.534	59.517	Slope angle	0.737	0.224
3	0.831	13.842	73.359	Unit weight	0.804	-0.153
4	0.728	12.139	85.498	Cohesion	0.638	-0.201
5	0.511	8.509	94.008	Friction angle	0.757	0.231
6	0.360	5.992	100.000	Pore water pressure	-0.036	0.936

The loss function $\vartheta(y, F(\vec{x}))$ is assumed to be differentiable and the function $\tau(\vec{x}, \vec{a})$ is fit by minimizing the k -class multinomial negative log-likelihood. The predicted value of output for \vec{x}_i at the m th iteration is given in the following equation:

$$\tilde{y}_{im} = - \left[\frac{\partial \vartheta(y_i, F(\vec{x}_i))}{\partial F(\vec{x}_i)} \right]_{F(\vec{x})=F_{m-1}(\vec{x})}. \quad (3)$$

The optimally fit $\tau(\vec{x}; \vec{a}_m)$ is used to find the optimal value of coefficient β_m

$$\beta_m = \arg \min_{\beta} \sum_{i=1}^N \vartheta(y_i, F_{m-1}(\vec{x}_i) + \beta \tau(\vec{x}_i; \vec{a}_m)). \quad (4)$$

The base learner $\tau(\vec{x}; \vec{a})$ is a decision tree where, in each iteration m , the tree segments the input slope feature \vec{x} space into Z -disjoint regions $\{R_{zm}\}_{z=1}^Z$ and

predicts a separate constant value in each one as described in the following equation:

$$\tau(\vec{x}; \{R_{zm}\}_1^Z) = \sum_{z=1}^Z y_{zm} 1(\vec{x} \in R_{zm}), \quad (5)$$

\dots \tilde{y}_{zm} is the majority class predicted in each region R_{zm} , i.e., the majority of the points in the R_{zm} region are predicted to belong to this class. It can also be considered as the class with the highest probability to be predicted in that region, i.e., $\tilde{y}_{zm} = \arg \max_{\tilde{y}_{zm}} \hat{p}_m(\tilde{y}_{zm})$. In decision trees, the parameters are the features/attributes/variables being split at node and the specific value at which the chosen variable is split. These two parameters define the region $\{R_{zm}\}_1^Z$ of the partitions at the m th iteration.

Step 3: Generate the GBDT:

Because the decision tree produces a constant value \hat{y}_{zm} within each region R_{zm} , the expansion coefficients and base learner function's value can be reduced to the following equation:

$$\omega_{zm} = \arg \min_{\omega} \sum_{\vec{x}_i \in R_{zm}} \vartheta(y_i, F_{m-1}(\vec{x}_i) + \omega). \quad (6)$$

The current approximation $F_{m-1}(\vec{x})$ is updated in each region as depicted in the following equation:

$$F_m(\vec{x}) = F_{m-1}(\vec{x}) + t \cdot \omega_{zm} 1(\vec{x} \in R_{zm}). \quad (7)$$

The shrinkage parameter $0 < t \leq 1$ controls the learning rate of the procedure [32, 34].

3.3. Xgboost Model. Xgboost is widely used in major machine-learning puzzle tasks and is fitting for landslide susceptibility prediction. Chen and Guestrin [35] added a regularization term to the loss function on Xgboost to control the complexity of the trees and prevent overfitting. The regularization term is represented as follows:

$$\hat{y}_i = \sum_{K=1}^K f_i(X_i), f_k \in F, \quad (8)$$

where x_i is the i th sample data of input; \hat{y}_i is the model prediction value of the i th sample; K is the number of trees; F is the set space of trees; and f_k is a function in set space F .

$$L(\varphi) = \sum_i l(y_i, \hat{y}_i^{(t-1)}) + \sum_k \Omega(f_k). \quad (9)$$

The objective function in the formula consists of two parts. The first calculates the error between the predicted value \hat{y}_i and the true value y_i . The second is the regularization term, which represents the sum of the complexity of each tree:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2, \quad (10)$$

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_i(X_i)) + \Omega(f_t).$$

For the t th round loss function $L(\varphi) = \sum_i l(y_i, \hat{y}_i^{(t-1)}) + \sum_k \Omega(f_k)$, a second-order Taylor function is taken at the point $\hat{y}_i^{(t-1)}$ and the loss function can be calculated for the accumulation of every leaf node as follows:

$$\tilde{L}^{(t)} = \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) \omega_j^2 \right], \quad (11)$$

$$g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}),$$

$$h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)}),$$

$$I_j = \{i \mid q(X_i) = j\},$$

where I_j represents the samples in leaf node j .

3.4. Hyperparameter Tuning. Classical machine learning prediction algorithms are more sensitive to hyperparameters, which are crucial for building high-accuracy prediction models for landslide susceptibility. To find the optimal hyperparameters, particle swarm optimization [36], genetic algorithm (GA) [37], artificial bee colony ABC [24], grid search [38], and firefly algorithm (FA) [39] have been adopted, amongst others. GBDT and Xgboost have many parameters that are tedious to adjust. Furthermore, each parameter can significantly impact the algorithm's prediction performance and needs to be optimized for tuning parameters. Based on GBDT and Xgboost integrated with multiple decision trees, the global search ability and flexibility of GA are used to compensate for the defects of the GBDT and Xgboost model, including the abundant tuning of parameters, slow convergence, and easily falling into local optimum.

3.5. GA Feature Selection. GA is a class of randomized search algorithms that draws on natural selection and natural genetic mechanisms in the biological world [40]. After each iteration, the termination criterion is checked to see if convergence has been reached or if the maximum number of iterations allowed has been completed. Convergence occurs when all chromosomes in the population have reached the same fitness level. This indicates that an optimal set of characteristics has been determined, and the process can be terminated. However, because this condition may not always be satisfied, a limit is placed on the number of iterations (the maximum number of iterations = 50). Thus, the algorithm is ended if convergence is achieved before the completion of 50 iterations. Otherwise, the maximum number of generations is executed. If neither of these termination criteria is met, the next chromosome population is generated from the previous population by applying tournament selection, mutation, and crossover operations. Figure 3 shows the principle and optimization flowchart of the GA.

3.6. Evaluation Criterion. This work evaluated the predictive performance of classification algorithms for landslide data using the area under the operating characteristic curve (AUC). The receiver operating characteristic curve (ROC) curve plots the relationship between sensitivity and specificity. It evaluates the performance of different classifiers. The ROC [31] is a two-dimensional plot of the false positive rate (FPR) (1-specificity) versus the true positive rate (TPR or sensitivity) on the horizontal and vertical axes. It is a quantitative metric based on which to assess the model's overall performance. The AUC represents the area under the ROC curve and is mainly used to measure the model's generalization performance, namely, the good or bad classification effect. It can quantitatively compare good or bad models. Different AUC values reflect different classification effects [41, 42]: 0.900–1.00 represents outstanding performance, 0.800–0.900 represents good performance, and 0.700–0.800 represents average performance. The destructive consequences of slope instability are more serious, so accuracy recall = TP/(TP + FN) is introduced to evaluate

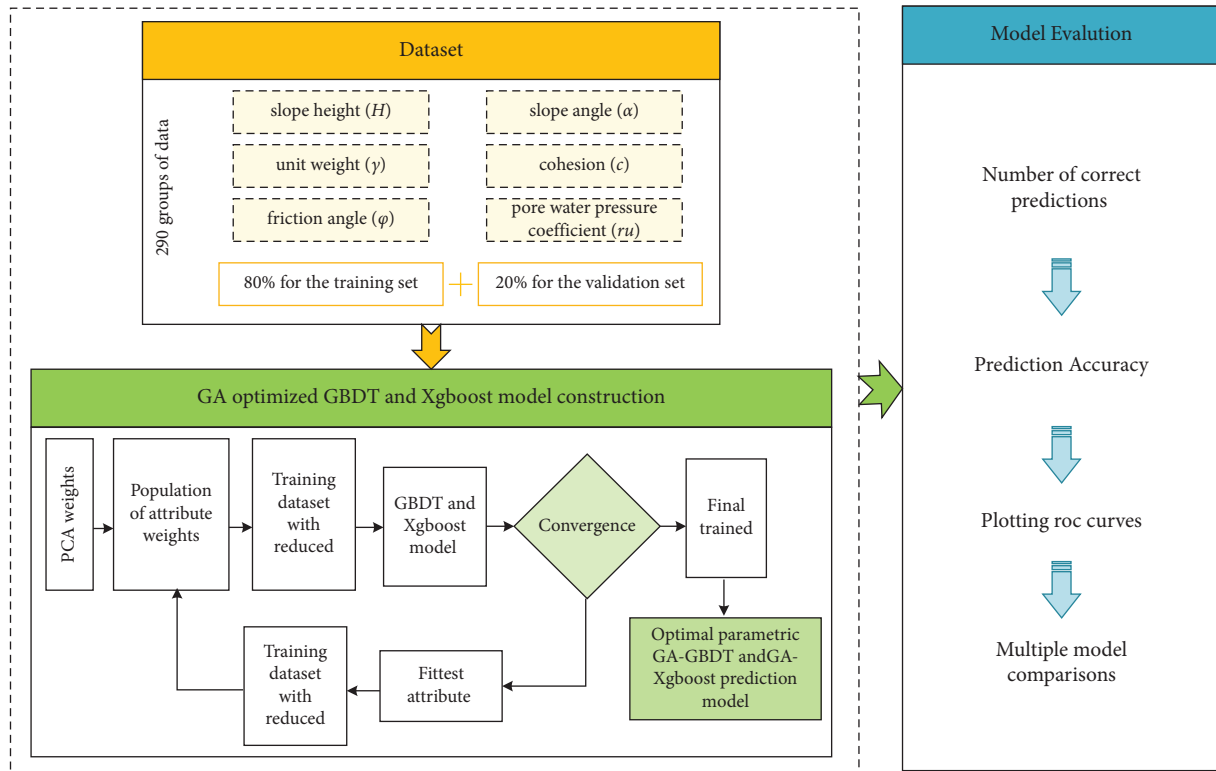


FIGURE 3: Flowchart of GA-GBDT and GA-Xgboost slope prediction model.

the model performance. When slope instability actually occurs, the model prediction is not wrong. Such a classifier is optimal, and the classifier's recall must be as high as possible under the premise of a certain correct rate.

4. Results

4.1. Comparison of Model Performance after GA Optimization. Height (H), cohesion (c), slope angle (β), unit weight (γ), friction angle (φ), and pore water pressure (ru) [10] are the input parameters of the GBDT model, and the slope stability state (failure/stability) is the output. In this paper, 80% of the 290-group database was used as training data to train the model, and the remaining 20% was used as test data to verify the model's accuracy. AUC and recall were used to evaluate the accuracy of the prediction model. The computation time was calculated to be 95.5s for GA-GBDT and 94.6s for GA-Xgboost. Another calculation shows that the CPU utilization is 3.1%, with 55.3% of memory used and 44.7% of memory available. The different parameters of different models lead to differences in model complexity, training effects, and results during the training process. Figures 4 and 5 illustrate the significant differences in the performance of Xgboost and GBDT with different hyperparameters, respectively. The different nodes in the figure represent the model prediction accuracy under the four parameters. The different parameters have different degrees of influence, which together determine the prediction performance of the model. Table 2 shows the hyperparameter search space table for GBDT and Xgboost. Table 3 shows the optimal values of the parameters after optimization by the GA algorithm. The n _ estimator is

a numerical parameter with a default value of 100, which specifies the number of weak classifiers. The max depth is numeric with a default value of 3, which is a parameter related to pruning. The learning rate is numeric, with a default value of 0.1, and is commonly used to specify a learning rate. The random state is a random seed that controls random mode as a parameter in any random class or function.

Table 4 shows the number of accurate predictions and recalls of different models for predicting landslide susceptibility and instability conditions. Figure 6 shows the AUC line graphs of different models before and after optimization. After comparison, before the optimization of the GA algorithm, both the GBDT and Xgboost algorithms have the same accuracy of 93% and recall of 85.2%, but after the optimization, both the model performance and accuracy improved. Overall, GA-GBDT has better accuracy and model performance.

4.2. Multivariate Statistical Logistic Regression Prediction Model. Multiple logistic regression is simpler and more convenient for analyzing multifactor models, and it can accurately measure the degree of correlation and fit between each factor [43]. The landslide data are dichotomous, so binary logistic regression was used to further investigate the effect of six factors on the slope. First, the overall validity of the model was analyzed. Table 5 shows that the original hypothesis is that the quality of the model is the same in both cases, whether the independent variables (slope height, slope angle, unit weight, cohesion, friction angle, and pore water pressure) are inputs or not. The p value is less than 0.05, which indicates that the original hypothesis is rejected, i.e.,

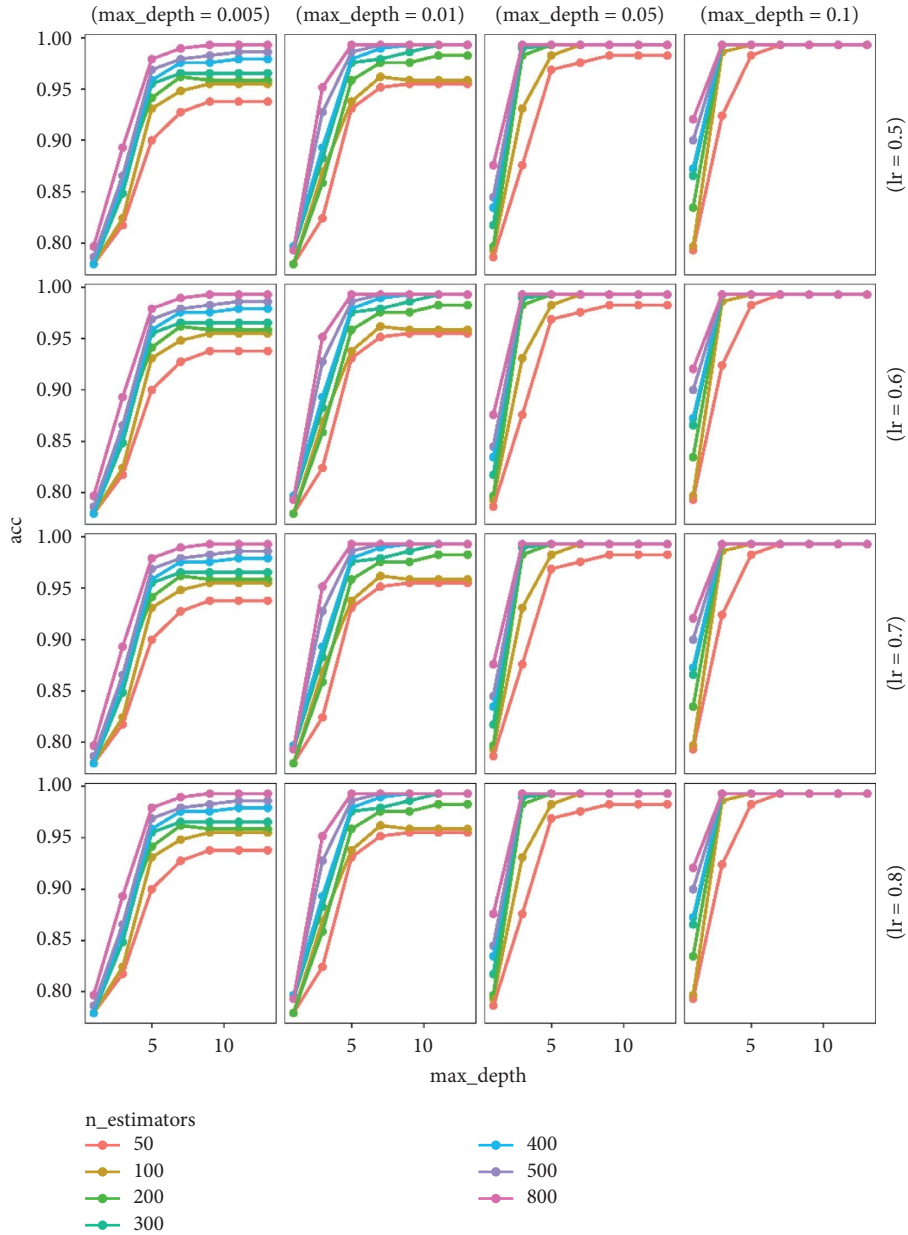


FIGURE 4: Performance of the Xgboost model with the different hyperparameters.

the independent variables are valid for this model construction and the model construction is meaningful.

Slope height, slope angle, unit weight, cohesion, friction angle, and pore water pressure were used as independent variables, while stability was a dependent variable for binary

logit regression analysis. Table 6 shows that slope height, slope angle, unit weight, cohesion, friction angle, and pore water pressure can explain the 0.22 variation in stability. The fitted model equation is

$$\ln\left(\frac{p}{1-p}\right) = -5.738 + 0.002H - 0.045\beta + 0.202\gamma + 0.002c + 0.086\varphi + 0.370ru, \quad (12)$$

where p represents the probability that stability is 1 and $1-p$ represents the probability that stability is 0. Based on the results of multiple logistic regression, the AUC and recall are

0.824 and 77.8%, respectively, which have lower accuracy and performance compared with GA-GBDT and GA-Xgboost.

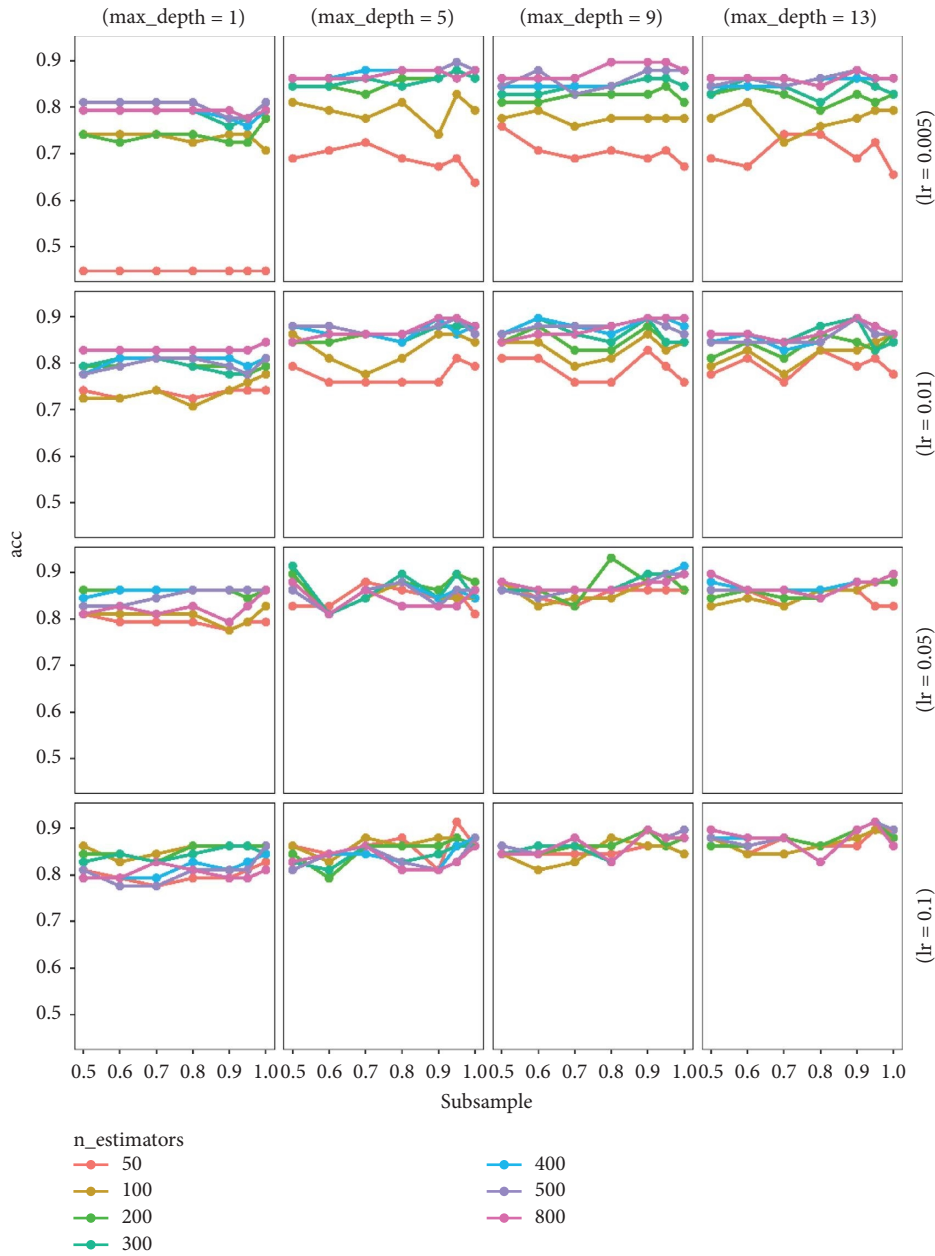


FIGURE 5: Performance of the GBDT model with the different hyperparameters.

TABLE 2: Hyperparametric search of spatial tables.

Algorithm	N_estimators	Learning_rate	Max_depth	Random_state
GA-Xgboost	(10, 1000)	(0.001, 1)	(1, 100)	(1, 100)
GA-GBDT	(10, 1000)	(0.001, 1)	(1, 100)	(1, 100)

TABLE 3: Optimized parameters in GA-Xgboost and GA-GBDT models.

Algorithm	N_estimators	Learning_rate	Max_depth	Random_state
GA-Xgboost	1000	0.89	90	64
GA-GBDT	404	0.59	8	82

Furthermore, the regression coefficient value of the slope height is 0.002, but it does not show significance ($z = 1.448, p = 0.148 > 0.05$), implying that slope height has

less effect on the stability. The regression coefficient value of the slope angle is -0.045 and shows a 0.01 level of significance ($z = -2.607, p = 0.009 < 0.01$), indicating that the

TABLE 4: Confusion matrix and recall before and after optimization of GBDT and Xgboost.

Model	True positive (TP)	True negative (TN)	Number of predictions	Recall (%)
Xgboost	23	40	58	85.2
GBDT	23	30	58	85.2
GA-Xgboost	24	30	58	88.9
GA-GBDT	26	28	58	96.3

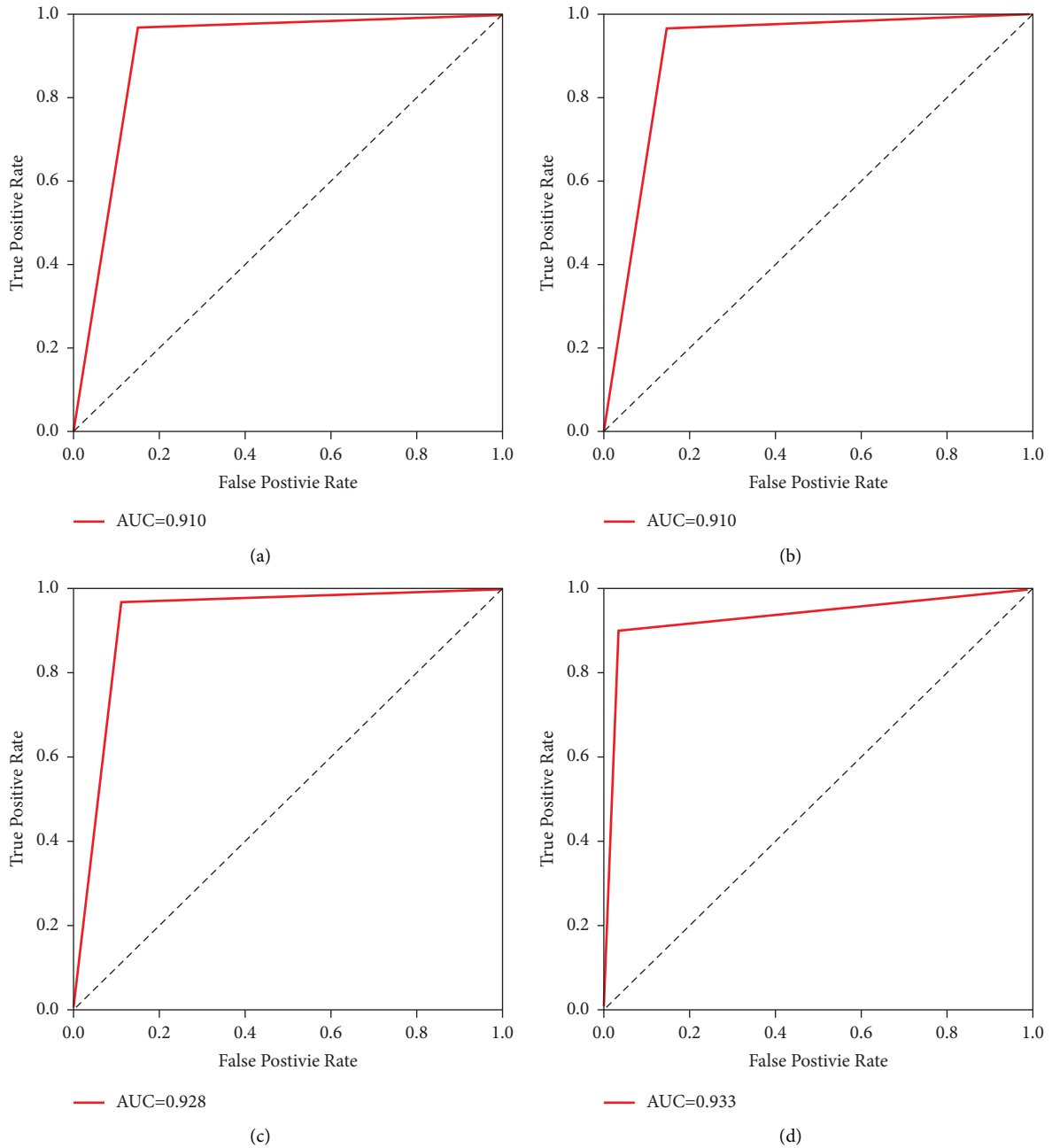


FIGURE 6: ROC curves before and after optimization of GBDT and Xgboost. (a) Xgboost ROC. (b) GBDT ROC. (c) GA-Xgboost ROC. (d) GA-GBDT ROC.

slope angle can significantly negatively influence stability. The regression coefficient value of unit weight was 0.202 and shows significance at 0.01 level ($z=3.849, p \leq 0.01$), which indicates that the unit weight will significantly positively

influences stability. Furthermore, the dominance ratio (OR value) is 1.224, indicating stability increases 1.224x when the unit weight increases by one. The regression coefficient value of cohesion is 0.002, but it does not show significance

TABLE 5: Likelihood ratio test results of binary logit regression model.

Models	-2 times the log-likelihood value	Cardinality	df	p	AIC	BIC
Intercept distance	387.648					
Final model	300.616	87.032	6	0.001	314.616	340.059

TABLE 6: Summary of results of binary logit regression analysis.

Items	Regression coefficient	Standard error	z	Wald χ^2	p	OR	OR 95% CI
Slope height	0.002	0.001	1.448	2.098	0.148	1.002	0.999 ~ 1.005
Slope angle	-0.045	0.017	-2.607	6.797	0.009	0.956	0.924 ~ 0.989
Unit weight	0.202	0.053	3.849	14.814	0.001	1.224	1.104 ~ 1.357
Cohesion	0.002	0.005	0.529	0.279	0.597	1.002	0.994 ~ 1.011
Friction angle	0.086	0.023	3.806	14.487	0.001	1.090	1.042 ~ 1.139
Pore water pressure	0.370	0.833	0.444	0.197	0.657	1.448	0.283 ~ 7.406
Intercept distance	-5.738	1.033	-5.557	30.877	0.001	0.003	0.001 ~ 0.024

Dependent variable: stability. McFadden R^2 : 0.225. Cox and Snell R^2 : 0.259. Nagelkerke R^2 : 0.352.

($z = 0.529, p = 0.597 > 0.05$), implying that cohesion has little influence on stability. The regression coefficient value of the friction angle is 0.086 and shows a significance at 0.01 level ($z = 3.806, p \leq 0.01$), implying that the friction angle significantly influences stability. The dominance ratio (OR value) is 1.090, implying that stability increases by 1.090x when the friction angle increases by one unit. The regression coefficient value of the pore water pressure is 0.370, but it does not show significance ($z = 0.444, p = 0.657 > 0.05$), implying that the pore water ratio does not have a significant influence on stability ($z = 0.444, p = 0.657 > 0.05$).

Overall, according to the results in Table 7, unit weight and friction angle significantly positively affect stability. The slope angle has a negative effect on stability. However, slope height, cohesion, and pore water ratio have less influence on slope stability.

4.3. Comparison of Statistical-Based Models with Multiple Machine Learning Algorithm Models. To compare the performance of different classification algorithms, SVM, logistic regression (LR), GBDT, K-nearest neighbor (KNN), RF, XGboost, Naive Bayes model (GaussianNB), GA-Xgboost, and GA-GBDT methods were used for landslide susceptibility prediction. The prediction results of the models are shown in Figure 7. The AUC of SVM is 0.746, LR is 0.824, GBDT is 0.894, KNN is 0.817, RF is 0.894, Xgboost is 0.910, GaussianNB is 0.824, GA-Xgboost is 0.928, and GA-GBDT is 0.933. Table 8 shows the recall of SVM, LR, KNN, RF, and GaussianNB. It is worth noting that the AUC determines classifier performance, with 1.0 representing an ideal performance. The ROC curve of the GA optimization model is closer to the left and upper axes than the other models. The AUC value of GA-GBDT is highest at 0.933, slightly higher than RF and GA-Xgboost, and significantly higher than SVM, LR, GaussianNB, GBDT, and Xgboost.

Figure 8 is the radar chart of AUC and recall distribution, which visually shows the prediction effects of different algorithm models. The ensemble learning algorithm is more suitable for landslide susceptibility prediction than the

TABLE 7: Results of binary logit regression analysis-simplified format.

Item	Regression coefficient
Slope height	0.002 (1.448)
Slope angle	-0.045** (-2.607)
Unit weight	0.202** (3.849)
Cohesion	0.002 (0.529)
Friction angle	0.086** (3.806)
Pore water pressure	0.370 (0.444)
Intercept distance	-5.738** (-5.557)
Likelihood ratio test	$\chi^2(6) = 87.032, p = 0.001$
Hosmer-Lemeshow test	$\chi^2(8) = 17.027, p = 0.030$

Dependent variable: stability. McFadden R^2 : 0.225. Cox and Snell R^2 : 0.259. Nagelkerke R^2 : 0.352. * $p < 0.05$ ** $p < 0.01$ z values in parentheses.

other algorithms, while the prediction performance of SVM is the lowest. Using GA to optimize Xgboost and GBDT significantly improves the AUC curve compared with the original model, indicating that the algorithm accuracy strongly correlates with the parameters. However, the GA-GBDT model has the best model accuracy and performance. The test results also show that the GA-GBDT model is more suitable for predicting landslide susceptibility.

Different modeling approaches can lead to different results. An evaluation of the predictive performance of numerous machine learning models shows that most models are more accurate than the traditional statistical models used for landslide susceptibility modeling. Machine learning methods can automatically identify the hidden complex relationships between valid variables. The results of this study are more applicable to the integrated learning algorithm for landslide susceptibility prediction when compared with the results of recent studies [27–29] and that the accuracy of the model is improved after the optimization of the heuristic algorithm. The results of the integrated algorithm model differ from other studies in different country regions of the world, widely showing excellent landslide susceptibility predictions ($AUC > 0.8$). It is worth noting that different geological conditions and regional influences have

TABLE 8: Confusion matrix and recall for the prediction model based on 290 slope cases.

Model	True positive (TP)	True negative (TN)	Number of predictions	Recall (%)
SVM	22	21	58	81.5
LR	21	27	58	77.8
KNN	18	30	58	66.7
RF	23	29	58	85.2
GaussianNB	21	27	58	77.8

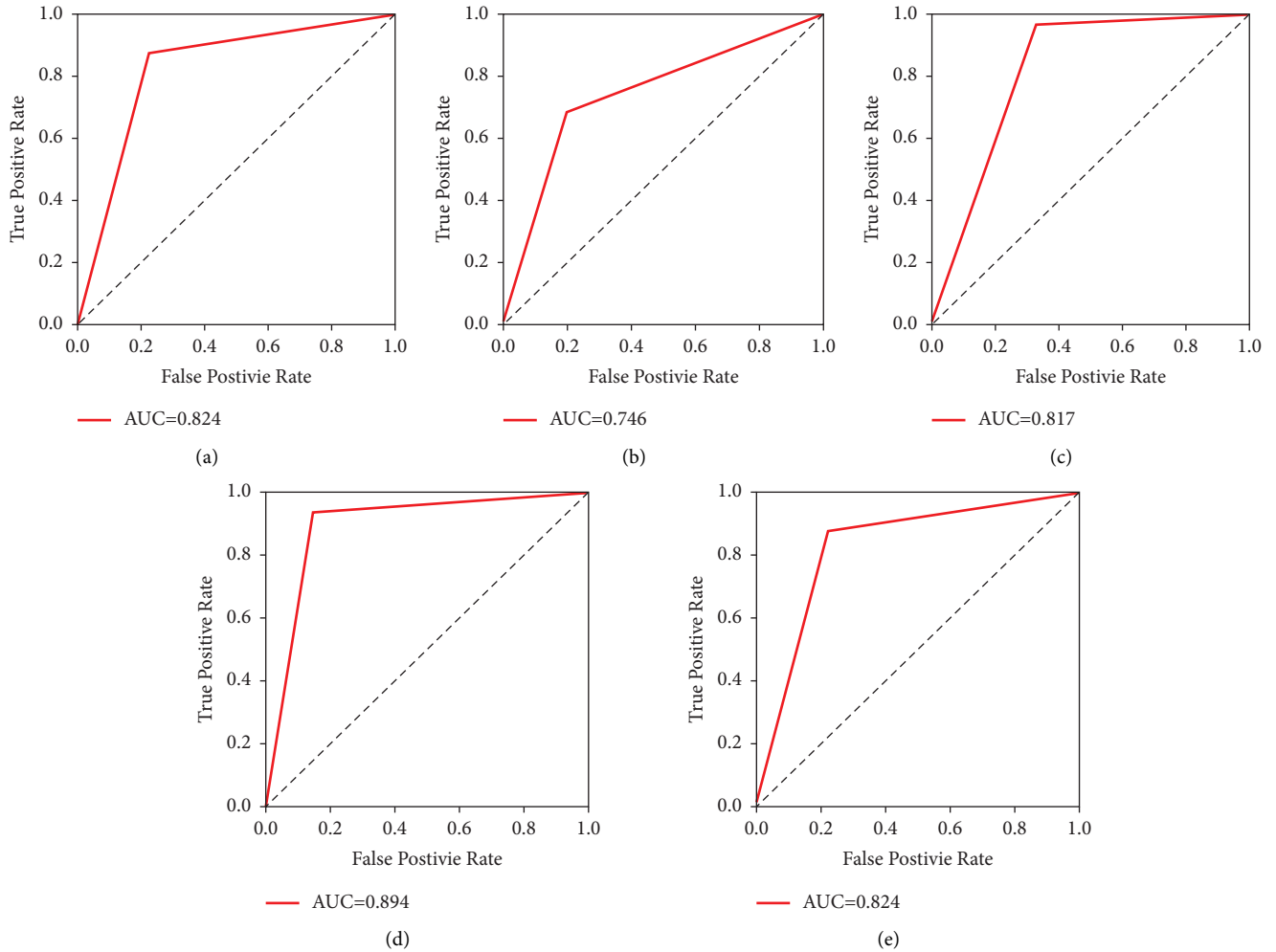


FIGURE 7: ROC curves of different algorithms. (a) LR ROC. (b) SVM ROC. (c) KNN RON. (d) RF ROC. (e) GaussianNB ROC.

important effects on the accuracy of the model prediction results. Therefore, selecting a rich and high-quality landslide dataset can increase the actual predictive power of the model.

4.4. Analysis of the Importance of Influencing Factors. It is crucial to determine the sensitivity of factors affecting landslide susceptibility to evaluate the landslide susceptibility and the design of support structures. The GA-GBDT algorithm has a good feature identification function and can output the strength of different parameters on landslide susceptibility. Figure 9 shows the importance of the six influencing factors ranked by the GA-GBDT model. This

study used relative importance scores to investigate the sensitivity based on the best prediction results (GA-GBDT). The method was selected based on the superior performance during the test setup. The results of the GA-GBDT feature selection were displayed to obtain the variable importance ranking. Figure 7 shows the normalized scores for the importance of the variables. Unit weight (score = 0.4593), friction angle (score = 0.1245), and cohesion (score = 0.1237) are the most sensitive factors for landslide susceptibility, which indicates the importance of the slope material variables. Therefore, the values of material unit weight, friction angle, and cohesion in artificial slopes must be selected reasonably and accurately based on specific indoor and field tests. The geological material cohesion and friction angle

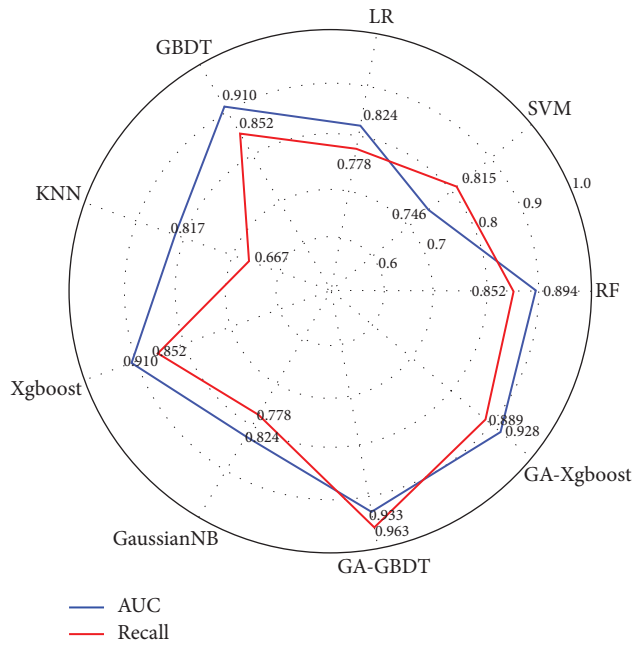


FIGURE 8: Radar plot of AUC vs recall distribution for different models.

should be increased when assessing a slope for landslide potential. The importance scores of slope angle and height are 0.1161 and 0.1026, respectively, indicating that geometric variables also affect landslide susceptibility. Optimizing these two variables in the actual design is a feasible approach to ensure the stability. Finally, pore water pressure (ru) (0.0737) has the lowest sensitivity.

5. Discussion

5.1. Variability in Model Performance. For landslide susceptibility prediction results compared with other models, the integrated learning models GBDT, RF, and Xgboost have higher accuracy than statistical models and some machine learning models such as SVM and GaussianNB. The models optimized by the heuristic algorithm GA had increased prediction accuracy. Pham et al. [27] strongly recommend implementing the chosen heuristic along with the machine learning model rather than directly applying only the machine learning model or its integrated form.

It is also essential to discuss the strengths and weaknesses of the applied models, as different models have their strengths and weaknesses. Usually, model performance depends on different research areas and related influencing factors [44]. In this study, GA-Xgboost and GA-GBDT have higher accuracy than machine learning models (ex., Xgboost and GaussianNB). GA has the advantages of fast random search capability and scalability with free attention to problem domains and is easy to combine with other algorithms, but the potential power of the algorithm's parallel mechanism is not fully exploited [45]. There are several other drawbacks as well, such as the dependence on fitness function. This is new and current research in GAs [46].

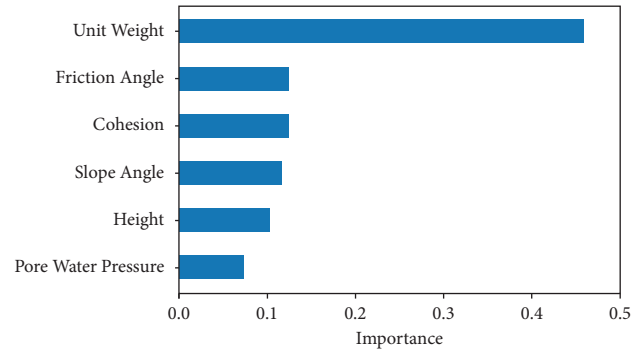


FIGURE 9: Variable importance plot generated by the GA-GBDT classifiers.

RF is highly tolerant to outliers and noise and can handle multidimensional data without overfitting, yet many trees may slow down the algorithm and prevent real-time prediction (Arabameri et al., 2019). KNN only needs to save training samples and tokens without estimating parameters and training. However, the categories of the new samples are biased towards the category with the dominant number in the training sample when the samples are unbalanced, which can easily lead to prediction errors [47]. The GaussianNB model can achieve high prediction accuracy in learning data with missing conditions [30]. It assumes the independence of attributes among data, which leads to difficulties in practical applications and low efficiency of model classification when facing more data attributes or stronger correlations among data [48]. The Xgboost method supports linear classifier and CART classifier, which can better prevent overfitting and reduce model complexity despite lacking smoothness [49]. Although SVM is a classical small-sample learning algorithm, which can reach a high level of learning accuracy with a small amount of data, its accuracy and computational speed both decrease when facing multidimensional variables and large data [50]. The performance of GBDT is a step up from RF, so its advantages are also obvious. It is flexible to handle various data types and has high prediction accuracy with relatively less tuning time. Because it is boosting, there is a serial relationship before the base learner, which makes it challenging to train data in parallel [51]. Therefore, it is necessary to apply the selected models to different datasets and scenarios of landslide susceptibility problems for training and comparative analysis concerning the merits and performance of each model.

5.2. Challenges in Landslide Susceptibility Assessment. With the deepening of research, many researchers have proposed optimization models with high prediction accuracy and strong generalization ability [52]. These techniques have a very strong self-learning ability and can process large amounts of data efficiently and accurately. In addition, these techniques are mainly used in landslide susceptibility evaluation to help make decisions on risk reduction [53]. However, there are still some problems and challenges to be solved in the future. (1) Incomplete consideration of

influencing factors when using machine learning techniques for landslide susceptibility modeling. In most of the current research focuses on geotechnical properties, slope, height, and so on. But in fact factors not covered often have significant impacts, such as climate and environmental changes. (2) Since machine learning is based on data-driven search for inherent hidden relationships as a way to achieve the purpose of landslide geohazard prediction interacting with the natural world. If the prediction process has a better machine learning model but is not able to obtain higher-quality data, it will not be able to obtain better prediction results. (3) Based on the data-driven machine learning prediction model is difficult to explain the mechanism of landslide occurrence, the model applicability is poor, which also affects the accuracy of the model prediction, so how to improve the interpretability of the machine learning algorithm is also an important research direction in the future.

6. Conclusions

This work successfully used GA-GBDT and GA-Xgboost methods to study landslide susceptibility prediction using 290 historical case records of slope conditions. The variability of model performance with different parameters was checked, and GA was used to optimize the four parameters of Xgboost and GBDT, which are `n_estimators` `learning_rate` `max_depth` `random_state`.

GA-GBDT and GA-Xgboost models were compared with SVM, LR, KNN, RF, GaussianNB, Xgboost, and GBDT to check the fit. It was found that Xgboost and GBDT optimized using GA obtained classification models with the highest AUC of 0.928 and 0.933, respectively. Relative variable importance analysis showed that the geometric slope design parameters (unit weight, friction angle, and cohesion) significantly affected landslide susceptibility. The results suggest that the GA-GBDT and GA-Xgboost can explore the nonlinear relationship between landslide susceptibility and its influencing factors.

Based on the database of this paper, the results of logistic regression to study the influence of different variables show that unit weight and friction angle significantly positively influences stability, slope angle has a significant negative influence on stability. However, slope height (m), cohesion, and pore water pressure have less influence on stability. The intrinsic effects were further explored.

However, although the analysis results are impressive and encouraging, there are still some outstanding issues. The impact of data imbalance on landslide susceptibility prediction will be discussed in future research. The developed model can be improved by analyzing more extensive datasets, and its applicability to other mining and geotechnical damage problems can be recommended when data are available.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

Binbin Zheng and Jiahe Wang worked together to complete the entire article. Wensong Wang and Tingting Feng conceptualized the study. Guangjin Wang and Yonghao Yang contributed to data curation. Yufei Wang, Wensong Wang, and Tingting Feng performed formal analysis. Wensong Wang, Binbin Zheng, and Guangjin Wang provided funding acquisition. Jiahe Wang, Yufei Wang, and Yonghao Yang performed investigation. Jiahe Wang contributed to the methodology. Wensong Wang provided project administration. Binbin Zheng performed supervision. Tingting Feng and Yufei Wang contributed to validation. Tingting Feng and Binbin Zheng performed visualization. Wensong Wang and Jiahe Wang wrote the original draft. Wensong Wang wrote, reviewed, and edited the manuscript.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (51804178 and 51804051), the Chongqing Natural Science Foundation (No. CSTB2022NSCQ-MSX0497), the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant no. KJQN202100717), and the Yangtze River Joint Research Phase II Program (No. 2022-LHYJ-02-0201).

References

- [1] H. Basahel and H. Mitri, "Probabilistic assessment of rock slopes stability using the response surface approach—a case study," *International Journal of Mining Science and Technology*, vol. 29, no. 3, pp. 357–370, 2019.
- [2] S. Lin, H. Zheng, B. Han, Y. Li, C. Han, and W. Li, "Comparative performance of eight ensemble learning approaches for the development of models of slope stability prediction," *Acta Geotechnica*, vol. 17, no. 4, pp. 1477–1502, 2022.
- [3] D. K. Yadav, S. Jayanthu, S. K. Das, S. Chinara, and P. Mishra, "Critical review on slope monitoring systems with a vision of unifying WSN and IoT," *IET Wireless Sensor Systems*, vol. 9, no. 4, pp. 167–180, 2019.
- [4] W. S. Wang, G. S. Cao, Y. Li et al., "Experimental study of dynamic characteristics of tailings with different reconsolidation degrees after liquefaction," *Frontiers in Earth Science*, vol. 10, Article ID 876401, 2022.
- [5] N. Kardani, A. Zhou, M. Nazem, and S. Shen, "Improved prediction of slope stability using a hybrid stacking ensemble method based on finite element analysis and field data," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 13, no. 1, pp. 188–201, 2021.
- [6] Z. Liu, J. Shao, W. Xu, H. Chen, and Y. Zhang, "An extreme learning machine approach for slope stability evaluation and prediction," *Natural Hazards*, vol. 73, no. 2, pp. 787–804, 2014.
- [7] B. Thiebes, R. Bell, T. Glade et al., "Integration of a limit-equilibrium model into a landslide early warning system," *Landslides*, vol. 11, no. 5, pp. 859–875, 2014.

- [8] H. Zheng, D. Liu, and C. G. Li, "Slope stability analysis based on elasto-plastic finite element method," *International Journal for Numerical Methods in Engineering*, vol. 64, no. 14, pp. 1871–1888, 2005.
- [9] M. Sakellariou and M. Ferentinou, "A study of slope stability prediction using neural networks," *Geotechnical & Geological Engineering*, vol. 23, no. 4, pp. 419–445, 2005.
- [10] S. Li, H. Zhao, and Z. Ru, "Slope reliability analysis by updated support vector machine and Monte Carlo simulation," *Natural Hazards*, vol. 65, no. 1, pp. 707–722, 2013.
- [11] F. Huang, J. Zhang, C. Zhou, Y. Wang, J. Huang, and L. Zhu, "A deep learning algorithm using a fully connected sparse autoencoder neural network for landslide susceptibility prediction," *Landslides*, vol. 17, no. 1, pp. 217–229, 2020.
- [12] S. Rukhaiyar, M. Alam, and N. K. Samadhiya, "A PSO-ANN hybrid model for predicting factor of safety of slope," *International Journal of Geotechnical Engineering*, vol. 12, no. 6, pp. 556–566, 2018.
- [13] L. Zhang, Z. Chen, T. Zhou, G. Nian, J. Wang, and Z. Zhou, "Multi-source information fusion and stability prediction of slope based on gradient boosting decision tree," *Journal of China Coal Society*, vol. 45, pp. 173–180, 2020.
- [14] B. Gordan, D. Jahed Armaghani, M. Hajihassani, and M. Monjezi, "Prediction of seismic slope stability through combination of particle swarm optimization and neural network," *Engineering with Computers*, vol. 32, no. 1, pp. 85–97, 2016.
- [15] X. Li and J. Kong, "Application of GA-SVM method with parameter optimization for landslide development prediction," *Natural Hazards and Earth System Sciences*, vol. 14, no. 3, pp. 525–533, 2014.
- [16] W. Zhang, X. Gu, L. Tang, Y. Yin, D. Liu, and Y. Zhang, "Application of machine learning, deep learning and optimization algorithms in geoengineering and geoscience: comprehensive review and future challenge," *Gondwana Research*, vol. 109, pp. 1–17, 2022.
- [17] C. Qi and X. Tang, "Slope stability prediction using integrated metaheuristic and machine learning approaches: a comparative study," *Computers & Industrial Engineering*, vol. 118, pp. 112–122, 2018.
- [18] D. Tien Bui, H. Moayed, M. Gör, A. Jaafari, and L. K. Foong, "Predicting slope stability failure through machine learning paradigms," *ISPRS International Journal of Geo-Information*, vol. 8, no. 9, p. 395, 2019.
- [19] Z. Chang, F. Catani, F. Huang et al., "Landslide susceptibility prediction using slope unit-based machine learning models considering the heterogeneity of conditioning factors," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 15, 2022.
- [20] S. Lin, H. Zheng, C. Han, B. Han, and W. Li, "Evaluation and prediction of slope stability using machine learning approaches," *Frontiers of Structural and Civil Engineering*, vol. 15, no. 4, pp. 821–833, 2021.
- [21] F. Huang, Z. Cao, J. Guo, S. H. Jiang, S. Li, and Z. Guo, "Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping," *Catena*, vol. 191, Article ID 104580, 2020.
- [22] Z. Chang, Z. Du, F. Zhang et al., "Landslide susceptibility prediction based on remote sensing images and GIS: comparisons of supervised and unsupervised machine learning models," *Remote Sensing*, vol. 12, no. 3, p. 502, 2020.
- [23] Y. Huang, X. Han, and L. Zhao, "Recurrent neural networks for complicated seismic dynamic response prediction of a slope system," *Engineering Geology*, vol. 289, Article ID 106198, 2021.
- [24] M. Koopialipoor, D. Jahed Armaghani, A. Hedayat, A. Marto, and B. Gordan, "Applying various hybrid intelligent systems to evaluate and predict slope stability under static and dynamic conditions," *Soft Computing*, vol. 23, no. 14, pp. 5913–5929, 2019.
- [25] S. H. Jiang, J. Huang, F. Huang, J. Yang, C. Yao, and C. B. Zhou, "Modelling of spatial variability of soil undrained shear strength by conditional random fields for slope reliability analysis," *Applied Mathematical Modelling*, vol. 63, pp. 374–389, 2018.
- [26] Y. Achour and H. R. Pourghasemi, "How do machine learning techniques help in increasing accuracy of landslide susceptibility maps?" *Geoscience Frontiers*, vol. 11, no. 3, pp. 871–883, 2020.
- [27] Q. B. Pham, Y. Achour, S. A. Ali et al., "A comparison among fuzzy multi-criteria decision making, bivariate, multivariate and machine learning models in landslide susceptibility mapping," *Geomatics, Natural Hazards and Risk*, vol. 12, no. 1, pp. 1741–1777, 2021.
- [28] Y. Achour, Z. Saidani, R. Touati et al., "Assessing landslide susceptibility using a machine learning-based approach to achieving land degradation neutrality," *Environmental Earth Sciences*, vol. 80, no. 17, p. 575, 2021.
- [29] I. Drid, Y. Achour, K. Zighmi et al., "Designing gully erosion susceptibility maps (GESM) in the Algerian Eastern Tell: a case study of the K'sob River watershed," *Arabian Journal of Geosciences*, vol. 15, no. 14, p. 1264, 2022.
- [30] X. Feng, S. Li, C. Yuan, P. Zeng, and Y. Sun, "Prediction of slope stability using naive Bayes classifier," *KSCE Journal of Civil Engineering*, vol. 22, no. 3, pp. 941–950, 2018.
- [31] J. Zhou, E. Li, S. Yang et al., "Slope stability prediction for circular mode failure using gradient boosting machine approach based on an updated database of case histories," *Safety Science*, vol. 118, pp. 505–518, 2019.
- [32] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378, 2002.
- [33] W. Zhang, R. Zhang, C. Wu et al., "State-of-the-art review of soft computing applications in underground excavations," *Geoscience Frontiers*, vol. 11, no. 4, pp. 1095–1106, 2020.
- [34] S. Sachdeva and B. Kumar, "A comparative study between frequency ratio model and gradient boosted decision trees with greedy dimensionality reduction in groundwater potential assessment," *Water Resources Management*, vol. 34, no. 15, pp. 4593–4615, 2020.
- [35] T. Chen and C. Guestrin, "Xgboost: a scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, San Francisco, CA, USA, August 2016.
- [36] Z. Luo, X.-N. Bui, H. Nguyen, and H. Moayed, "A novel artificial intelligence technique for analyzing slope stability using PSO-CA model," *Engineering with Computers*, vol. 37, no. 1, pp. 533–544, 2021.
- [37] X. Xue, Y. Li, X. Yang, X. Chen, and J. Xiang, "Prediction of landslide susceptibility based on GA-BP hybrid algorithm," *Neural Network World*, vol. 25, no. 2, pp. 189–202, 2015.
- [38] S. Kostić, N. Vasović, and D. Sunarić, "A new approach to grid search method in landslide susceptibility analysis using Box-Behnken statistical design," *Applied Mathematics and Computation*, vol. 256, pp. 425–437, 2015.
- [39] A. H. Gandomi, A. R. Kashani, M. Mousavi, and M. Jalalvandi, "Slope stability analyzing using recent swarm

- intelligence techniques,” *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 39, no. 3, pp. 295–309, 2015.
- [40] J. H. Ang, E. J. Teoh, C. H. Tan, K. C. Goh, and K. C. Tan, “Dimension reduction using evolutionary support vector machines,” in *Proceedings of the 2008 IEEE congress on evolutionary computation (IEEE world congress on computational intelligence)*, pp. 3634–3641, IEEE, Hong Kong, China, June 2008.
- [41] F. Huang, J. Yan, X. Fan et al., “Uncertainty pattern in landslide susceptibility prediction modelling: effects of different landslide boundaries and spatial shape expressions,” *Geoscience Frontiers*, vol. 13, no. 2, Article ID 101317, 2022.
- [42] W. Li, X. Fan, F. Huang et al., “Uncertainties analysis of collapse susceptibility prediction based on remote sensing and GIS: influences of different data-based models and connections between collapses and environmental factors,” *Remote Sensing*, vol. 12, no. 24, p. 4134, 2020.
- [43] A. Nandi and A. Shakoor, “Application of logistic regression model for slope instability prediction in Cuyahoga River Watershed, Ohio, USA,” *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, vol. 2, no. 1, pp. 16–27, 2008.
- [44] Y. Wang, D. Sun, H. Wen, H. Zhang, and F. Zhang, “Comparison of random forest model and frequency ratio model for landslide susceptibility mapping (LSM) in Yunyang County (Chongqing, China),” *International Journal of Environmental Research and Public Health*, vol. 17, no. 12, p. 4206, 2020.
- [45] E. Alba, G. Luque, and S. Nesmachnow, “Parallel meta-heuristics: recent advances and new trends,” *International Transactions in Operational Research*, vol. 20, no. 1, pp. 1–48, 2013.
- [46] R. Yadav, I. Sreedevi, and D. Gupta, “Bio-inspired hybrid optimization algorithms for energy efficient wireless sensor networks: a comprehensive review,” *Electronics*, vol. 11, no. 10, p. 1545, 2022.
- [47] F. Bao, Y. Deng, Y. Kong, Z. Ren, J. Suo, and Q. Dai, “Learning deep landmarks for imbalanced classification,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 8, pp. 2691–2704, 2020.
- [48] B. Twala, “Multiple classifier application to credit risk assessment,” *Expert Systems with Applications*, vol. 37, no. 4, pp. 3326–3336, 2010.
- [49] J. Yang, C. Zhao, H. Yu, and H. Chen, “Use GBDT to predict the stock market,” *Procedia Computer Science*, vol. 174, pp. 161–171, 2020.
- [50] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, “A comprehensive survey on support vector machine classification: applications, challenges and trends,” *Neurocomputing*, vol. 408, pp. 189–215, 2020.
- [51] G. Yu, S. Zhang, M. Hu, and Y. K. Wang, “Prediction of highway tunnel pavement performance based on digital twin and multiple time series stacking,” *Advances in Civil Engineering*, vol. 2020, Article ID 8824135, 21 pages, 2020.
- [52] T. Kavzoglu, I. Colkesen, and E. K. Sahin, “Machine learning techniques in landslide susceptibility mapping: a survey and a case study,” *Landslides: Theory, practice and modelling*, pp. 283–301, 2019.
- [53] E. Kutlug Sahin and I. Colkesen, “Performance analysis of advanced decision tree-based ensemble learning algorithms for landslide susceptibility mapping,” *Geocarto International*, vol. 36, no. 11, pp. 1253–1275, 2021.