Hindawi

Research Article

A Hybrid Random Forest and Least Squares Support Vector Machine Model Based on Particle Swarm Optimization Algorithm for Slope Stability Prediction: A Case Study in Sichuan–Tibet Highway, China

Tao Shu [ID],[1,2] Wei Tao [ID],[3] Haotian Lu [ID],[4,5,6] Hao Li [ID],[4,5,6] and Jingxuan Cao [ID][4,5,6]

[1]*Science Center for Deep Ocean Multispheres and Earth System, Key Lab of Submarine Geosciences and Prospecting Techniques, Ministry of Education and College of Marine Geosciences, Ocean University of China, Qingdao 266100, China*
[2]*International Center for Submarine Geosciences and Geoengineering Computing, Ocean University of China, Qingdao 266100, China*
[3]*Tibet University, Lhasa 850000, China*
[4]*University of Chinese Academy of Sciences, Beijing 100049, China*
[5]*Key Laboratory of Mountain Hazards and Earth Surface Process, Chinese Academy of Sciences, Chengdu 610041, China*
[6]*Institute of Mountain Hazards and Environment, Chinese Academy of Sciences and Ministry of Water Conservancy, Chengdu 610041, China*

Correspondence should be addressed to Wei Tao; zroot@utibet.edu.can

Slope stability estimation is an engineering problem that involves several parameters. The problems of low accuracy of the model and blind data preprocessing are commonly existent in slope stability prediction research. To address these problems, 10 quantitative indicators are selected from 135 field cases to improve the accuracy of the model. These indicators were analyzed and visualized to examine their reliabilities after preprocessing. Combined with random forest (RF), particle swarm optimization (PSO), and least squares support vector machine (LSSVM) algorithms, a hybrid prediction model that the RF–PSO–LSSVM model is proposed for identifying slope stability, and its reliability is verified by other prediction models that SVM, logistic regression, decision trees, k-nearest neighbor, naive Bayes, and linear discriminant analysis. Besides, the importance score of each indicator in the prediction of slope stability is discussed by employing the RF algorithm. The research results show that the proposed hybrid model exhibits the best accuracy and superiority in slope stability prediction than other models in this paper, which its values of the best fitness, area under the curve, T-measure, and accuracy are 98.15%, 96.4%, 96.55%, and 95.82%, respectively. The most influential factors affecting slope stability are precipitation and gravity, and the slope type and pore water ratio are identified as the least significant factors in this paper. The results provide a novel approach toward slope stability prediction in the field of geotechnical engineering.

1. Introduction

Landslides, resulting from highway slope instability, are one of the profoundly destructive natural disasters that threaten the operation of highways and the safety of people's lives [1–3]. Timely and accurate slope stability analysis and prediction are of great importance in highland mountain highway construction and maintenance to prevent or reduce the occurrence of landslides, ensure the safe and stable passage of highways, and reduce highway maintenance costs [4].

Slope engineering is often considered an uncertain, nonlinear, complex, dynamic system, and its stability is affected by many factors [5, 6]. Although the traditional slope stability determination methods (e.g., the Swedish slice method [7], the Bishop method [8–10], and the Janbu slices method [11]) can obtain more accurate determination results, the
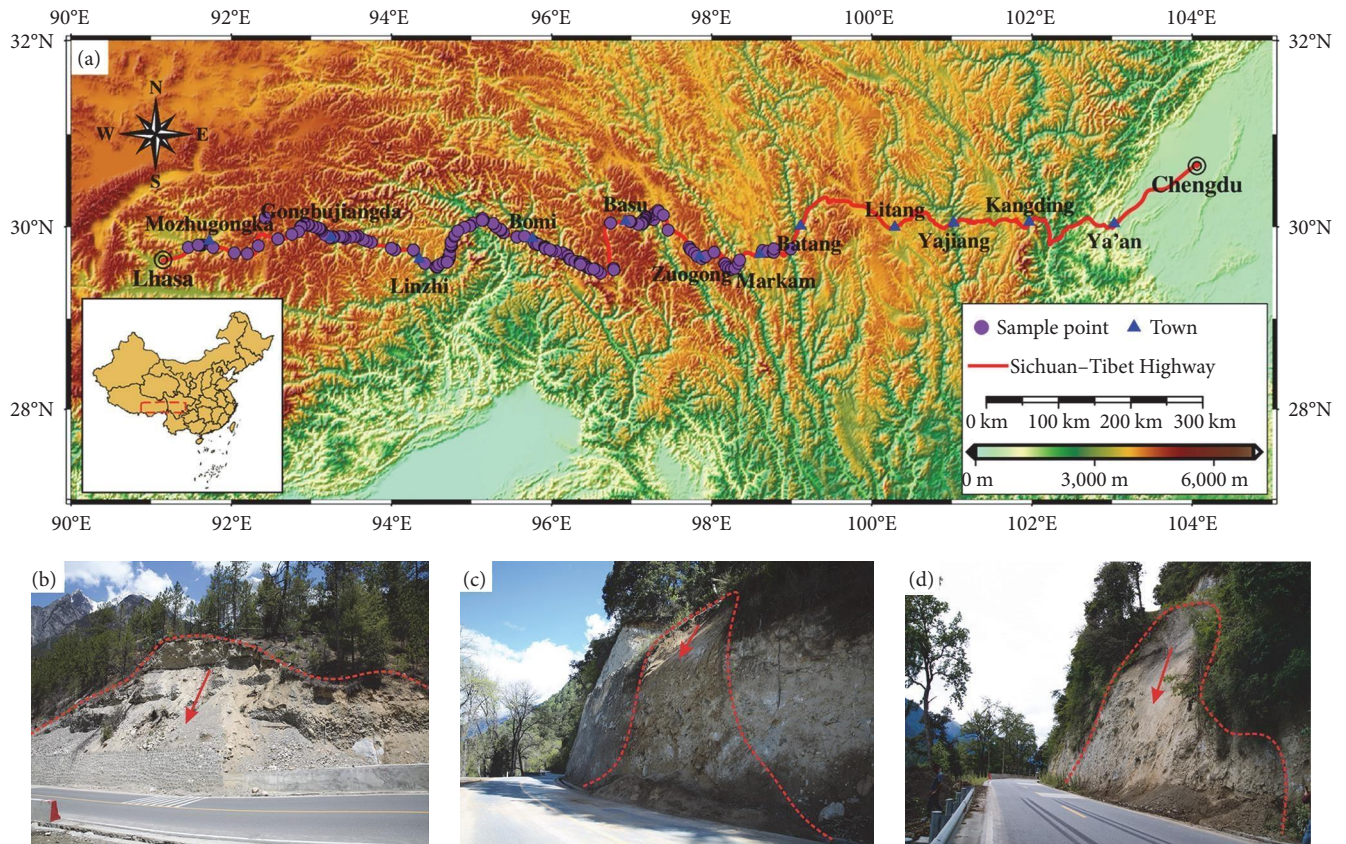
FIGURE 1: (a) Field data investigation area. (b–d) Field investigation of slopes.

complexity of the material parameters of the slope body makes these methods require large and complicated calculation processes [12]. Furthermore, these methods often focus on analyzing specific types of slope stability influencing factors, making it challenging to capture the nonlinearity among influencing factors of slope stability. To reflect the nonlinear problem between the slope stability influencing factors and prediction results, nonlinear machine learning methods are applied to slope stability analysis [13–16]. These machine learning methods overcome the shortcomings of the traditional methods but require a large amount of slope sample data to learn the training model, thus they have certain limitations. To overcome these limitations, a least squares support vector machine (LSSVM) model [17, 18], which is suitable for small sample machine learning models, has been widely applied in slope stability prediction due to its advantages of better prediction accuracy and generalization ability [19–21]. The accuracy of LSSVM is determined by its penalty terms and kernel functions. To address these problems, optimization algorithms such as the genetic algorithm [22], firefly algorithm [23], metaheuristics [24], and germinal center optimization [25], among others, have been proposed to obtain the optimal combination of parameters. These algorithms can optimize the LSSVM model well, but they are prone to fall into the local optimal solution [26–28]. Slope stability identification requires a comprehensive, representative, and practical set of influencing factors. However, this does not imply that every evaluation factor must be included or incorporated into the evaluation process [29]. The inclusion of excessive factors in the evaluation process may reduce the efficiency and accuracy of the evaluation model, as these factors often exhibit linear correlations that could result in erroneous results [30].

To address the above problems, this paper proposed the random forest (RF)–particle swarm optimization (PSO)–LSSVM model for predicting slope stability and studying the suitability of this model. The RF algorithm is employed to calculate the importance score of slope stability influencing factors, which effectively reduces the dimensionality of factors and eliminates certain linear correlations among influencing factors of slope stability. The PSO algorithm is employed for obtaining the optimal parameter combination of the LSSVM model. Section 2 displays that the research methods and the data processing and data analysis of slope data are briefly introduced. Section 3 presents the process of the prediction model of slope stability based on the RF, PSO, and LSSVM algorithms. Section 4 displays the results and discussion discusses of this study. Section 5 presents the conclusion of this study.

2. Data Sources and Methods

2.1. Data Sources and Processing

2.1.1. Study Area. The Sichuan–Tibet Highway (G318) has 2,142 km long and starts from Chengdu in Sichuan Province and ends at Lhasa of Tibet [31] (Figure 1(a)). The collision between the Indian and Eurasian tectonic plates has resulted
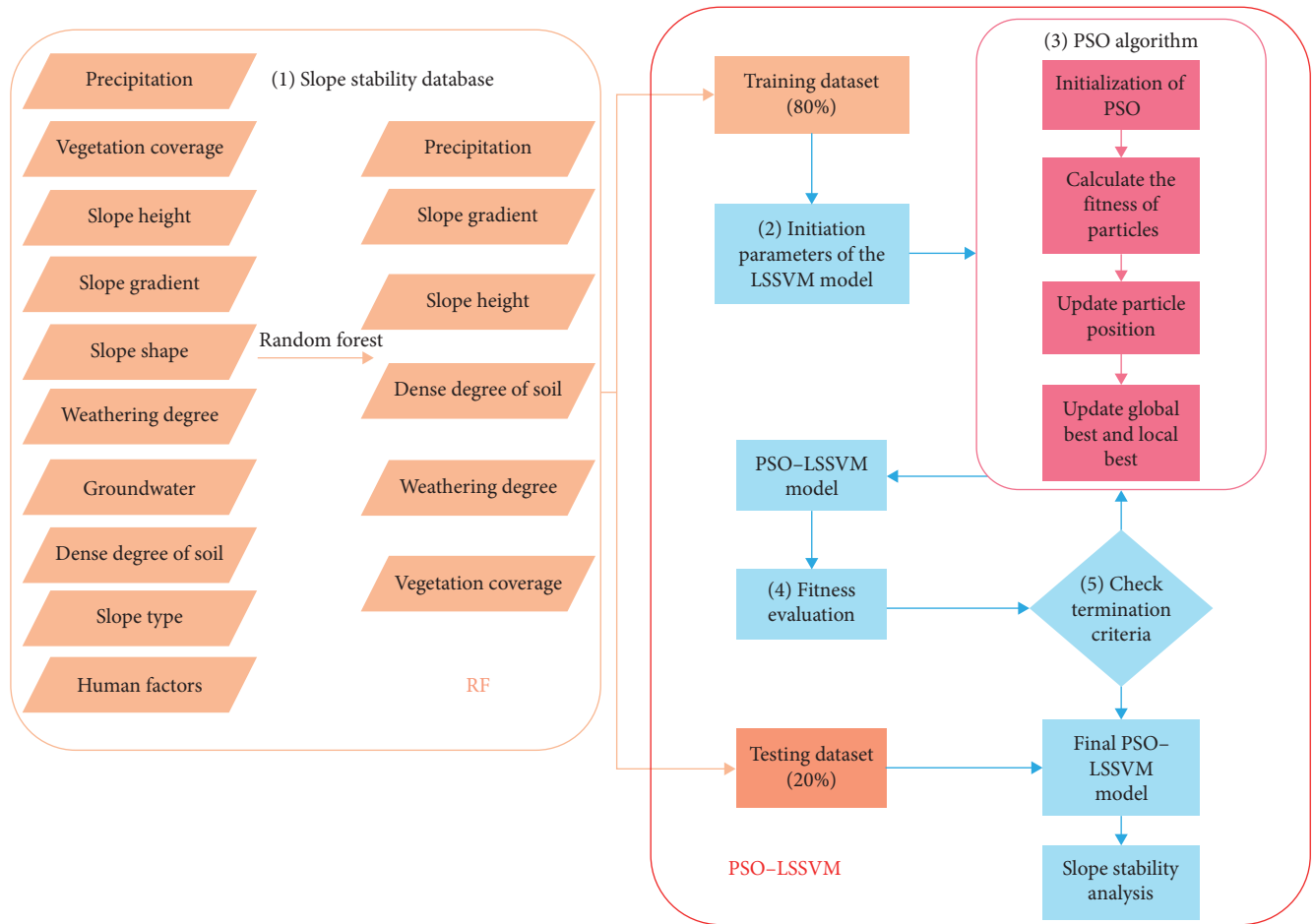
FIGURE 2: Flowchart of RF–PSO–LSSVM model.

in intensive tectonic activity in the area, leading to a high frequency of geological hazards in the region [32]. In this paper, the study area is the Sichuan–Tibet Highway (G318) from Jinsha River to Lhasa (Figure 1(a)).

In this paper, from June 2019 to July 2021, a comprehensive investigation of slopes along 1,365 km of the Jinsha River to the Lhasa Qushui portion of the Sichuan–Tibet Highway was conducted. A total of 250 sets of slope data were collected along the highway. By data preprocessing (eliminating missing data items and data collection error samples), the final valid slope data for the remaining 153 sets of slope data. The slope data investigation area is shown in Figure 1. The investigated slopes are distributed along the Sichuan–Tibet Highway and are mainly concentrated in mountainous canyons with abundant precipitation. The slope is mostly between 30° and 40°, which is relatively unstable (Figure 1 (b)–1(d)).

2.1.2. Selection of Influencing Factors. Selecting influencing factors is necessary to identify slope stability. In this paper, through field investigation and a review of relevant literature, the main factors affecting the stability of slopes are screened and divided into four main categories, namely, topographic features, geological conditions, precipitation, and other factors (vegetation, hydrogeology, etc.). The 10 selected factors

affecting slope stability from these four categories are input into the RF–PSO–LSSVM model as independent variables for slope stability prediction, namely, slope type ($X1$), precipitation ($X2$), vegetation coverage ($X3$), slope height ($X4$), slope gradient ($X5$), slope shape ($X6$), groundwater ($X7$), weathering degree ($X8$), dense degree of soil ($X9$), and human factors ($X10$). Due to space limitations, the raw data of slope are shown in the supplementary material.

The slope gradient and height indicate the topographic condition of the slope, and the slope shape and slope type represent the geological condition of the slope to some extent. Precipitation is one of the main factors that induce landslides [33]. The vegetation coverage could represent the forest vegetation of affection for slope stability identification [34]. Understanding the density and strength of the soil is necessary for slope stability studies, and the dense degree of soil indirectly represents these soil characteristics [35]. Human factors have a substantial impact on slope stability, such as deforestation, and slope cutting, which will accelerate the deformation and instability of the slope, hence have been considered. The slope stability is calculated based on the qualitative characteristics of the site slope, which include encompassing observations such as the presence of substantial debris accumulation at the slope's base, the occurrence of landslide scars on the surface, and the visibility of fissures
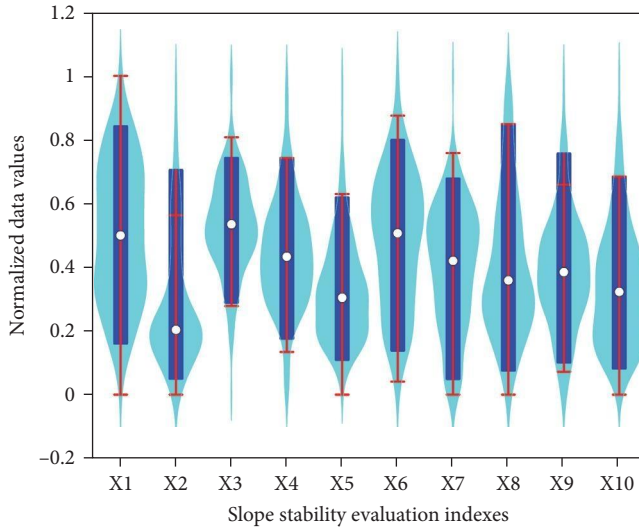
FIGURE 3: Violin plots with different indices.

along the slope. The results of the slope stability (I1) are divided into two types, stable and failure, represented by 1 and 0, respectively, as the dependent variable for the RF–PSO–LSSVM model.

2.1.3. Data Preprocessing. The primary purpose of data pre-processing is to eliminate the potential influence of features with larger numerical ranges over those with smaller numerical ranges, thus enabling the extraction of comprehensive and valuable information from the raw data [29]. Since the evaluation indices are of different magnitudes, they are normalized, and all data are normalized to between [0,1]. The main advantage of normalization processing is preventing calculation difficulties caused by different dimensions and improving LSSVM model accuracy [36]. Due to space limitations, the quantitation and standardization of influencing factors of slope stability and the quantified slope sample data are shown in the supplementary material.

2.2. Research Methods. In this study, a hybrid model that the RF–PSO–LSSVM was employed predicting slope stability. The workflow of the RF–PSO–LSSVM model is illustrated in Figure 2. In this paper, firstly, the collected slope data from the field undergo an initial preprocessing and are normalized to eliminate the impact of different data scales on the prediction model. Then, exploratory data analysis is conducted to gain a deeper understanding of the data. The objectives of exploratory data analysis include the following: (1) conducting a comprehensive review of the data's integrity and validity; (2) analyzing the correlations among the characteristic and post-processing the highly relevant characteristic with an RF algorithm. The PSO algorithm is employed to optimize the LSSVM model, which is then utilized to predict slope stability. Finally, the effectiveness of the proposed model is validated by an analysis and comparison with the slope stability prediction results obtained from other models.

2.2.1. RF Algorithm. Slope stability estimation is an engineering problem that involves several parameters. Given that the

influence factors that the prediction of slope stability are generally high dimensional and include a vast number of irrelevant features. Therefore, it becomes crucial to select the feature of influence factors. The RF algorithm is extensively employed to extract nonlinear structure information and dimensionality reduction.

The RF algorithm is an integrated machine-learning algorithm proposed by Breiman [37]. RF can avoid overfitting by sampling both samples and their features, is suitable for handling high-dimensional data, and can simultaneously give variable importance scores in the decision–making process [36, 37]. In this study, we use the variable importance score of RF to feature extraction from influencing factors of slope stability. The process is to calculate the ground contribution of each decision tree (DT) in an RF for each evaluation index and then take the average and compare and rank the contribution between each feature [38, 39].

2.2.2. LSSVM Algorithm. The LSSVM algorithm is an improved method proposed by Suykens and Vandewalle [40] based on SVM. Since the traditional SVM requires a quadratic solution for inequality constraints, the LSSVM converts inequality constraints into equation constraints for a solution, turning the quadratic solution into a linear system of equation solutions. This improvement simplifies the computation process and reduces the computation time. Therefore, the LSSVM has SVM characteristics and the advantages of a shorter model training time and more accurate results. Since the slope data are multidimensional nonlinear data, LSSVM is selected as the slope stability prediction model for a more accurate prediction of slope stability [41]. Using different kernel functions will constitute different LSSVM models, and the accuracy of the models may also vary. Therefore, kernel function selection is discussed in this paper.

2.2.3. PSO Algorithm. The PSO algorithm, developed by Kennedy and Eberhart [42], is an influential optimization algorithm. In contrast to traditional optimization algorithms, the PSO algorithm demonstrates remarkable proficiency in global search, enabling it to efficiently find the global optimal solution [40, 41]. Furthermore, it exhibits notable advantages in terms of achieving rapid convergence and calculating the global optimal solution.

The selection of hyperparameters significantly affects the accuracy of the results in the LSSVM model. Traditionally, these parameters are set empirically, which can lead to a locally optimal solution. To overcome this limitation, this study employs the PSO algorithm to adjust the hyperparameters and address this issue effectively.

The inertia factor ω is a crucial parameter that controls the balance between global and local search within the algorithm, making it highly significant [43]. Setting a fixed value of ω restricts the algorithm's ability to achieve global optimization and convergence speed. To resolve this problem, this paper employs Equation (1) to dynamically and nonlinearly adjust ω to enhance the performance of the PSO algorithm.

$$\omega = \omega_{\max} - (\omega_{\max} - \omega_{\min}) \cdot \arcsin\left(\frac{t}{t_{\max}} \cdot \frac{2}{\pi}\right), \qquad (1)$$

where $\omega_{\max}$ is the upper weight factor, $\omega_{\min}$ is the lower weight factor, and $t_{\max}$ is the upper limit of iterations.

3. Results

3.1. Exploratory Data Analysis. Before establishing the slope stability prediction models, it is crucial to conduct preliminary data analysis. This aims to unveil fundamental information, assess data integrity, examine the distribution characteristics of slope data, and explore factor correlations that these crucial steps form the foundation for the selection of an appropriate prediction model.

3.1.1. Data Integrity. To directly verify the integrity of the data processing for a total of 153 slopes, a data integrity analysis was performed and visualized using a violin chart. In Figure 3, the median of the character is represented by a white circle in each violin chart. In each blue box of the violin chart, the range represents the lower quartile and the upper quartile, while the thin red line inside the box indicates the 95% confidence interval. The shape of each violin in the plot depicts the kernel density estimation of the characteristic of the influencing factor. Based on the results, it can be indicated that the slope data after data preprocessing are integrated and obey the normal distribution.

3.1.2. Data Distribution Characteristics. Normalizing the data does not alter the characteristics of the data distribution. Therefore, the normalized data can be assessed to determine the reasonability of the dataset. The distributions of these data characteristics for the influencing factors on slope stability are shown in Figure 4. To visually demonstrate the normal distribution of different indices, a combination of a box plot and a normal distribution curve is plotted for these features, as depicted in Figure 5. According to Figures 4 and 5, it can be observed that a few features exhibit a right-skewed distribution, while the remaining features demonstrate a reasonable normal distribution. The results indicate that the quantification of the raw slope data is deemed reasonable, and the resulting quantified data exhibits a certain level of generalizability and predictability.

3.1.3. Data Correlation Analysis. Before the ultimate establishment of prediction models for slope stability, it is crucial to conduct a comprehensive analysis of the correlation among the characteristics of 10 influencing factors of slope stability. The existence of a strong correlation among the characteristics has the potential to significantly influence the precision of the prediction models and may result in erroneous conclusions that contradict the facts [44]. Pearson's correlation coefficient is commonly employed to quantify the correlation between two factors, which can be calculated by Equation (2) as follows:

$$P(m_1, m_2) = \frac{\mathrm{cov}(m_1, m_2)}{\sqrt{\mathrm{Var}[m_1]\mathrm{Var}[m_2]}} \qquad (2)$$

where m_1 and m_2 present independent factors; $P(m_1, m_2)$ presents the Pearson's correlation coefficient of m_1 and m_2; cov (m_1, m_2) denotes the covariance of m_1 and m_2 and Var $[m_1]$ and Var $[m_2]$ denote the variances of m_1 and m_2, respectively.

Pearson's correlation coefficient takes values within the range of -1 to 1. Figures 6 displays the correlation matrix of the influencing factors of slope stability. When Pearson's correlation coefficient approaches $+1$, it indicates a strong correlation between two factors. Conversely, if the correlation coefficient is far from $+1$, it suggests a weak correlation between the two factors. From Figure 6, it can be shown that the groundwater (X7) and the slope shape (X6) exhibit the strongest correlation, with a Pearson's correlation coefficient of 0.86. Furthermore, the slope type (X1) and the human factors (X10) also display a strong correlation, which is 0.608. Conversely, the remaining factors exhibit weak correlations, as evidenced by their lower values of Pearson's correlation coefficient. For a more direct representation of the distributions and correlations among the variables, Given the presence of strong correlations among certain factors, it becomes essential in subsequent chapters to employ the RF algorithm to identify the significant factors and their corresponding eigenvalues in the analysis of slope stability.

3.1.4. Determination of Influencing Factors on Slope Stability. The selection of high-quality and highly correlated datasets plays a crucial role in training model accuracy. To obtain a high-quality slope stability dataset, we employ the RF algorithm to calculate the importance score of 10 evaluation indices (153 sets of slope data) and rank the importance score. The out-of-bag (OOB) error is often employed as a measure of generalization error in RFs. In this paper, OOB is used to evaluate the performance of the RF algorithm. The influencing factors on slope stability with low importance scores can be eliminated, resulting in a high-quality dataset. In this study, the RF algorithm was trained by using the training set, and the training results are depicted in Figures 7(a) and 7(b).

As observed from the OOB error rate depicted in Figure 7(a), it can be inferred that the model exhibits a strong generalization ability. It stabilizes with an error of approximately 0.21% when the DT is 600 trees. The correlation coefficient (R^2) is 0.91 for the training set in Figure 7(b), demonstrating that the model is well-trained and meets the training requirements. Combined Figures 7(a) and 7(b) illustrate that the importance scores of the RF for the slope stability influencing factors have high confidence and meet the requirements.

The ranking of importance scores for the slope stability influencing factors was obtained through eigenvalue calculation employing the RF algorithm (as in Figure 7(c)), which are precipitation, slope gradient, slope height, dense degree of soil, weathering degree, vegetation cover, human factors, slope shape, groundwater, and slope type. In this paper, the six influencing factors with the top importance scores are selected as the dataset for slope stability prediction of the
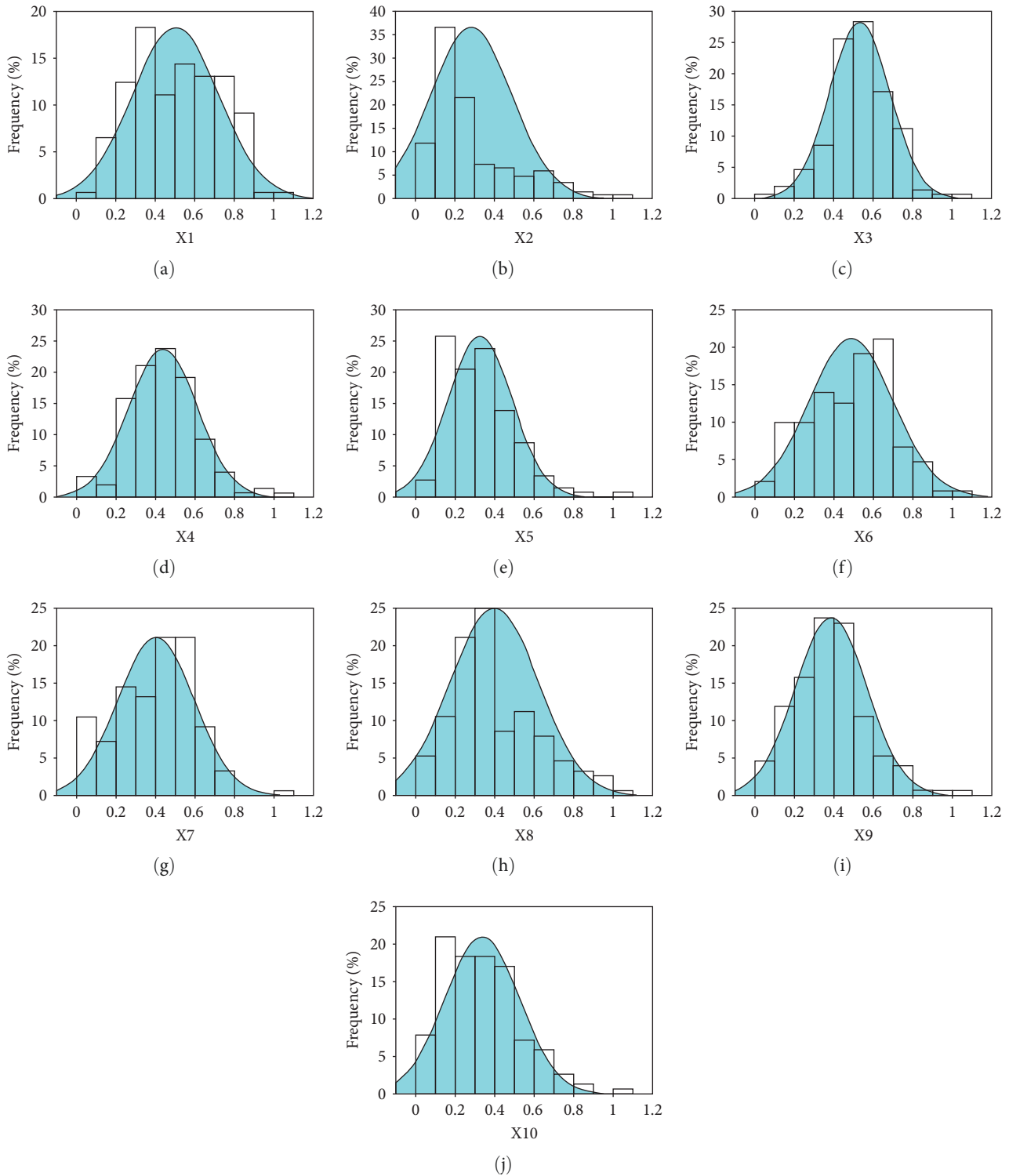
FIGURE 4: Distribution histogram of different indices: (a) slope type; (b) precipitation; (c) vegetation coverage; (d) slope height; (e) slope gradient; (f) slope shape; (g) groundwater; (h) weathering degree; (i) dense degree of soil; (j) human factors.

PSO–LSSVM model and the remaining four influencing factors are excluded. After the RF analysis, the dimension of the influencing factors is decreased from 10 to 6, eliminating the linear correlation present in the raw data.

3.2. Selection of the Kernel Function for the LSSVM Model. The accuracy of the LSSVM model depends not only on the quality of the selected input data but also on factors such as the penalty term (c) and the kernel function along with its
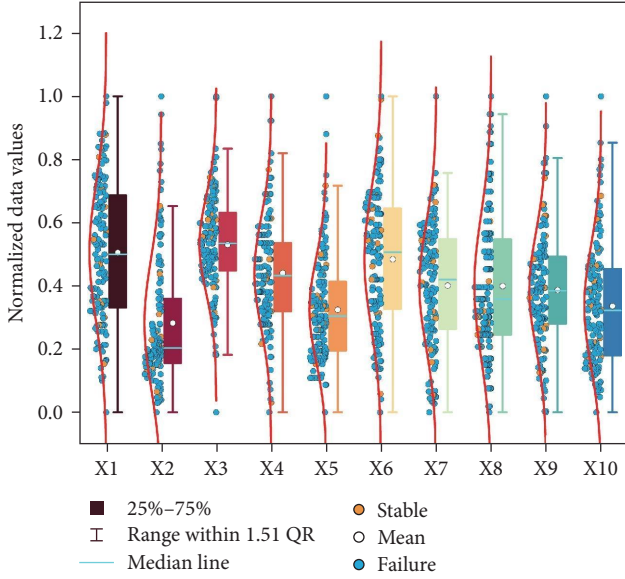
FIGURE 5: Box plots and normal distribution curves of different indices.

parameters within the LSSVM model. The LSSVM basic principle is as follows:

$$\begin{cases} \min J(\omega, \zeta) = \dfrac{1}{2}\|\omega\|^2 + \dfrac{1}{2}c\sum_{i=1}^{n}\zeta_i^2 \\ \text{subject to } y_i = \omega^T\varphi(x_i) + b + \zeta_i\, i = 1, 2, \cdots, n \end{cases}, \quad (3)$$

where $\|\omega\|$ is the normal constant of the hyperplane, b is a scalar basis, and c is a penalty term.

To succinctly expound upon the optimization principle governing kernel function parameter (σ) and penalty term (c) in the LSSVM model, we employ a simplified figure as depicted in Figure 8. For the multidimensional nonlinear multiclassification problems, the penalty term (c) in the LSSVM model is employed to determine the optimal decision boundary, aiming to effectively distinguish between the different samples to the greatest extent possible, namely, the location of the red dotted line in Figure 8(a)–8(c). The kernel function parameter (σ) determines the level of refinement of the decision boundary, namely, the complexity of the red dotted line in Figure 8(d)–8(f). Deviation from the optimal value, whether excessively large or too small, can significantly impact the accuracy of the LSSVM model. Thus, a suitable penalty term (c) and the kernel function are crucial for an LSSVM model. To assess the impact of kernel function selection on the accuracy of the LSSVM model, four distinct kernel functions are chosen for a comparison test while keeping other parameters consistent [45]. These kernel functions include linear, polynomial (poly), and Gaussian (RBF) kernel functions, and their formulas and kernel parameters are shown in Table 1.

The process is as follows: the data of each of the six evaluation indices are used as a single training dataset for

the model and input to the LSSVM model for learning. The LSSVM model's penalty factors and kernel function parameters were optimized by employing the PSO algorithm. In the PSO algorithm, the parameters were uniformly configured as follows: the swarm size was set to 20, the inertia weight coefficient was 1, the personal and social learning factors (c_1 and c_2) were 1.0 and 1.5, respectively, and the maximum number of iterations was limited to 200. To evaluate the accuracy of the LSSVM model with different kernel functions, the evaluation indicators employed are the mean absolute error R_{MAPE} (Equation (4)) and the correlation coefficient R^2 (Equation (5)). R_{MAPE} is a commonly used model evaluation metric that quantifies the magnitude of the error between predicted and actual values, reflecting the accuracy of the prediction model. R^2 indicates the correlation coefficient, with higher values indicating better regression performance and evaluation accuracy.

$$R_{\mathrm{MAPE}} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2}. \quad (5)$$

In Equations (4) and (5), y_i and $\hat{y}_i$ are the actual and predicted values in the test set data samples, respectively. n is the number of test set samples.

Figure 9 illustrates the comparison of the test results of the three selected kernel functions. The R_{MAPE} curves show that the linear kernel function has a high error for the six evaluation indicators, while the polynomial kernel function has a low error for rainfall and vegetation cover but a high error for the other evaluation indices. The RBF kernel function shows a low error for all six evaluation indices. Additionally, from the correlation coefficient R^2, it can show that the regression value of the RBF is closer to the actual value. Therefore, the RBF has been adopted as the kernel function for the LSSVM model in this study. Where σ denotes the width parameter.

3.3. Parameter Adjustment of Hyperparameters. The predictive accuracy of the LSSVM model is determined not only solely by the raw data but also largely depends on the hyperparameters of the optimization algorithm [46]. The particle size significantly impacts the PSO algorithm of suitability, identification accuracy, and convergence rate. The search stability and accuracy improve with increasing particle size, but the convergence rate decreases [45, 46]. There is currently no reliable method to accurately select the proper particle size. In this paper, the fitness function is employed to evaluate the optimal combination of parameter values for the LSSVM model. The definition of the fitness function is provided as follows:
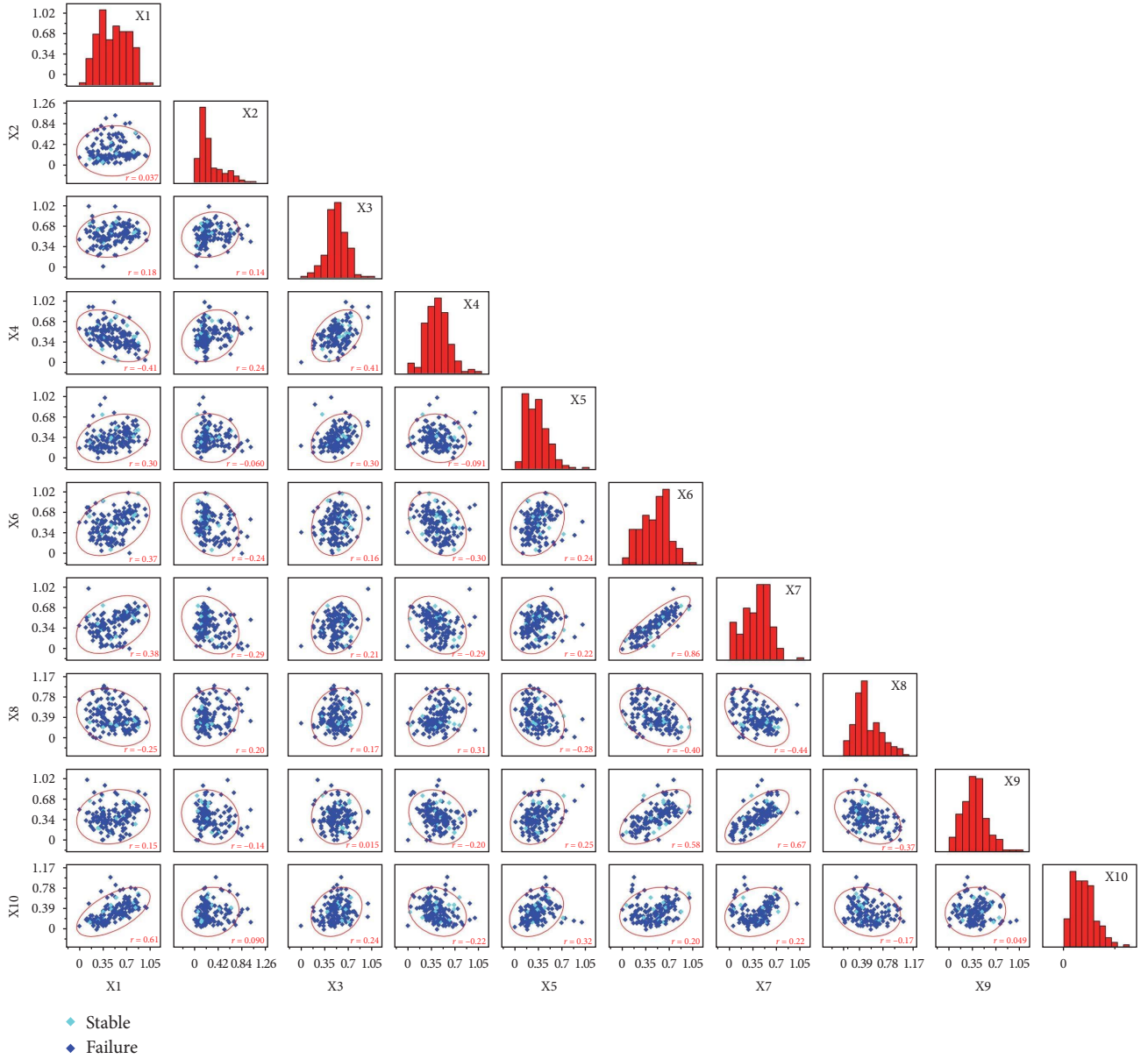
FIGURE 6: Correlation matrix.

$$f_{\text{Fitness}} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{F_i - F_i'}{F_i} \right|, \tag{6}$$

where F_i represents the actual slope stability degree, F_i' represents the validation slope stability degree, and n denotes the swarm size.

In this paper, we adopt particle sizes of 10, 20, 25, 30, 40, and 60 (as shown in Figure 10). All other hyperparameters of the PSO are the same, namely, the maximum number of iterations is 200, the inertia weight coefficient is 1, and the personal and social learning factors are 1.0 and 1.5, respectively. Additionally, the value of c is set to 2.0, σ is set to 4.0, and the selected kernel function is RBF. The results showed that the identification accuracy was 74% for the particle size of 10. Moreover, when the particle size exceeds 20, the identification accuracy has a slight variation. Furthermore, the particle size influences the average fitness, thereby affecting the optimal parameter combination. As shown in the figure below, the average fitness obtained from different particle sizes is different, while the highest average fitness is obtained when the particle size is 20, and the average fitness is, however, lower as the particle size increases. Therefore, the optimal particle size chosen in the RF–PSO–LSSVM model is 20.

3.4. Parameter Optimization of LSSVM Based on PSO. The study demonstrated the influence of the parameters c and σ of the RBF kernel function on the predictive accuracy of the LSSVM model to a certain extent [46]. The methods such as the enumeration method and the genetic algorithm are used to solve such problems. However, these methods exhibit
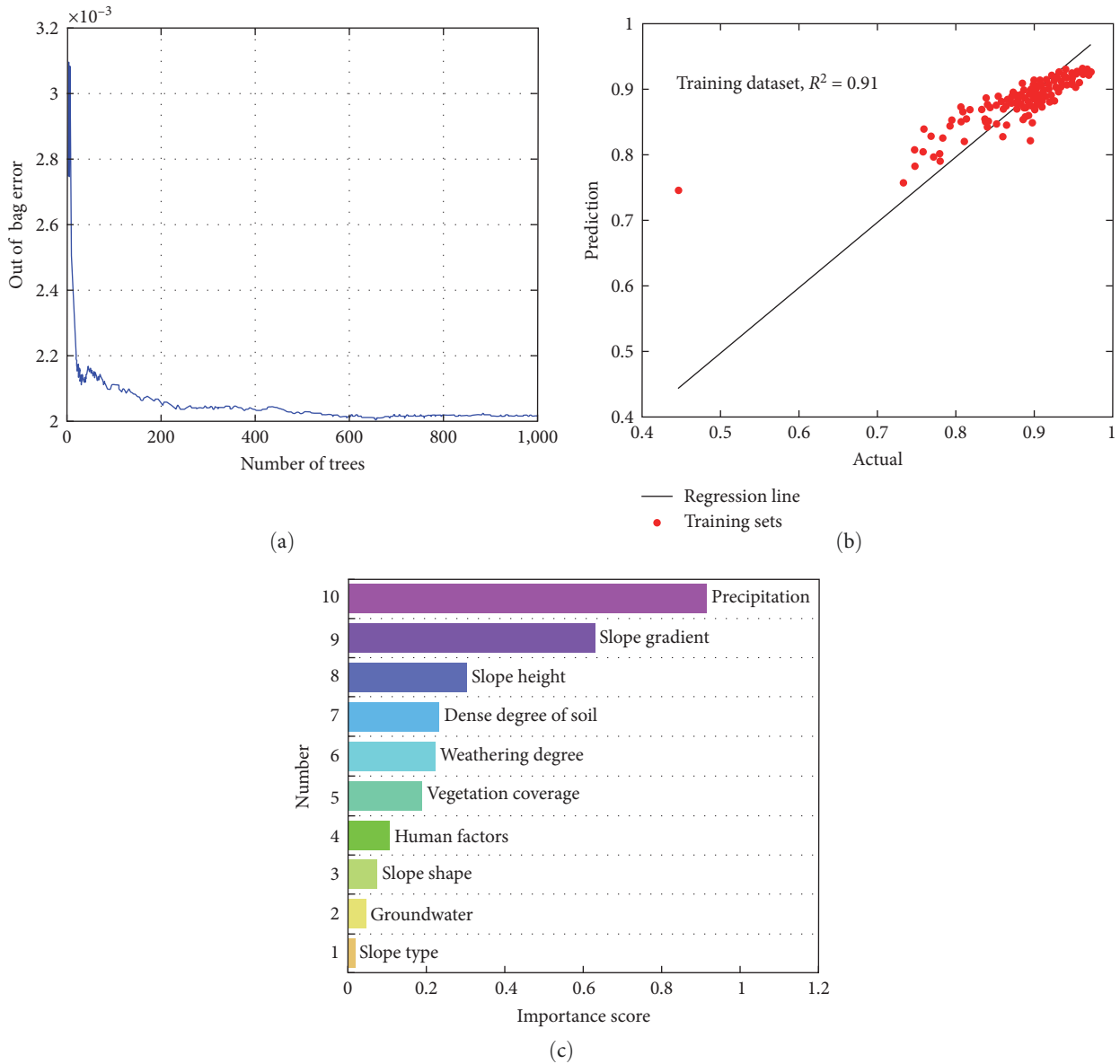
(a)

(b)

(c)

FIGURE 7: (a) Out-of-bag error of training results for random forest model. (b) Regression effect of random forest model. (c) Importance score of slope stability influencing factors.

some defects, such as computationally complex and unsolvable optimal solutions. Therefore, PSO is implemented to search for the optimal combination of parameter values for the LSSVM model. The following choices are made for the parameters of the model: the parameters c and σ are varied within the ranges of $(10^{-2}, 10^{-4})$ and $(10^{-1}, 10^{-5})$, respectively. Furthermore, the initial parameters of the PSO are as follows: the swarm size is set to 20, the inertia weighting factor (ω) is set to 0.6, and improvement ω according to Equation (1), personal and social learning factors c_1 and c_2 are 1.2 and 1.6, respectively, and the maximum number of iterations is 200.

The search results of PSO for the optimal combination of parameter values for the LSSVM model are $c = 2.89$, $\sigma = 4.89$, and the best fitness is 98%, while the average fitness is 90%, which is shown in Figure 11. This demonstrates that the RF–PSO–LSSVM model shows excellent performance in slope

stability discrimination. Therefore, the RF–PSO–LSSVM is feasible and effective for identifying slope stability in this study area.

3.5. Reliability Evaluation Indicators of the Prediction Model. Evaluating machine learning algorithms is crucial for addressing practical problems. It is only through accurate evaluation that algorithms can be optimized in later stages. The commonly employed evaluation indicators for prediction models include accuracy, precision, recall, the receiver operating characteristic (ROC) curve, and the area under the curve (AUC) value [44]. In this paper, these evaluation indicators are also utilized to evaluate and gauge the performance of the model. In Equations (7)–(12), TP represents true positive, FP denotes false positive, TN denotes true negative, and FN denotes false negative.
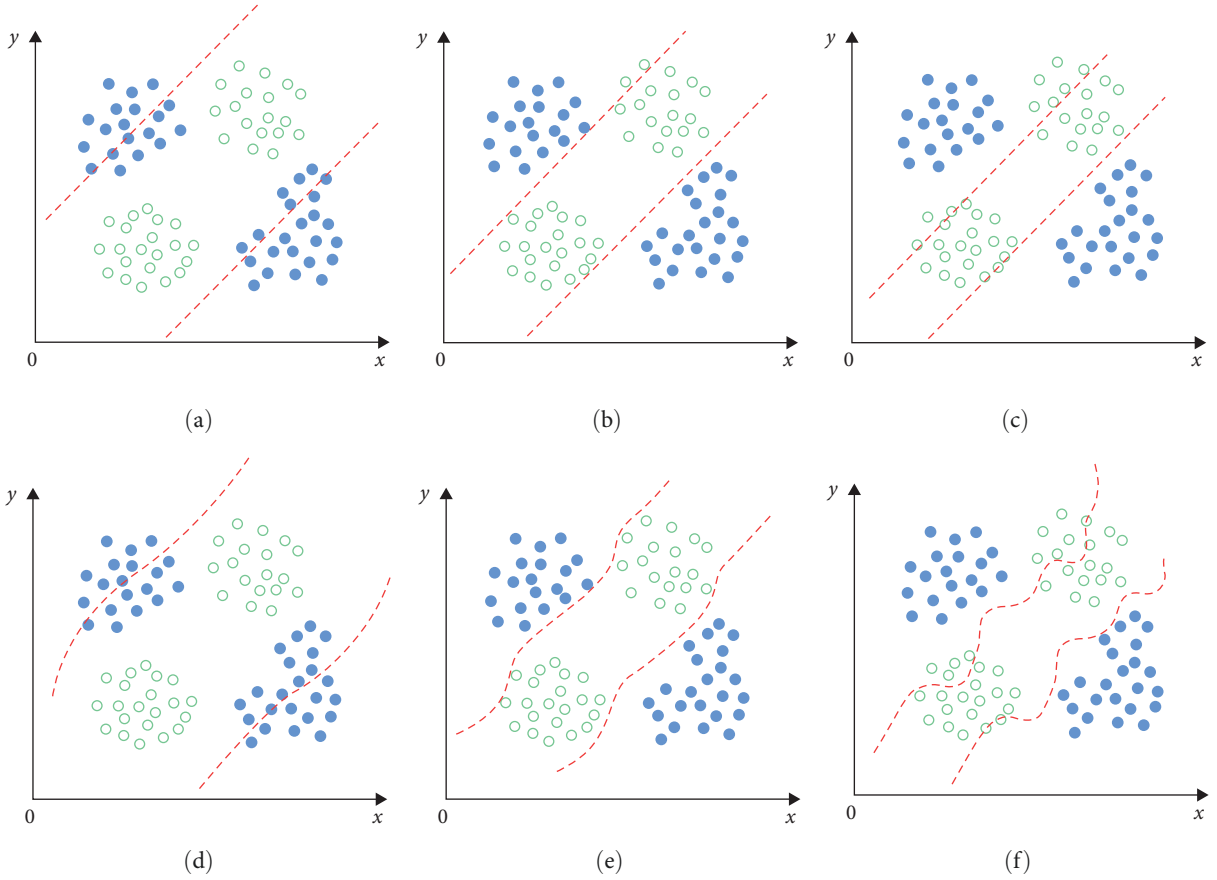
(a)

(b)

(c)

(d)

(e)

(f)

FIGURE 8: Penalty parameter (c) and kernel function parameter (σ) of the LSSVM model: (a) Underestimation of c value; (b) optimal c value; (c) overestimation of c value; (d) overestimation of σ value; (e) optimal σ value; (f) underestimation of σ value.

TABLE 1: Kernel functions of the LSSVM model and their parameters.

Kernel name	Kernel function	Kernel parameters
Linear	$K(x_i, x_j) = \langle x_i, x_j \rangle$	None
Poly	$K(x_i, x_j) = \left(\frac{1}{\sigma^2} \langle x_i, x_j \rangle + r \right)^d (\sigma > 0)$	d, r, σ
RBF	$K(x_i, x_j) = \exp\left(-\frac{1}{\sigma^2} \|x_i - x_j\|^2 \right) (\sigma > 0)$	σ

Equation (7) is employed to calculate the accuracy.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (7)$$

Precision is calculated by Equation (8).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (8)$$

Recall, which is the opposite of precision and can be calculated by Equation (9).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

The F-measure serves as a valuable metric for evaluating and comparing the predictive performance of two models. It

overcomes the limitations of solely considering precision or recall individually. Equation (10) provides the mathematical expression for calculating the F-measure.

$$F - \text{measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FN} + \text{FP}} \quad (10)$$

When constructing the ROC curve, the true positive rate (TPR) is denoted on the vertical axis, and the false positive rate (FPR) is represented on the horizontal axis. The AUC value, which represents the area under the ROC curve, serves as a valuable metric for comparing the predictive performance of two models. When evaluating prediction models, a higher AUC value indicates superior performance. TPR and FPR are calculated by Equations (11) and (12), respectively.
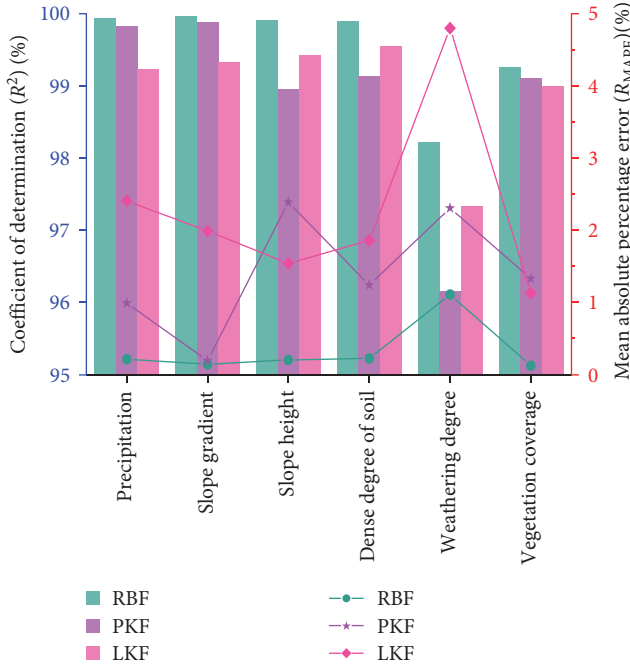
FIGURE 9: R_{MAPE} and R^2 results of LSSVM with different kernel functions. (The line chart represents the R_{MAPE} of different kernel functions and uses the right coordinate axis. The bar chart represents the R^2 of different kernel functions and uses the left coordinate axis. RBF = Gaussian kernel function, PKF = polynomial kernel function, LKF = linear kernel function).

$$TPR = \frac{TP}{TP + FN} \tag{11}$$

$$FPR = \frac{FP}{FP + TN} \tag{12}$$

3.6. Slope Stability Prediction Results and Model Performance Analysis. In the RF–PSO–LSSVM model, RF is employed to extract nonlinear feature information and resolve the linear correlations within the input data. PSO is implemented to optimize the selection of the penalty term (c) and the kernel function parameter (σ) of the LSSVM classifier in the solution space. The extracted new features and the optimal parameters for the LSSVM classifier are employed to predict the test data. Compared to other existing models, the proposed model could prevent blindness in data processing, more effectively select the primary influencing factors of slope stability from slope data, and improve prediction accuracy.

To verify the reliability of the RF–PSO–LSSVM model for slope stability identification, three models, the GA–SVM model, PSO–SVM, and PSO–LSSVM model, were constructed in this paper for comparison testing. The SVM and LSSVM models use RBF as their kernel function, and the maximum number of iterations and the swarm size in the GA algorithm are the same as the PSO algorithm. In this study, the 153 sets of slope data are applied to the GA–SVM, PSO–SVM, PSO–LSSVM, and RF–PSO–LSSVM models, respectively, each containing ten evaluation indices, of which sets 1–129 are employed as the training set to train the

model, and sets 130–153 are adopted as the testing set to verify the model, as shown in Table 2.

Table 3 lists the accuracy, F-measure, and AUC values obtained using Equations (6)–(11) for the four models on the training and testing sets. From the aspect of accuracy, both the PSO–LSSVM and RF–PSO–LSSVM models have 100% training accuracy. The RF–PSO–LSSVM model exhibits the best prediction performance, which is 96.55% in testing sets, while the PSO–LSSVM model also demonstrates good prediction results, with an accuracy of 89.66% in testing sets. These results demonstrate the strong generalization capability of the LSSVM. Accuracy serves as one of the fundamental indicators for evaluating the performance of a model. However, it may not be reliable when evaluating imbalanced samples. Therefore, to assess the performance of prediction models effectively, it is crucial to consider both the values of F-measure and AUC. As can be shown in Figure 12, the models of RF–PSO–LSSVM and PSO–LSSVM are closer to the upper left part of the coordinate axis, indicating their higher overall performance compared to other models. The AUC values of the two models in the testing sets are also the highest, which values of 0.964 and 0.944, respectively. The RF–PSO–LSSVM model achieved the best scores in testing sets for all three metrics: an accuracy of 95.82%, an F-measure of 96.55%, and an AUC of 0.964. Furthermore, the performance of the RF–PSO–LSSVM model remained consistent between the training and testing sets, which indicates that the model has a strong generalization ability without overfitting or underfitting. Based on the results presented in Table 3 and Figure 12, it is observed that the RF–PSO–LSSVM model exhibits the highest F-measure value and AUC value. Consequently, the RF–PSO–LSSVM model is considered to possess the best prediction performance among the other models evaluated.

By comparing the fitness curves of four models, GA–SVM, PSO–SVM, PSO–LSSVM, and RF–PSO–LSSVM (Figure 12), as can be shown that the RF–PSO–LSSVM is significantly better than the other three models in terms of algorithm convergence speed and accuracy. The RF–PSO–LSSVM model converged at six times of evolution, and the tested fitness values were above 92.5% overall. The GA–SVM, PSO–SVM, and PSO–LSSVM models were all clearly at local extremes, and the final maximum fitness obtained was 74.3%, 84.1%, and 85.5%, respectively. The average fitness of the RF–PSO–LSSVM model tends to have a stable fluctuation range after six evolutions. While GA–SVM, PSO–SVM, and PSO–LSSVM need to evolve: 120, 51, and 45 times, respectively, to reach this condition, the RF–PSO–LSSVM model outperforms the other three models in terms of convergence speed and accuracy. Hence, the RF–PSO–LSSVM model has been deemed the most effective approach for slope stability prediction.

By comparing the prediction results of four model test sets, which are shown in Figure 13, GA–SVM, PSO–SVM, PSO–LSSVM, and RF–PSO–LSSVM, it can be found that in the 24 test sets, the RF–PSO–LSSVM model had only one slope grade identification error, while the GA–SVM model, PSO–SVM model, and PSO–LSSVM model have 6, 4, and 3 identification errors, respectively. The RF–PSO–LSSVM
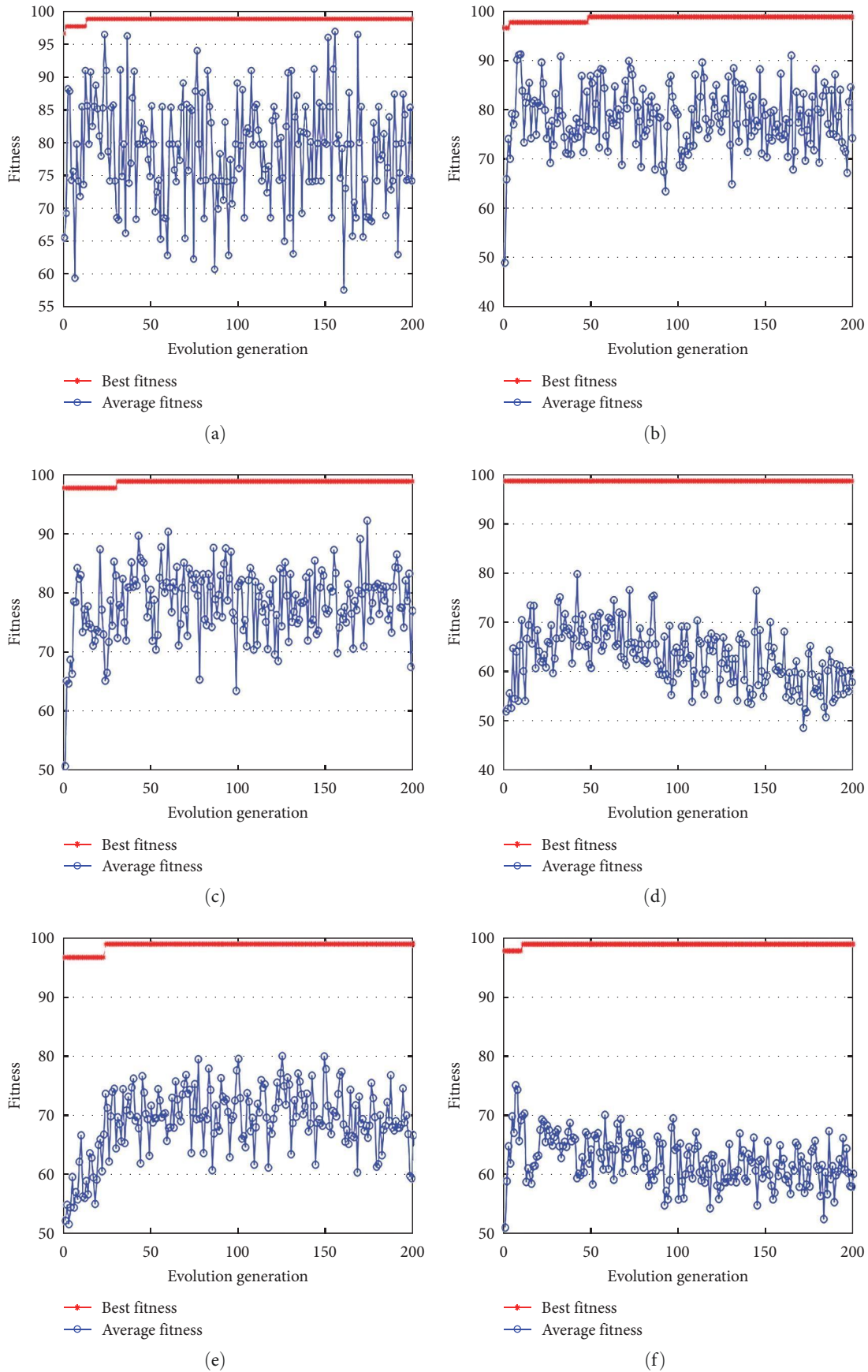
FIGURE 10: Fitness curves of different particle sizes by PSO: (a) particle size 10; (b) particle size 20; (c) particle size 25; (d) particle size 30; (e) particle size 40; (f) particle size 60.
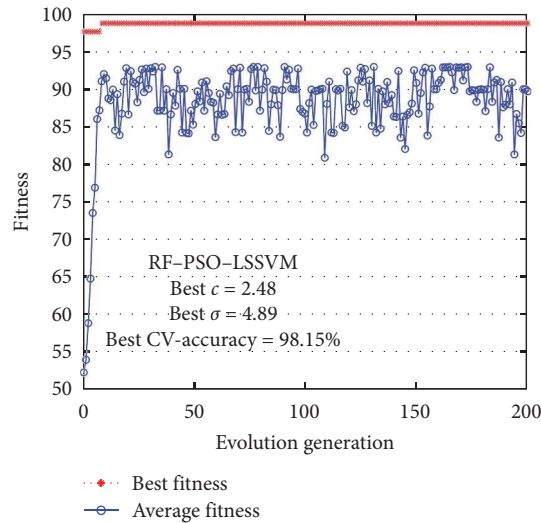
FIGURE 11: Fitness curves of the RF–PSO–LSSVM mode.

TABLE 2: Data from 24 test samples were used for the GA–SVM model, PSO–SVM, PSO–LSSVM, and RF-PSO-LSSVM models.

Sample	X1	X2	X3 (%)	X4	X5	X6	X7	X8	X9	X10	I1
1	0.81	640.1	88	38	41	0.61	0.12	0.12	0.13	0.61	1
2	0.64	244.3	35	20	35	0.87	0.64	0.31	0.68	0.32	1
3	0.32	414.1	71	87	29	0.35	0.34	0.10	0.48	0.13	0
4	0.21	542.1	54	43	32	0.45	0.25	0.36	0.24	0.51	0
5	0.35	504.4	24	56	18	0.67	0.34	0.23	0.47	0.34	0
6	0.42	612.3	41	31	25	0.51	0.42	0.15	0.52	0.42	0
7	0.29	462.1	48	38	29	0.45	0.38	0.29	0.63	0.48	1
8	0.33	524.4	53	28	21	0.37	0.29	0.27	0.39	0.34	0
9	0.35	601.9	18	42	36	0.56	0.31	0.32	0.45	0.53	1
10	0.24	445.1	82	36	28	0.42	0.36	0.31	0.58	0.44	0
11	0.41	536.8	59	49	58	0.31	0.35	0.39	0.48	0.33	1
12	0.40	472.5	42	29	35	0.26	0.47	0.28	0.34	0.21	1
13	0.34	509.1	35	27	39	0.41	0.25	0.43	0.58	0.29	0
14	0.28	374.7	68	23	27	0.79	0.58	0.49	0.67	0.21	0
15	0.34	492.5	33	45	41	0.39	0.32	0.50	0.51	0.56	0
16	0.43	489.3	20	42	38	0.47	0.14	0.46	0.31	0.43	0
17	0.78	379.5	17	25	28	0.63	0.27	0.62	0.18	0.38	1
18	0.53	403.6	59	38	34	0.81	0.70	0.67	0.58	0.82	0
19	0.37	602.4	23	54	29	0.47	0.39	0.43	0.38	0.52	0
20	0.24	535.2	29	48	38	0.32	0.44	0.38	0.28	0.41	0
21	0.21	505.2	21	32	25	0.22	0.49	0.18	0.18	0.11	0
22	0.16	127	10	17	32	0.42	0.13	0.56	0.37	0.31	0
23	0.93	387	41	128	27	0.83	0.68	0.86	0.10	0.59	1
24	0.41	444	51	45	35	0.91	0.70	0.65	0.63	0.81	1

exhibits superior identification accuracy compared to the other three models. Therefore, the RF–PSO–LSSVM model is the best in slope stability prediction.

4. Discussion

4.1. Comparison of Model Identification Results with Field Data. The training and prediction results of the RF–PSO–LSSVM model are presented in Table 3 and Figure 13(d). Compared

with the field survey results, it is evident that the RF–PSO–LSSVM model demonstrates a high level of prediction accuracy, achieving an accuracy of 95.7%. Only no. 10 (group 140) slope exhibits ordinary correspondence; namely, the stable ("1") is classified as failure ("0").

The determination of the slope stability condition is based on field investigations, which include evaluating factors such as the presence of slope fissures, accumulation at the foot of the slope, and whether it has collapsed. The

TABLE 3: Comparison of the performances of the different algorithms on the training and testing sets.
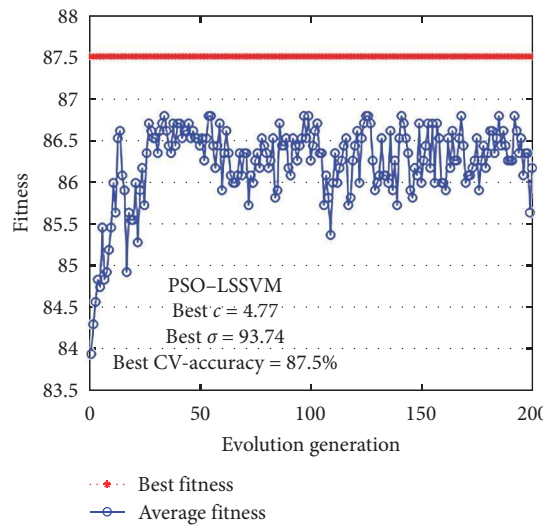
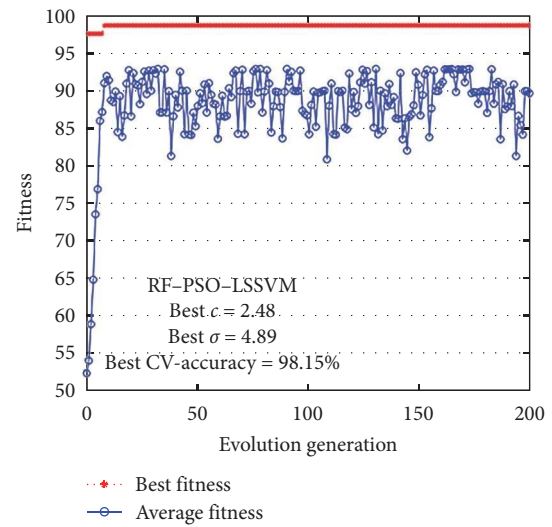| Dataset	Model	Model parameters		Accuracy (%)	F-measure (%)	AUC
		c	σ			
Training	GA–SVM	0.1	0.01	93.74	94.12	0.937
	PSO–SVM	0.1	14.46	95.12	94.85	0.941
	PSO–LSSVM	4.77	93.74	100	100	1.00
	RF–PSO–LSSVM	2.48	4.89	100	100	1.00
Testing	GA–SVM	0.1	0.01	70.83	75.86	0.865
	PSO–SVM	0.1	14.46	83.33	86.67	0.887
	PSO–LSSVM	4.77	93.74	87.50	89.66	0.944
	RF–PSO–LSSVM	2.48	4.89	95.82	96.55	0.964
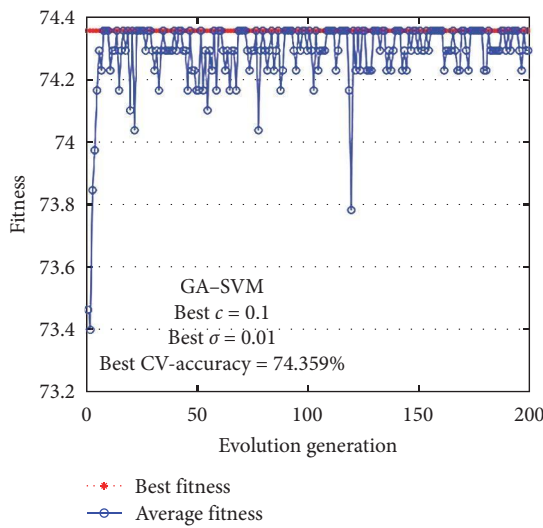
(a)
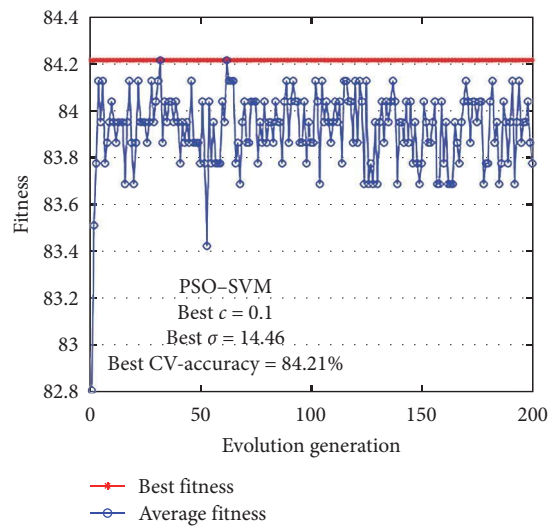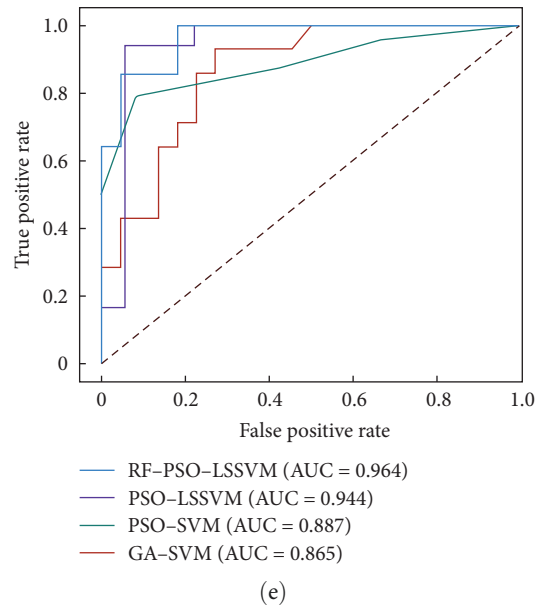
(b)

(c)

(d)

FIGURE 12: Continued.

(e)

FIGURE 12: Fitness curves of the four models: (a) GA–SVM model; (b) PSO–SVM3 model; (c) PSO–LSSVM model; (d) RF–PSO–LSSVM model; (e) ROC curves and AUC values of the four models.
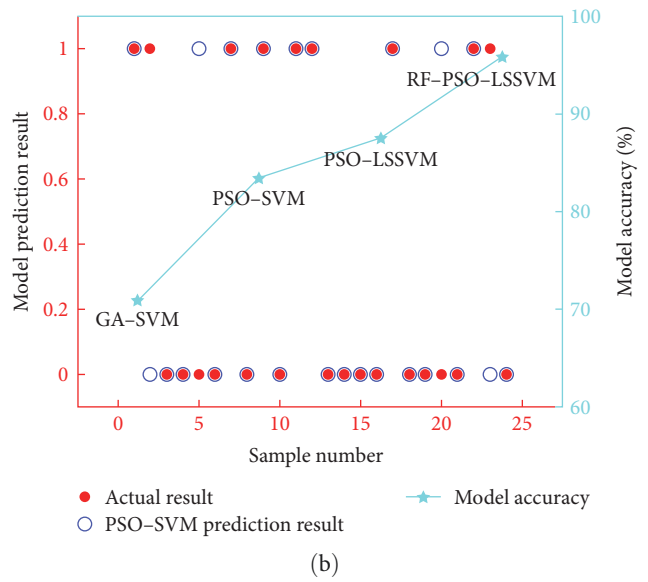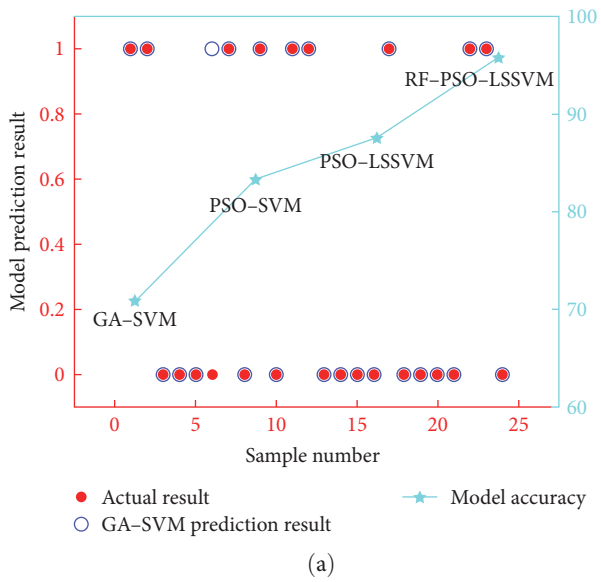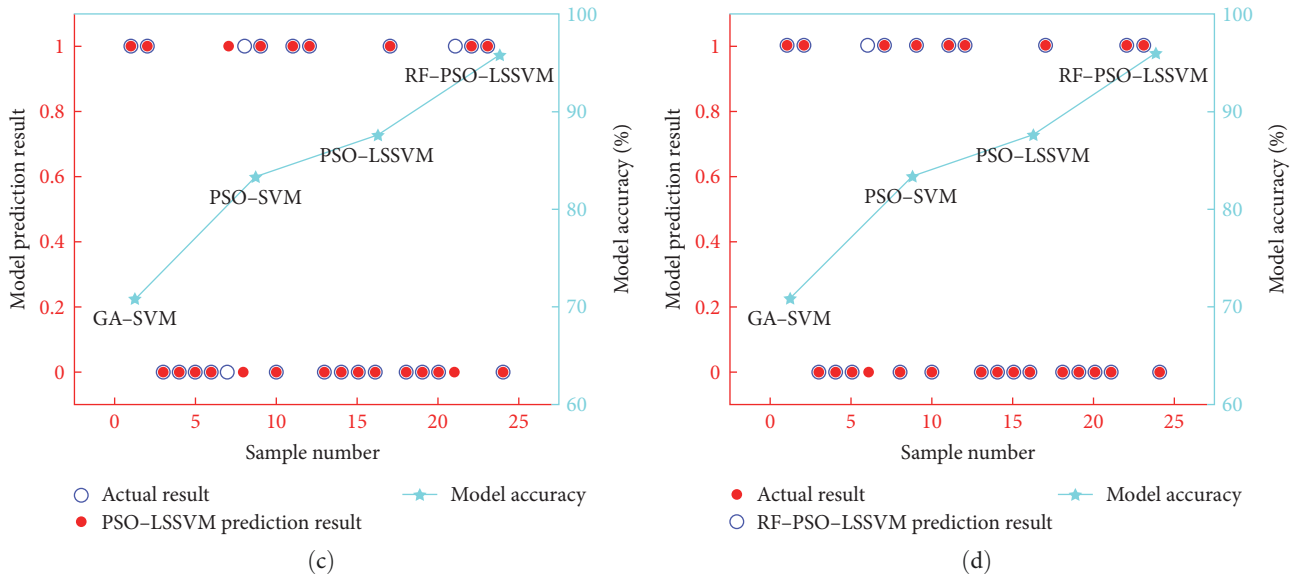

FIGURE 13: Continued.

FIGURE 13: Test set identification results of four models: (a) GA–SVM model; (b) PSO–SVM model; (c) PSO–LSSVM model; (d) RF–PSO–LSSVM model.

selection of the 10 influencing factors in this study may reflect slope stability. Proper selection and discrimination could reflect a given influencing factor of slope stability to a certain extent. However, in practice investigation, field investigation data is often processed using average processing methods [48]. Moreover, some infrequently noted local field data are ignored, for example, the depth of slope fissures and the presence of retaining walls, which leads to errors in identifying slope stability situations due to subjective factors. In the following paragraphs, slope no. 10 is analyzed to determine the causes of such inaccuracies.

The slope gradient of slope no. 10 is 42°, the slope height is 22 m, the slope width is 114 m, the slope length is 22 m, the slope shape is flat and straight, the main composition of the slope body is gravel soil, the dense degree of soil is medium, and the vegetation coverage is relatively high. Based on the site survey, the slope body exhibits the following deformation characteristics: there is a landslide scarp near the top of the slope, spanning the width of the slope. In the middle position, localized sliding is observed, and the trees on the slope are leaning or tilted. The above situation indicates that the stability of slope no. 10 should be a failure, that is, "0." Compared with the field investigations conducted in May 2020 and October 2019, it can be observed that there is no significant deformation in slope no. 10 (Figure 14). Therefore, the slope remained relative stability throughout this period of almost a year, that is "1". Slope no. 10 had experienced a collapse before the field survey, resulting in the recorded slope data representing the postcollapse condition rather than the precollapse state, to lead to the slope data being inconsistent with slope stability. Therefore, this may be the reason for the error in model identification.

4.2. Model Applicability Validation Analysis. In this study, we have chosen 26 groups of data from the slope data provided by Lin et al. [49] to verify the suitability and applicability of the RF–PSO–LSSVM model in other studies areas, as shown in Table 4. In Table 4, r presents gravity, C presents cohesion, φ denotes internal friction angle, β presents slope angle, H presents slope height, and r_u presents pore water pressure ratio.

By employing the RF to calculate the variable importance of the six influencing factors of slope stability in Table 4. The weight of each indicator is depicted in Figure 15(a); it is evident from the results that gravity exerts the most significant influence on slope stability, followed by cohesion, while the pore water ratio exhibits the least significance, which is 0.0014. This conclusion aligns with the sensitivity analysis conducted on various factors during slope stability analysis using the limit equilibrium method. Therefore, excluding the influencing factor of pore water ratio, the data of the other five influencing factors are implemented to validate the trained PSO–LSSVM model.

To further verify the accuracy and reliability of the RF–PSO–LSSVM model, compared and analyzed its prediction performance with other models, including SVM, logistic regression (LR), DT, k-nearest neighbor, naive Bayes, and linear discriminant analysis (LDA). The outputs of all models, which is "1" indicates stable status, and "0" indicates failure status. The predictive performance of the seven prediction models is illustrated in Figure 15(b). Through the comparison, it is evident that the RF–PSO–LSSVM model proposed in this study exhibits outperforms other models in predicting slope stability, and the worst-performing method is the LDA. Furthermore, it can also be seen that the RF–PSO–LSSVM, SVM, and DT models have relatively stable performance, as their values of accuracy, F-measure, and AUC show slight discrepancies. Therefore, in this study, we have demonstrated the feasibility of employing the RF–PSO–LSSVM model for identifying the slope stability of the Sichuan–Tibet Highway and its potential for generalization to slopes in other geological backgrounds. Practical engineering applications have shown that the proposed

FIGURE 14: Actual field view of slope no. 10: (a) and (b) are the field survey of slope no. 10 in October 2019; (c) and (d) refer to the field survey of slope no. 10 reviewed in May 2020.

TABLE 4: Slope stability prediction database.

| No.	r (kN/m³)	C (kPa)	φ (°)	β (°)	H (m)	r_u	Status	Location
1	20.41	33.52	11	16	10.67	0.35	Stable	Seven Sisters Landslide, UK
2	21.82	8.62	32	28	12.8	0.49	Failure	Selset Landslide, Yorkshire, UK
3	20.41	33.52	11	16	45.72	0.2	Failure	Saskatchewan Dam, Canada
4	18.84	14.36	25	20	30.5	0.45	Failure	Sudbury Slide, UK
5	22	29	15	18	400	0	Failure	Qing River Area Landslide, China
6	23	24	19.8	23	380	0	Failure	Qing River Area Landslide, China
7	20	8	20	10	10	0	Failure	Slope in Tailie Elementary School
8	27	32	33	42.6	301	0.29	Failure	Heishe Landslide, China
9	27.3	10	39	41	511	0.29	Stable	Shijiapo Landslide, China
10	27.3	10	39	40	470	0.29	Stable	Tanggudong Landslide, China
11	25	46	35	47	443	0.29	Stable	Tianbao Landslide, China
12	18.8	8	26	21.8	40	0.4	Failure	Lingli Reservoir, China
13	27	27.3	29.1	21	565	0.26	Failure	Guzhang Gaofeng Slope, China
14	20.45	16	15	30	36	0.25	Stable	Baijiagou Earth Slope, China
15	22.4	10	35	45	10	0.4	Failure	Rockfill Slope
16	20.5	6.5	12.5	42	70	0	Failure	Yuan–Mo Highway
17	21.4	7.1	16.7	44	70	1	Failure	Yuan–Mo Highway
18	21.5	9.5	11.5	40	75	0	Failure	Yuan–Mo Highway
19	20.6	6.7	9.4	45	30	0	Failure	Yuan–Mo Highway
20	21.96	34.77	14.15	28	60	0	Stable	Yuan–Mo Highway
21	22.93	32.33	19.73	30	50	1	Stable	Yuan–Mo Highway
22	19.18	12.8	9.45	45	20	0	Failure	Yuan–Mo Highway
23	17.8	22.2	6.05	40	51.6	1	Failure	Yuan–Mo Highway
24	19.6	15.53	15.88	35	97	1	Failure	Yuan–Mo Highway
25	19.81	33.75	19.46	20	120	1	Stable	Yuan–Mo Highway
26	17.96	24.01	28	40	60	1	Failure	Yuan–Mo Highway
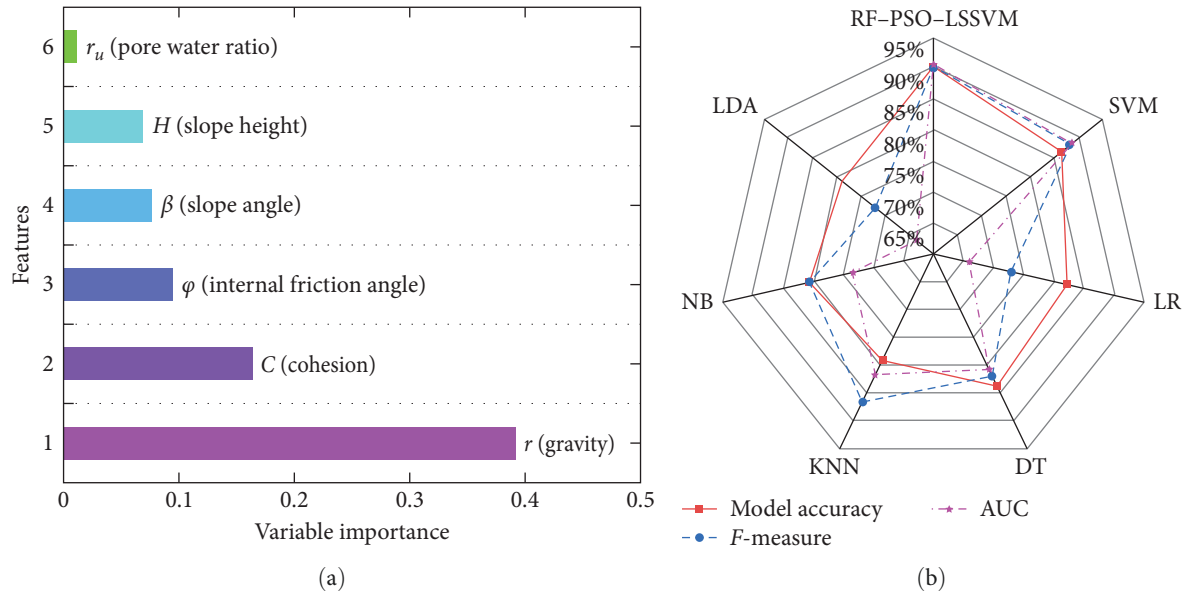
FIGURE 15: (a) Variable importance plot by RF algorithm. (b) Prediction efficiency of different models.

model could effectively reduce the risk of engineering accidents and has instructional importance for identifying slope stability. However, it cannot be considered the standard for assessing slope stability.

From a statistical standpoint, the stability states of slopes, namely stable and failure, are not absolute and can transform between states as the slope evolves. Consequently, a previous assessment of slope stability as stable may not be entirely inaccurate, but this result is of informative value. For slopes exhibiting failure of stability, it is crucial to conduct field investigations to accurately determine their stability status.

5. Conclusions

In this study, a hybrid model of RF–PSO–LSSVM was proposed to solve the low accuracy of the model and blind data preprocessing in the slope stability study. We investigate the slopes located along the Sichuan–Tibet highway in China as our research example. The processed data demonstrated scientific soundness and predictability, and conformed to a normal distribution through data preprocessing and exploratory data analysis of the slope data. Through a comparative analysis between the prediction outcomes of our proposed model and the GA–SVM, PSO–SVM, and PSO–LSSVM models, we validate the effectiveness of our approach. The values of the RF–PSO–LSSVM model for the best fitness, AUC, T-measure, and accuracy are 98.15%, 96.4%, 96.55%, and 95.82%, respectively. To demonstrate the practicality of the model, we employed 26 diverse sets of slope data obtained from various regions. The research results show the RF–PSO–LSSVM model still exhibits high accuracy. Thus, the proposed model could become a practical tool for predicting slope stability in limited samples in the future.

It is worth noting that there are other qualitative factors, such as existing joints and pore pressure, which also have significant effects on slope stability. Transforming these qualitative factors into quantitative ones presents a greater challenge. Thus, the future focus and difficulty in further studies lie in selecting more objective and reasonable indicators for evaluating slope stability.

Data Availability

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

Conflicts of Interest

The authors declare that there is no conflict of interest.

Acknowledgments

Supplementary Materials

The slope data are divided into two categories: quantitative data and qualitative data. The data are shown in Table S1. The quantitation and standardization of influencing factors of slope stability of the Sichuan–Tibet Highway are shown in Table S2. The influencing factors become dimensionless values by quantifying, as shown in Table S3. The evaluation indices were made dimensionless by normalization, as shown in Table S4. Table S1: partial sample original data. Table S2: slope stability evaluation indices quantitative classification table of Sichuan–Tibet Highway. Table S3: quantified slope sample data. Table S4: part of slope data after normalization. (*Supplementary Materials*)

References

[1] H. B. Wang, W. Y. Xu, and R. C. Xu, "Slope stability evaluation using back propagation neural networks," *Engineering Geology*, vol. 80, no. 3-4, pp. 302–315, 2005.

[2] D. Petley, "Global patterns of loss of life from landslides," *Geology*, vol. 40, no. 10, pp. 927–930, 2012.

[3] B. Wen, S. Wang, E. Wang, and J. Zhang, "Characteristics of rapid giant landslides in China," *Landslides*, vol. 1, pp. 247–216, 2004.

[4] Z.-W. Li, X.-L. Yang, and T.-Z. Li, "Static and seismic stability assessment of 3D slopes with cracks," *Engineering Geology*, vol. 265, Article ID 105450, 2020.

[5] S. Qin, J. J. Jiao, and S. Wang, "A nonlinear dynamical model of landslide evolution," *Geomorphology*, vol. 43, no. 1-2, pp. 77–85, 2002.

[6] H. Chen, Z. Zeng, and H. Tang, "Landslide deformation prediction based on recurrent neural network," *Neural Processing Letters*, vol. 41, pp. 169–178, 2015.

[7] W. Fellenius, "Calculation of stability of earth dams," in *Proceedings of the Transactions 2nd Congress on large dams*, pp. 445–462, Government Printing Office, America, Washington D.C, 1936.

[8] A. W. Bishop, "The use of the slip circle in the stability analysis of slopes," *Géotechnique*, vol. 5, no. 1, pp. 7–17, 1955.

[9] A. W. Bishop and N. Morgenstern, "Stability coefficients for earth slopes," *Géotechnique*, vol. 10, no. 4, pp. 129–153, 1960.

[10] N. R. Morgenstern and V. E. Price, "The analysis of the stability of general slip surfaces," *Géotechnique*, vol. 15, no. 1, pp. 79–93, 1965.

[11] N. Janbu, "Soil compressibility as determined by oedometer and triaxial tests," *Proceedings of European Conference on Soil Mechanics and Foundation Engineering*, vol. 1, pp. 19–25, 1963.

[12] P. Lu and M. S. Rosenbaum, "Artificial neural networks and grey systems for the prediction of slope stability," *Natural Hazards*, vol. 30, pp. 383–398, 2003.

[13] J. Kumar and P. Samui, "Stability determination for layered soil slopes using the upper bound limit analysis," *Geotechnical & Geological Engineering*, vol. 24, pp. 1803–1819, 2006.

[14] M. G. Sakellariou and M. D. Ferentinou, "A study of slope stability prediction using neural networks," *Geotechnical & Geological Engineering*, vol. 23, pp. 419–445, 2005.

[15] J. Du, K. Yin, and S. Lacasse, "Displacement prediction in colluvial landslides, Three Gorges Reservoir, China," *Landslides*, vol. 10, pp. 203–218, 2013.

[16] W. Zhang, R. Xiao, B. Shi, H.-H. Zhu, and Y.-J. Sun, "Forecasting slope deformation field using correlated grey model updated with time correction factor and background value optimization," *Engineering Geology*, vol. 260, Article ID 105215, 2019.

[17] V. N. Vapnik, "Support vector (SV) machines," in *The Nature of Statistical Learning Theory*, pp. 138–146, Springer, Verlag, New York, America, 2nd edition, 1995.

[18] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[19] P. Samui, "Slope stability analysis: a support vector machine approach," *Engineering Geology*, vol. 56, pp. 255–267, 2008.

[20] H.-B. Zhao, "Slope reliability analysis using a support vector machine," *Computers and Geotechnics*, vol. 35, no. 3, pp. 459–467, 2008.

[21] Z. Liu, J. Shao, W. Xu, H. Chen, and C. Shi, "Comparison on landslide nonlinear displacement analysis and prediction with computational intelligence approaches," *Landslides*, vol. 11, pp. 889–896, 2014.

[22] X. Z. Li and J. M. 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.

[23] 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.

[24] J. Ma, D. Xia, Y. Wang et al., "A comprehensive comparison among metaheuristics (MHs) for geohazard modeling using machine learning: insights from a case study of landslide displacement prediction," *Engineering Applications of Artificial Intelligence*, vol. 114, Article ID 105150, 2022.

[25] D. Xia, H. Tang, S. Sun, C. Tang, and B. Zhang, "Landslide susceptibility mapping based on the germinal center optimization algorithm and support vector classification," *Remote Sensing*, vol. 14, no. 11, Article ID 2707, 2022.

[26] C. Zhou, K. Yin, Y. Cao, and B. Ahmed, "Application of time series analysis and PSO–SVM model in predicting the Bazimen landslide in the Three Gorges Reservoir, China," *Engineering Geology*, vol. 204, pp. 108–120, 2016.

[27] F. Miao, Y. Wu, Y. Xie, and Y. Li, "Prediction of landslide displacement with step-like behavior based on multialgorithm optimization and a support vector regression model," *Landslides*, vol. 15, pp. 475–488, 2018.

[28] O. Kisi and K. S. Parmar, "Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution," *Journal of Hydrology*, vol. 534, pp. 104–112, 2016.

[29] X. Qian, J.-P. Chen, L.-J. Xiang, W. Zhang, and C.-C. Niu, "A novel hybrid KPCA and SVM with PSO model for identifying debris flow hazard degree: a case study in Southwest China," *Environmental Earth Sciences*, vol. 75, Article ID 991, 2016.

[30] C. Zhao, C. Zheng, M. Zhao, Y. Tu, and J. Liu, "Multivariate autoregressive models and kernel learning algorithms for classifying driving mental fatigue based on electroencephalographic," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1859–1865, 2011.

[31] C.-M. Ye, R.-L. Wei, Y.-G. Ge, Y. Li, J. M. Junior, and J. Li, "GIS-based spatial prediction of landslide using road factors and random forest for Sichuan–Tibet Highway," *Journal of Mountain Science*, vol. 19, pp. 461–476, 2022.

[32] S. Ma, C. Xu, and X. Shao, "Spatial prediction strategy for landslides triggered by large earthquakes oriented to emergency response, mid-term resettlement and later reconstruction," *International Journal of Disaster Risk Reduction*, vol. 43, Article ID 101362, 2020.

[33] E. Jamalinia, P. J. Vardon, and S. C. Steele-Dunne, "The impact of evaporation induced cracks and precipitation on temporal slope stability," *Computers and Geotechnics*, vol. 122, Article ID 103506, 2020.

[34] M. T. Löbmann, C. Geitner, C. Wellstein, and S. Zerbe, "The influence of herbaceous vegetation on slope stability—a review," *Earth-Science Reviews*, vol. 209, Article ID 103328, 2020.

[35] H. H. Bui, R. Fukagawa, K. Sako, and J. C. Wells, "Slope stability analysis and discontinuous slope failure simulation by elasto-plastic smoothed particle hydrodynamics (SPH)," *Géotechnique*, vol. 61, no. 7, pp. 565–574, 2011.

[36] J. Zhou, X. Li, and X. Shi, "Long-term prediction model of rockburst in underground openings using heuristic algorithms and support vector machines," *Safety Science*, vol. 50, no. 4, pp. 629–644, 2012.

[37] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[38] X. Hu, S. Wu, G. Zhang et al., "Landslide displacement prediction using kinematics-based random forests method: a case study in Jinping Reservoir Area, China," *Engineering Geology*, vol. 283, Article ID 105975, 2021.

[39] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, Springer, New York, America, 2013.

[40] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, pp. 293–300, 1999.

[41] T. Wen, H. Tang, Y. Wang, C. Lin, and C. Xiong, "Landslide displacement prediction using the GA–LSSVM model and time series analysis: a case study of Three Gorges Reservoir, China," *Natural Hazards and Earth System Sciences*, vol. 17, no. 12, pp. 2181–2198, 2017.

[42] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95-International Conference on Neural Networks*, pp. 1942–1948, IEEE, Perth, WA, Australia, 1995.

[43] J. Kennedy, "Particle swarm optimization," in *Encyclopedia of Machine Learning*, pp. 760–766, Springer, US, 2011.

[44] G. Wang, B. Zhao, B. Wu, C. Zhang, and W. Liu, "Intelligent prediction of slope stability based on visual exploratory data analysis of 77 in situ cases," *International Journal of Mining Science and Technology*, vol. 33, no. 1, pp. 47–59, 2023.

[45] M. Khandelwal, "Evaluation and prediction of blast-induced ground vibration using support vector machine," *International Journal of Rock Mechanics and Mining Sciences*, vol. 47, no. 3, pp. 509–516, 2010.

[46] C.-L. Huang and J.-F. Dun, "A distributed PSO–SVM hybrid system with feature selection and parameter optimization," *Applied Soft Computing*, vol. 8, no. 4, pp. 1381–1391, 2008.

[47] N. Dong, C.-H. Wu, W.-H. Ip, Z.-Q. Chen, C.-Y. Chan, and K.-L. Yung, "An opposition-based chaotic GA/PSO hybrid algorithm and its application in circle detection," *Computers & Mathematics with Applications*, vol. 64, no. 6, pp. 1886–1902, 2012.

[48] M. Liu, Y. He, J. Wang, H. P. Lee, and Y. Liang, "Hybrid intelligent algorithm and its application in geological hazard risk assessment," *Neurocomputing*, vol. 149, pp. 847–853, 2015.

[49] 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, pp. 1477–1502, 2022.