



Slope Stability Analysis Based on Machine Learning Optimized with Sparrow Search

Fangfang Zhou¹, Jing Wang², Rongfang Yan^{1,3} and Jiandong Zhang^{1,3,*}

¹ College of Mathematics and Statistics, Northwest Normal University, Lanzhou 730070, China

² College of Artificial Intelligence and Computer Science, Northwest Normal University, Lanzhou 730070, China

³ Gansu Provincial Research Center for Basic Disciplines of Mathematics and Statistics, Lanzhou 730070, China

Abstract

Slope stability prediction is a critical challenge in geotechnical engineering. This challenge arises from the complex nonlinear interactions among multiple influencing factors such as soil properties, slope geometry, and hydraulic conditions. Traditional analytical methods rely on simplified assumptions that fail to capture these complex relationships. To address these limitations, this study proposes a hybrid predictive framework that integrates machine learning models with intelligent optimization algorithms. Six key parameters, including unit weight, cohesion, internal friction angle, slope angle, slope height, and pore water pressure ratio, are selected as input features. Three machine learning models, namely Random Forest, XGBoost, and Support Vector Machine, are optimized using four intelligent optimization algorithms, including the Sparrow Search Algorithm, Lion Swarm Optimization, Whale Optimization Algorithm, and Genetic Algorithm, resulting in twelve combined

configurations. Among all combinations, the Sparrow Search Algorithm-optimized Random Forest model achieves the best performance, with an accuracy of 95.56%, precision of 95.24%, recall of 95.24%, F1 score of 95.24%, and AUC value of 0.9850. Validation using ten independent engineering cases confirms that the predictions of this model are consistent with actual slope conditions, demonstrating strong engineering applicability. This study provides a systematic comparative framework for hybrid model selection and offers practical guidance for intelligent slope stability assessment.

Keywords: Slope stability prediction, Machine learning, Intelligent optimization algorithms, SSA-RF.

1 Introduction

Slope stability is a fundamental concern in geotechnical engineering, particularly as mining activities and infrastructure construction continue to intensify worldwide. Once slope instability occurs, it poses serious threats to engineering safety, human lives, property, and the ecological environment. Therefore, accurate and reliable prediction of slope stability is a critical prerequisite for ensuring long-term safety in geotechnical engineering projects.

Traditional slope stability analysis methods, including



Submitted: 24 July 2026

Revised: 07 September 2026

Accepted: 08 September 2026

Published: 25 October 2026

Vol. 2, No. 4, 2026.

10.62762/TSSR.2026.749493

*Corresponding author:

✉ Jiandong Zhang

jdzhang@nwnu.edu.cn

Citation

Zhou, F., Wang, J., Yan, R., & Zhang, J. (2026). Slope Stability Analysis Based on Machine Learning Optimized with Sparrow Search. *ICCK Transactions on Systems Safety and Reliability*, 2(4), 235–249.

© 2026 ICCK (Institute of Central Computation and Knowledge)

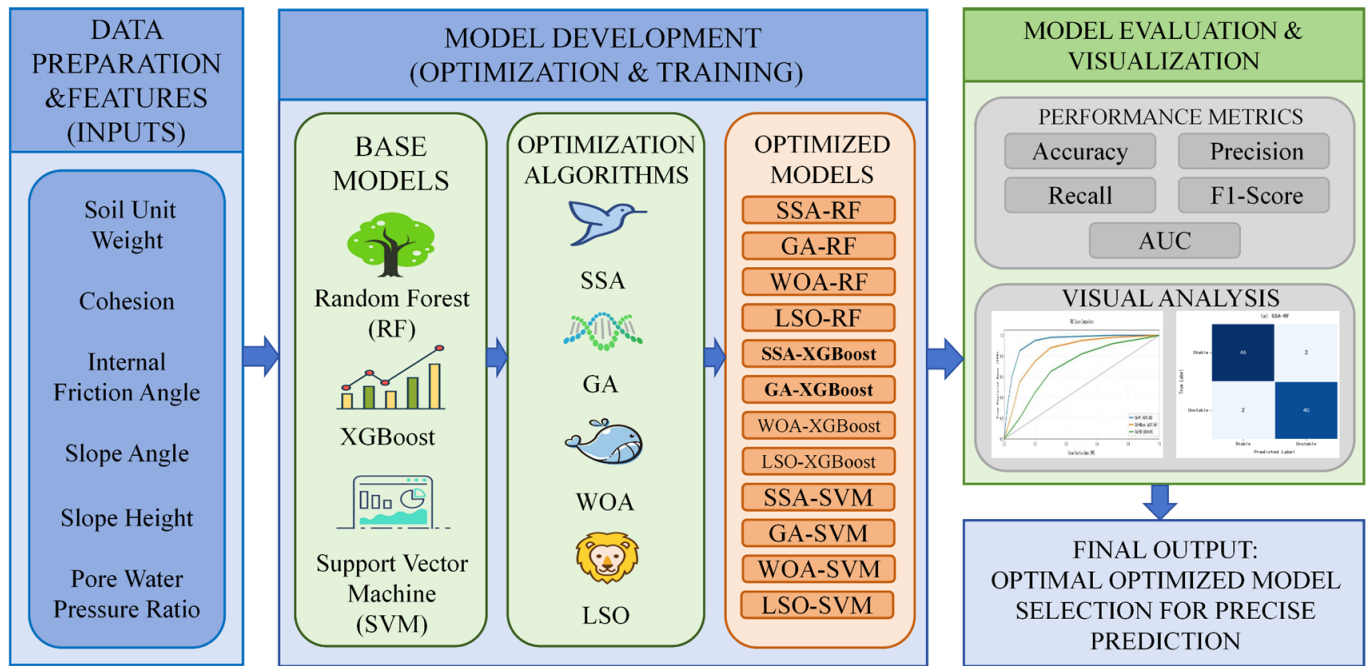


Figure 1. Research framework diagram.

the limit equilibrium method and numerical simulation, have been widely used in practice. However, these approaches exhibit significant limitations when dealing with complex, nonlinear, and multi-factor interactive systems. The limit equilibrium method requires a predefined slip surface and fails to capture progressive failure processes [1], while numerical simulation methods demand substantial computational resources and require precise specification of constitutive models, boundary conditions, and material parameters, which limits their applicability in complex slope systems [2, 3].

To address these limitations, machine learning methods have been increasingly adopted for slope stability prediction due to their powerful nonlinear mapping capabilities. Models such as Random Forest (RF), Support Vector Machine (SVM), and XGBoost have demonstrated promising accuracy in classifying slope conditions [4, 5]. The theoretical foundation of SVM was established by [6], and its applications in slope stability have been extensively validated [7]. However, the performance of these machine learning models heavily depends on proper hyperparameter settings, and traditional tuning methods, such as manual adjustment and grid search, are time-consuming and often fail to locate the global optimum. To overcome these challenges, intelligent optimization algorithms have been developed to automatically search for optimal hyperparameter combinations. For instance, [8] proposed an Improved

Pelican Optimization Algorithm (IPOA) to optimize RF, achieving significant improvement in prediction accuracy. [9] developed a combined ANN-PSO model for slope stability assessment, demonstrating the effectiveness of hybrid intelligent approaches. [10] applied a hybrid ANN and heuristic model for slope stability prediction, and [11] demonstrated that machine learning paradigms, including ensemble and kernel-based methods, can effectively predict slope stability failure across diverse geological conditions. These studies collectively demonstrate that integrating intelligent optimization algorithms with machine learning models can effectively enhance predictive performance. Therefore, this study proposes a hybrid predictive framework integrating three machine learning models (RF, XGBoost, and SVM) with intelligent optimization algorithms, systematically evaluates their performance, and identifies the most effective combination for slope stability prediction.

2 Methods

Taking the natural unit weight, cohesion, internal friction angle, slope angle, slope height, and pore water pressure ratio as input features, the goal is to predict the state of the slope. Random Forest (RF) [12], XGBoost [13], and Support Vector Machine (SVM) [6] were selected as the basic machine learning models. Four intelligent optimization algorithms, namely the Sparrow Search Algorithm (SSA) [14], Genetic Algorithm (GA) [15], Whale

Table 1. Sample dataset containing input parameters and corresponding slope stability outputs.

Natural Unit Weight (kN/m ³)	Input Parameters					Output Parameters	
	Cohesion (kPa)	Internal Friction Angle (°)	Slope Angle (°)	Slope Height (m)	Pore Water Pressure Ratio	Safety Factor	Slope State
18.5	25	0	30	32.5	0.70	1.01	Stable
10	39.81	20.36	0.98	36	0.25	0.79	Failed
50	45	20	0	120	0.25	1.80	Stable
26.89	150	33	52	50	0.25	0.79	Failed
20	0	36	45	443	0.25	1.28	Stable
25	46	35	47	299	0.25	1.52	Stable
25	55	36	45.5	292	0.25	1.15	Stable
25	46	35	46	20	0.32	0.98	Failed
...	...	...	...	...	...	...	...
18	0	30	20	8	0.30	2.05	Stable

Optimization Algorithm (WOA) [16], and Lion Swarm Optimization (LSO) [17], were used to optimize the hyperparameters of the above three machine learning models, resulting in 12 combinations. Finally, the optimal combination was selected based on five evaluation indicators, namely accuracy, precision, recall, F1 score, and AUC value, to achieve precise prediction of slope stability. The schematic diagram of the research framework is shown in Figure 1.

2.1 Data Collection

2.1.1 Selection of slope stability prediction indicators

The input indicators of the slope stability prediction model are directly related to the reliability of the model, so it is crucial to select scientific and reasonable input indicators. Based on geotechnical mechanics theory, combined with engineering practice experience, and a systematic review of the literature, this study finally establishes six core indicators: unit weight (γ), cohesion (c), internal friction angle (ϕ), slope angle (α), slope height (H), and pore water pressure ratio (r_u).

Among them, γ is the weight per unit volume of the rock and soil mass; the larger its value, the greater the self-weight stress and sliding force of the slope, resulting in worse slope stability. c is the cementation force between rock and soil particles; the higher its value, the higher the shear strength of the rock and soil mass, leading to better slope stability. ϕ is the friction resistance angle between rock and soil particles; the larger its value, the stronger the interlocking effect of particles, resulting in better slope stability. α is the angle between the slope surface and the horizontal plane; the larger the angle, the worse the slope stability.

H is the vertical height; the greater the height, the worse the slope stability. r_u is the ratio of pore water pressure to total stress; the larger its value, the smaller the effective stress of the rock and soil mass, leading to worse slope stability.

2.1.2 Establishment of the sample dataset

In the process of slope stability prediction, data quality directly affects the performance of the model. To avoid the impact of sample singleness on prediction results, it is necessary to cover multiple types of slope data to enhance the persuasiveness of the prediction results. This study integrates slope engineering cases compiled from the dataset reported in [8], which itself was assembled from multiple published engineering case histories, to construct a slope stability prediction dataset. The dataset contains 210 slope engineering cases. Each case has six input indicators, namely unit weight (γ), cohesion (c), internal friction angle (ϕ), slope angle (α), slope height (H), and pore water pressure ratio (r_u). The output variables are the safety factor and the slope state. The slope state has two classes: stable and failed. Among the 210 samples, 39 cases are failed slopes, and the remaining 171 cases are stable slopes. Table 1 shows some sample data.

To eliminate the impact of different feature dimensions on model training, the six input features are standardized using Z-score normalization. The calculation formula is:

$$x_{\text{std}} = \frac{x - \mu_{\text{train}}}{\sigma_{\text{train}}}, \quad (1)$$

where μ_{train} and σ_{train} are the mean and standard deviation of the feature on the training set, respectively.

After standardization, each feature has a mean of 0 and a variance of 1, ensuring that all features have equal opportunities for weight updates in the model, which helps improve convergence speed and prediction stability.

It should be noted that the standardization process strictly follows the principle of using only training set statistics: first, calculate the mean and standard deviation of each feature on the training set, then use these statistics to standardize the training set and test set separately, ensuring that the test set data distribution is consistent with the training set and avoiding data leakage. The standardized feature matrix serves as the direct input for the subsequent SSA optimization and various machine learning models.

2.2 Analysis Methods

2.2.1 Random Forest Algorithm (RF)

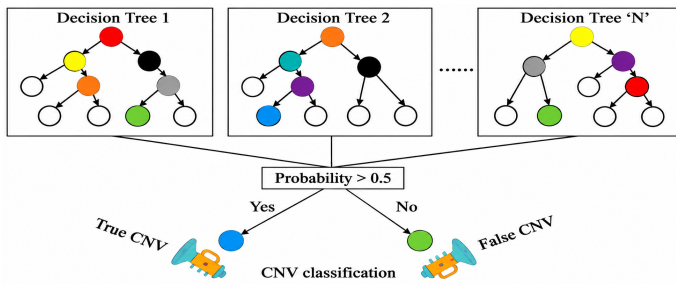


Figure 2. Schematic diagram of Random Forest (RF).

The core of Random Forest (RF) [12] consists of K decision trees $\{h(X, \beta_k), k = 1, 2, \dots, K\}$, where β_k refers to the independent random vector corresponding to each tree, and X refers to the input attribute vector. RF has also been widely applied to slope stability analysis in complex geological settings [18]. In this study, RF is adopted as one of the base classifiers, and the process of model construction is shown in Figure 2. The prediction and classification process is as follows:

- (1) Generate training sets through bootstrap sampling. In the original dataset containing N samples, an independent training set containing N samples is generated for each decision tree through the bootstrap method with replacement. The purpose of this is to make the training data seen by each decision tree slightly different, thereby introducing randomness and avoiding each decision tree learning the same pattern.
- (2) Generate a single decision tree through random feature splitting. For each dataset obtained through bootstrap, when each node of the decision tree splits, m features are randomly selected from all M features,

where $m < M$, and the optimal split point is selected through the Gini index. Its mathematical expression is as follows:

$$Gini(N) = 1 - \sum_{i=1}^k p_i^2, \quad (2)$$

where p_i refers to the probability of category i appearing in the sample set N . The tree is not pruned during the growth process, allowing it to reach maximum complexity.

- (3) Repeatedly build the forest. Repeat the steps described above K times to generate K different decision trees.

- (4) Output results through a voting decision. For the input sample, all decision trees independently give classification results, and the output is finally determined through a simple majority-voting rule. The category with the most votes is taken as the final classification. Its mathematical expression is as follows:

$$H(x) = \arg \max_Y \sum_{k=1}^K I(h_k(x) = Y), \quad (3)$$

where $I(\cdot)$ refers to the indicator function, $h_k(\cdot)$ is the classifier of the k -th tree, and Y refers to the target classification set.

2.2.2 Sparrow Search Algorithm (SSA)

SSA [14] is a metaheuristic intelligent optimization algorithm that solves complex optimization problems by simulating the foraging and anti-predation behaviors of sparrow populations in nature. Based on differences in individual energy levels and fitness, the sparrow population is divided into three roles: producers, scroungers, and alerters.

Among them, producers have stronger energy storage and environmental adaptability, enabling them to find food locations first and guide the foraging direction of the entire sparrow population. During the algorithm iteration process, the position update of producers follows the rule:

$$x_{i,j}^{t+1} = \begin{cases} x_{i,j}^t \cdot \exp\left(-\frac{t}{T}\right), & R_2 < S_T, \\ x_{i,j}^t + Q \cdot L, & R_2 \geq S_T, \end{cases} \quad (4)$$

where $x_{i,j}^t$ and $x_{i,j}^{t+1}$ denote the position of the i -th sparrow in the j -th dimension at the t -th and $(t+1)$ -th

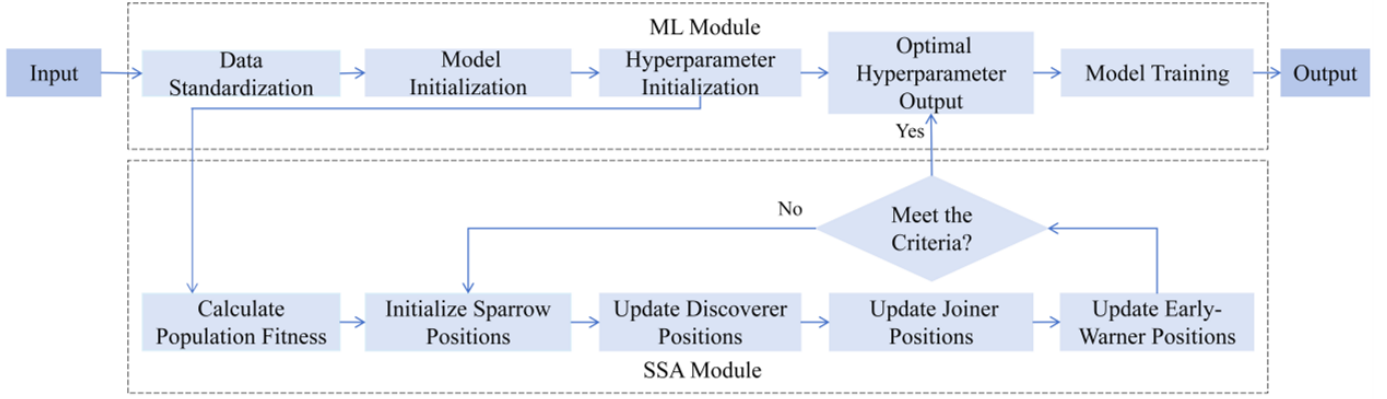


Figure 3. Overall framework of the SSA-optimized ML modeling.

iterations, respectively; T is the maximum number of iterations; Q is a random number following a standard normal distribution; L is an all-ones matrix whose dimension is consistent with the number of problem variables; $R_2 \in [0, 1]$ is the warning value; and $S_T \in [0.5, 1]$ is the safety threshold.

When the warning value $R_2 < S_T$, the current environment is safe, and producers can search within a wide range. Conversely, if $R_2 \geq S_T$, predators are approaching, and the sparrow population immediately moves to a safe area.

Scroungers closely follow producers to obtain food and continuously adjust their positions around the producers. The position update formula is:

$$x_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{x_{worst}^t - x_{i,j}^t}{i^2}\right), & i > n/2, \\ x_P^{t+1} + |x_{i,j}^t - x_P^{t+1}| \cdot A^+, & \text{otherwise,} \end{cases} \quad (5)$$

where x_P^{t+1} is the current best position occupied by the producers, x_{worst}^t is the current global worst position, and A^+ is a randomly assigned $1 \times d$ matrix (each element 1 or -1) with $A^+ = A^T(AA^T)^{-1}$. When $i > n/2$, the i -th scrounger has low fitness, has not obtained food, and needs to fly elsewhere to forage again.

Alerters play the role of sensing danger signals in the environment. When they detect danger in their surroundings, they can quickly respond and adjust their positions. The position update rule is:

$$x_{i,j}^{t+1} = \begin{cases} x_{best}^t + \eta \cdot |x_{i,j}^t - x_{best}^t|, & f_i > f_g, \\ x_{i,j}^t + K \cdot \frac{|x_{i,j}^t - x_{worst}^t|}{(f_i - f_w) + \epsilon}, & f_i = f_g, \end{cases} \quad (6)$$

where $x_{i,j}^t$ is the current position of the i -th alerter in dimension j , x_{best}^t and x_{worst}^t are the current global best and worst positions, η is a step-size control parameter, K is a constant coefficient, f_i is the fitness of the current sparrow, f_g and f_w are the global best and worst fitness, respectively, and ϵ is a very small constant that avoids a zero denominator.

Prior to optimization, all input features are standardized by Z-score normalization, as already defined in Eq. (1).

2.2.3 Overall Framework and Optimization Process of SSA-ML

SSA is a metaheuristic algorithm for optimizing ML hyperparameters. The optimization process works as follows. We first build the ML model and define the hyperparameter search ranges. SSA then generates a random population, where each sparrow represents one set of hyperparameters. For each set, the algorithm trains the model and evaluates its performance. This performance score becomes the fitness value. After calculating all fitness values, SSA ranks the population accordingly. The high-fitness sparrows become producers and lead the global search. The low-fitness sparrows become scroungers and search locally around the producers. A few sparrows act as alerters to monitor risks and guide the population away from danger. Through repeated updates, the algorithm gradually converges to the optimal hyperparameter combination. The overall framework of the SSA-optimized ML modeling is

illustrated in Figure 3. The final model is then ready for validation and engineering applications.

When using SSA to tune RF, we focus on three key issues: hyperparameter encoding, fitness function definition, and search space design. Based on RF characteristics, this paper designs a customized strategy. The next subsection provides the detailed implementation.

2.2.4 Parameter Settings and Optimization Process of the SSA-RF Model

This subsection presents the specific implementation of SSA for RF hyperparameter optimization. It covers encoding, the fitness function, parameter settings, search space, initialization, workflow, and pseudo-code.

Hyperparameter Encoding and Fitness Function

In the SSA-RF model, each sparrow's position encodes a set of RF hyperparameters. We optimize four hyperparameters: the number of decision trees (`n_estimators`), the maximum tree depth (`max_depth`), the minimum samples for splitting (`min_samples_split`), and the minimum samples at a leaf node (`min_samples_leaf`). The position vector is a four-dimensional integer vector:

$$X_i = [\text{n_estimators}, \text{max_depth}, \text{min_samples_split}, \text{min_samples_leaf}]. \quad (7)$$

We define a fitness function to evaluate each hyperparameter combination. The slope stability data are typically imbalanced, with far more stable cases than failed ones, and the AUC metric is insensitive to both class imbalance and the classification threshold. This study therefore uses the average AUC from 5-fold stratified cross-validation as the fitness function. The calculation steps are: (1) split the training set into five subsets using stratified sampling; (2) for each fold $k = 1, \dots, 5$, train the model on four subsets and compute AUC_k on the remaining subset; and (3) average the five AUC values:

$$f(X_i) = \frac{1}{5} \sum_{k=1}^5 \text{AUC}_k. \quad (8)$$

This 5-fold cross-validation uses the data fully, ensures an adequate training size, provides a stable evaluation, and prevents overfitting.

SSA Parameter Settings Table 2 lists the SSA parameter settings. A population size of 30 balances diversity and computational cost. Fifty iterations allow the algorithm to reach convergence. The safety threshold $S_T = 0.8$ controls the producer update switching. The warning value R_2 is random in $[0, 1]$ at each iteration. Producers are 20% of the population (6 individuals), and alerters are 10% (3 individuals).

Table 2. SSA algorithm parameter settings.

Parameter	Symbol	Value
Population size	N	30
Maximum iterations	T	50
Safety threshold	S_T	0.8
Warning value	R_2	Random in $[0, 1]$
Producer proportion	p_{pro}	20% (6)
Alerter proportion	p_{ale}	10% (3)

Hyperparameter Search Space Table 3 shows the search ranges. The number of decision trees controls the scale of base learners: too few cause underfitting, and too many increase cost. The maximum depth limits complexity to avoid overfitting. A larger minimum-samples-for-splitting value makes smoother models, while a smaller minimum-samples-at-leaf value captures subtle patterns but may overfit.

Table 3. Search space of hyperparameters.

Hyperparameter	Symbol	Range	Step
No. of trees	$n_estimators$	$[10, 200]$	1
Max. depth	max_depth	$[3, 50]$	1
Min. split samples	$min_samples_split$	$[2, 30]$	1
Min. leaf samples	$min_samples_leaf$	$[1, 20]$	1

Initialization Strategy This study uses random initialization for the population. Each parameter is generated uniformly within its range. This covers the search space broadly and prevents artificial bias toward any region. Dispersing individuals across the domain allows diverse exploration, which is critical early on: it helps find promising regions before refinement, supporting efficient convergence.

Figure 4 presents the detailed workflow of the SSA-RF optimization process.

Pseudo-code of SSA-RF Algorithm 1 presents the SSA-RF pseudo-code. It starts with population initialization and fitness definition, then updates the producer, scrounger, and alerter positions iteratively. It stops when reaching the maximum number of

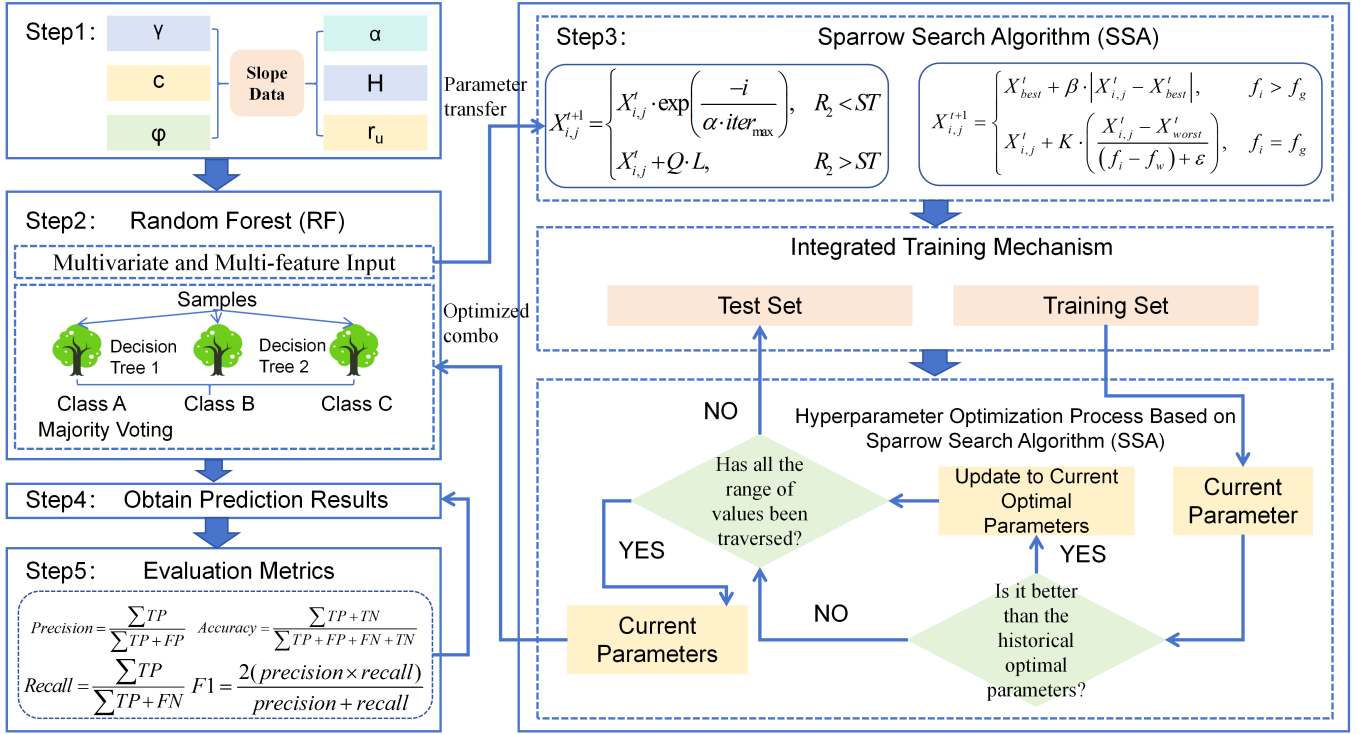


Figure 4. Detailed workflow of the SSA-RF optimization process.

iterations and outputs the best hyperparameters and the trained RF model. Parallel computing accelerates the fitness evaluation, so the optimization finishes in an acceptable time.

2.2.5 Model Evaluation Metrics

Evaluating the performance of the model is a core step in verifying the effectiveness of the slope stability classification model. For the binary classification problem of stable versus failed, the evaluation metrics selected in this study include accuracy, precision, recall, F1 score, and AUC value, while visualization is presented using the confusion matrix and the ROC curve.

Confusion Matrix The confusion matrix is a visual analysis tool for classification models. It cross-compares the model's prediction results with the actual results, intuitively presenting the prediction outcomes. Based on the confusion matrix, we can calculate the following indicators to characterize the model's performance from different dimensions. In this study, "Failed" is defined as the positive class. These indicators include accuracy, precision, recall, and the F1 score, each offering unique insights into the model's predictive capabilities. A detailed explanation of these metrics, along with their calculation formulas, is elaborated below.

1. Accuracy refers to the ratio of the number of samples correctly predicted by the model to the total number of samples, which reflects the overall prediction level of the model. The calculation formula is:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad (9)$$

2. Precision refers to the proportion of samples predicted as failed by the model that are actually failed, which measures the reliability of the model's prediction results for failure cases. The calculation formula is:

$$Precision = \frac{TP}{TP + FP}. \quad (10)$$

3. Recall refers to the proportion of actually failed samples that are correctly predicted as failed, which measures the model's ability to identify failed slopes. The calculation formula is:

$$Recall = \frac{TP}{TP + FN}. \quad (11)$$

4. F1 score refers to the harmonic mean of precision and recall, which balances the trade-off between the

Algorithm 1: SSA-RF: Optimizing Random Forest Hyperparameters with the Sparrow Search Algorithm

Input: Training set $D_{train} = \{(\mathbf{x}_i, y_i)\}_{i=1}^M$; population size $N = 30$; maximum iterations $T = 50$; safety threshold $S_T = 0.8$; producer ratio $p_{pro} = 0.2$; alerter ratio $p_{ale} = 0.1$; lower bound $\mathbf{lb} = [10, 3, 2, 1]$; upper bound $\mathbf{ub} = [200, 50, 30, 20]$

Output: Optimal hyperparameter combination $\mathbf{X}_{best}$; optimal RF model $\mathcal{M}^*$ trained with $\mathbf{X}_{best}$

begin

Generate an initial population of N individuals;

$\mathbf{X}_i = [n_estimators, max_depth, min_samples_split, min_samples_leaf]$, each component uniformly sampled from $[lb_j, ub_j]$;

Define fitness $f(\mathbf{X})$ as the average 5-fold stratified CV AUC: $f(\mathbf{X}) = \frac{1}{5} \sum_{k=1}^5 AUC_k$;

for $i \leftarrow 1$ **to** N **do**

$f_i \leftarrow f(\mathbf{X}_i)$;

$\mathbf{Pbest}_i \leftarrow \mathbf{X}_i, pbest_i \leftarrow f_i$;

if $f_i > gbest$ **then**

$\mathbf{Gbest} \leftarrow \mathbf{X}_i, gbest \leftarrow f_i$;

end

end

for $t \leftarrow 1$ **to** T **do**

 Sort individuals by fitness in descending order;

$N_{pro} = \lceil N \times p_{pro} \rceil$ as producers;

 Randomly select $N_{ale} = \lceil N \times p_{ale} \rceil$ as alerters;

 Remaining individuals as scroungers;

 Update producer positions according to Eq. (4);

 Update scrounger positions according to Eq. (5);

 Update alerter positions according to Eq. (6);

 Clip each hyperparameter to $[lb_j, ub_j]$ and round to integers;

for $i \leftarrow 1$ **to** N **do**

$f_i \leftarrow f(\mathbf{X}_i)$;

if $f_i > pbest_i$ **then**

$\mathbf{Pbest}_i \leftarrow \mathbf{X}_i, pbest_i \leftarrow f_i$;

end

if $f_i > gbest$ **then**

$\mathbf{Gbest} \leftarrow \mathbf{X}_i, gbest \leftarrow f_i$;

end

end

end

$\mathbf{X}_{best} \leftarrow \mathbf{Gbest}$;

Train the final RF model $\mathcal{M}^*$ on the full training set using $\mathbf{X}_{best}$;

return $\mathbf{X}_{best}, \mathcal{M}^*$;

end

two and reflects the comprehensive performance of the model. The calculation formula is:

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (12)$$

The schematic diagram of the confusion matrix is shown in Figure 5.

ROC Curve In addition to the confusion matrix mentioned above, AUC (the area under the ROC curve) is also an important indicator for evaluating the performance of classification models. The vertical axis of the ROC curve represents the True Positive Rate (TPR), and the horizontal axis represents the False Positive Rate (FPR). The closer the ROC curve is to the upper-left region of the graph, the larger the corresponding AUC value, indicating that the

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) Type II Error	Sensitivity $\frac{TP}{TP + FN}$
	Negative	False Positive (FP) Type I Error	True Negative (TN)	Specificity $\frac{TN}{TN + FP}$
		Precision $\frac{TP}{TP + FP}$	Negative Predictive Value $\frac{TN}{TN + FN}$	Accuracy $\frac{TP + TN}{TP + TN + FP + FN}$

Figure 5. Confusion matrix diagram.

model has a better ability to distinguish between stable and failed samples. The corresponding calculation formulas are:

$$TPR = \frac{TP}{TP + FN}. \quad (13)$$

$$FPR = \frac{FP}{FP + TN}. \quad (14)$$

3 Results and Analysis of Slope Stability Prediction

3.1 Descriptive Statistics

To clearly demonstrate the characteristics of the prediction indicators in the selected sample dataset, this study employs a combination of violin plot, box plot, and scatter plot to visualize the six selected prediction indicators, as shown in Figure 6. In these plots, the violin plot shows the probability density distribution of the data, with a wider belly indicating a larger number of samples in that numerical interval; the box plot displays the quantiles of the data, from bottom to top representing the minimum, first quartile, median, third quartile, and maximum; the scatter plot represents the original data of each sample, showing the dispersion and overlap of the data; and the top of the plot annotates the sample mean and sample standard deviation for each prediction indicator.

From Figure 6, it can be observed that γ , c , ϕ , α , and r_u are distributed across all intervals, and the shapes of the violin plots are symmetric. The belly of γ is widest in the interval of 18–20, the belly of c is widest in the interval of 0–30, the belly of ϕ is widest in the interval of 25–32, and the belly of r_u is widest in the interval of 0.2–0.5. This indicates that the slopes in the dataset range from gentle to steep, covering the typical parameter intervals of soil and rock slopes, which can

provide reliable data support for model generalization in complex slope scenarios.

3.2 Sample Correlation Analysis

This study employs the Pearson correlation coefficient (Pearson's r) for correlation analysis, as shown in Figure 7. The Pearson correlation coefficient is a statistical measure that quantifies the linear relationship between two continuous variables, ranging from -1 to $+1$, where values close to ± 1 indicate strong linear correlations, while values near 0 suggest a weak or no linear relationship. In the figure, the flatness of the confidence ellipse represents the strength of correlation between two variables; the red straight line is the linear regression line fitted by the least-squares method, which can intuitively show the correlation between two variables; and the dispersion and overlap of scatter points can also indicate the strength of correlation between variables.

From Figure 7, it can be observed that the absolute values of the correlation coefficients between γ and other variables are all less than 0.15, and those between r_u and other variables are all less than 0.25. Only ϕ and α ($r = 0.595$), ϕ and H ($r = 0.464$), and α and H ($r = 0.402$) show moderate positive correlations. These moderate correlations are theoretically reasonable, as steeper slopes (α) often involve greater heights (H) in natural terrain, and both factors are intrinsically linked to the internal friction angle (ϕ) through the geometric and mechanical properties of the slope.

Overall, there is no high-strength correlation ($|r| > 0.8$) among the variables, indicating that the selected features are reasonably independent, possess complex nonlinear characteristics, can effectively avoid multicollinearity interference, and are suitable for the construction of slope stability prediction models. The absence of strong multicollinearity ensures that each input feature contributes uniquely to the model, enhancing both interpretability and predictive stability.

3.3 Main Findings

3.3.1 Optimization Effects of Different ML Models by SSA

To evaluate the performance of different ML models optimized by SSA for slope stability prediction, this study selects accuracy, precision, recall, F1 score, and AUC value as the core evaluation metrics. These metrics comprehensively reflect the classification capability of the models from various perspectives, with AUC being particularly important for imbalanced datasets such as slope stability cases. From Figure 8,

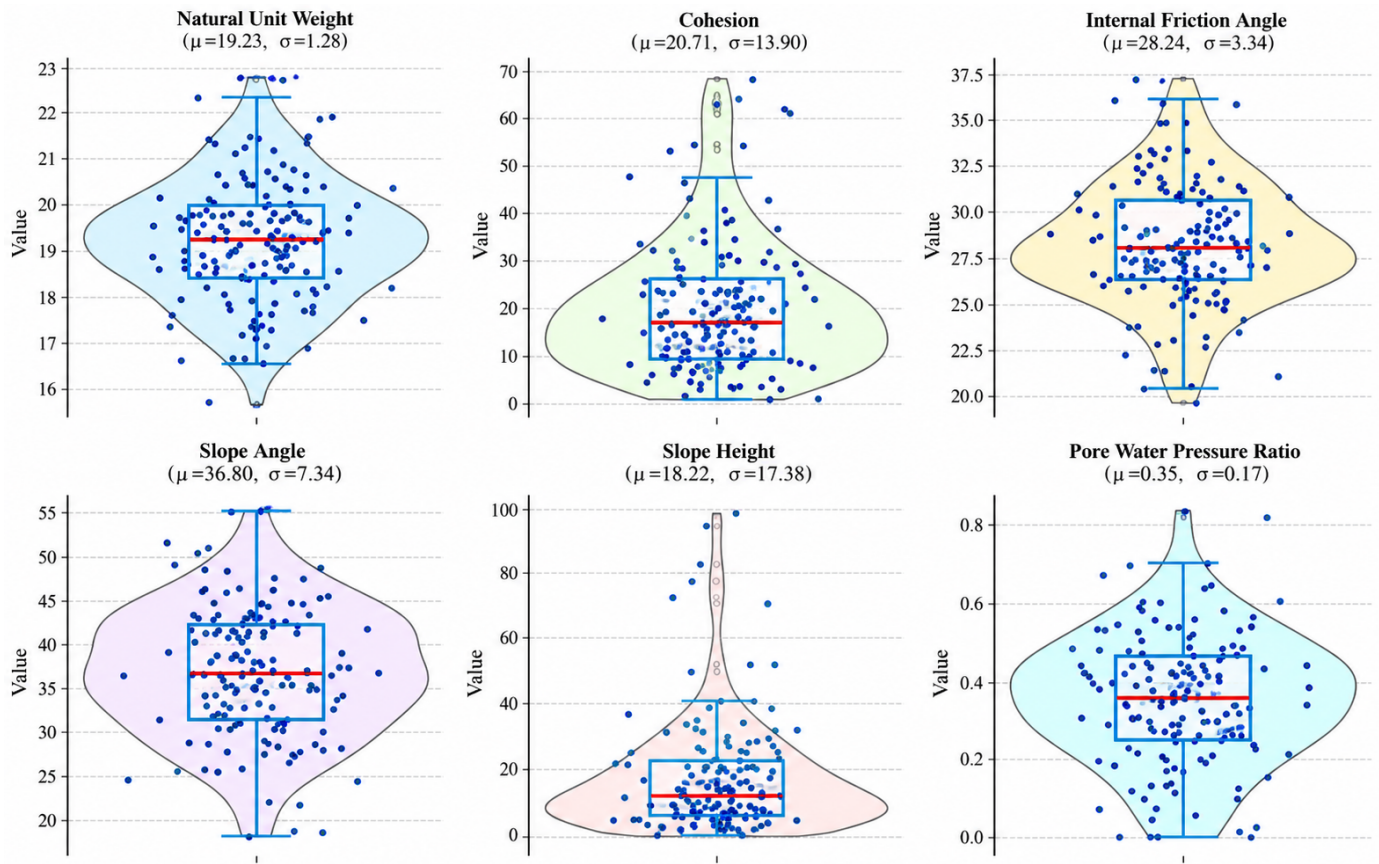


Figure 6. Violin plot, box plot, and scatter diagram of the prediction indicators.

it can be seen that the SSA-RF combination performs best, with only four samples misclassified. Among them, two samples with an actual stable state were misclassified as failed, and two samples with an actual failed state were misclassified as stable. This combination has the largest area under the ROC curve ($AUC = 0.985$), indicating the most reliable classification results. The superior performance of SSA-RF can be attributed to the ensemble nature of Random Forest, which effectively handles nonlinear relationships and reduces overfitting, while SSA successfully identifies the optimal hyperparameters that balance model complexity and generalization.

The performance of the SSA-XGBoost combination is slightly lower than that of SSA-RF, with a total of five samples misclassified. XGBoost, as a gradient boosting framework, also demonstrates strong predictive capability, but its performance may be more sensitive to certain hyperparameters that were not fully explored within the given search space or iteration limits. Nevertheless, SSA-XGBoost still achieves competitive results, outperforming the SVM-based counterpart.

Finally, the SSA-SVM combination has the highest

number of misclassified samples, reaching eleven, and the smallest area under the ROC curve ($AUC = 0.890$), showing the weakest performance among the three combinations. This may be due to the inherent limitations of SVM when dealing with complex, high-dimensional slope data, or because the SSA-optimized hyperparameters for SVM could not fully compensate for the need for more sophisticated kernel functions or feature transformations. Overall, these results highlight the advantage of ensemble methods (RF and XGBoost) over single classifiers such as SVM in slope stability prediction, and demonstrate that SSA is particularly effective when paired with RF.

3.3.2 Comparison of Model Performance Before and After SSA Optimization

As shown in Figure 9, after introducing SSA to optimize the three ML models, all key evaluation metrics for slope stability prediction were improved, demonstrating that SSA can effectively enhance the accuracy of slope stability analysis. Specifically, for Random Forest and XGBoost, accuracy, precision, and recall all increased by varying degrees, with recall showing the most notable improvement. For SVM, the evaluated variants also demonstrated consistent

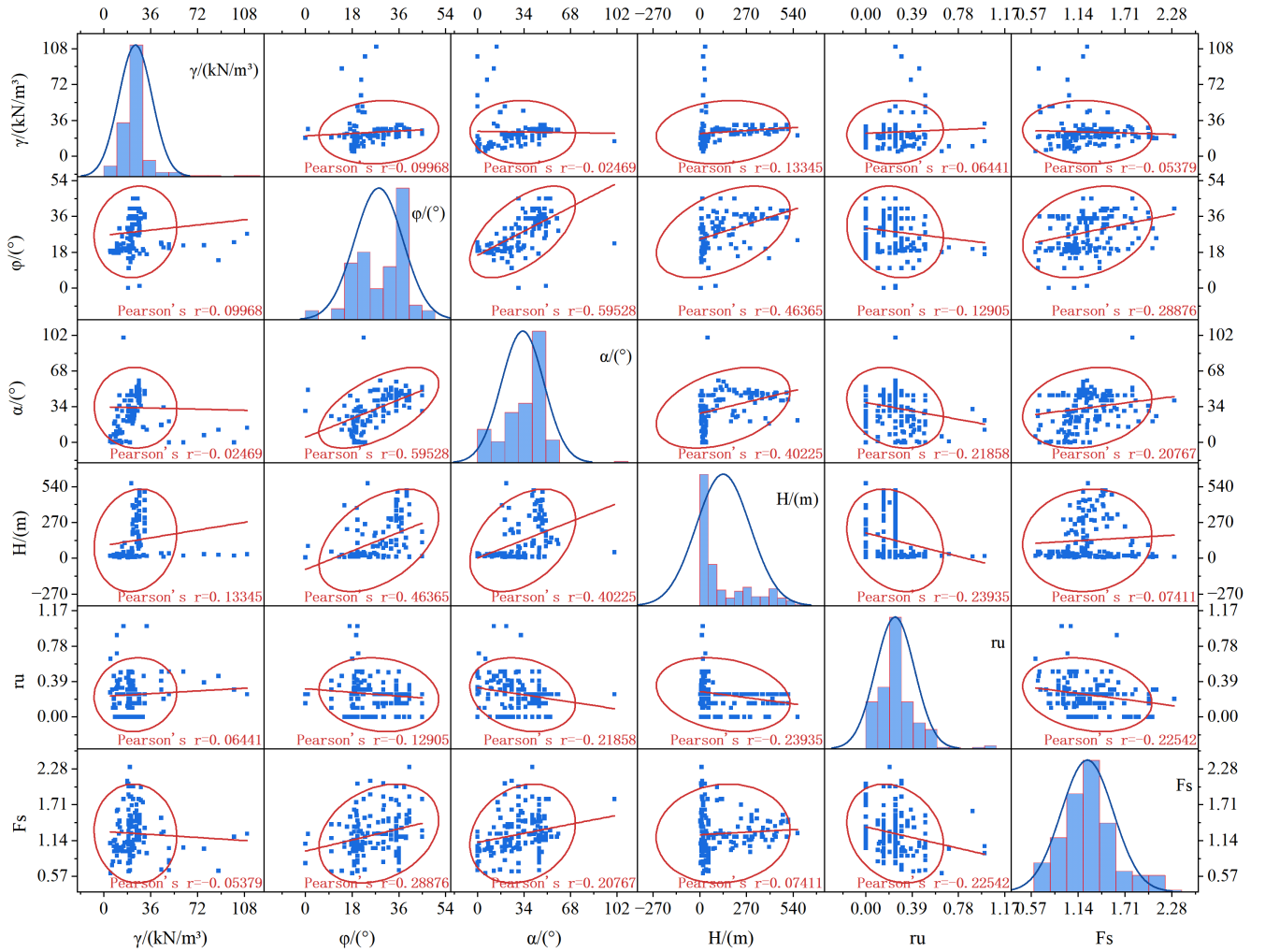


Figure 7. Sample correlation analysis matrix.

improvements across all metrics, particularly in accuracy.

These results indicate that SSA optimization not only enhances overall model performance but also improves the stability and reliability of predictions. The consistent gains across different model architectures further confirm the robustness and generalizability of SSA as an effective hyperparameter optimization tool. Overall, the findings confirm that SSA is a reliable and effective method for optimizing machine learning models in slope stability analysis.

3.3.3 Comparison of the Performance of RF Optimized by Different Algorithms

As shown in Figure 10, RF is a highly advantageous algorithm. When combined with different intelligent optimization algorithms, all performance evaluation metrics remain above 0.8000. Among them, the LSO-RF combination achieves an accuracy of 0.8730, precision of 0.8770, recall of 0.9800, F1 score of 0.9260,

and AUC value of 0.8680; the GA-RF combination achieves an accuracy of 0.8410, precision of 0.8480, recall of 0.9800, F1 score of 0.9090, and AUC value of 0.9040; and the WOA-RF combination achieves an accuracy of 0.8730, precision of 0.8770, recall of 0.9800, F1 score of 0.9260, and AUC value of 0.8750, demonstrating excellent predictive capability and serving as a reliable algorithm for slope stability prediction tasks.

3.4 Summary of Findings

This study evaluates the performance of four intelligent optimization algorithms (SSA, LSO, WOA, and GA) in optimizing three machine learning models (RF, XGBoost, and SVM) for slope stability prediction. The main findings are as follows:

1. The SSA-RF combination performs best among all 12 combinations, with an accuracy of 95.56%, precision of 95.24%, recall of 95.24%, F1 score of 95.24%, and AUC value of 0.9850. Only

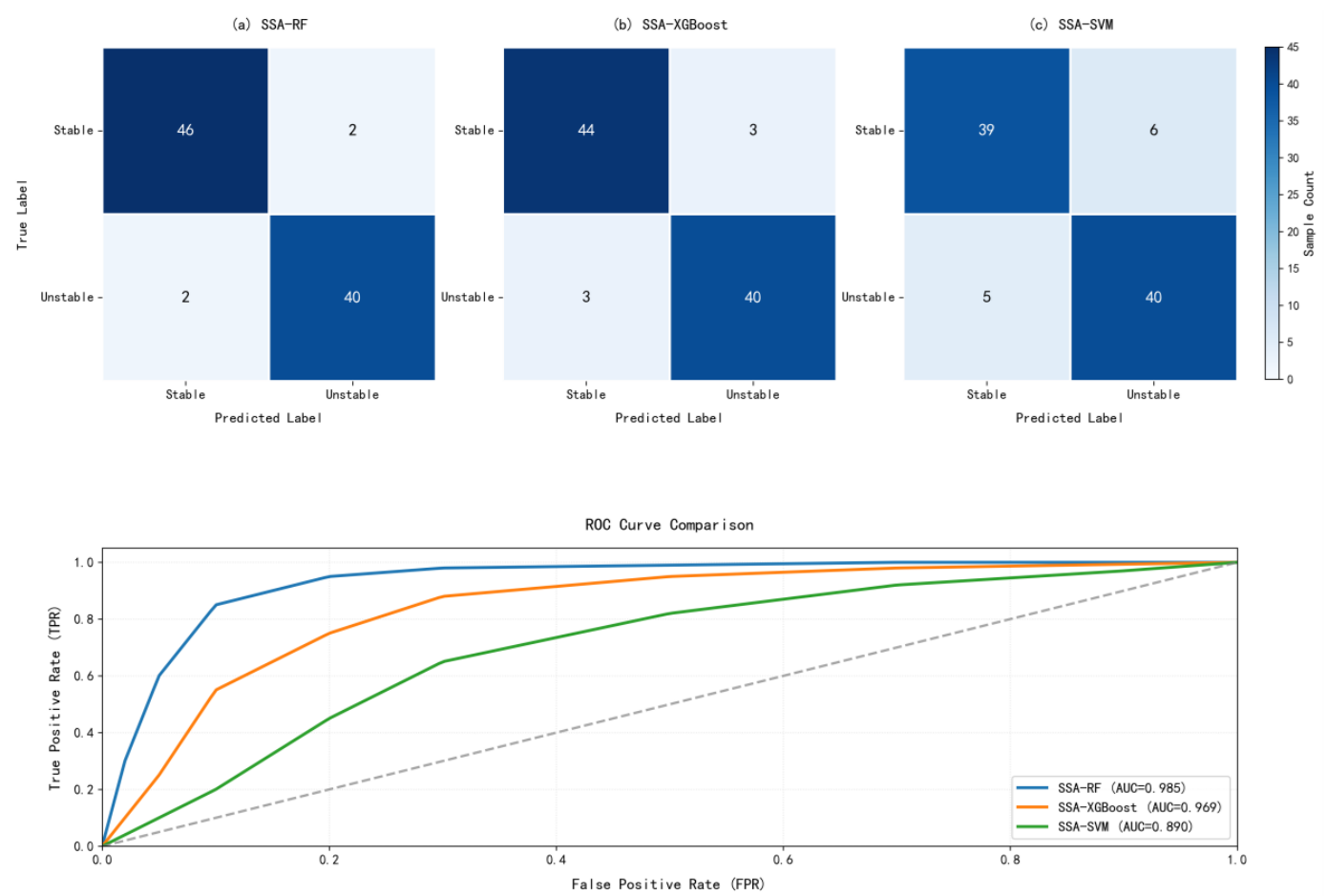


Figure 8. Comparison of performance indicators for three ML models optimized by SSA.

Table 4. Validation results of the SSA-RF model on 10 slope cases.

Input Parameters						Validation Results	
Natural Unit Weight (kN/m ³)	Cohesion (kPa)	Internal Friction Angle (°)	Slope Angle (°)	Slope Height (m)	Pore Water Pressure Ratio	Actual State	Predicted State
20.45	16	15	30	36	0.250	Failed	Failed
27.00	70	22.8	45	60	0.320	Stable	Stable
22.00	10	35	45	10	0.403	Stable	Stable
20.00	20	36	45	30	0.503	Failed	Failed
20.00	0.1	36	45	50	0.290	Stable	Stable
20.00	0.1	36	45	50	0.503	Failed	Failed
22.00	0	40	33	8	0.393	Failed	Failed
24.00	0	40	33	8	0.303	Failed	Failed
20.00	0	24.5	20	8	0.350	Stable	Stable
18.00	0	30	33	8	0.303	Failed	Failed

- four samples are misclassified, demonstrating excellent classification performance.
2. After introducing SSA to optimize the three ML models, all key evaluation metrics are improved, indicating that SSA can effectively enhance the accuracy of slope stability analysis.

3. RF is a highly advantageous algorithm. When combined with different intelligent optimization algorithms, all performance metrics remain above 0.8000, making it a reliable choice for slope stability prediction tasks.

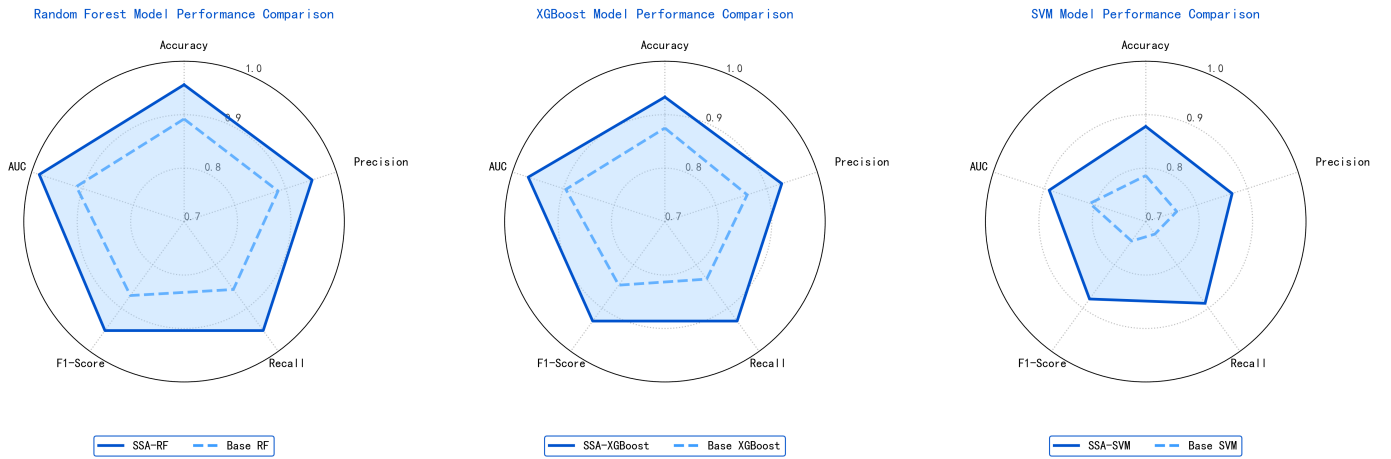


Figure 9. Performance comparison radar chart before and after SSA optimization.

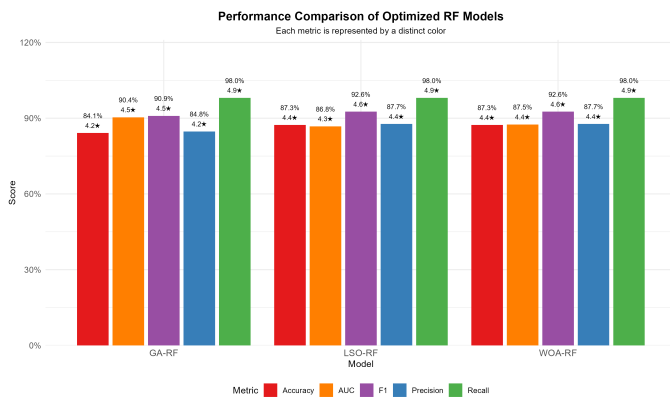


Figure 10. Performance comparison of RF optimized by different algorithms.

4 Discussion

4.1 Verifying the Engineering Applicability of the SSA-RF Combination

To verify the engineering applicability of the SSA-optimized RF model, this study selects 10 slope cases from the references for validation. The validation results show that the prediction results of the selected slope cases are consistent with the actual states, which fully demonstrates that the SSA-RF model has engineering applicability in slope stability prediction and is a reliable prediction model. The specific validation results are shown in Table 4.

4.2 Comparative Analysis of the Improvements in This Study and Existing Research

Compared with existing studies, the improvements of this study in applying SSA to optimize RF are mainly reflected in the following two aspects:

1. **Difference in optimization objects.** Most existing studies have employed SSA to optimize Support Vector Machine or Kernel Extreme

Learning Machine. In contrast, this study combines SSA with RF for slope stability prediction. As an ensemble learning model, RF exhibits stronger robustness to high-dimensional nonlinear data, and the global search capability of SSA effectively addresses the challenge of tuning numerous RF parameters.

2. **Setting of the fitness function.** This study adopts the average AUC under 5-fold stratified cross-validation as the fitness function. Compared with directly using training-set accuracy or a single validation, this approach provides a more stable evaluation of hyperparameter generalization ability and avoids overfitting. Compared with single-metric cross-validation approaches used in prior studies [8], AUC is more friendly to imbalanced data and more suitable for slope stability prediction tasks.

5 Conclusion

1. SSA significantly improves the performance of ML models. After introducing the intelligent optimization algorithm SSA to optimize three ML models (RF, XGBoost, and SVM), all key evaluation metrics were enhanced, especially the AUC value (area under the ROC curve). The AUC values of SSA-RF, SSA-XGBoost, and SSA-SVM are 0.985, 0.969, and 0.890, respectively, and the SSA-RF combination misclassified only four samples, indicating that SSA can improve the accuracy and reliability of slope stability prediction.
2. RF exhibits excellent performance in slope stability prediction. After optimization with four intelligent optimization algorithms (SSA, LSO,

GA, and WOA), all key performance metrics of RF remained above 0.8000, demonstrating that RF is a reliable choice for slope stability prediction models.

3. The SSA-RF combination is the optimal prediction model. Among the three ML models optimized by SSA, the SSA-RF combination achieved the highest AUC value of 0.985 with only four misclassified samples. In the engineering applicability validation, the prediction results of 10 slope cases were consistent with the actual states, demonstrating that the SSA-RF combination has excellent engineering applicability and reliability for slope stability prediction.

Data Availability Statement

The dataset used in this study is compiled from the published literature cited in this work and is available from the corresponding author upon reasonable request.

Funding

This work was supported by the Northwest Normal University Undergraduate Research Funding Project under Grant CXCYKT2026143, the National Natural Science Foundation of China under Grant 12526628, the Youth Science and Technology Talent Special Fund Project of Lanzhou Science and Technology Bureau under Grant 2025-QN-101, the Young Doctor Fund Project of Gansu Provincial Department of Education under Grant 2026QB-016, the College Teachers Innovation Foundation Project of Gansu Provincial Education Department under Grant 2024A-002, the Funds for Innovative Fundamental Research Group Project of Gansu Province under Grant 25JRRA002, and the Industrial Support Fund Program of Gansu Provincial Department of Education under Grant 2025CYZC-016.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable. This study does not involve any experiments on humans or animals; all data are derived from previously published engineering case histories.

References

- [1] Cheng, Y. M., Lansivaara, T., & Wei, W. B. (2007). Two-dimensional slope stability analysis by limit equilibrium and strength reduction methods. *Computers and geotechnics*, 34(3), 137-150. [CrossRef]
- [2] Griffiths, D. V., & Lane, P. A. (1999). Slope stability analysis by finite elements. *Geotechnique*, 49(3), 387-403. [CrossRef]
- [3] Liu, S. Y., Shao, L. T., & Li, H. J. (2015). Slope stability analysis using the limit equilibrium method and two finite element methods. *Computers and Geotechnics*, 63, 291-298. [CrossRef]
- [4] Nanekharan, Y. A., Licai, Z., Chengyong, J., Chen, J., Anwar, S., Azarafza, M., & Derakhshani, R. (2023). Comparative analysis for slope stability by using machine learning methods. *Applied Sciences*, 13(3), 1555. [CrossRef]
- [5] Zhou, J., Li, E., Yang, S., Wang, M., Shi, X., Yao, S., & Mitri, H. S. (2019). Slope stability prediction for circular mode failure using gradient boosting machine approach based on an updated database of case histories. *Safety Science*, 118, 505-518. [CrossRef]
- [6] Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144-152). [CrossRef]
- [7] Huang, F., Xiong, H., Chen, S., Lv, Z., Huang, J., Chang, Z., & Catani, F. (2023). Slope stability prediction based on a long short-term memory neural network: comparisons with convolutional neural networks, support vector machines and random forest models. *International Journal of Coal Science & Technology*, 10(1), 18. [CrossRef]
- [8] Li, M., Li, K., Qin, Q., & Yue, R. (2023). Slope stability prediction based on IPOARF algorithm: A case study of Lala Copper Mine, Sichuan, China. *Expert Systems with Applications*, 229, 120595. [CrossRef]
- [9] Gordan, B., Jahed Armaghani, D., Hajihassani, M., & Monjezi, M. (2016). Prediction of seismic slope stability through combination of particle swarm optimization and neural network. *Engineering with computers*, 32(1), 85-97. [CrossRef]
- [10] Khajehzadeh, M., Taha, M. R., Keawsawasvong, S., Mirzaei, H., & Jebeli, M. (2022). An effective artificial intelligence approach for slope stability evaluation. *IEEE Access*, 10, 5660-5671. [CrossRef]
- [11] Tien Bui, D., Moayedi, H., Gör, M., Jaafari, A., & Foong, L. K. (2019). Predicting slope stability failure through machine learning paradigms. *ISPRS International Journal of Geo-Information*, 8(9), 395. [CrossRef]

- [12] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. [[CrossRef](#)]
 - [13] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794). [[CrossRef](#)]
 - [14] Xue, J., & Shen, B. (2020). A novel swarm intelligence optimization approach: sparrow search algorithm. *Systems science & control engineering*, 8(1), 22-34. [[CrossRef](#)]
 - [15] Holland, J. H. (1992). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. MIT Press. [[CrossRef](#)]
 - [16] Mirjalili, S., & Lewis, A. (2016). The whale optimization algorithm. *Advances in engineering software*, 95, 51-67. [[CrossRef](#)]
 - [17] Yazdani, M., & Jolai, F. (2016). Lion optimization algorithm (LOA): A nature-inspired metaheuristic algorithm. *Journal of Computational Design and Engineering*, 3(1), 24-36. [[CrossRef](#)]
 - [18] Lin, S., Zheng, H., Han, C., Han, B., & Li, W. (2021). Evaluation and prediction of slope stability using machine learning approaches. *Frontiers of Structural and Civil Engineering*, 15(4), 821-833. [[CrossRef](#)]
- Fangfang Zhou** is currently pursuing a Bachelor's degree in Statistics at the College of Mathematics and Statistics, Northwest Normal University. Her research focuses on statistical machine learning methods for system safety and slope stability. (Email: xcwdcy111@163.com)
- Jing Wang** is currently pursuing a Bachelor's degree in Data Science and Big Data Technology at the College of Artificial Intelligence and Computer Science (School of Software), Northwest Normal University. Her research focuses on the application of statistical machine learning theory in system stability and safety. (Email: wj3538759346@126.com)
- Rongfang Yan** is currently a professor from Northwest Normal University. His research interests include reliability, mathematical statistics, reliability theory, statistical dependence. (Email: yanrf@nwnu.edu.cn)
- Jiandong Zhang** is currently an associate professor from Northwest Normal University. His research interests include reliability theory and dependent risk models (Email: jdzhang@nwnu.edu.cn)