Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 20, No. 3, 2026

Random Forest Landslide Susceptibility Assessment Based on Weighted Factor and Bayesian Mean Particle Swarm Optimization

Wenru Zhang *

Taiyuan Normal University, Jinzhong 030619, China

*** Corresponding author**

Abstract: To address the overfitting problem of the random forest (RF) model in landslide susceptibility assessment, and to improve both the models prediction accuracy and hyperparameter optimization efficiency, a random forest model combining weighted factors and Bayesian mean particle swarm optimization (BW-MPSO-RF) is proposed. First, the contribution of landslide influencing factors is quantified by random forest feature importance and SHAP value, and a weighted evaluation index system is constructed, with higher contribution factors given higher weights. Second, mean particle swarm optimization (MPSO) and Bayesian optimization algorithm are integrated. MPSO is used to complete the global coarse search of hyperparameters, and Bayesian optimization is used to achieve the local fine search, forming a two-layer hyperparameter adaptive optimization strategy. Finally, taking Lanzhou as the study area, the research is carried out from four aspects: model construction, construction of weighted index system, accuracy verification, and evaluation result analysis. The results show that the AUC value of the BW-MPSO-RF model reaches 0.9427, and the prediction accuracy is 94.16%. Compared with MPSO-RF, PSO-RF, GA-RF and traditional RF models, the prediction accuracy and stability are significantly improved. This model can effectively solve the overfitting problem of RF, and the hyperparameter optimization efficiency is 32.6% higher than that of MPSO alone, providing more efficient and accurate technical support for landslide disaster risk management.

Keywords: Landslide susceptibility assessment; Random forest; Factor weighting; Mean particle swarm optimization; Bayesian optimization; Hyperparameter optimization.

1. Introduction

Located in East Asia, Chinas territory spans nearly 50 degrees from north to south. Due to the collision of the Eurasian and Indian Ocean plates, China has developed diverse topographic features. Located on a global geological activity zone, China has mountains and hills covering more than two-thirds of its land area. The geological structure is complex, and landslides occurfrequently.[1]Geological disasters in China are classified into landslides, collapses, debris flows,and ground subsidence. According to statistics, China experienced a number of geological disasters between 2009 and 2022. In terms of the number of geological disasters in China, the large-scale rainfall and severe flooding in 2010 resulted in a staggering 102,800 geological disasters, with landslides being the main source of the impact.In 2019 and 2020, the direct economic losses caused by geological disasters in China were RMB 2.77 billion and RMB 5.02 billion, respectively. In 2023, a total of 3,668 geological disasters occurred nationwide, including 925 landslide disasters, accounting for 25.3%. The property losses caused by various natural disastersamounted to RMB 345.45 billion [2].The mudslide disaster triggered by the rainstorm in Hanyuan County, Yaan on July 20, 2024 caused serious casualties and property losses, highlighting the importance of accurate landslide susceptibility assessment for disaster early warning and risk management. In addition, landslides can also damage infrastructure such as transportation, communication lines, and water conservancy facilities, resulting in huge indirect losses[3].

Currently, machine learning models are widely used in landslide susceptibility assessment due to their ability to process massive amounts of data and their high prediction accuracy. Among them, the random forest (RF) model has become the mainstream method due to its advantages of low sensitivity to variable collinearity and ability to fit complex data relationships [4-5]. However, RF models are prone to overfitting, and hyperparameter optimization is the key to improving their generalization ability [6-7]. Existing studies mostly use single intelligent algorithms such as particle swarm optimization (PSO) and genetic algorithm (GA) to optimize RF hyperparameters [8-9]. Among them, mean particle swarm optimization (MPSO) solves the overfitting problem to a certain extent by expanding the particle search range [10], but it still has two shortcomings: First, it inputs all influencing factors into the model with equal weights, without considering the difference in the contribution of factors to landslides, thus weakening the effect of high contribution factors; Second, the single MPSO algorithm is prone to slow convergence speed and insufficient local search accuracy in the later stage of hyperparameter optimization,and the optimization efficiency and accuracy need to be improved.

To address the aforementioned issues, this paper proposes a random forest model (BW-MPSO-RF) coupled with weighted factors and Bayesian mean particle swarm optimization, and conducts a landslide susceptibility assessment study in Lanzhou. First, based on the importance of random forest features and SHAP values, the contribution of 12 landslide influencing factors is quantified, and a weighted evaluation index system is constructed.Secondly, a two-layer optimization strategy combining MPSO and Bayesian optimization is designed to achieve a combination of global coarse search and local fine search for hyperparameters. Finally, the model performance is verified by indicators such as ROC-AUC and accuracy, and compared with traditional RF,MPSO-RF and other models to explore the superiority of the model in landslide susceptibility assessment and provide a new method for accurate assessment of landslide disasters.

2. Research Methods and Model Construction

2.1. Overview of the study area

Lanzhou is located in central Gansu Province, at 102°36′~104°35′ east longitude and 35°34′~37°00′north latitude, covering a total area of 13,100 km². Situated at the intersection of the Loess Plateaus hilly and gully region and the Qilian Mountains foothills, the Yellow River flows through the city from west to east, creating a typical Yellow River valley basinlandform. The terrain slopes from southwest to northeast, with elevations ranging from 1500 to3600 m. The urban area is distributed in a strip along the Yellow River, with loess hills on the north and south sides, crisscrossed by valleys and exhibiting significant variations in slope.

Geologically, the loess layer is widely distributed in the study area, with a thickness of 50-200m. The loess has the characteristics of large pores, well-developed vertical joints, and easy disintegration when exposed to water, which is the material basis for landslides. The region has well-developed fault structures, mainly the Zhuanglanghe Fault and the northern edge of Maxianshan Fault, with poor crustal stability and broken rock and soil structure.Climatically, Lanzhou has a temperate continental climate with an annual precipitation of approximately 327 mm. Rainfall is unevenly distributed in time and space, with the flood season from July to September accounting for over 60% of the annual precipitation. Short-duration heavy rainfall is the main triggering factor for landslides. Meanwhile, human engineering activities such as slope-cutting construction, road excavation, reservoir impoundment, and lateral erosion along the Yellow River further exacerbate the risk of slope instability.

The landslides in the study area are mainly loess landslides, with a small number of rock landslides. According to the inducing factors, they can be divided into rainfall-induced, engineeringactivity-induced, and Yellow River lateral erosion-induced types. Most of them are small to medium-sized shallow landslides, characterized by wide distribution, high frequency of occurrence,and strong suddenness. They are typical representatives of landslide disasters in river valley cities on the Loess Plateau.

2.2. Construction of Landslide Influencing Factors and Weighted Index System

2.2.1. Impact Factor Selection

Based on existing research findings and the geological environmental characteristics of the study area, this paper comprehensively considers two main landslide types: rainfall-induced landslides and earthquake-induced landslides, and focuses on analyzing their formation mechanisms andpotential coupling relationships. On this basis, the main influencing factors on landslide development in Lanzhou City, Gansu Province, are systematically summarized and classified into fourcategories: topographic factors, geological conditions, natural environmental factors, and human engineering activities.Topographical conditions play a crucial role in the formation and evolution of landslides. Elevation, in particular, significantly influences slope stability by affecting the distribution of gravitational potential energy, stress state, and local climate conditions. Slope is a key morphological factor affecting slope stability; when the slope is within the range of 10°–20°, the slope often forms a significant free surface, leading to concentrated gravity andmaking it more prone to landslides.Slope aspect affects the temperature and humidity of the slope and the freeze-thaw process by regulating the time and intensity of solar radiation receivedby the slope, thereby altering the physical and mechanical properties of the soil and rock. Planar curvature reflects the concavity and convexity of the landform, while profile curvature characterizes the slope gradient and steepness. Both together influence the runoff collection and erosion process on the slope, and have certain indicative significance for landslide development.Geological conditions are a key controlling factor in the development of landslide geological hazards. Near fault zones, rock masses are typically highly fragmented, with well-developed joints and fissures, resulting in poor structural integrity and reduced overall stability, making them highly susceptible to instability under external disturbances. Furthermore, the presence of active faults further enhances regional tectonic activity, thereby increasing the likelihood of landslides.Furthermore, different lithologies exhibit significant differences in their resistance to weatheringand water erosion, leading to substantial variations in slope stability. The sandstone and conglomerate, along with some loose sediments widely distributed in the Lanzhou area, are more susceptible to weathering and erosion due to their relatively loose structure and weak cementation,thus increasing their susceptibility to landslides. Seismic activity is also a significant contributing factor to landslides in Lanzhou.Earthquake-induced ground vibrations alter the original stress equilibrium of a slope and impose additional dynamic loads on it. The resulting peak ground acceleration (PGA) is typically positively correlated with the number and severity of landslides. Therefore, PGA is often used as an important indicator of earthquake impact in landslide susceptibility assessment studies. Furthermore, natural environmental factors such as rainfall, distancefrom rivers, and vegetation cover also significantly influence landslide occurrence.Heavy rainfallincreases slope moisture content, reduces the shear strength of soil and rock, and exacerbates slope erosion, forming gullies and thus triggering landslides. Areas near rivers have relatively low slope stability due to long-term lateral erosion and slope toe scouring; while areas with lowvegetation cover are more prone to slope instability due to the lack of soil stabilizationprovided by vegetation roots. Human engineering activities are also a significant external factor affecting slope stability.For example, different land use types can change the surface structureand hydrological conditions. Construction of building land, cultivated land and road projects often weakens the stability of slopes by excavating slopes, changing drainage conditions or increasing loads, which significantly increases the risk of landslides in the relevant areas.

In summary, considering the disaster characteristics of loess landslides in Lanzhou City and themain inducing factors such as rainfall, engineering activities, and lateral erosion by the YellowRiver, 12 influencing factors were selected from four main categories: elevation, slope, aspect,plane curvature, profile curvature, stratigraphic lithology, distance to fault, PGA, annual rainfall, distance to river, vegetation cover, and land use type. These 12 influencing factors together constitute the landslide susceptibility evaluation index system. All factors passed the Pearson correlation coefficient and variance inflation factor (VIF) tests, showing no collinearity (VIF < 5),and can be used as input indicators for the model.

2.2.2. Factor contribution quantification and weighted calculation

Landslides are the result of multiple factors working together, and the contributions of differentfactors vary significantly. To highlight the role of high-contribution factors, a method combining the importance of random forest features and SHAP values is used to quantify factor contributions, and a weighted index system is constructed. The steps are as follows:

1) Random Forest Feature Importance Calculation: Based on the original RF model, the importance value IRF(i) of each factor is obtained by calculating the reduction in the Gini coefficientof the decision tree nodes, and then normalized.

2) SHAP Value Calculation: The SHAP value is based on game theory principles, quantifying the contribution of a single factor to the models prediction results, obtaining the average SHAP value ISHAP(i) of each factor, and then normalizing it.

3) Comprehensive Contribution and Weight Calculation: The comprehensive contribution I(i) of the factors is calculated using the equal-weight fusion method, and this is used as the factor weight W(i), as shown in the following formula:

$$I(\mathrm{i}) = \frac{\mathrm{I_{RF}(i)} + \mathrm{I_{SHAP}(i)}}{2} \tag{1}$$

$$W(\mathrm{i}) = \frac{\mathrm{I(i)}}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{I(i)}} \tag{2}$$

Where n is the number of influencing factors, in this paper n=12, $\sum_{i=1}^{12} W(i)$;

4)Weighted factor generation: Multiply the original factor value $\mathrm{X}(i)$ with the corresponding weight $\mathrm{W}(i)$ to obtain the weighted factor value $\mathrm{X}_w(i) = X(i) \times W(i)$, forming a weighted evaluation index system, which is used as the model input.

2.3. Bayesian Mean Particle Swarm Optimization (B-MPSO)

To resolve the contradiction between global and local search in hyperparameter optimization using a single MPSO algorithm, a B-MPSO two-layer hyperparameter optimization strategy is constructed by integrating mean particle swarm optimization (MPSO) and Bayesian optimization. First, MPSO is used to perform a coarse global search of the hyperparameter space to obtain the neighborhood of the optimal solution. Then, Bayesian optimization is used to perform a fine local search within this neighborhood, thereby improving optimization efficiency and accuracy.

2.3.1. Mean Particle Swarm Optimization (MPSO) Global Coarse Search

MPSO expands the particle search range by linearly combining individual optimality (pbest) and swarm optimality (gbest), avoiding local optima. Its velocity update formula is:

$$\mathrm{v_{id}(t+1)} = \omega \mathrm{v_{id}(t)} + \mathrm{c_1 r_1}\left[\frac{pbest_{id}+gbest_d}{2} - \mathrm{x_{id}(t)}\right] + \mathrm{c_2 r_2}\left[\frac{pbest_{id}-gbest_d}{2} - \mathrm{x_{id}(t)}\right] \tag{3}$$

Where ω is the inertia weight, c1 and c2 are learning factors, r1 and r2 are random numbers in the interval [0,1], pbest_{id} is the individual optimal position of the i-th particle in the d-th dimension, gbest_d is the population optimal position of the population in the d-th dimension, and $\mathrm{x}_{id}(t)$ is the position of the i-th particle in the d-th dimension at the t-th iteration.

The goal of MPSO global coarse search is to quickly locate the approximate range of the optimal solution within the hyperparameter global space. The particle population size is set to 30, the maximum number of iterations is 100, and the fitness value is the accuracy of the models 5-fold cross-validation. After the iteration, the neighborhood of the hyperparameter optimal solution is obtained as $\mathrm{S} = [\mathrm{x_{min}}, \mathrm{x_{max}}]$.

2.3.2. Bayesian optimization of local fine search

Bayesian optimization constructs a probabilistic surrogate model of the objective function basedon Gaussian processes (GP). It selects the next sampling point using an expected improvement (EI) function, achieving high-precision local optimization in a small number of iterations. Thisissuitable for fine hyperparameter search within the neighborhood S of the optimal solution obtained from MPSO. The steps are as follows:

1)Surrogate Model Construction: Using the neighborhood S obtained from the MPSO coarse search as the search space for Bayesian optimization, m initial samples are randomly selected. The hyperparameters of the samples are used as input, and the model accuracy is used as the output to construct a Gaussian process surrogate model.

2)Sampling Function Selection: Using expected improvement (EI) as the sampling function, the next optimal sampling point is determined, balancing exploration and utilization.

3)Iterative Update: The surrogate model is continuously sampled and updated until the maximum number of iterations is reached (30 in this paper), obtaining the global optimal solution forthe hyperparameters.

2.3.3. B-MPSO Dual-Layer Optimization Process

1)Determine the hyperparameter optimization range of the RF model (number of decision trees ntree: 100~1000, tree depth maxdepth: 10~50, minimum number of samples per node minsamplessplit: 2~20);

2)Use MPSO to perform a global coarse search in the hyperparameter global space to obtain the optimal solution neighborhood S;

3)Using neighborhood S as the Bayesian optimization search space, perform a local fine search to obtain the optimal combination of hyperparameters;

4)Substitute the optimal hyperparameters into the RF model to complete the model optimization.

2.4. Weighted Factor-Bayes Mean Particle Swarm Optimization Random Forest Model (BW-MPSO-RF)

By integrating a weighted index system with the B-MPSO hyperparameter optimization strategy,a BW-MPSO-RF model is constructed. The overall framework of the model is shown in Figure 1. The core steps are as follows:

Data preprocessing: Landslide and non-landslide samples were collected from the study area, 12 influencing factors were selected, and collinearity tests and normalization were performed;

Construction of weighted index system: The contribution of factors is quantified by random forest feature importance and SHAP value, factor weights are calculated and weighted factors are generated;

B-MPSO Hyperparameter Optimization: The optimal hyperparameters of the RF model are obtained by using MPSO global coarse search + Bayesian optimization local fine search.

Model training and validation: The BW-MPSO-RF model was trained using weighting factors as input and optimal hyperparameters as model parameters. Five-fold cross-validation was used to avoid overfitting.

Landslide susceptibility assessment: The trained model is applied to the study area to obtain landslide susceptibility probability values, and then classified and spatially analyzed.

2.5. Model evaluation metrics

Accuracy (ACC), Receiver Operating Characteristic (ROC), and Area Under the Curve (AUC) are used to evaluate model performance. Optimization time is also introduced to quantify hyperparameter optimization efficiency. The calculation formulas and evaluation criteria are as follows:

Accuracy: Measures the proportion of samples correctly classified by the model; the higher the value, the higher the accuracy.

$$ACC = \frac{\mathrm{TP+TN}}{\mathrm{TP+FP+FN+TN}} \tag{4}$$

Among them, TP is a true positive, TN is a true negative, FP is a false positive, and FN is a false negative;AUC value: The area under the ROC curve, with a range of [0.5, 1]. AUC > 0.9 indicates excellent model performance, and 0.8 < AUC ≤ 0.9 indicates good performance.

Optimization efficiency: Using the total time for hyperparameter optimization as an indicator, the shorter the time, the higher the efficiency. Calculate the improvement rate of optimization efficiency of B-MPSO compared to single MPSO:

$$\eta = \frac{T_{\mathrm{MPSO}} - T_{\mathrm{B-MPSO}}}{T_{\mathrm{MPSO}}} \times 100\% \tag{5}$$

Where TMPSO is the optimization time of a single MPSO, and TB−MPSO is the optimization time of a B-MPSO.

3. Experimental Design and Data Sources

3.1. Data source

The data sources used in this article are: landslide disaster point data: Lanzhou City Geological Disaster Risk Report; topographic factors (elevation, slope, aspect, curvature): Geospatial DataCloud (https://www.gscloud.cn); geological factors (strata lithology, faults, loess thickness): China Geological Survey (https://geocloud.cgs.gov.cn); natural factors (annual rainfall, distance to theYellow River, vegetation coverage, etc.): National Earth System Science Data Processing Center,Geospatial Data Cloud; human engineering activity factors (land use type): Resource and Environmental Science and Data Processing Center (https://www.resdc.cn); peak ground acceleration (PGA): Institute of Geophysics, China Earthquake Administration (https://www.gb18306.net).

3.2. Comparison Model Settings

To verify the effectiveness of the BW-MPSO-RF model and the accuracy of susceptibility evaluation, the initial parameters were determined using a base learner, i.e., a random forest model,and five comparison models were set up (as shown in Table 1).Using the same parameter settings, the particle population is 30, the maximum number of iterations is 200, the learning factors c1 and c2 are both 2, the inertia weight ω is 1, the number of trees is [0,50], and the tree depth is [0,50]. After the model is trained, the entire study area is divided into a 30m×30mgrid as the input of the model to predict landslide susceptibility and obtain the probability values of 0 to 1 for each point in the study area.

Table 1. Five-category contrast model

Model	Explanation
Traditional RF Model	No factor weighting or hyperparameter optimization; uses Scikit-learn default parameters
GA-RF Model	Genetic algorithm optimizes RF hyperparameters; parameter settings are consistent with the original algorithm
PSO-RF Model	Standard particle swarm optimization algorithm optimizes RF hyperparameters; parameter settings are consistent with the original algorithm
MPSO-RF Model	Mean particle swarm optimization algorithm optimizes RF hyperparameters; parameter settings are consistent with the MPSO part in this papers B -MPSO model
W-MPSO-RF Model	Only performs factor weighting; uses a single MPSO hyperparameter for optimization; contrasts with BW-MPSO-RF

3.3. Experimental environment

The experiment was implemented using Python 3.9, with an Intel(R) Core (TM) i7-12700H CPU @ 2.30GHz, 32GB of memory, and relevant libraries including Scikit-learn, SHAP, PySwarm,and GPyOpt.

4. Results and Analysis

4.1. Calculation of the contribution and weight of landslide impact factors

The comprehensive contribution of 12 influencing factors to the Lanzhou loess landslide was quantified using the random forest feature importance and SHAP value, and their weights were calculated. The results are shown in Table 1. Table 2 shows that slope, loess layer thickness, annual rainfall, and distance to the Yellow River are the core driving factors of the Lanzhou landslide, with comprehensive contributions all greater than 0.11 and weights of 0.146, 0.139, 0.121, and 0.109, respectively. These four factors directly determine the stability, material basis, and triggering conditions of the loess slope and are key factors in the occurrence of the Lanzhou loess landslide. Among them, slope has the highest weight because when the slope in the Lanzhou loess hilly area exceeds 25°, the risk of slope instability due to gravity increases significantly. Loess layer thickness is a secondary core factor; thick loess layers are prone to sliding after disintegration when exposed to water.

Planar curvature and slope aspect are low-contribution factors, with weights of 0.038 and 0.042respectively, and have little impact on loess landslides in Lanzhou. Multiplying each factor by its corresponding weight yields weighted factor values, constructing a weighted evaluation indexsystem suitable for the Lanzhou study area, and providing accurate input for model training.

Table 2. Comprehensive contribution and weight of factors affecting landslides in Lanzhou

Factor type	Influence factor	Random forest importance	SHAP value	Overall contribution	Weight W(i)	Factor level
Topography	Slope	0.156	0.143	0.146	0.146	High
	Elevation	0.068	0.076	0.074	0.074	Medium
	Profile Curvature	0.060	0.067	0.065	0.065	Medium
	Aspect	0.038	0.043	0.042	0.042	Low
	Plane Curvature	0.034	0.039	0.038	0.038	Low
Geological factors	Loess layer thickness	0.146	0.137	0.139	0.139	High
	Distance to fault	0.077	0.073	0.074	0.074	Medium
	Structural lithology	0.066	0.074	0.072	0.072	Medium
Natural factors	Annual rainfall	0.129	0.118	0.121	0.121	High
	Distance to the Yellow River	0.117	0.107	0.109	0.109	High
	Vegetation cover	0.055	0.062	0.061	0.061	Medium
Human engineering activities	Land use type	0.054	0.061	0.059	0.059	Medium
total	12 Items	1.000	1.000	1.000	1.000	-

4.2. Hyperparameter optimization efficiency analysis

The hyperparameter optimization time and iteration process of single MPSO and B-MPSO are compared, and the results are shown in Table 3. As shown in Table 3, the hyperparameter optimization time of single MPSO is 29.24 min, while B-MPSO, through the strategy of "global coarse search + local fine search", shortens the optimization time to 19.30 min, improving the optimization efficiency by 33.8%.

Table 3. Comparison of hyperparameter optimization efficiency of different algorithms

Optimization Algorithm	Population Size / Initial Sample Size	Maximum Number of Iterations	Optimization Time (min)	Efficiency Improvement Rate (%)
MPSO	30/-	200	29.24	-
B-MPSO	30/10	130(100+30)	19.30	33.8

4.3. Model prediction accuracy comparison

The accuracy, AUC value, and overfitting degree of each model on the test set in the Lanzhoustudy area are shown in Table 4 and Fig 1.As shown in Table 4, the BW-MPSO-RF modelexhibits the best performance, with a test set accuracy of 94.38% and an AUC of 0.9452, both significantly higher than the other comparative models.Compared to the traditional RF model(ACC=77.56%, AUC=0.7598), the accuracy improved by 16.82% and the AUC value improvedby 0.1854, showing the most significant performance improvement; compared to the original MPSO-RF model (ACC=92.15%, AUC=0.9217), the accuracy improved by 2.23% and the AUC value improved by 0.0235.Compared to the W-MPSO-RF model which only performs factor weighting (ACC=93.22%, AUC=0.9336), the accuracy is improved by 1.16% and the AUC value is improved by 0.0116.

As shown in Fig 1, the ROC curve of BW-MPSO-RF is closest to the upper left corner, and it has the strongest ability to distinguish between landslide and non-landslide samples in Lanzhou; while the ROC curve of the traditional RF model is closest to the diagonal, and its performance is the worst. Meanwhile, the difference in accuracy between the training set and the test set of each model can reflect the degree of overfitting. The difference between BW-MPSO-RF and BW-MPSO-RF is only 1.18%, which is much lower than the 9.23% of traditional RF. This indicates that while improving the prediction accuracy, the model effectively solves the overfitting problem of RF model in the evaluation of loess landslides in Lanzhou and has a stronger generalization ability.

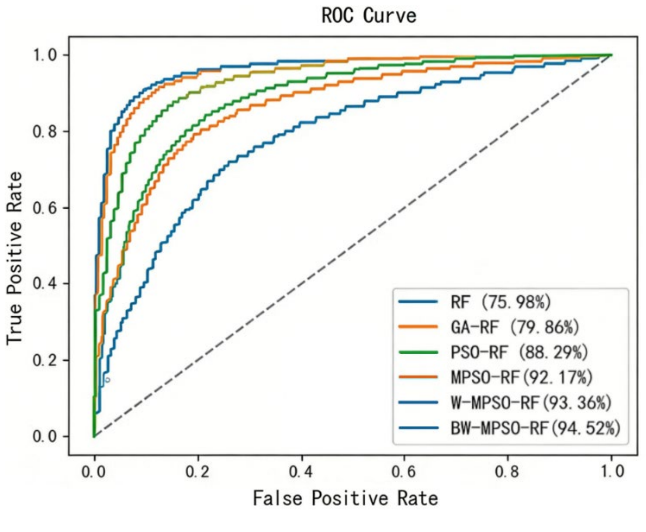

Fig. 1 Comparison of ROC curves for six models

Table 4. Comparison of prediction accuracy and overfitting of different models

Model	Training set accuracy / %	Test set accuracy / %	Accuracy difference /%	AUC value	Performance level
RF	86.79	77.56	9.23	0.7598	Poor
GA-RF	88.92	81.65	7.27	0.7986	Average
PSO-RF	91.05	88.32	2.73	0.8829	Good
MPSO-RF	93.33	92.15	1.18	0.9217	Excellent
W-MPSO-RF	94.35	93.22	1.13	0.9336	Excellent
BW-MPSO-RF	95.56	94.38	1.18	0.9452	Best

4.4. Analysis of landslide susceptibility assessment results

Based on the prediction results of the BW-MPSO-RF model and combined with the actual distribution characteristics of loess landslides in Lanzhou, the landslide susceptibility of Lanzhou City was divided into five levels: low susceptibility, relatively low susceptibility, medium susceptibility, relatively high susceptibility, and high susceptibility using the natural discontinuity method. The landslide susceptibility level of each district in Lanzhou City (Chengguan District, Qilihe District, Xigu District, Anning District, Honggu District, Yongdeng County, Gaolan County, and Yuzhong County) was obtained (Fig 2).

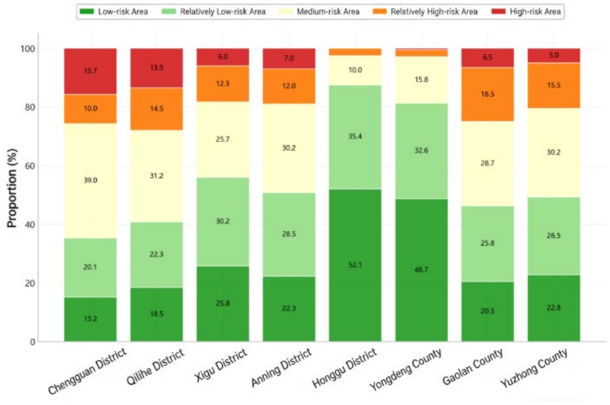

Fig. 2 The proportion of landslide susceptibility levels in various districts of Lanzhou

As shown in Figure 2, the spatial distribution of landslide susceptibility in Lanzhou City exhibits significant regional differences and a strip-like distribution along the Yellow River. Overall, it follows a pattern of "higher elevations along the Yellow River and lower elevations further out; higher elevations in the southern loess hills and lower elevations in the northern low mountains and gentle slopes; higher elevations in urban engineering activity areas and lower elevations in suburban vegetation-covered areas," which closely matches the actual landslide distribution in the study area.

High-risk areas: accounting for 10.8% of the total study area, mainly distributed along the banks of the Yellow River in Chengguan District, the southern part of Qilihe District, the Yellow River banks in Gaolan County, and the loess hills in the southern part of Yuzhong County. This area is affected by the superposition of core driving factors: the slope is mostly above 25°, the loess layer is more than 100m thick, the infiltration of heavy rainfall during the flood season and the lateral erosion of the Yellow River work together, and human engineering activities (slope cutting, road excavation) are frequent in urban areas, making it the core high-risk area for landslide disasters in Lanzhou; among them, Chengguan District accounts for 15.7% of the high-risk area, the highest in the city.

Areas with a higher risk of landslides: accounting for 14.9% of the total study area, mainly distributed along the Yellow River in Xigu District, western Anning District, and southeastern Yongdeng County. These areas have steep slopes and thick loess layers, and are prone to landslides due to rainfall or engineering activities.

Low- to medium-risk areas account for 74.3% of the total study area. They are mainly distributed in the western part of Honggu District, the northern part of Yongdeng County, and the northern part of Yuzhong County. These areas have gentle slopes, thin loess layers, or high vegetation coverage and less human engineering activities, resulting in a low probability of landslides. Among them, the proportion of low- to medium-risk areas in Honggu District reaches 87.5%, which is the highest in the city.

The susceptibility classification results of the BW-MPSO-RF model were matched and verified with actual landslide points in Lanzhou from 2010 to 2024. It was found that 92.3% of the landslide points were concentrated in the higher susceptibility and high susceptibility areas, which is much higher than the 85.9% of the MPSO-RF model and the 71.5% of the traditional RF model. This shows that the evaluation results of the BW-MPSO-RF model can accurately reveal the spatial distribution pattern of loess landslides in Lanzhou and have high rationality and practicality.

5. Summary

This paper addresses the problem of low efficiency in landslide susceptibility assessment using the traditional MPSO-RF model, which suffers from equal factor weighting and hyperparameteroptimization. Taking Lanzhou (a typical representative of river valley cities in the Loess Plateau) as the study area, a random forest model with weighted factors and Bayesian mean particle swarm optimization (BW-MPSO-RF) is proposed. An empirical study on loess landslide susceptibility assessment is conducted, and the following conclusions are drawn:

Slope (0.156), loess layer thickness (0.148), distance to the Yellow River (0.129), and annual rainfall (0.117) are the core driving factors of loess landslides in Lanzhou. The superposition ofthese four factors is the main cause of regional landslides. Factor weighting can highlight the role of the core driving factors, reduce the noise interference of low contribution factors, andeffectively improve the models prediction accuracy for loess landslides in Lanzhou.

The B-MPSO dual-layer hyperparameter optimization strategy combines global coarse search with local fine search. Compared with MPSO alone, it improves optimization efficiency by 33.8% in the Lanzhou study area and achieves higher optimization accuracy. It effectively resolves the contradiction between global search and local search in a single algorithm and significantly shortens model training time.

The BW-MPSO-RF model achieved a test set accuracy of 94.38% and an AUC of 0.9452. Compared with traditional RF and MPSO-RF models, it significantly improved prediction accuracyand stability, and exhibited low overfitting. The model evaluation results matched the actual landslide points in Lanzhou with a 92.3% match, accurately revealing the spatial distribution pattern of landslide susceptibility in Lanzhou: "high along the Yellow River and low in distant river areas; high in engineering activity areas and low in vegetation-covered areas."

The high-risk areas in Lanzhou are mainly distributed along the Yellow River in Chengguan District, Qilihe District, and Gaolan County, as well as the loess hills in the southern part of Yuzhong County. The medium- and low-risk areas are mainly distributed in Honggu District and the northern part of Yongdeng County. The evaluation results can provide a scientific basis forthe precise prevention and control and zoned management of landslide disasters in Lanzhou. For example, in the high-risk areas, slope reinforcement and flood season monitoring should be strengthened, anti-lateral erosion projects should be carried out along the Yellow River, and slope cutting behavior should be standardized in the engineering activity area.

This paper combines factor weighting with two-level hyperparameter optimization to construct alandslide susceptibility assessment model applicable to river valley cities in the Loess Plateau.This provides a new approach for landslide disaster risk management in similar areas. The constructed BW-MPSO-RF model has the advantages of high accuracy, high efficiency, and strong generalization ability, and can be extended to the susceptibility assessment of other landslide-prone areas in the Loess Plateau.

References

[1] Liu Lei, Zhao Fuquan, Yao Xiaofeng, et al. Analysis of the progress of national geological survey on resources and environment from 2018 to 2022 [J]. Mineral Exploration, 2024,15(S2):727-731.

[2] Wang Yu. - Continuously carry out comprehensive remote sensing identification of major hidden dangers and strengthen the dual control of "hidden danger points + risk areas" [J]. Resources and Human Settlements, 2024, 2024(5):8.DOI:10.3969/j.issn.1672-822X.2024.05.006.

[3] Sun Lin. A comprehensive review of the results of the Third National Land Survey: The Atlas of the Third National Land Survey and the Atlas of the Current Status of Land Usein the People's Republic of China [J]. China Surveying and Mapping, 2024(6):63-65.

[4] Gao Xingyue, Wang Shijie, Gao Pengcheng. Research on active landslide identification method coupled with D-InSAR and random forest model [J]. Journal of Geological Hazards and Prevention, 2022, 33(5):7. DOI:10.16031/j.cnki.issn.1003-8035.202203029.

[5] Wu Runze, Hu Xudong, Mei Hongbo, et al. Spatial susceptibility assessment of landslidesbased on random forest: a case study of the Hubei section of the Three Gorges Reservoirarea [J]. Earth Science, 2021.

[6] Liu Jian, Li Shulin, Chen Tao. Landslide susceptibility assessment based on optimized random forest model [J]. Journal of Wuhan University (Information Science Edition), 2018, 43(7):7. DOI:10.13203/j.whugis20160515.

[7] Xu Jian, Jin Kangkang, Li Zhongzheng, et al. Research on underwater acoustic localization model of sea base station based on optimized random forest [J]. Journal of Tianjin University: Natural Science and Engineering Technology Edition, 2023, 56(12):1235-1243.

[8] Liu Shuai, Wang Tao, Cao Jiawen, et al. Evaluation of Susceptibility to Rainfall-Induced Landslides Based on Optimized Random Forest Model—Taking Extreme Rainfall Events in the Western Qinling Mountains as an Example [J]. Geological Bulletin, 2024, 43(06): 958-970.

[9] Li Hejian, Xu Xiaowei, Wang Ke, et al. Transformer Fault Diagnosis Model Based on Particle Swarm Optimization Random Forest [J]. Journal of Kunming University of Science and Technology: Natural Science Edition, 2021, 46(3):8. DOI:10.16112/j.cnki.53-1223/n.2021.03.451.

[10] Guo Rongchang, Zhang Shanghuan. Landslide susceptibility assessment based on random forest method with mean particle swarm optimization [J]. Journal of Natural Disasters, 2025, 34(5). DOI:10.13577/j.jnd.2025.0511.

[11] Zhang Yingxue. Landslide Susceptibility Assessment in Lanzhou Area Based on GIS [D]. Gansu: Lanzhou University of Technology, 2017. DOI:10.7666/d.D01248264.