Landslide Hazard Assessment Based On Improved Stacking Model

Rongchang Guo[1*], Lingyan Yu[1], Rui Zhang[1], Chao Yuan[1], and Pan He[2]

[1] *chool of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou, 730070, China*
[2] *Baoji Electric Service Section of Xi'an Railway Bureau Xi'an 710054, China*
[*] *Corresponding author. E-mail: 2485307443@qq.com*

Received: Apr. 17, 2023; Accepted: July 12, 2023

The early warning of landslides is crucial in mitigating the losses caused by frequent and abrupt landslide disasters along the railway. The scientific construction of an evaluation model is pivotal in conducting a comprehensive landslide hazard assessment. Using a railway section in Ya'an City as a case study, an improved Stacking model was developed to assess landslide hazard by selecting eight evaluation factors and employing support vector machines, random forests, K-neighborhood, and naive Bayesian learning. Logical regression was utilized as a meta learning tool to evaluate the model's performance. To address the issue of a limited number of input samples for the meta learner, the proposed approach incorporates reduced dimensionality data from the original dataset as input for the meta learner. This is based on the output of the base learner, resulting in the establishment of an improved Stacking model. The ROC curve is used to verify the accuracy of the model, compare the accuracy of the Stacking model and the single model before and after the improvement, and generate the risk zoning map of the study area. The results show that the AUCs of support vector machines, random forests, and stacking models are 0.8068, 0.8203, and 0.8368 , respectively, with good performance, while the accuracy of the improved stacking model reaches 0.8806 . A reference for the prevention and management of geological catastrophes, the accuracy of the landslide hazard zoning map created using ArcGIS in the research area has reached 0.853 , which is essentially compatible with the real distribution.

Keywords: Landslide; Support vector machines; Random forest; Stacking model

http://dx.doi.org/10.6180/jase.202405_27(5).0002

1. Introduction

The abrupt and extremely devastating natural disaster known as a landslide is very common in southwest China. There is currently no unifying standard for the prevention and control of landslide disasters because the incidence of landslides is influenced by several different causes and is highly unknown. Southwest China's varied ecology, undulating topography, and heavy rainfall all contribute to the formation of landslides. Based on the available statistics, it has been found that in the year 2022, a total of 5,659 geological disasters were recorded, out of which 3,919 were identified as landslides, comprising a significant percentage of 69.25% of the overall disasters. Therefore, efficient and accurate hazard assessment is of great significance for the prevention and control of geological disasters.

Currently, professionals and academics both domestically and internationally frequently utilize machine learning approaches to analyze the risk of landslides, and they have had success [1]. Among them, the more commonly used machine learning methods include Neural Networks [2, 3], Support Vector Machines [4, 5], Random Forests [6, 7] et al. In order to divide Wenchuan County's landslide susceptibility, Wen et al. [8] used both LR and a neural network, and the findings reveal that both models are accurate. However, the LR effect is superior. To create the landslide susceptibility zoning map of Hanzhong City, Lin et al. [9] used the information from evaluation elements with the objective weight calculated using Random forest.

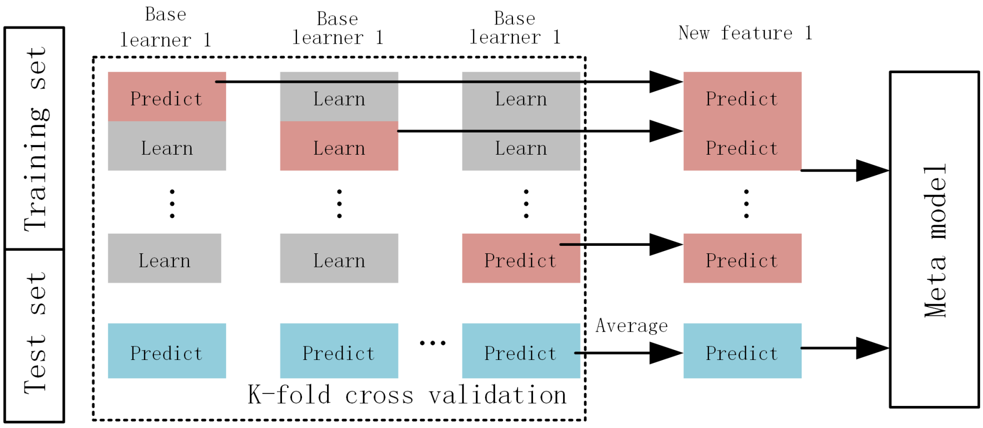

Fig. 1. The framework of the Stacking model

ZHANG et al. [10] proposed the RF-AHP approach to estimate the landslide risk in the northern mountainous area of Guide, Qinghai. and the results show that the accuracy of the RF-AHP model is higher than that of the single model. However, traditional machine learning models have certain limitations and do not have absolute advantages. How to achieve complementary advantages of various machine learning models is significant for improving model accuracy. In summary, to give full play to the advantages of each model, a Stacking model based on the integration of multiple machine learning algorithms is proposed, and improvements are made on the original model to evaluate the landslide hazard in the study area. Through stratified cross validation, the data is split into a training set and a testing set before the model is created. Four algorithms - SVM, RF, KNN, and naïve Bayesian - are then chosen as the first layer of the Stacking model after carefully weighing their differences. Finally, the output of the first layer and the principle components of the original data after dimensionality reduction by PCA are used as inputs to the second layer. An improved Stacking model is established to compare and analyze the accuracy of each model. In Sichuan province, where there is a high concentration of human activity and fault zones, frequent landslides with a flaky distribution pose a serious threat to public safety. Based on this as a starting point, this article selects the evaluation factors that have the greatest impact on landslides based on grid elements, and generates a map of landslide hazard zones using ArcGIS 10.6, and further validates the efficacy of the improved Stacking model, serving as a guide for the prevention and management of geological disasters.

2. Principle of stacking model

2.1. Introduction to Stacking

The stacking method is an integrated method that can combine various machine learning models and enhance their precision and generalizability [11]. The Stacking model is divided into two layers, the first layer being a base learner and the second layer being a meta learner [12]. To avoid overfitting, the first layer usually uses K-fold cross-validation to divide the training set into K pieces and predict 1 of them. The cross-validation results of the first layer and the average value of the test set are input into the second layer. The Stacking model framework is shown in Fig. 1.

2.2. Improved Stacking Model

When partitioning data sets, traditional stacking models frequently overlook the proportion of positive and negative samples, which to some extent reduces the model's accuracy. Therefore, when dividing data, the percentage of positive and negative samples should be taken into full consideration. The outcomes of the Stacking model depend on how well base learners can learn. Since models with similar principles typically learn similarly, the first layer should consist of strong learners with significant differences in principles, while the second layer typically employs a straightforward linear model [13]. Since the input of the second layer of the traditional Stacking module depends on the number of base learners, the learning features are limited, and the training ability is poor. Therefore, the number of features should be increased to improve the learning ability of the meta-learners. In summary, to improve the accuracy and generalization ability of the Stacking model, improvements have been made based on the traditional Stacking model:

1. When dividing the dataset, a stratified cross-validation method is adopted, fully considering the proportion of positive and negative samples;
2. When selecting base learners, it is necessary to comprehensively consider the differences in their principles and select two or more strong learners with significant differences in their principles;
3. Based on the output of the first layer model, the principal components of the original dataset after PCA dimensionality reduction are added as input to the second layer.

The improved Stacking model framework is shown in Fig. 2

2.3. Selection of base learners

Comprehensively considering the principle of the model and the application of commonly used Stacking models [14–17], support Vector Machines, Random Forests, K-Neighborhood algorithms, and Naive Bayes are selected as the base learners, and logical regression is used as the meta-learner to construct the Stacking model.

2.3.1. *Support Vector Machines*

Support Vector Machine (SVM) is a binary classification model that can find a hyperplane based on the eigenvalues of samples to maximize the geometric distance between two types of data from the plane. Its advantage lies in handling small samples and high-dimensional binary classification problems [18]. The hyperplane equation for the SVM classification problem is:

$$f(x) - \omega^T x - b = 0 \tag{1}$$

Where ω is the normal vector and b is the intercept. Introducing Relaxation Variables ξ And the penalty coefficient c convert it into a constraint problem:

$$\left\{ \begin{array}{c} \min \frac{1}{2}\|\omega\|^2 + c\sum_{i=1}^{N} \xi_i \\ \text{s.t. } 1 - y_i\left(\omega^T x_i + b\right) \leq \xi_i \end{array} \right. \tag{2}$$

Introducing Lagrange multipliers λ, convert it to a Lagrange dual problem:

$$L(\omega, b, \lambda) = \frac{1}{2}\|\omega\|^2 + \sum_{i=1}^{N} \lambda_i \left[1 - y_i\left(\omega^T x_i + b\right)\right] \tag{3}$$

Respectively set the partial derivatives of $L(\omega, b, \lambda)$ with respect to ω and b to 0, and then substitute them into Equation (5) to obtain the best advantage. Finally, use a kernel function to map the best advantage to a high-dimensional space, and the decision function becomes:

$$f(x) = \operatorname{sgn}\left(\sum_{i,j=1}^{N} \lambda_i y_i k\left(x_i, x_j\right) + b\right) \tag{4}$$

Where $k\left(x_i, x_j\right)$ is a kernel function, the commonly used Gaussian kernel function (RBF) formula is as follows:

$$k\left(x_i, x_j\right) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right) \tag{5}$$

2.3.2. *Random Forest*

Random Forest (RF) is a classifier that integrates multiple decision trees through ensemble learning ideas [19]. The bootstrap method is used to randomly and selectively extract m samples to generate a new sample set. Each sample is used to train a tree and repeatedly generate m trees, namely, a random forest [20]. Each tree in a random forest randomly selects a portion of the sample, which can effectively avoid overfitting, and has a good noise resistance ability, capable of processing high-dimensional data. The main steps for RF are as follows:

1. Using the bootstrap method, m samples are randomly and randomly selected as training subsets, and the unselected samples are used as out of pocket errors;
2. Randomly select x variables and m training subsets from all samples to generate a corresponding decision tree, randomly select n features at nodes, and then select the features with the best classification ability to split;
3. Repeat the above steps k times to generate k decision trees, which then constitute a random forest.

2.3.3. *K-Nearest Neighbor*

$\mathrm{K}-$ Nearest Neighbor (KNN) algorithm is one of the simplest data mining algorithms. The principle is: find the K known samples closest to the sample to be tested, and divide the sample to be tested into the same category as the K known sample labels based on the principle of minority obeying majority [21]. The method is as follows:

1. Calculate the distance between the sample to be tested and each point in the known sample;
2. Select the k points closest to the sample to be tested;
3. The category label with the most frequent occurrence among the k points is counted as the prediction label for the sample to be tested.

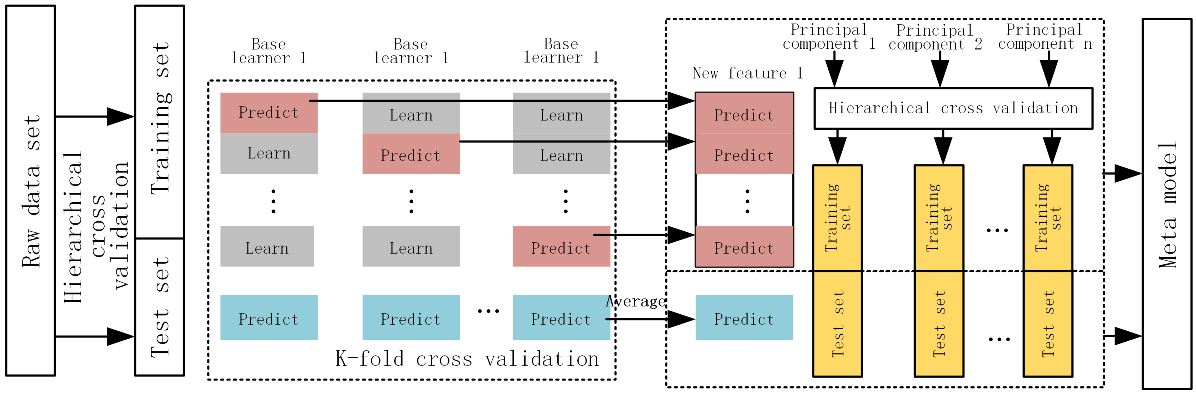

Fig. 2. The framework of the improved Stacking model

2.3.4. *Naive Bayes*

Naive Bayes (NB) is a model that uses a prior probability of a sample and calculates a posterior probability through Bayesian equations to determine it. The calculation method is as follows:

Let each group of data contain n features $X = (x_1, x_2, \cdots, x_n)$ that are independent of each other and have m states $Y = (y_1, y_2, \cdots, y_m)$. First, calculate a priori probability, that is, the proportion of landslide points to non landslide points in the landslide data set:

$$P(y_i) = \frac{N_i}{N} \tag{6}$$

The Bayesian formula is:

$$P(y_i \mid X) = \frac{P(X \mid y_i) P(y_i)}{P(X)} \tag{7}$$

Since each feature is independent of each other and P(X) is a fixed constant, there is a posterior probability:

$$\begin{aligned} P(X \mid y_i) \cdot P(y_i) &= P(x_1 \mid y_i) \cdot P(x_2 \mid y_i) \cdots P(x_n \mid y_i) \cdot P(y_i) \\ &= P(y_i) \cdot \prod_{j=1}^{n} P\left(x_j \mid y_i\right) \end{aligned} \tag{8}$$

Substitute Equation (8) into Equation (7) to obtain:

$$P(y_i \mid X) = \frac{\prod_{j=1}^{n} P\left(x_j \mid y_i\right) P(y_i)}{P(X)} \tag{9}$$

According to the maximum inspection rule, X is assigned to the $\hat{y}$ category:

$$\hat{y} = \arg\max_{y_i} \prod_{j=1}^{n} P\left(x_j \mid y_i\right) P(y_i) \tag{10}$$

3. Method

The landslide hazard assessment process based on the improved Stacking model is shown in Fig. 3. As can be seen from Fig. 3, the steps for landslide hazard assessment based on the improved Stacking model are as follows:

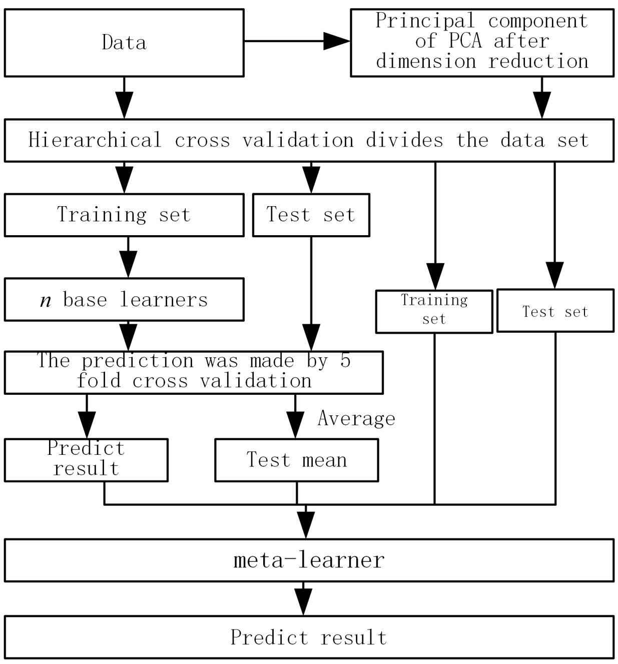

Fig. 3. Evaluation process for landslide hazard based on the Stacking model

Step 1 performs stratified cross validation on the samples outside the model to divide the training set and test set, and imports the divided optimal data into n base learners;

Step 2 performs a 5-fold cross validation on the training set again, and predicts the test set. The cross-validation results are used as the new training set, and the average of the prediction results is used as the new test set;

Step 3 Reduce the dimensions of the original dataset through PCA, reasonably retain the principal components, and perform stratified cross validation on the retained principal components to divide the training set and test set.

Step 4 As the training set for the meta learner, combine the new training set and the reduced dimension training set. As the test set for the meta learner, combine the new test set and the reduced dimension test set.

3.1. Evaluation factor data set

Based on previous studies [22–24] and in combination with the actual situation of the study area, eight evaluation factors were selected, including elevation, rainfall, slope, land use type, stratum lithology, slope aspect, distance from the river, and vegetation coverage. the data source is shown in Table 1. To extract each evaluation factor layer, ArcGIS 10.6 was used, as demonstrated in Fig. 4. In Fig. 4, the Natural Breaks method was used to grade the continuous evaluation factors, while the artificial assignment method was employed to quantify the discrete evaluation factors. The stratum lithology layer was quantified using the artificial assignment method, where lithologies were grouped into five categories. The formation lithology is divided into five groups. the first group includes carbonate, granite, and clastic rock; the second group includes limestone, diorite, and gabbro; the third group includes dolomite, pebbly sandstone, and Laban basalt; the fourth group includes fine sandstone, felsic sandstone, acid tuff, and slate, and the fifth group includes mudstone, siltstone, ice water accumulated sand, and clay.

To avoid linear relationships between factors that affect the accuracy of the evaluation model, a collinearity diagnosis of the evaluation factors is performed by calculating the tolerance (T) and variance expansion factor (VIF) of each evaluation factor. The calculation formula is as follows:

$$VIF = \frac{1}{1 - A^2} = \frac{1}{T} \tag{11}$$

Where A^2 is the variance between factors.

3.2. Stratified cross validation

To determine the level of compliance of the model parameters, the training set is divided into K parts, with one remaining as the test set. This process is known as K fold cross validation. However, the traditional cross validation method of randomly dividing samples can lead to a situation where all the samples are positive, which affects the accuracy of the model. Stratified cross validation can overcome the shortcomings of traditional cross validation, and data can be divided according to the proportion of positive and negative samples of the original data. The schematic diagram of stratified cross validation is shown in Fig. 5.

Table 1. Data source

Factor Name	source
Elevation, Slope, Aspect	Geospatial Data Cloud
Rainfall	China Meteorological Data
Sistance from the river, Land use type	Resource and Environment Science and Data Center
NDVI	Geospatial Data Cloud
Stratum lithology	1 : 250 000 geological map

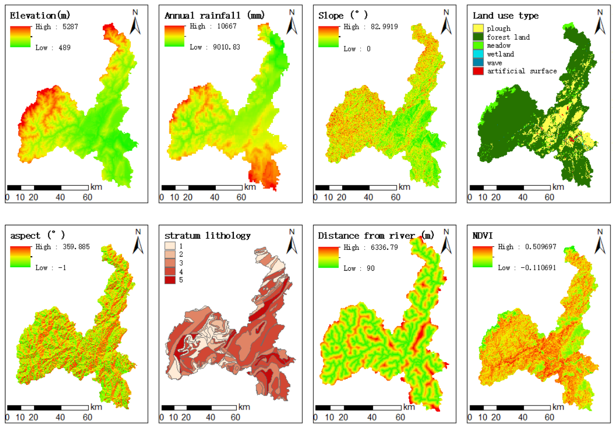

Fig. 4. Graph of landslide risk assessment factors

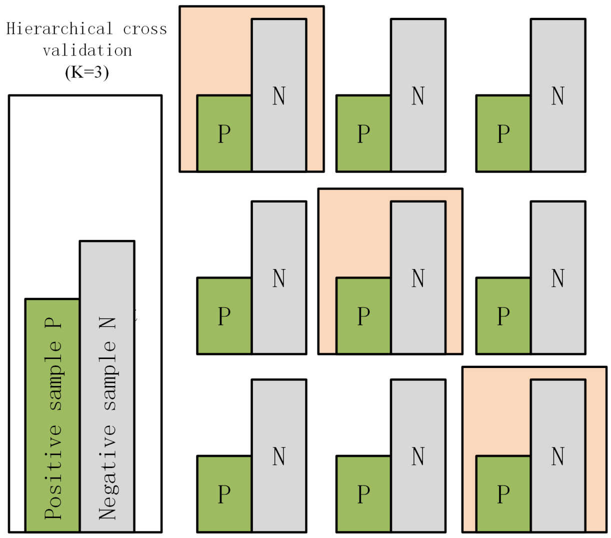

Fig. 5. Schematic diagram of stratified cross validation

3.3. PCA dimensionality reduction

Principal Component Analysis (PCA) is a commonly used dimensionality reduction method [25]. The main idea is to map an n-dimensional feature vector to a k-dimensional orthogonal vector ($k < \mathrm{n}$), which preserves k principal components. The calculation method is as follows:

1. There are m pieces of data, each of which has n characteristics, forming a matrix $X = \left(x_{ij}\right)_{m\times n}$, where x_{ij} represents the j-th evaluation factor in the i-th group of data.
2. Centralize matrix X to obtain $X^* = \left(x_{ij}^*\right)_{m\times n}$:
$$x_{ij}^* = \frac{1}{m}\sum_{i=1}^{m} x_{ij} \tag{12}$$
3. Calculate covariance matrix $C_{m\times n}$:
$$C = \begin{bmatrix} \operatorname{cov}(x_{11}, x_{11}) & \cdots & \operatorname{cov}(x_{1n}, x_{1n}) \\ \vdots & \ddots & \vdots \\ \operatorname{cov}(x_{m1}, x_{m1}) & \cdots & \operatorname{cov}(x_{mn}, x_{mn}) \end{bmatrix} \tag{13}$$
Where, $\operatorname{cov}(x_1, x_2) = \frac{\sum_{i=1}^{m}(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2))}{m-1}$
4. Calculate the eigenvalue λ_j and eigenvector $\boldsymbol{\alpha}_j$ of $C_{m\times n}$, rank λ_j from the largest to the smallest, and select the first k groups $\left(\lambda_j, \boldsymbol{\alpha}_j\right)$ Calculate new features $Y = \alpha X$ after projection:
$$\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1k} \\ y_{21} & y_{22} & \cdots & y_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mk} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_k \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \tag{14}$$

3.4. Accuracy verification

The ROC curve is used to test the accuracy of the Stacking model. ROC curve, also known as receiver operating curve, is widely used in accuracy verification of various models. Area under curve (AUC) is often used to measure model performance. The closer the AUC is to 1, the better the model performance. The ROC curve takes false positive rate (FPR) as the abscissa and true positive rate (TPR) as the ordinate. The formula for calculating FPR and TPR is as follows:

$$FPR = \frac{FP}{TN + FP} \tag{15}$$

$$TPR = \frac{TP}{TP + FN} \tag{16}$$

"FP" refers to the prediction being true and the actual being false; TP means that the prediction is true and the actual value is also true; TN means that the prediction is false and the actual value is also false; FN means that the prediction is false and the actual value is true.

4. Landslide hazard assessment based on improved stacking model

Taking Yucheng District, Tianquan County, and Lushan County in Ya'an, Sichuan Province as the study areas, a Stacking model was established for landslide hazard assessment. With a high terrain in the west and a low terrain in the east, and an altitude that ranges mostly from 500 to 5000 meters, the research area is situated in the central and western parts of Sichuan Province. It is a transitional zone from the Qinghai-Tibet Plateau to the Chengdu Plain, as shown in Fig. 6. Among them, the formation lithology of Lushan County and Yucheng District is mainly composed of siltstone and mudstone, while that of Tianquan County is mainly composed of carbonate and gravelly sandstone. With the occurrence of rainfall, the hardness of siltstone and mudstone decreases, which is easy to form landslides. The research area has a relatively substantial relative elevation difference, mostly in hilly areas, which makes the geomorphic structure unstable and vulnerable to landslides under other adverse conditions. 243 landslide data were selected as samples, and a grid unit of 30 m $\times$ 30 m was used as the evaluation unit, totaling 3231 $\times$ 4333 grid cells.

4.1. Factor collinearity diagnosis

SPSS 24 was used to calculate the tolerance (T) and variance inflation factor (VIF), and the calculation results are shown in Table 2. The conditions of T $>$ 0.1 and VIF $<$ 10 are met, so there is no linear relationship between the factors, which can be directly input into the model.

Table 2. Tolerance and variance inflation factor

Factor Name	T	VIF
Rainfall	0.958	1.044
Stratum lithology	0.967	1.034
Land use type	0.857	1.167
Slope	0.784	1.275
Elevation	0.779	1.284
Aspect	0.919	1.088
NDVI	0.942	1.062
Distance from river	0.933	1.072

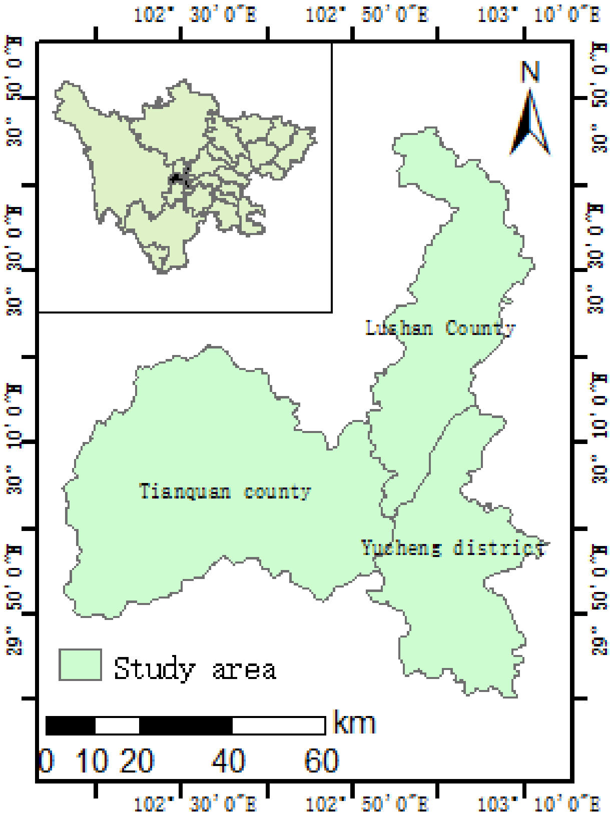

Fig. 6. Landslide zoning map of study area

4.2. Hyperparametric optimization and accuracy verification

Machine learning models usually contain multiple hyperparameters. The most significant hyperparameters are often chosen for tuning to produce the best model results [26]. The grid search method is used to optimize the super parameters of SVM, RF, and KNN models. The optimization results are shown in Table 3.

Table 3. The result of hyperparameter optimization

Model	hyperparameter
SVM	Kernel = rbf, gamma =0.05, C=10
RF	n_estimators =38, max_depth =4
KNN	n_neighbors =5

Import the divided data into the optimized SVM, RF, KNN, NB, and traditional Stacking models to generate ROC curves, as shown in Fig. 7. From Fig. 7, SVM, RF, and traditional stacking model all have AUCs greater than 0.8, and they perform well compared to single models. The eight evaluation factors are reduced by PCA, and four principal components are retained. They are input into the meta learner together with the output of the base learner. The improved Stacking model is established. The ROC of the improved Stacking model is shown in Fig. 8, and its AUC reaches 0.8806, which is 0.0438 higher than that of the traditional Stacking model. In addition, Accuracy, Precision, Recall, and F1 values were used to verify the accuracy of each model. The results of the accuracy verification are shown in Table 4. From Table 4, the evaluation indicators of the improved Stacking model are relatively high, which verifies the effectiveness of the proposed method in improving the traditional Stacking model.

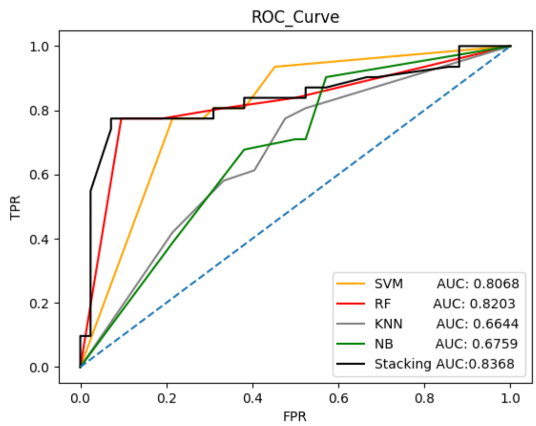

Fig. 7. ROC curve of each model

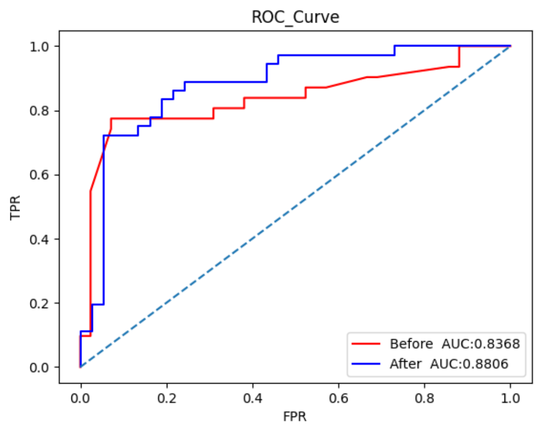

Fig. 8. ROC curve of the improved Stacking model Tab. 4 Verification of model accuracy

4.3. Landslide Hazard Assessment

Use ArcGIS 10.6 to generate a hazard zoning map for the study area, as shown in Fig. 9. The Natural Breaks method is used to divide the landslide hazard level into five levels, namely, higher hazard area, high hazard area, medium hazard area, low hazard area, and lower hazard area. The

Table 4. Verification of model accuracy

Model	Accuracy	Precision	Recall	F1
SVM	0.781	0.778	0.778	0.778
RF	0.808	0.806	0.806	0.806
KNN	0.616	0.643	0.500	0.563
NB	0.671	0.676	0.639	0.657
Traditional Stacking	0.822	0.848	0.778	0.812
Improved Stacking	0.836	0.875	0.778	0.824

232 landslide points that were not involved in model training and testing were imported into ArcGIS. According to statistics, 85.34% of the landslide points were distributed in higher hazard areas and high hazard areas, and 12.07% of the landslide points were distributed in medium hazard areas, which is relatively reasonable and can be considered as having a model accuracy of 0.853 .

According to Fig. 9 and field analysis, higher hazard areas and high hazard areas are mostly distributed near rivers and cultivated land, mainly due to strong human activities and the great impact of rainfall and rivers on the geological environment, which contributes to the development of landslides. The main impacts include river erosion of the riverbed, resulting in changes in the lithology of the riverbed formation; Most of the annual rainfall is concentrated in summer, and long-term continuous rainfall reduces the hardness of the slope, making it more prone to landslides; At the same time, strong human activities on both sides of the river, land reclamation, and the construction of houses and roads have created favorable conditions for the occurrence of landslides.

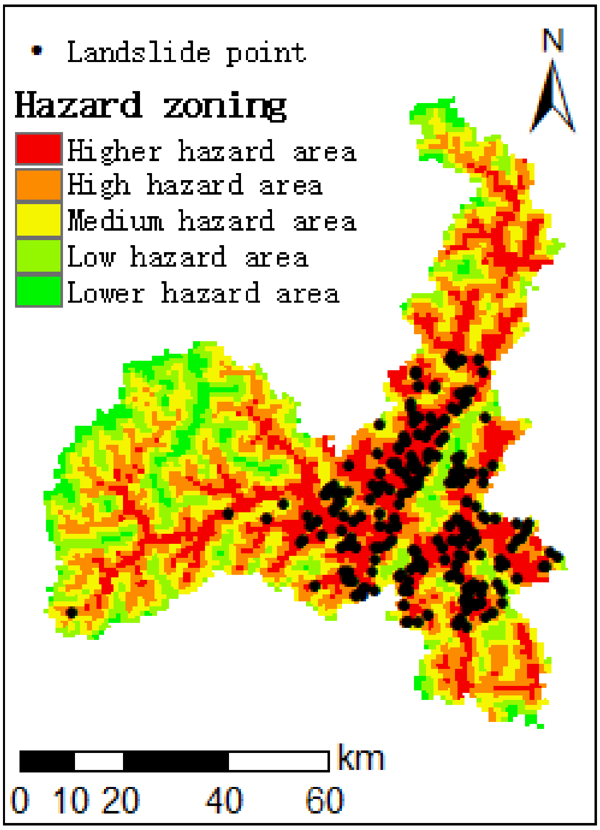

Fig. 9. Zoning map of risk in study area

5. Conclusion

Taking Yucheng District, Tianquan County, and Lushan County of Sichuan Province as the study areas, eight evaluation factors, including elevation, rainfall, slope, land use type, stratum lithology, aspect, distance from river, and NDVI, were selected to establish an improved Stacking model based on SVM, RF, KNN, and NB for landslide risk assessment, and the following conclusions were obtained:

1. Among the four machine learning models, the accuracy of SVM model and RF model is higher than 0.8, with good performance; The accuracy of KNN model and NB model is lower than 0.7, resulting in poor performance. The accuracy of the four models measured using AUC is from high to low: RF (0.8203) > SVM(0.8068) > NB(0.6759) > KNN(0.6644), which is basically consistent with the results of previous landslide risk assessment models.

2. Using the traditional Stacking algorithm to integrate four machine learning models, the accuracy of the model reached 0.8368, which was 0.03 and 0.0165 higher than that of SVM and *RF*, respectively, and can improve the accuracy of the evaluation model to a certain extent. The accuracy of the improved Stacking model has reached 0.8806 , which is 0.0438 higher than the traditional model, and can provide a reference for the prevention and control of geological disasters.

3. Based on ArcGIS, a landslide hazard zoning map was generated for the study area. 85.34% of the landslide points were distributed in higher hazard areas and high hazard areas, with high accuracy, further verifying the rationality of the model.

4. The method proposed in this paper has certain limitations: supervised learning is usually affected by parameters, and the author will explore the optimization of hyperparameter in the next step.

References

[1] R. Fang, Y. Liu, and Z. Huang, (2021) *"A review of the methods of regional landslide hazard assessment based on machine learning"* **The Chinese Journal of Geological Hazards and Control *32***: 1–8.

[2] F. Liu, L. Wang, and D. Xiao, (2021) *"Application of machine learning model in landslide susceptibility evaluation"* **The Chinese Journal of Geological Hazards and Control *32***: 98–106.

[3] Z. Jiang, M. Wang, and K. Liu, (2023) *"Comparisons of Convolutional Neural Network and Other Machine Learning Methods in Landslide Susceptibility Assessment: A Case Study in Pingwu"* **Remote Sensing *15***(3): 798.

[4] W. Wang, P. Liu, and L. Gong, (2019) *"Landslide susceptibility mapping of sichuan province based on support vector machine"* **Journal of Railway Science and Engineering *16***(5): 7.

[5] Z. Chen, H. Zhou, F. Ye, B. Liu, and W. Fu, (2022) *"Landslide Susceptibility Mapping along the Anninghe Fault Zone in China using SVM and ACO-PSO-SVM Models"* **Lithosphere *2022***(1): 2–16.

[6] K. Mu, W. Xie, Q. Liu, M. Yan, H. Yang, H. Li, Y. Huang, and R. Zhu, (2022) *"Research on Landslide Susceptibility Evaluation Based on Logistic Regression and LR Coupling Model"* **Journal of Catastrophology** (003): 037.

[7] E. K. Sahin, (2023) *"Implementation of free and open-source semi-automatic feature engineering tool in landslide susceptibility mapping using the machine-learning algorithms RF, SVM, and XGBoost"* **Stochastic Environmental Research and Risk Assessment *37***(3): 1067–1092.

[8] H. Wen, D. Hu, and G. Wang, (2014) *"A comparative study on the susceptibility mapping of earthquake triggered landslide by neural network and logistic regression model, Wenchuan County"* **China Civil Engineering Journal *47***: 17–23.

[9] F. Lin, J. Liu, S. Xu, M. Liu, M. Zhang, and E. Liang, (2020) *"Evaluation method of landslide susceptibility based on random forest weighted information"* **Science of Surveying and Mapping *45***(12): 8.

[10] X. ZHANG, C. ZHANG, H. MENG, et al., (2018) *"Landslide hazard evaluation in the northern mountainous area of Guide County based on Random Forest and AHP"* **Hydrogeology & engineering geology *45***(4): 142–149.

[11] Y. Wang, Z. Fang, and R. Niu, (2021) *"Prediction of Landslide Susceptibility in Three Gorges Reservoir Area Based on Integrating Deep Neural Network"* **Resources Environment & Engineering *35***(5): 652.

[12] W. Wu, (2022) *"Research and application of tax reduction and fee reduction risk prediction model based on stacking integrated learning"*:

[13] G. Yang, S. Liu, D. Wang, W. Wang, and J. Liu, (2022) *"Short-term wind power forecasting based on Attention-GRU wind speed correction and stacking"* **Acta Energiae Solaris Sinica *43***: 273–281.

[14] Z. Yuan, X. Zhang, Y. Guo, C. Yin, and Xiaoyan-Ning, (2022) *"Research on Cost Forecasting of Power Grid Transmission Project Based on Stacking Integrated Model"* **Shandong Electric Power *49***(12): 14–19.

[15] J. Xu, B. Liu, G. Zhang, and J. Zhu, (2023) *"State-of-health estimation for lithium-ion batteries based on partial charging segment and stacking model fusion"* **Energy Science and Engineering *11***(1): 383–397.

[16] W. Yan, H. Rao, and H. Duan, (2022) *"Prediction Method on Harmful Bird Density Based on Stacking Multi-Model Fusion Algorithm"* **Industrial Control Computer *35***(12): 20–22.

[17] G. Pan, M. Gong, M. He, C. Wu, X. Tang, L. Yang, and J. Ouyang, (2021) *"Stacking"* **Dianli Zidonghua Shebei/Electric Power Automation Equipment *41***(1): 152–158.

[18] L. Li, X. Zhang, H. Deng, and L. Han, (2020) *"Assessment of Landslide Susceptibility Based on SVM-LR Model: A Case of Shanyang County"* **Science and Technology and Engineering *20***(26): 8.

[19] Z. Luan, X. Zhang, and Y. Wang, (2022) *"Prediction of surface subsidence coefficientin mining zrea based on RF-GWO-LSSVM"* **Beijing Surveying and Mapping *007***: 036.

[20] R. Wu, X. Hu, H. Mei, J. He, and J. Yang, (2021) *"Spatial susceptibility assessment of landslides based on random forest: a case study from Hubei section in the three gorges reservoir area"* **Earth Science *46***(1): 321–330.

[21] F. Liu. *Research on K-means algorithm cluster center optimization*. 2021.

[**22**] Z. Sun, Y. Li, Y. Jiang, and Z. Wang, (2023) *"Sea temperature forecast in the northern South China Sea base on Stacking machine learning model"* **Marine Forecasts *40***: 39–45.

[**23**] H. Peng, W. Yan, S. Lin, H. Lai, and Z. Zhang, (2022) *"Point cloud simplification improved algorithm integrating k-means clustering and Hausdorff distance"* **Geospatial Information *008***: 020.

[**24**] P. He, R. Guo, R. Zhang, and L. Yu, (2022) *"Risk assessment of landslide along railway based on different combination of evaluation factors"* **Journal of Lanzhou Jiaotong University *41***(5): 8.

[**25**] J. Liu, E. Liang, S. Xu, M. Liu, Y. Wang, F. Zhang, and A. Luo, (2022) *"Multi-kernel support vector machine considering sample optimization selection for analysis and evaluation of landslide disaster susceptibility"* **Acta Geodaetica et Cartographica Sinica *51***(10): 2034–2045.

[**26**] Z. Bai, Q. Liu, and Y. Liu, (2022) *"Large-scale landslide susceptibility evaluation based on entropy index-random forest coupling model"* **Yangtze River *53***(10): 95–102.