land

MDPI

Article

Predicting the Spatial Distribution of Geological Hazards in Southern Sichuan, China, Using Machine Learning and ArcGIS

Ruizhi Zhang [1,*], Dayong Zhang [2], Bo Shu [3] and Yang Chen [3]

[1] School of Architecture, Southwest Jiaotong University, Chengdu 611756, China
[2] School of Architecture and Urban Planning, Lanzhou Jiaotong University, Lanzhou 730070, China; 11230665@stu.lzjtu.edu.cn
[3] School of Design, Southwest Jiaotong University, Chengdu 611756, China; shubo@swjtu.edu.cn (B.S.); cheny@my.swjtu.edu.cn (Y.C.)
* Correspondence: zhangruizhi@my.swjtu.edu.cn

Abstract: Geological hazards in Southern Sichuan have become increasingly frequent, posing severe risks to local communities and infrastructure. This study aims to predict the spatial distribution of potential geological hazards using machine learning models and ArcGIS-based spatial analysis. A dataset comprising 2700 known geological hazard locations in Yibin City was analyzed to extract key environmental and topographic features influencing hazard susceptibility. Several machine learning models were evaluated, including random forest, XGBoost, and CatBoost, with model optimization performed using the Sparrow Search Algorithm (SSA) to enhance prediction accuracy. This study produced high-resolution susceptibility maps identifying high-risk zones, revealing a distinct spatial pattern characterized by a concentration of hazards in mountainous areas such as Pingshan County, Junlian County, and Gong County, while plains exhibited a relatively lower risk. Among different hazard types, landslides were found to be the most prevalent. The results further indicate a strong spatial overlap between predicted high-risk zones and existing rural settlements, highlighting the challenges of hazard resilience in these areas. This research provides a refined methodological framework for integrating machine learning and geospatial analysis in hazard prediction. The findings offer valuable insights for rural land use planning and hazard mitigation strategies, emphasizing the necessity of adopting a "small aggregations and multi-point placement" approach to settlement planning in Southern Sichuan's mountainous regions.

Keywords: rural settlements; potential geological hazards; machine learning; hazard prevention and mitigation; Southern Sichuan region

check for updates

Academic Editors: Giulio Iovine, Gan Feng and Qiao Lyu

Received: 7 January 2025
Revised: 4 March 2025
Accepted: 5 March 2025
Published: 10 March 2025

Citation: Zhang, R.; Zhang, D.; Shu, B.; Chen, Y. Predicting the Spatial Distribution of Geological Hazards in Southern Sichuan, China, Using Machine Learning and ArcGIS. *Land* **2025**, *14*, 577. https://doi.org/10.3390/land14030577

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In recent years, with the development of China's social economy and the gradual improvement in people's quality of life, people have begun to realize the importance of disaster prevention and mitigation for production and living spaces, and their awareness of preventing geological hazards has also been gradually enhanced. Compared with cities, rural areas with relatively weak supporting facilities for public services and infra-structure are facing more severe challenges when dealing with natural hazards [1]. As an important urban-intensive area in Sichuan Province, Southern Sichuan has a large and widely distributed regional population and a relatively high level of urbanization. How-ever, affected by regional topography, geology, human factors, and other factors, this region has also

become one of the areas in Sichuan Province that suffer the most economic losses due to geological hazards. Therefore, clarifying the distribution characteristics of geological hazards in Southern Sichuan, summarizing the relationship between them and the spatial distribution of rural and township residents, and putting forward relevant suggestions and prevention measures for hazard prevention and mitigation in rural and township areas in Southern Sichuan to reduce the losses of life and property caused by natural hazards to local residents have become the most urgent tasks in the process of rural and township construction in this region.

Geological hazard prediction has been a critical research topic, with numerous methodologies developed to improve predictive accuracy. Early studies mainly relied on traditional statistical models such as logistic regression and analytical hierarchy processes (AHPs) [2–6]. These methods provided some interpretability but were often limited by linear assumptions.

With advancements in machine learning, various models have been applied to hazard prediction, including decision trees [7–11], random forests [12–15], and support vector machines [16–21]. More recently, ensemble learning approaches such as gradient boosting trees [22–27], multi-layer perceptron model [28,29] and CatBoost [30–33] have gained attention due to their ability to capture complex, nonlinear relationships. However, different machine learning models exhibit varying performances depending on hyperparameters and feature selection [34,35].

To address these challenges, hyperparameter optimization techniques, such as genetic algorithms, Bayesian optimization, and particle swarm optimization, have been introduced [36–39]. Among them, the Sparrow Search Algorithm (SSA) has shown superior exploration capabilities and optimization efficiency [40,41]. Despite these advancements, existing studies have rarely compared different learning models under uniform conditions or systematically evaluated the impact of hyperparameter tuning on prediction accuracy.

Thus, this study aims to fill this gap by systematically comparing multiple learning models, optimizing hyperparameters using SSA, and applying them to the geological hazard prediction of Southern Sichuan. This approach will provide a more robust and generalized methodology applicable to other hazard-prone regions.

Geological hazards, such as landslides and debris flows, pose significant threats to human settlements worldwide, particularly in mountainous and geologically unstable regions. Rural communities in many developing countries are particularly vulnerable due to weak infrastructure and limited disaster prevention measures. To address this issue, this study focuses on Southern Sichuan, a region highly prone to geological hazards, aiming to identify the spatial distribution patterns of potential hazards and develop a machine learning-based prediction model. Given that similar geological conditions exist in other mountainous areas globally, the findings of this study have broad applicability, contributing to improved hazard prediction and mitigation strategies beyond the studied region. By analyzing their relationship with residential settlements, this study seeks to reveal the spatial patterns and influencing factors of geological hazards. Furthermore, using machine learning models and optimization techniques, this research predicts potential geological hazards and provides a more refined scientific basis for future rural planning and hazard prevention efforts. The findings will serve as a valuable reference for formulating policies that minimize geological risks, ensuring the safety and sustainability of rural living environments in Southern Sichuan.

2. Study Area

2.1. General Overview

Yibin City is located in the southwest of the Southern Sichuan region in Sichuan Province (Figure 1). Among the four cities in Southern Sichuan (Yibin, Luzhou, Zigong, and Neijiang), it has the largest territorial area and the most potential geological hazards (Table 1). The whole city is mainly composed of mountains and hills. Specifically, mountains with an altitude ranging from 500 to 2000 m account for approximately 46.6%, hills account for 45.3%, and flatland areas account for 8.1%. The complexity of the topography has led to a large number and wide distribution of potential geological hazards throughout Yibin City. There are approximately 2700 potential geological hazards in total across the city, including 752 collapse sites, 282 unstable slope sites, 1456 landslide sites, and 206 sites of other types of geological hazards. The density of hazard sites across the whole region is 0.2 per square kilometer, and the average number of hazards per 10,000 people is 5.85 (Figure 2). There are a total of 20,913 rural settlements in the whole region, mainly distributed in Xuzhou District, Nanxi District, Junlian County, and Cuiping District (Figure 3). The distribution range has a relatively high degree of coincidence with that of potential geological hazards.

Table 1. Basic data table of potential geological hazards in cities of Southern Sichuan.

City Name	Number of Hazards	Population (in Ten Thousands)	Area (km^2)	Density of Hazards (per Square Kilometer)	Number of Hazards (per 10,000 People)
Yibin	2700	461.8	13,271	0.2034	5.85
Zigong	336	323.94	4381	0.0767	1.04
Luzhou	303	508.54	12,236	0.0248	0.60
Neijiang	792	411.77	5385	0.1471	1.92

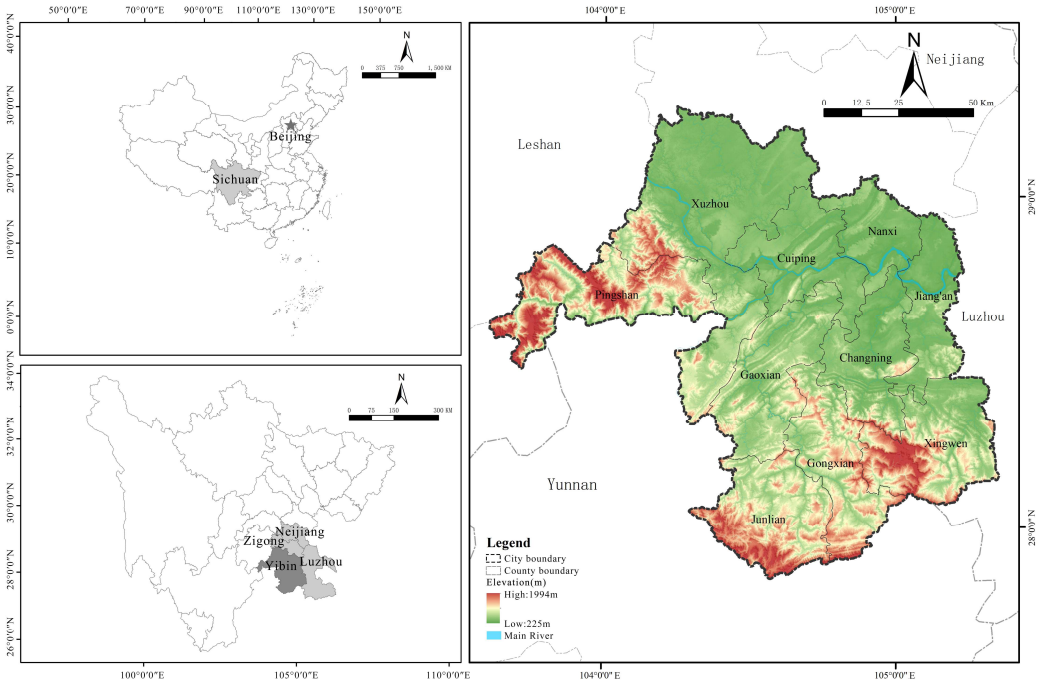


Figure 1. Location of the study area.

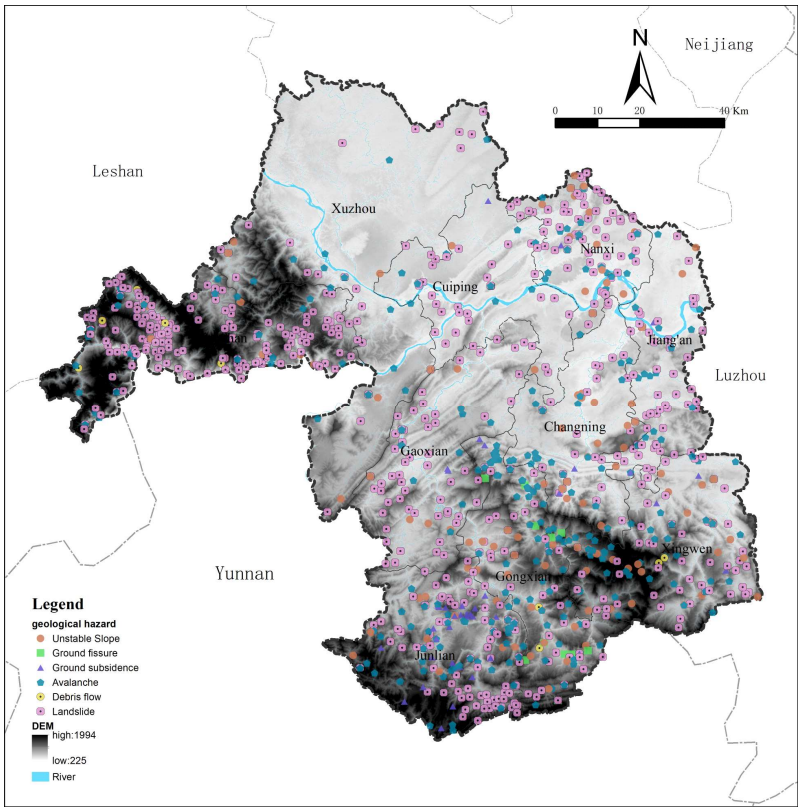


Figure 2. Spatial location map of potential geological hazards in Yibin City.

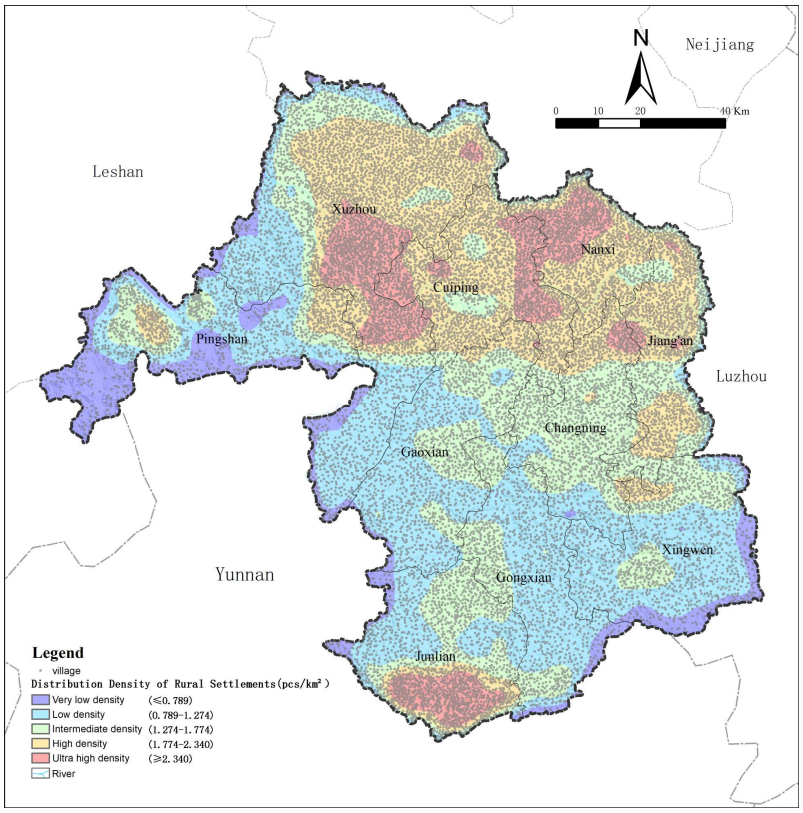


Figure 3. Distribution density map of rural settlements in Yibin City.

2.2. Spatial Distribution Characteristics of Potential Geological Hazards

The potential geological hazards in Yibin City exhibit obvious spatial differences. On the whole, this city can be generally summarized as "sparse in plains and dense in mountainous areas". Pingshan County, Junlian County, and Gong County, which have the highest altitudes in the city and where mountains are concentrated, are the areas with a high concentration of potential geological hazards, with a density ranging from 0.57 to 0.93 per square kilometer. Analyzed by different types of geological hazards, debris flows are mainly distributed in Pingshan County, Xingwen County, Gong County, and Junlian County; landslides are mainly distributed in Pingshan County, Nanxi District, Junlian County, and Jiang'an County; ground collapses are mainly distributed in Junlian County and Gao County; ground fissures are mainly distributed in Gong County; unstable slopes are mainly distributed in Changning County, Gong County, Nanxi District, and Xingwen County; and collapses are mainly distributed in Junlian County, Gong County, and Changning County (Figure 4).

Through an overlay analysis of the buffer zones of geological hazards in Yibin City and the spatial distribution of rural and township settlements, it is found that 59.63% of the settlements have one or more potential geological hazards within a range of 2.5 km, and 19.8% of the settlements have one or more potential geological hazards within a range of 1 km. This is consistent with the actual research situation; that is, geological hazards occur frequently in the rural and township areas in Southern Sichuan. Although the settlements have taken potential geological hazards into consideration during the site selection process, they still "coexist side by side with potential geological hazards and live in symbiosis with hazards" (Figure 5).

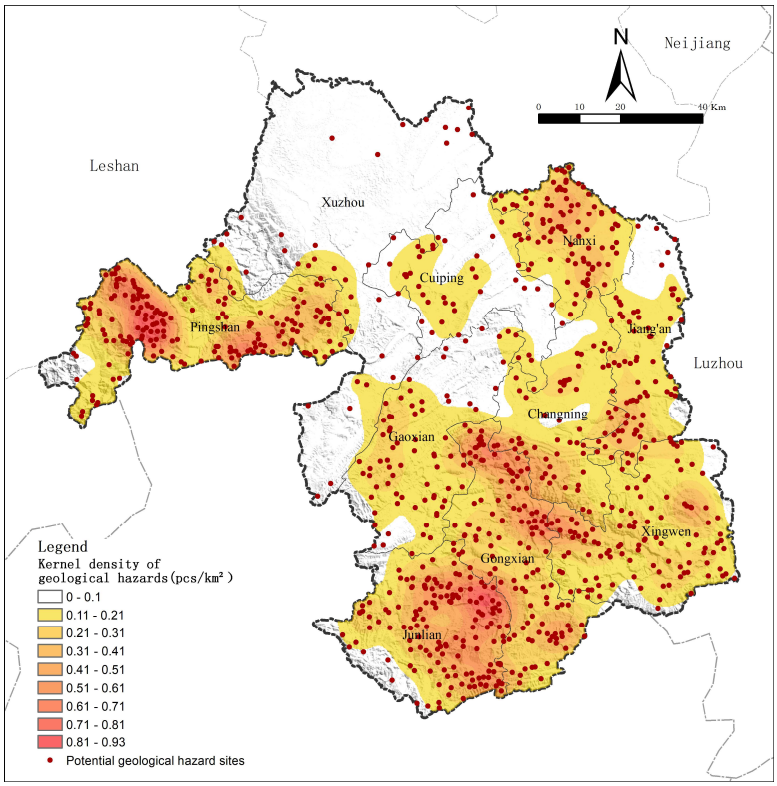


Figure 4. Kernel density of geological hazards.

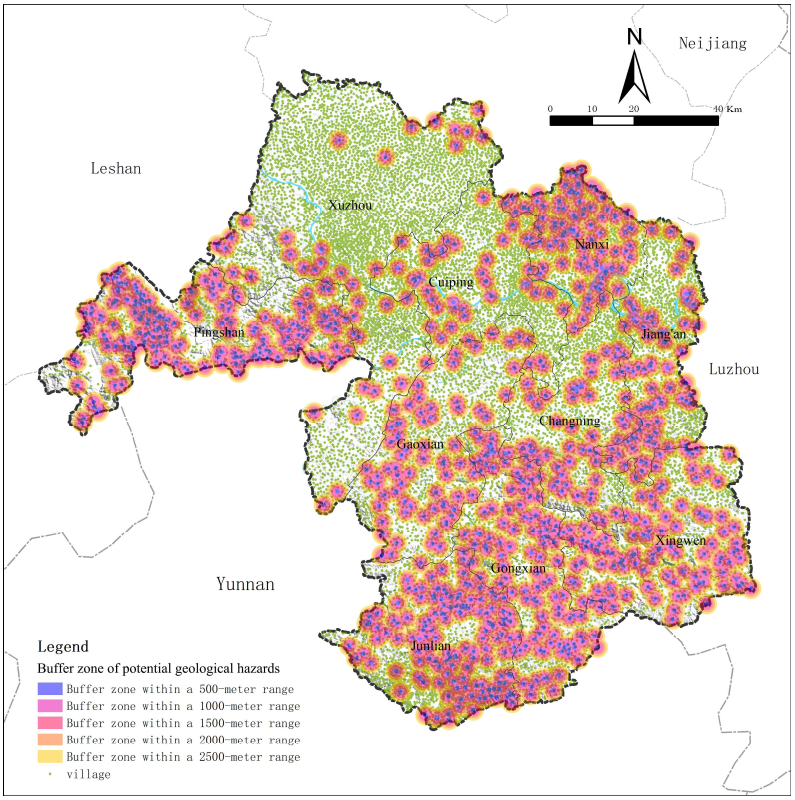


Figure 5. Relationship map between the buffer zones of potential geological hazards and the settlement space.

3. Data and Methodology

3.1. Data

(1) The data on potential geological hazards across Yibin City were given by the Sichuan Provincial Department of Natural Resources. For this study, the data on geological hazards during the season with the highest frequency of geological hazards were selected, with a total of 2700 potential geological hazards.

(2) The visual data of rural settlements, water systems, and roads in the study were sourced from the "National Science and Technology Infrastructure Platform—National Earth System Science Data Center". Data selection and processing were carried out according to the research objects.

(3) The topographic and geomorphic data of Yibin City were sourced from the "Geospatial Data Cloud", and these are GDTM data with a resolution of 30 m.

(4) Data on population, area, etc., were sourced from the *Yibin Statistical Yearbook*.

3.2. Methodology

Machine learning is an advanced computational technique that enables systems to identify patterns and make predictions based on data. It has been widely applied in various scientific domains, including natural hazard assessment, climate prediction, and risk management. In the field of geological hazard prediction, machine learning can process large datasets, analyze complex spatial relationships, and generate highly accurate hazard susceptibility maps.

By using methods combining the Pearson correlation coefficient method, machine learning method, statistical index evaluation method, ROC curve statistical method, and on-site investigation, an overlay analysis was conducted on spatial elements such as potential geological hazards, geological conditions, rural and township settlements, and rivers and roads in Yibin City. The spatial distribution characteristics of potential geological hazards

in rural and township settlements in Southern Sichuan and their main influencing factors were summarized. Then, relying on the machine learning prediction method and based on the main influencing factors, predictions were made for the areas in the Yibin region where potential geological hazards may occur in the future, with the aim of achieving the goal of hazard prevention and mitigation (Figure 6). In terms of software and program implementation details, we employed the Visual Studio Code editor (version 1.97.2) along with Python (version 3.11) to analyze the influencing factors of geological hazard risk points. Additionally, predictive modeling was conducted to obtain more reliable forecasting results.

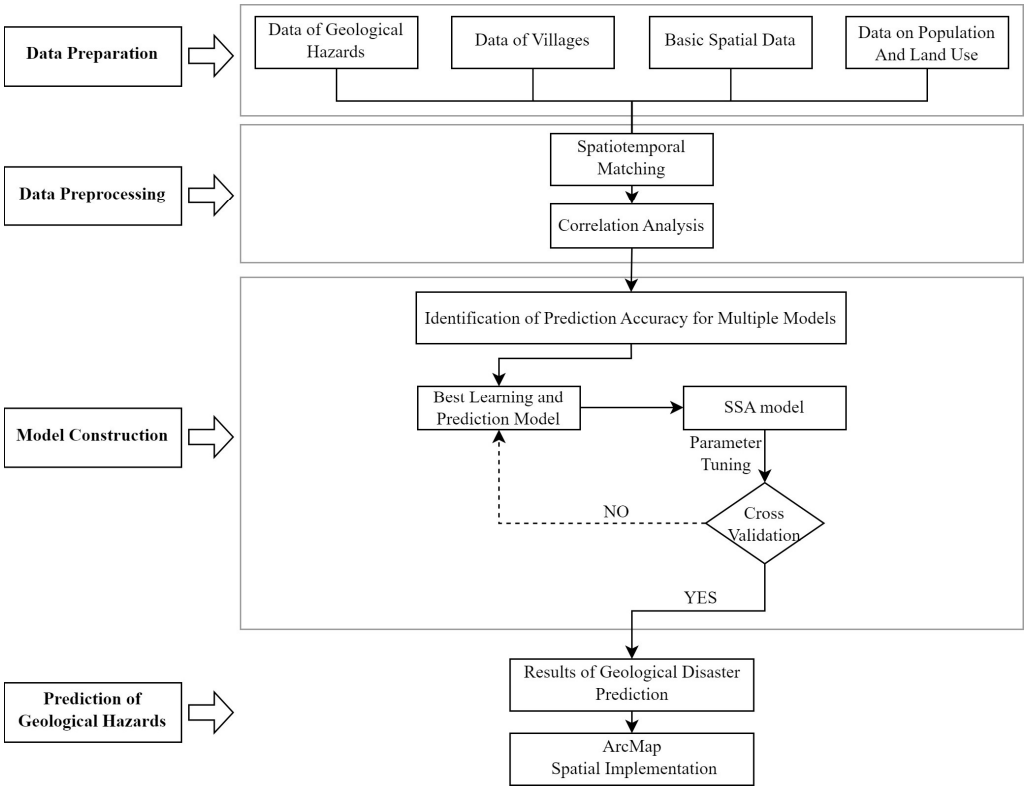


Figure 6. Research flow chart.

3.2.1. Correlation Analysis Model

The Pearson correlation coefficient is designed to measure the strength of a linear association between two variables x and y, where the value r = 1 denotes a perfect positive correlation and the value r = −1 denotes a perfect negative correlation, At the same time, the significance of the Pearson correlation analysis should also be tested, that is, the *p*-value should be greater than 0.05. The following equation is used to find out whether variable x and variable y are correlated [42].

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \quad (1)$$

In the formula, $\overline{x}$ is the average value of the samples x_1, x_2, x_3, …, x_n, and $\overline{y}$ is the average value of the samples y_1, y_2, y_3, …, y_n.

3.2.2. Decision Tree Model

The core algorithm of the decision tree model is a top-down tree-growth algorithm [9]. Firstly, the dataset is divided into two samples, namely the training set and the test set. And the information gain is taken as the criterion for selecting and testing attributes.

The decision tree model is constructed through the samples in the training set. Secondly, "pruning" work is carried out through the samples in the test set to further optimize the decision tree. Therefore, the quality of its decision results is correlated to a certain extent with the selection of attributes of the data test set, the pruning of nodes, and the control of tree parameters (Figure 7).

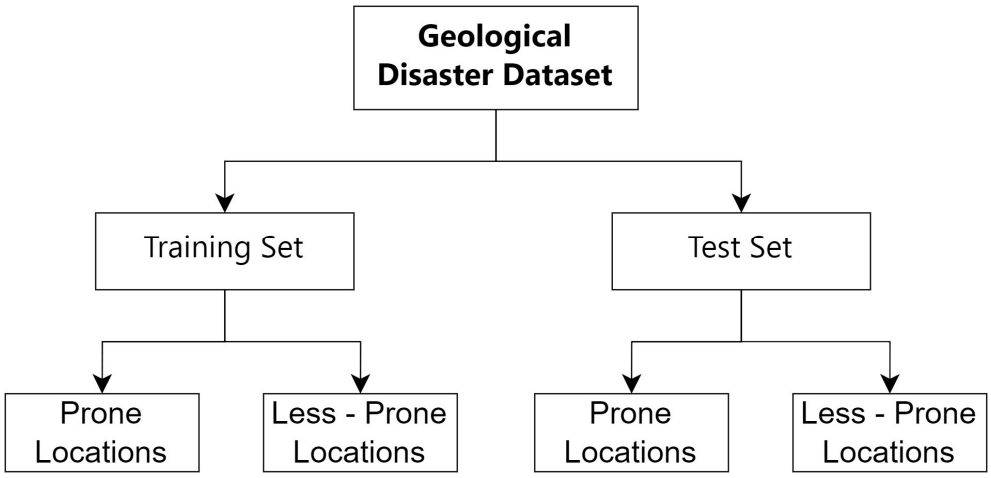


Figure 7. Calculation flow chart of the decision tree learning and prediction model.

The calculation formula for information gain is as follows:

$$IG(S, A) = H(S) - H(S|A) \tag{2}$$

$$H(S) = -\sum_{j=1}^{j=m} P_j \log_2 P_j \tag{3}$$

In the formula, S represents the dataset; A represents the feature of the dataset; H(S) is the entropy of the dataset S; and Pj is the probability of a specific value appearing in the dataset S.

3.2.3. Random Forest Model

Random forest is an integrated algorithm model based on multiple decision trees [12]. It reduces the correlations of the dataset by obtaining random samples and features through bootstrap sampling, thereby improving the generalization ability of the model. Therefore, random forest can effectively control the overfitting problem and has relatively high robustness to noise and outliers (Figure 8).

The construction process of random forest is as follows:

(1) Bootstrap sampling: Randomly draw M samples from the training set samples with replacements to form a training set K for one decision tree.

(2) Feature selection: Randomly select *n* features from N features as the alternative feature subset for decision-making under the current node, and then use the method of information gain to select the feature with the largest information gain among the n features as the splitting attribute of this node.

(3) Decision tree generation: Generate decision trees by using the process of random feature selection.

(4) Random forest construction: Construct multiple decision trees through the steps of bootstrap sampling, feature selection, and decision tree generation to form a random forest.

(5) Prediction result generation: Form the final prediction result by classifying the prediction results of all decision trees.

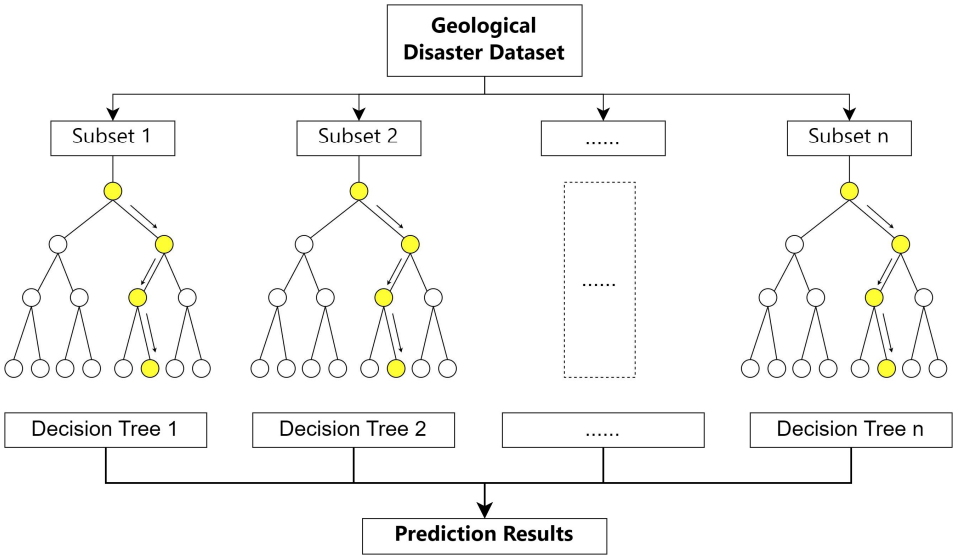


Figure 8. Calculation flow chart of the random forest model.

3.2.4. Gradient Boosting Tree Model

The Gradient Boosting Tree Model is a type of boosting algorithm [22]. It consists of three parts: a regression tree, gradient iteration, and shrinkage. The first tree is generated based on the original data, and the prediction accuracy is improved through data learning. At the same time, regularization methods can be used to prevent the occurrence of model overfitting.

The construction process of gradient boosting is as follows:

(1) Data input: Input the training set T = {(X_1,y_1), (X_2,y_2), . . ., (X_N,y_N)}, where X_i is the independent variable of the dataset and y_i is the dependent variable of the dataset. The maximum number of iterations is M, and the loss function is L(y,f(X)).

(2) Initialization:

$$f_0(X) = \text{argmin}\sum_{i=1}^{n} L(y_i, c) \tag{4}$$

(3) Iteration: m = 1, 2, . . ., M

(a) Calculate for i = 1, 2, 3, . . ., N:

$$r_{mi} = -\left\{\frac{\partial L[y_i, f(x_i)]}{\partial f(x_i)}\right\} f(x) = f_{m-1^{(x)}} \tag{5}$$

(b) Fit a regression tree to r_{mi} to obtain the regional nodes R_{mj} of the m-th tree, where j = 1, 2, 3, . . ., J.

(c) Calculate for j = 1, 2, 3, . . ., J.

(d) θ represents the set of samples that fall into the current node.

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J} c_{mj}\theta(x \in R_{mj}) \tag{6}$$

(4) Generate a regression tree model.

$$\hat{f}(x) = f_M(x) = \sum_{m=1}^{M}\sum_{j=1}^{J} c_{mj}\theta(x \in R_{mj}) \tag{7}$$

3.2.5. Support Vector Machine Model

The support vector machine model is a machine learning method based on statistical learning theory [16]. Its basic theory is to seek the optimal hyperplane that maximizes the margin in the feature space. This hyperplane can maximize the sample margin between the potential geological hazard points and the non-potential geological hazard points in the training dataset.

The construction process of the support vector machine (SVM) is as follows:

Data input: Input the training set T = {(X_1,y_1), (X_2,y_2), . . ., (X_N,y_N)}, where X_i is the independent variable of the dataset and y_i is the dependent variable of the dataset, and $y_i \in (-1,1)$.

(1) Generate the decision function:

$$g(X) = \omega^T \phi(X) + b \tag{8}$$

(2) Generate the minimization function:

$$g(\omega, \xi) = \| \omega \|^2 + C \sum_{i=1}^{N} \xi_i \tag{9}$$

(3) Generate the constraint conditions:

$$y_i((\omega, \phi(X)) + b) \geq 1 - \xi_i, \xi_i \geq 0 \tag{10}$$

In the formula, ϕ(x) is the mapping of samples from the input space to the high-dimensional feature space. ω is the normal vector; b is the displacement term; X_i and y_i are the sample quantities of the training set; C is the regularization parameter; and ξ_i is the slack variable.

3.2.6. Multilayer Perceptron Model

The multilayer perceptron model is a type of neural network, and it has outstanding nonlinear prediction capabilities. The model consists of three layers, namely the input layer, the hidden layer, and the output layer [20]. The neurons between different layers determine the corresponding relationship between the input and output vectors through changes in weights, thus forming a network structure with decision-making capabilities.

The construction formulas for the multilayer perceptron are as follows:

$$net_j = \sum_{t=1}^{i} w_{ij} p_i \tag{11}$$

In the formula, W_{ij} represents the weight value; i and j respectively denote the number of the input neurons and the hidden neurons; p_i represents the variable of the input neurons; t represents the number of input neurons.

3.2.7. CatBoost Model

CatBoost is a machine learning algorithm based on Gradient Boosting Decision Trees (GBDT), which can be applied to different machine learning tasks such as classification and regression [30]. Compared with other machine learning algorithms based on GBDT, the improvement in its algorithm lies in the fact that the model adopts a symmetric tree structure and a ranking-based target encoding method. That is to say, it not only improves the efficiency of model training and prediction but also avoids the common problems of the high dimensionality of categorical features and the risk of overfitting.

The calculation formulas of the CatBoost model are as follows:

$$X_{ik} = \frac{\sum_{j=1}^{n} \left[x_{jk} = x_{ik} \right] \cdot Y_j}{\sum_{j=1}^{n} \left[x_{jk} = x_{ik} \right]} \tag{12}$$

$$X_{\delta k,k} = \frac{\sum_{j-1}^{p-1} \left[x_{\delta jk} = x_{\delta pk} \right] \cdot Y_{\delta j} + \alpha \cdot p}{\sum_{j-1}^{p-1} \left[x_{\delta jk} = x_{\delta pk} \right] + \alpha} \tag{13}$$

In the formula, X_{ik} represents the mean value of the dataset label values; x_{ik} denotes the value of the I–th categorical feature of the k–th training sample; Y_j represents the label value (0 and 1) of the target variable; $X_{\delta k,k}$ represents the value used for replacement; δ is the random permutation sequence generated in the randomly permuted dataset; and p and α are the prior value and the prior value weight of the sample label, respectively.

3.2.8. Comparison of Model Accuracy

In the process of machine learning and prediction, the accuracy of learning and prediction is often used as the main way to evaluate performance [34]. For the evaluation of accuracy, this paper uses indicators such as True Positive (TP), True Negative (TN), recall, accuracy (ACC), F1 score, AUC value, and Receiver Operating Characteristic (ROC) curve. True Positive (TP) refers to the number of potential geological hazard points that can be accurately identified by the model during the learning and prediction process. True Negative (TN) refers to the number of non-potential geological hazard points that can be accurately identified by the model during the learning and prediction process. Recall refers to the measure of how well the learning and prediction model correctly detects True Positives. Accuracy (ACC) is used to evaluate the probability that samples in the learning and prediction model can be correctly identified in classification problems. The F1 score refers to the harmonic mean of recall and precision in the learning and prediction model. In the calculation, false positives and false negatives are also taken into account simultaneously to make the model more accurate. The specific formulas are as follows:

$$\text{Recall} = \text{TP}/\text{TP} + \text{FN} \tag{14}$$

$$\text{ACC} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \tag{15}$$

$$\text{F1 Score} = 2\text{TP}/(2\text{TP} + \text{FP} + \text{FN}) \tag{16}$$

The ROC curve is widely used in the evaluation of the performance of classification models. The closer the ROC curve is to the upper left corner of the graph, the better the performance of the model is. The AUC (area under the curve) value is the area under the curve calculated based on the ROC curve, and it is used to quantify the overall performance of the learning and prediction model. Its value range is from 0 to 1, and the closer the value is to 1, the better the learning and prediction effect is.

3.3. Model Optimization

The Intelligent Sparrow Search Algorithm (SSA) is adopted to adaptively optimize the parameters of the learning model in order to improve the accuracy of model learning. The best relevant parameters of the prediction model for the distribution characteristics of potential geological hazard points in Yibin area are updated [40], providing support for further improving the prediction results of potential geological hazard points. The Sparrow Intelligent Algorithm can be abstracted as the "Discoverer–Joiner" model, and an early warning mechanism is added on this basis.

4. Analysis and Results

4.1. Correlation Analysis

In the Southern Sichuan region, the main types of geological hazards are landslides and collapses. Previous studies have confirmed that landslides and collapses are mainly related to factors such as water systems, geology, topography, and engineering excavation. Considering that there may be differences in the occurrence mechanisms of geological hazards in different regions, the Pearson correlation coefficient is used to reveal the correlations between different types of geological hazards and relevant influencing factors in the research area. The results are shown in Figure 9.

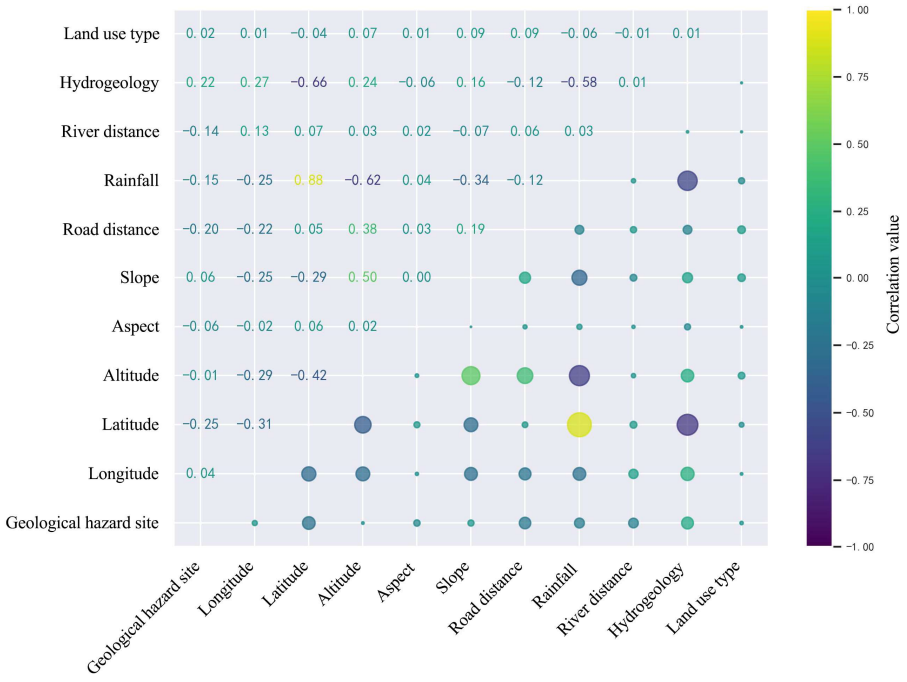


Figure 9. Correlation analysis diagram of each influencing factor and geological hazards.

Based on the above correlation analysis and combined with the current situation, the following conclusions can be drawn:

(1) The influencing factors on potential geological hazards, ranked from strong to weak, are latitude, hydrogeology, distance to roads, distance to rivers, rainfall, slope gradient, slope aspect, latitude, land use type, and elevation. Among them, hydrogeology, slope gradient, longitude, and land use type are positively correlated with potential geological hazards, which is consistent with the distribution characteristics of potential geological hazards in this area. From the analysis of the hydrogeology, it can be seen that the main hydrogeological type in this area is clastic rock aquifers (including moderate and weak types), accounting for 73.21% of the whole region. Among these types, there are 970 landslides, 274 collapses, and 140 unstable slopes, accounting for 51.26% of all potential geological hazards (Figure 10). From the analysis of slope gradients, it can be seen that the number of potential geological hazards first increases and then decreases with an increase in slope gradient. Among them, there are 402 gentle slopes, 1144 slopes and 862 steep slopes, accounting for 89.19% of all potential geological hazards (Figure 11). From the analysis of land use types, it can be seen that the land use types with the most distributed potential geological hazards are grassland, cultivated land, and forest land, respectively. Among them, there are 390 potential geological hazards in grassland, 1728 in cultivated land, and 314 in forest land, accounting for 90.07% of all potential geological hazards (Figure 12).

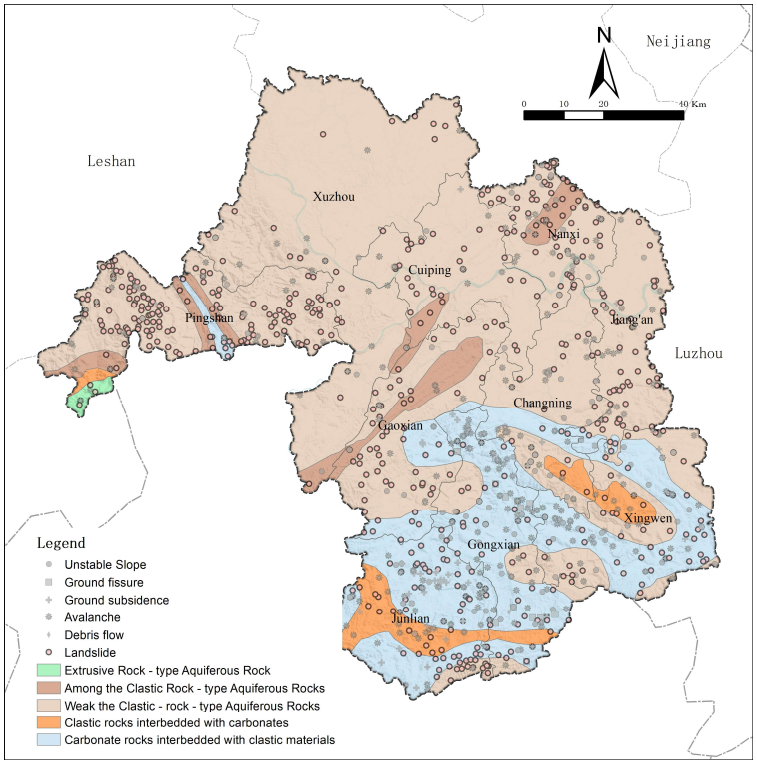


Figure 10. Relationship diagram between potential geological hazards and hydrogeological types.

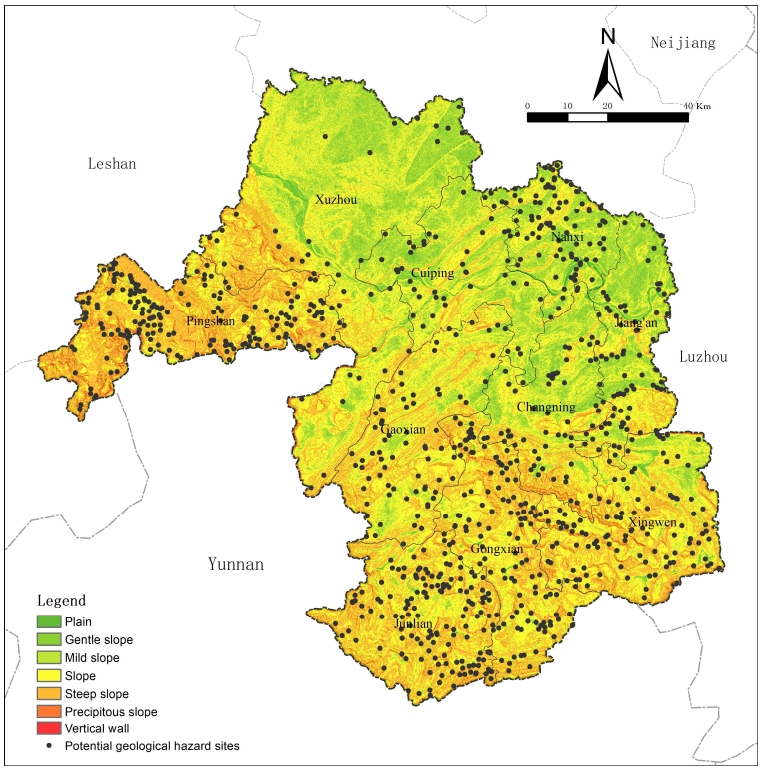


Figure 11. Relationship diagram between potential geological hazards and slope gradient.

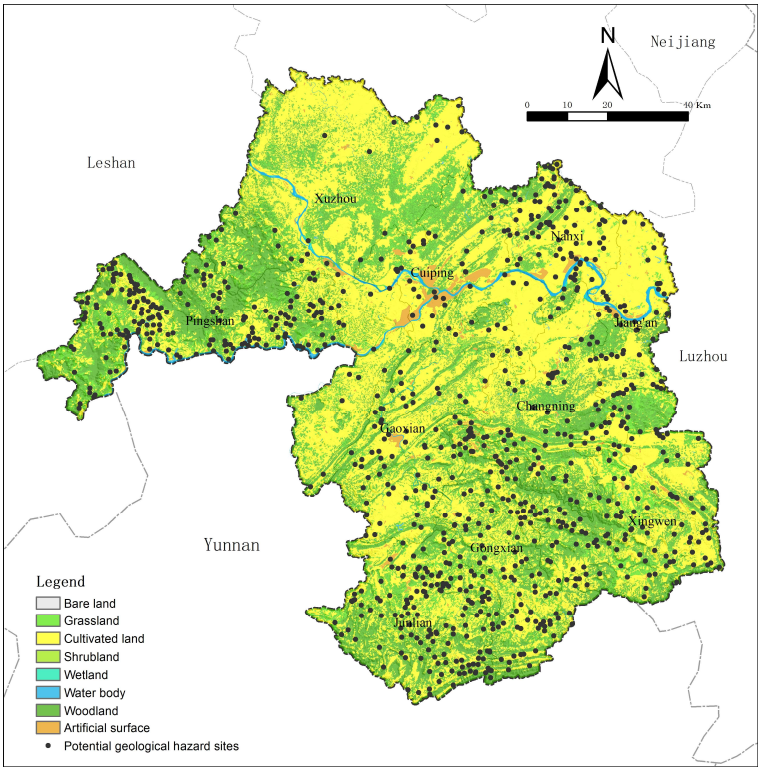


Figure 12. Relationship diagram between potential geological hazards and land use types.

(2) Latitude, distance to roads, distance to rivers, rainfall, slope aspect, and elevation are negatively correlated with potential geological hazards. As latitude, distance to roads, and distance to rivers increase and rainfall decreases, the number of potential geological hazards gradually decreases, which is consistent with the analysis results shown in Figures 13–15. Since the statistics of potential geological hazards are related to the distribution of residential areas, as elevation increases, the number of residential areas decreases, and the statistical number of hazard potential sites will also decrease accordingly. Therefore, it is not analyzed as a main influencing factor.

(3) The overall terrain of Yibin City is higher in the south and lower in the north. Moreover, the distribution of carbonate rocks interbedded with clastics is wider, and there are more high-slope areas in the south than in the north. Therefore, although the rainfall gradually decreases as the latitude decreases, the number of potential geological hazards gradually increases. The total number of potential geological hazards in Junlian County, Gong County, and Xingwen County is 1166, accounting for 43.19% of the whole area, indicating a large number of such sites.

(4) Through the Pearson correlation analysis, it can be obtained that there is no significant correlation among the above-mentioned influencing factors. Among them, the correlation between latitude and potential geological hazards is the highest, but only 0.25. Therefore, when predicting potential geological hazards, the above-mentioned influencing factors can all be included as input variables for research.

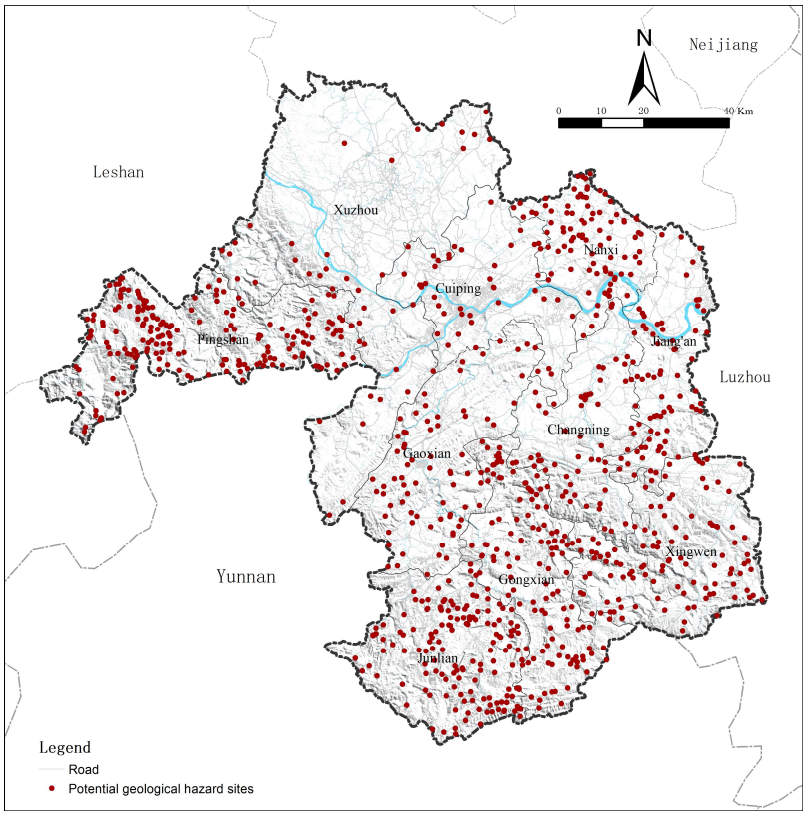


Figure 13. Relationship diagram between potential geological hazards and distance to roads.

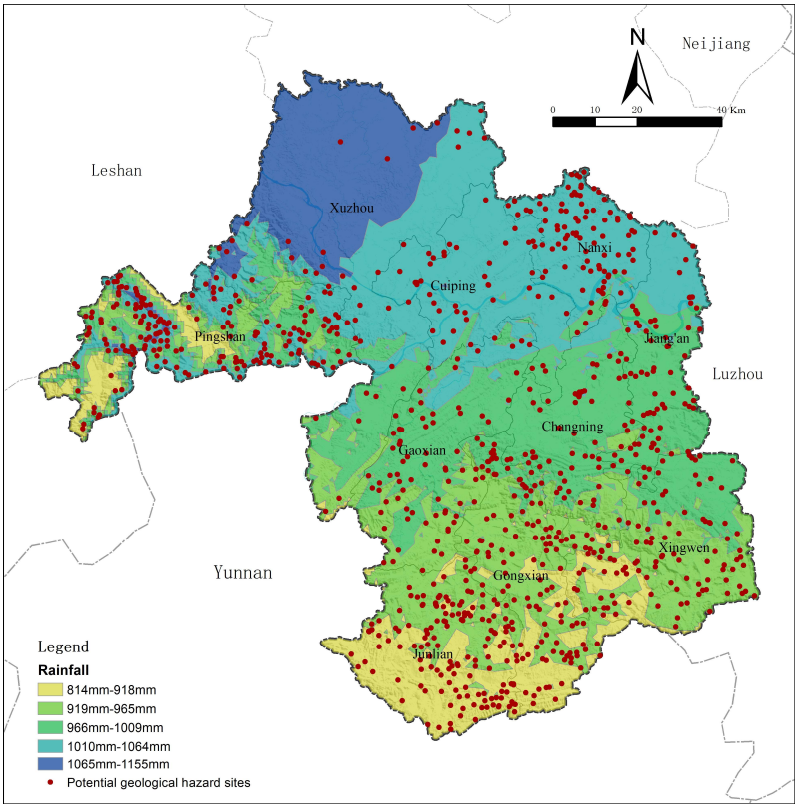


Figure 14. Relationship diagram between potential geological hazards and rainfall.

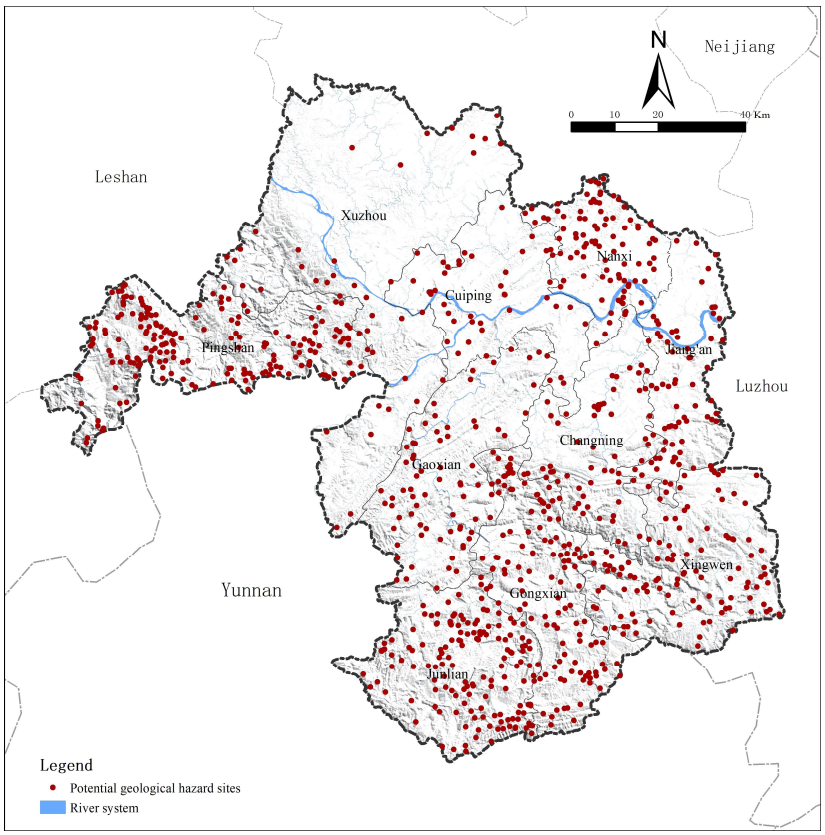


Figure 15. Relationship diagram between potential geological hazards and distance to rivers.

4.2. Model Verification and Analysis

Based on the Pearson correlation analysis and the spatial distribution characteristics of geological hazard points and urban–rural settlements in the study area, it is found that the occurrence of potential geological hazard points in Yibin City is positively correlated with hydrogeology, slope, longitude, and land use type and negatively correlated with latitude, distance to roads, distance to rivers, rainfall, slope aspect, and elevation. Moreover, there is no significant correlation among the influencing factors (p value > 0.5). Therefore, in this study, variables such as land use type, hydrogeological type, distances to rivers and roads, rainfall, slope, slope aspect, elevation, and latitude were determined as the input data for the model construction process to participate in the model calculation.

In this study, Yibin City was divided into grid units of 2.5 km × 2.5 km, and a total of 4347 pieces of data were generated. Among them, 2700 pieces of data were the data of actual potential geological hazard points, and 2327 pieces of data were the data of predicted potential geological hazard points. The dataset was then divided into a training set and a test set at a ratio of 7:3. Through the calculation of True Positive, True Negative, recall, and F1 score for the training sets and test sets of various machine learning models, it was concluded that the CatBoost learning and prediction model had the best performance (Table 2). Therefore, based on the CatBoost model, this paper conducts learning and prediction for the potential geological hazard points in Yibin City by means of optimizing the SSA model to improve the prediction accuracy.

Table 2. Statistical table of the degree of fitting of various machine learning models.

Model	Training Set					Test Set				
	TP	TN	Recall	ACC	F1 Score	TP	TN	Recall	ACC	F1 Score
Decision Tree	1894	1148	1.00	1.00	1.00	762	345	0.95	0.86	0.89
Random Forest	1894	1148	1.00	1.00	1.00	760	378	0.94	0.86	0.90
Gradient Boosting	1804	844	0.95	0.86	0.90	748	320	0.93	0.81	0.86
CatBoost	1868	1057	1.00	1.00	1.00	754	385	0.97	0.93	0.95
Support Vector Machine	1742	689	0.92	0.79	0.85	730	258	0.91	0.75	0.82
Multilayer Perceptron	1713	755	0.90	0.81	0.86	699	268	0.87	0.75	0.81

Through optimization and improvement by applying the SSA–CatBoost model, the learning and prediction performance of CatBoost has been significantly enhanced. The optimized parameter settings are as follows: iterations = 1000, learning-rate = 0.2, depth = 5, random-seed = 100. After optimization, the recall rate of the CatBoost learning set is 1.00, the recall rate of the test set is 0.98, the precision rate of the learning set is 1.00, the precision rate of the test set is 0.94, the F1 score of the learning set is 1.00, the F1 score of the test set is 0.96, the AUC of the learning set is 1.00, and the AUC of the test set is 0.99, showing excellent performance, as shown in Figure 16 and Table 3.

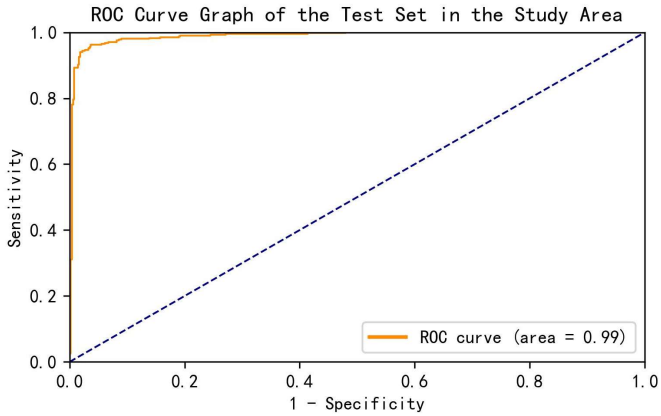


Figure 16. ROC curve diagram of the test set in the study area.

Table 3. Statistical table of the degree of fitting of the SSA–CatBoost model.

Evaluation Parameters	Training Set	Test Set
Recall	1.00	0.98
ACC	1.00	0.94
F1 Score	1.00	0.96
AUC	1.00	0.99

4.3. Prediction of the Susceptibility of Potential Geological Hazard Points

The SSA–CatBoost model is used for optimized predictions. Combined with the natural breaks method, the generated geological hazard susceptibility map is classified into different levels to produce a geological hazard susceptibility zoning map (Figure 17). Based on previous experience and the specific situation of geological hazards in Yibin City, the geological hazard susceptibility in the study area is divided into four levels, namely an extremely low-susceptibility area (ELSA), low-susceptibility area (LSA), medium-susceptibility area (MSA), and high-susceptibility area (HSA) (Table 4). It can be seen from the figure that the predicted potential geological hazard points are mainly distributed in the areas with higher terrain and steeper slopes on the south side of Yibin City, while there

are fewer potential geological hazard points in the areas with relatively lower terrain and gentler slopes. Moreover, the distribution of the number of potential geological hazard points in each district and county is uneven.

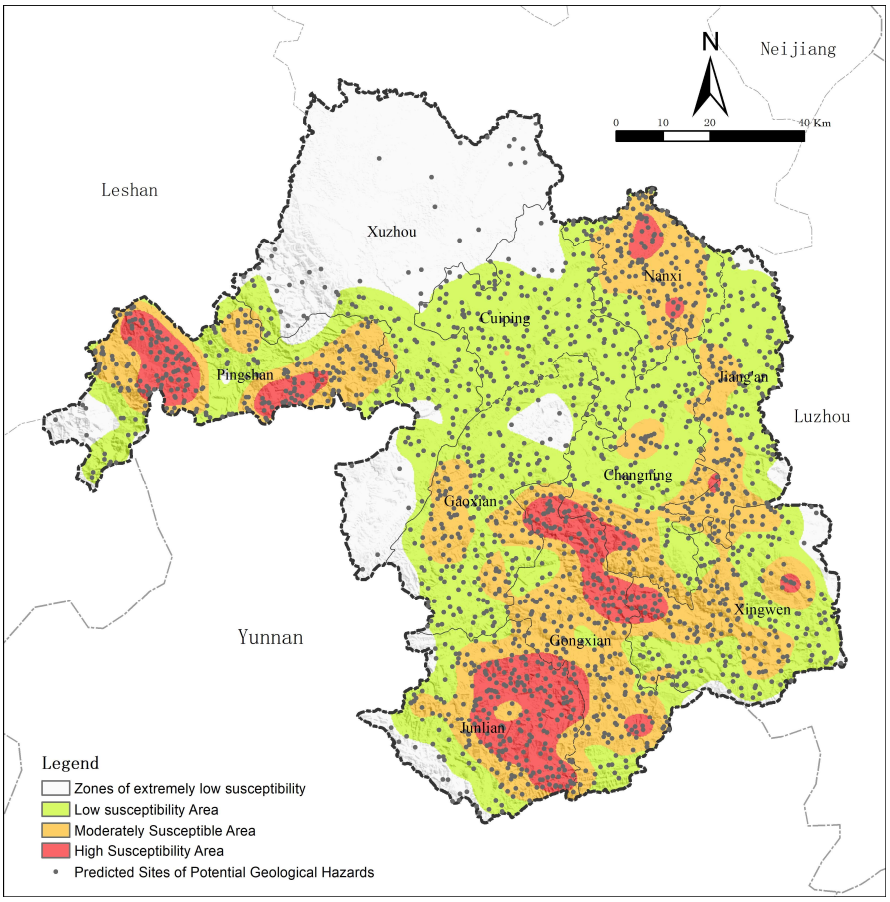


Figure 17. Prediction map of geological hazard susceptibility in Yibin City.

Combining this with the prediction statistical table of potential geological hazard points in Yibin City and the prediction map of geological hazard susceptibility, it can be seen that the potential geological hazard points in Yibin City are mainly distributed in Pingshan County, Junlian County, and Gong County, accounting for 46.21% of the total number of potential geological hazard points. Among them, Pingshan County accounts for 17.62%, Junlian County accounts for 15.71%, and Gong County accounts for 12.88%. According to the analysis of different geological hazard susceptibility zones, in the extremely low-susceptibility area, the numbers of potential geological hazard points in Xuzhou District, Junlian County and Cuiping District are relatively large, accounting for 47.73%, 15.91%, and 12.12%, respectively. In the low-susceptibility area, the numbers of potential geological hazard points in Gao County, Cuiping District, and Xingwen County are relatively large, accounting for 17.60%, 17.34%, and 13.39%, respectively. In the medium-susceptibility area, the numbers of potential geological hazard points in Pingshan County, Gong County, and Nanxi District are relatively large, accounting for 18.49%, 15.41%, and 14.71%, respectively. In the high-susceptibility area, the numbers of potential geological hazard points in Junlian County, Pingshan County, and Gong County are relatively large, accounting for 35.18%, 26.93%, and 20.44%, respectively (Table 4).

Table 4. Statistical table of the prediction of potential geological hazard points in Yibin City.

County Name	Geological Hazard Susceptibility Zoning									
	ELSA	Proportion	LSA	Proportion	MSA	Proportion	HSA	Proportion	Total	Proportion
Xu Zhou	63	47.73%	119	10.21%	3	0.21%	0	0.00%	185	4.79%
Cui Ping	16	12.12%	202	17.34%	1	0.07%	0	0.00%	219	5.67%
Nan Xi	0	0.00%	63	5.41%	210	14.71%	45	3.95%	318	8.23%
Jiang'an	6	4.55%	118	10.13%	140	9.80%	3	0.26%	267	6.91%
Chang Ning	0	0.00%	109	9.36%	141	9.87%	54	4.74%	304	7.87%
Ping Shan	10	7.58%	100	8.58%	264	18.49%	307	26.93%	681	17.62%
Xing Wen	8	6.06%	156	13.39%	186	13.03%	95	8.33%	445	11.51%
Gao Xian	8	6.06%	205	17.60%	126	8.82%	2	0.18%	341	8.82%
Gong Xian	0	0.00%	45	3.86%	220	15.41%	233	20.44%	498	12.88%
Jun Lian	21	15.91%	48	4.12%	137	9.59%	401	35.18%	607	15.71%
aggregate	132	100.00%	1165	100.00%	1428	100.00%	1140	100.00%	3865	100.00%

5. Discussion

5.1. Application of Machine Learning in Disaster Prediction

This study applies machine learning techniques to analyze the spatial distribution characteristics of geological hazard risk points in Yibin City and their relationship with rural settlements. Compared with traditional geological hazard prediction methods, machine learning offers advantages in handling large-scale spatial data and identifying key influencing factors with a higher accuracy. The results highlight that CatBoost, optimized with the Sparrow Search Algorithm (SSA), significantly enhances the prediction performance, offering a more reliable approach for hazard prevention and mitigation planning.

In this study, CatBoost showed the best performance and was compared with other studies in terms of similarities and differences (Table 2), because many studies have obtained better results from random forests or other models.

5.2. Spatial Correlation Between Geological Hazards and Rural Settlements

One of the critical findings is the clear spatial correlation between geological hazard risk points and rural settlements, particularly in Pingshan County and Nanxi District. The study identifies key environmental factors influencing the hazard distribution, including slope steepness, high road and river density, and hydrogeological conditions dominated by clastic rock aquifers. These findings suggest that hazard prevention strategies should prioritize monitoring and mitigation in areas with similar characteristics. Meanwhile, the strong spatial correlation between geological hazard risk points and rural settlements in Pingshan County and Nanxi District suggests that these regions are particularly vulnerable. Given that these areas exhibit steep slopes, high precipitation, and complex hydrogeological conditions, targeted intervention strategies, such as improved drainage systems and slope stabilization measures, should be prioritized. Moreover, this study highlights the importance of integrating high-resolution spatial data with machine learning to enhance hazard prediction accuracy.

Previous studies on geological hazard prediction have primarily focused on traditional statistical models, which often struggle to handle nonlinear relationships in large-scale spatial datasets [43–46]. In contrast, this study demonstrates that machine learning models, particularly CatBoost optimized with SSA, outperform traditional models by improving the classification accuracy and reducing false positives. The integration of SSA optimization in geological hazard prediction has not been extensively explored in previous research, marking a significant methodological contribution.

5.3. Policy Implications and Future Research Directions

From a policy perspective, these findings suggest that local governments should prioritize hazard mitigation strategies in high-risk rural areas by implementing stricter land use regulations and promoting sustainable construction practices. Additionally, the application of machine learning in hazard risk assessment can support real-time monitoring systems, enabling authorities to develop early-warning systems tailored to high-risk zones.

While this study provides valuable insights into geological hazard prediction, it has certain limitations. First, the model relies on currently available geospatial data, which may not capture the long-term effects of climate change on hazard susceptibility. Future research should incorporate climate projections and analyze temporal variations in hazard risk. Second, this study does not fully account for human-induced factors such as land use changes and deforestation, which can significantly alter hazard susceptibility. Integrating remote sensing data and land use change models would enhance the robustness of the predictions. Finally, rural settlements are undergoing rapid transformation due to urban migration and economic shifts. Developing a dynamic hazard prediction model that accounts for evolving settlement patterns would provide more accurate long-term risk assessments.

5.4. Limitations and Future Research Recommendations

This study has some other limitations. The influencing factors selected for susceptibility evaluation do not fully encompass all geological hazards due to data availability constraints. Future research should integrate additional variables, such as economic factors, proximity to earthquake fault zones, settlement-related industries, and climate characteristics, to develop a more comprehensive hazard evaluation model. Additionally, rural settlements are undergoing dynamic changes due to urban migration trends, which necessitates the inclusion of settlement reduction factors as a resistance variable in future hazard prediction models. Establishing a temporal framework for hazard risk governance will enhance long-term hazard prevention strategies.

6. Conclusions

Yibin City has the highest number of potential geological hazard points among the four major cities in Southern Sichuan, posing significant risks to local economic development and residents' safety. The dataset used in this study covers 2700 known geological hazard risk points in Yibin City as of 2020. These geological hazard risk points cover cities, towns, and their surrounding areas. The spatial distribution of the data includes multiple data samples, such as different slopes and different precipitation areas, ensuring the adaptability of the model to diverse geographical features. This study demonstrates the feasibility of using machine learning for hazard prediction and proposes strategies for risk mitigation. The key conclusions include the following:

(1) Effectiveness of the CatBoost Model: Among the various machine learning models tested, CatBoost outperforms other models in recall rate, precision, and F1 score. The optimized model (SSA–CatBoost) achieves a recall rate of 0.98, a precision rate of 0.94, and an F1 score of 0.96, indicating its strong predictive capabilities.

(2) Geospatial Distribution of Geological Hazards: The predicted hazard risk points exhibit a distribution pattern of being "sparse in plains and dense in mountainous regions", with high-risk areas concentrated in Pingshan County, Junlian County, and Gong County. Furthermore this study found that 59.63% of the settlements have at least one geological hazard risk point within a 2.5 km radius, while 19.8% of the settlements have at least one geological hazard risk point within a 1 km radius. These data confirm a significant overlap

between hazard points and rural settlements, highlighting the urgency of targeted risk mitigation efforts.

(3) Implications for Rural Settlement Planning: Given the complex terrain of Southern Sichuan, large-scale centralized resettlement should be avoided to prevent disturbing the natural environment and triggering new geological hazards. Instead, a "small aggregation, multi-point resettlement" strategy is recommended for rural settlement planning, aligning with local geographical and environmental conditions.

By integrating advanced machine learning techniques with spatial analysis, this study provides a robust theoretical framework for future geological hazard research and hazard prevention planning in Yibin City and beyond. Future studies should continue refining prediction models by incorporating more comprehensive hazard-related factors and analyzing long-term settlement trends to enhance risk-governance strategies.

Author Contributions: Conceptualization, R.Z.; Methodology, R.Z.; Software, D.Z.; Formal analysis, R.Z. and D.Z.; Investigation, R.Z. and Y.C.; Writing—original draft, R.Z.; Writing—review & editing, B.S. and Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Soft Science Special Project of the Gansu Basic Research Plan (Grant No. 24JRZM003), 2025 Gansu Province College Teachers Innovation Fund Project (Grant No. 2025A-203), Gansu Provincial Science and Technology Program (Grant No. 25YFFM003), and Sichuan Provincial Science and Technology Department Key Project (Grant No. 2020YFS0309).

Data Availability Statement: The Hazards data were derived from the following resources available in the public domain: [https://topic.scol.com.cn/2017/17dzzhcx/pc/, accessed on 7 January 2025]. Other data in this article are not readily available because of an ongoing study.

Acknowledgments: We would like to thank the open-source machine learning library of the Yandex team and the Sparrow Search Algorithm proposed by Xue and Shen from Donghua University; this research has been carried out smoothly and achieved the desired results.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhang, K.; Chen, Q.; Zhang, M. Priority evaluation of emergency facilities configuration in coastal rural area. *J. Nat. Hazards* **2022**, *6*, 56–66. (In Chinese) [CrossRef]
2. Ma, R.; Lyu, Y.; Chen, T.; Zhang, Q. Preliminary risk assessment of geological disasters in Qinglong Gorge Scenic Area of Taihang Mountain with GIS based on analytic hierarchy process and logistic regression model. *Sustainability* **2023**, *15*, 15752. [CrossRef]
3. Zhang, S.; Tan, S.; Liu, L.; Ding, D.; Sun, Y.; Li, J. Slope Rock and Soil Mass Movement Geological Hazards Susceptibility Evaluation Using Information Quantity, Deterministic Coefficient, and Logistic Regression Models and Their Comparison at Xuanwei, China. *Sustainability* **2023**, *15*, 10466. [CrossRef]
4. Zhang, H.; Zhang, G.; Jia, Q. Integration of analytical hierarchy process and landslide susceptibility index based landslide susceptibility assessment of the Pearl river delta area, China. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 4239–4251. [CrossRef]
5. Barman, J.; Das, J. Assessing classification system for landslide susceptibility using frequency ratio, analytical hierarchical process and geospatial technology map** in Aizawl district, NE India. *Adv. Space Res.* **2024**, *74*, 1197–1224. [CrossRef]
6. Orejuela, I.P.; Toulkeridis, T. Evaluation of the susceptibility to landslides through diffuse logic and analytical hierarchy process (AHP) between Macas and Riobamba in Central Ecuador. In *2020 Seventh International Conference on eDemocracy & eGovernment (ICEDEG), Buenos Aires, Argentina, 22–24 April 2020*; IEEE: Piscataway, NJ, USA, 2020; pp. 201–207.
7. Huang, J.; Zhang, Z. Study on the mining geological hazard susceptibility assessment in alpine areas using machine learning models. *J. China Univ. Min. Technol.* **2024**, *5*, 960–976. (In Chinese) [CrossRef]
8. Zhang, L.; He, S.; Zhang, J.; Peng, G.; Xu, Z.; Xie, J.; Tang, K.; Bu, J. Identification of landslide hazards based on multi-sourceremote sensing technology: A case study of the Changliarea in Hunan Province. *Remote Sens. Nat. Resour.* **2024**, *2*, 173–187. (In Chinese)
9. Moradmand, R.; Ahmadi, H.; Moeini, A.; Motamedvaziri, B.; Nazari Samani, A.A. Enhancing landslide susceptibility mapping through advanced hybridization of bootstrap aggregating based decision tree algorithms. *Earth Sci. Inform.* **2025**, *18*, 111. [CrossRef]

10. Saxena, V.; Singh, U.; Sinha, L.K. Comparative study of decision tree and its advance ensemble techniques for landslide susceptibility map. *J. Earth Syst. Sci.* **2025**, *134*, 11. [CrossRef]
11. Bai, X.; Yang, Z.; Luo, W.; Wang, J.; Tian, P.; Dai, Y. An Emergency Evaluation Approach to Road Disruption Risk Influenced by Earthquake-induced Landslide Using GIS and Multivariate Decision Tree. *J. Seismol. Res.* **2023**, *3*, 343–353. (In Chinese)
12. Lu, M.; Tay, L.T.; Mohamad-Saleh, J. Landslide susceptibility analysis using random forest model with SMOTE-ENN resampling algorithm. *Geomat. Nat. Hazards Risk* **2024**, *15*, 2314565. [CrossRef]
13. Jiangang, L.; Yi, H.; Lifeng, Z.; Qing, Z.; Binghai, G.; Hesheng, C.; Yumin, F. Ensemble learning landslide susceptibility assessment with optimized non-landslide samples selection. *Geomat. Nat. Hazards Risk* **2024**, *15*, 1.
14. Cui, T.; An, X.; Sun, D.; Chen, D.; Zhu, Y. Explainable Machine Learning Models for Landslide Susceptibility Mapping Based on SHAP. *J. Chengdu Univ. Technol. (Nat. Sci. Ed.)* **2025**, *52*, 153–172. (In Chinese)
15. Zhao, P.; Wen, G.; He, Z.; Wang, G.; Chen, L.; Shen, X.; Wang, K.; Tang, H. Shallow landslide susceptibility assessment in Jinsha River Basin based on machine learning models. *Water Resour. Hydropower Eng. (Chin. Engl.)* **2024**, *10*, 53–70. (In Chinese) [CrossRef]
16. Li, X.; Li, S. Large-scale landslide displacement rate prediction based on multi-factor support vector regression machine. *Appl. Sci.* **2021**, *11*, 1381. [CrossRef]
17. Wang, Z.; Brenning, A. Active-learning approaches for landslide mapping using support vector machines. *Remote Sens.* **2021**, *13*, 2588. [CrossRef]
18. Gong, W.; Tian, S.; Wang, L.; Li, Z.; Tang, H.; Li, T.; Zhang, L. Interval prediction of landslide displacement with dual-output least squares support vector machine and particle swarm optimization algorithms. *Acta Geotech.* **2022**, *17*, 4013–4031. [CrossRef]
19. Cai, S.; Yongbo, T.; Zhijie, N.; Wei, X.; Xiaohui, X. Landslide susceptibility mapping in Xide County, Sichuan Province based on frequency ratio-support vector machine coupling model. *Sediment. Tethyan Geol.* **2024**, *3*, 547–559. (In Chinese)
20. Teng, F.; Mao, Y.; Li, Y.; Qian, S.; Nanehkaran, Y.A. Comparative models of support-vector machine, multilayer perceptron, and decision tree predication approaches for landslide susceptibility analysis. *Open Geosci.* **2024**, *16*, 20220642. [CrossRef]
21. Hong, H. Assessing landslide susceptibility based on hybrid multilayer perceptron with ensemble learning. *Bull. Eng. Geol. Environ.* **2023**, *82*, 382. [CrossRef]
22. Yang, K.; Niu, R.; Song, Y.; Dong, J.; Zhang, H.; Chen, J. Dynamic Hazard Assessment of Rainfall-Induced Landslides Using Gradient Boosting Decision Tree with Google Earth Engine in Three Gorges Reservoir Area, China. *Water* **2024**, *16*, 1638. [CrossRef]
23. Dong, Z.; Zhang, J.; Peng, P.; Wang, Y.; Yang, Z. Landslide Susceptibility Evaluation Based on Coupling of GBDT-LR Model and Information Model. *Bull. Soil Water Conserv.* **2023**, *1*, 149–157+166. (In Chinese) [CrossRef]
24. Chen, T.; Zhu, L.; Niu, R.Q.; Trinder, C.J.; Peng, L.; Lei, T. Mapping landslide susceptibility at the Three Gorges Reservoir, China, using gradient boosting decision tree, random forest and information value models. *J. Mt. Sci.* **2020**, *17*, 670–685. [CrossRef]
25. Liu, Z.; Duan, G. Landslide Susceptibility Mapping based on Swarm Optimization Gradient Boosting Decision Tree in Sanshui District (China). In *2024 IEEE 4th International Conference on Software Engineering and Artificial Intelligence (SEAI), Xiamen, China, 21–23 June 2024*; IEEE: Piscataway, NJ, USA; pp. 331–336.
26. Rong, G.; Alu, S.; Li, K.; Su, Y.; Zhang, J.; Zhang, Y.; Li, T. Rainfall induced landslide susceptibility mapping based on Bayesian optimized random forest and gradient boosting decision tree models—A case study of Shuicheng County, China. *Water* **2020**, *12*, 3066. [CrossRef]
27. Zhou, C.; Gan, L.; Wang, Y.; Wu, H.; Yu, J.; Cao, Y.; Yin, K. Landslide Susceptibility Prediction based on Non-Landslide Samples Selection and Heterogeneous Ensemble Machine Learning. *J. Geo-Inf. Sci.* **2023**, *8*, 1570–1585. (In Chinese)
28. Zeng, Y.; Zhang, Y.; Zhang, Z.; Liu, J.; Zhu, H. Landslide Susceptibility Evaluation Based on Coupled X-Multilayer Perceptron Model—A Case Study of Songtao Autonomous County of Guizhou Province, China. *Mt. Res.* **2023**, *2*, 280–294. (In Chinese) [CrossRef]
29. Hu, J. Prediction of the internal corrosion rate for oil and gas pipelines and influence factor analysis with interpretable ensemble learning. *Int. J. Press. Vessel. Pip.* **2024**, *212*, 105329. [CrossRef]
30. Shi, Y.; Cai, Y.; Lou, S.; Chen, Y. Explainable prediction of deposited film thickness in IC fabrication with CatBoost and SHapley Additive exPlanations (SHAP) models. *Appl. Intell.* **2024**, *54*, 246–263. [CrossRef]
31. Ajin, R.S.; Segoni, S.; Fanti, R. Optimization of SVR and CatBoost models using metaheuristic algorithms to assess landslide susceptibility. *Sci. Rep.* **2024**, *14*, 24851. [CrossRef]
32. Yao, Z.; Chen, M.; Zhan, J.; Zhuang, J.; Sun, Y.; Yu, Q.; Yu, Z. Refined landslide susceptibility mapping by integrating the SHAP-CatBoost model and InSAR observations: A case study of Lishui, Southern China. *Appl. Sci.* **2023**, *13*, 12817. [CrossRef]
33. Jin, B.; Zeng, T.; Gui, L.; Kunlong, Y.; Yuhang, Z.; Yang, L. Mapping the Landslide Susceptibility Considering Future Land Use Dynamics Scenario. *J. Geo-Inf. Sci.* **2024**, *6*, 1486–1499. (In Chinese)
34. Huang, J.; Ling, S.; Wu, X.; Deng, R. GIS-based comparative study of the Bayesian network, decision table, radial basis function network and stochastic gradient descent for the spatial prediction of landslide susceptibility. *Land* **2022**, *11*, 436. [CrossRef]

35. Bennett, N.D.; Croke, B.F.; Guariso, G.; Guillaume, J.H.; Hamilton, S.H.; Jakeman, A.J.; Andreassian, V. Characterising performance of environmental models. *Environ. Model. Softw.* **2013**, *40*, 1–20. [CrossRef]
36. Hu, S.; Lu, Y.; Liu, X.; Huang, C.; Wang, Z.; Huang, L.; Li, X. Stability prediction of circular sliding failure soil slopes based on a genetic algorithm optimization of random forest algorithm. *Electron. Res. Arch.* **2024**, *32*, 6120–6139. [CrossRef]
37. Xu, J.; Zhao, Y. Stability Analysis of Geotechnical Landslide Based on GA-BP Neural Network Model. *Comput. Math. Methods Med.* **2022**, *2022*, 3958985. [CrossRef]
38. Sun, K.; Li, Z.; Wang, S.; Hu, R. A support vector machine model of landslide susceptibility map** based on hyperparameter optimization using the Bayesian algorithm: A case study of the highways in the southern Qinghai–Tibet Plateau. *Nat. Hazards* **2024**, *120*, 11377–11398. [CrossRef]
39. Wang, S.; Lin, X.; Qi, X.; Li, H.; Yang, J. Landslide susceptibility analysis based on a PSO-DBN prediction model in an earthquake-stricken area. *Front. Environ. Sci.* **2022**, *10*, 912523. [CrossRef]
40. Xue, F.; Liu, Y.; Chen, Y. Comparative Study on Several New Swarm Intelligence Algorithms in Solving Clustering Problems. *Comput. Simul.* **2023**, *3*, 370–376. (In Chinese)
41. Wang, H. Rural Ageing and Rural Spatial Evolution: Developments in Ageing of Rural China at Township Level. *Popul. Res.* **2019**, *43*, 66–80. (In Chinese)
42. Chen, J.; Zeng, Z.; Jiang, P.; Tang, H. Application of multi-gene genetic programming based on separable functional network for landslide displacement prediction. *Neural Comput. Appl.* **2016**, *27*, 1771–1784. [CrossRef]
43. Wei, Y.; Huang, Z. Variations in Microstructure and Collapsibility Mechanisms of Malan Loess across the Henan Area of the Middle and Lower Reaches of the Yellow River. *Appl. Sci.* **2024**, *14*, 8220. [CrossRef]
44. Zhang, J.; Qian, J.; Lu, Y.; Li, X.; Song, Z. Study on Landslide Susceptibility Based on Multi-Model Coupling: A Case Study of Sichuan Province, China. *Sustainability* **2024**, *16*, 6803. [CrossRef]
45. Sharma, A.; Sajjad, H.; Bhuyan, N.; Rahaman, M.H.; Ali, R. Climate change-induced landslide vulnerability: Empirical evidence from Shimla district, Himachal Pradesh, India. *Int. J. Disaster Risk Reduct.* **2024**, *110*, 104657. [CrossRef]
46. Prathom, K.; Sujitapan, C. Performance of logistic regression and support vector machine conjunction with the GIS and RS in the landslide susceptibility assessment: Case study in Nakhon Si Thammarat, southern Thailand. *J. King Saud Univ.-Sci.* **2024**, *36*, 103306. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.