

Article

Regional Landslide Susceptibility Assessment and Model Adaptability Research

Zhiqiang Zhang  and Jichao Sun *

School of Water Resource and Environment, China University of Geosciences, Beijing 100083, China;
2105210093@email.cugb.edu.cn

* Correspondence: sunjc@cugb.edu.cn; Tel.: +86-13120258076

Abstract: Landslide susceptibility denotes the likelihood of a disaster event under specific conditions. The assessment of landslide susceptibility has transitioned from qualitative to quantitative methods. With the integration of information technology in geological hazard analysis, a range of quantitative models for assessing landslide susceptibility has emerged and is now widely used. To compare and evaluate the accuracy of these models, this study focuses on Xupu County in Hunan Province, applying several models, including the CF model, FR model, CF-LR coupled model, FR-LR coupled model, SVM model, and RF model, to assess regional landslide susceptibility. ROC curves are used to evaluate the reliability of the model's predictions. The evaluation results reveal that the CF model (AUC = 0.756), FR model (AUC = 0.764), CF-LR model (AUC = 0.776), FR-LR model (AUC = 0.781), SVM model (AUC = 0.814), and RF model (AUC = 0.912) all have AUC values within the range of 0.7–0.9, indicating that the overall accuracy of the models is good and can provide a reference for landslide susceptibility zoning in the study area. Among these, the Random Forest model demonstrates the best accuracy for landslide susceptibility zoning in the study area. By extracting the extremely high susceptibility zones from the landslide susceptibility zonings obtained by six models, a comparative analysis of model adaptability was conducted. The results indicate that the Random Forest model has the best adaptability under specific conditions in Xupu County.

Keywords: landslide susceptibility; traditional statistical models; logistic regression models; machine learning models



Citation: Zhang, Z.; Sun, J. Regional Landslide Susceptibility Assessment and Model Adaptability Research. *Remote Sens.* **2024**, *16*, 2305. <https://doi.org/10.3390/rs16132305>

Academic Editor: Michele Saroli

Received: 23 April 2024

Revised: 18 June 2024

Accepted: 18 June 2024

Published: 24 June 2024



Copyright: © 2024 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

Landslide disasters are among the most prevalent natural hazards, characterized by the movement of rock or soil masses on slopes due to the combined influence of internal and external factors, as well as the force of gravity, leading to the displacement of the entire or part of the mass along a weak plane [1]. These disasters are marked by their rapid occurrence, swift pace, and large scale, resulting in significant and often underestimated losses of life and property annually. As illustrated in Figure 1, Xupu County is susceptible to such calamities [2]. In the picture, there is a sign with the symbol of the Chinese character “stop”, and a number “29 m”, indicating that there is danger ahead, and the vehicle stops here, 29 m away from the disaster point ahead. Landslide occurrence is influenced by a multitude of factors, and various methods exist for prediction and research. One important direction involves the study of landslides caused by water-sediment transport [3,4] combined with fault effects [5], while remote sensing image identification is currently the mainstream approach. The former approach focuses on solving problems through mechanism research, while the latter belongs to image perception methods. This paper primarily focuses on the identification of disasters through remote sensing images, aiming to guide and evaluate disaster response. The assessment of landslide susceptibility forms the cornerstone of landslide risk management, entailing a comprehensive analysis of the impact factors to evaluate the likelihood of landslide occurrence within a region. This

provides essential technical support for the prevention of landslide disasters [6,7]. Currently, the primary factors influencing landslide susceptibility are selected from categories such as topographical and morphological factors [8], hydrological factors [9,10], land use factors, and lithological factors. By identifying landslide-prone areas, relevant departments can take targeted measures such as strengthening monitoring, implementing engineering protection, or adjusting land use planning to reduce the probability and potential losses of landslide occurrence. This study aims to provide a scientific basis for landslide disaster prevention and mitigation in Xupu County by comparing and analyzing various landslide susceptibility assessment models and enriching the theoretical and practical knowledge of landslide disaster prevention and mitigation in the region.



Figure 1. Landslide Disasters in Xupu County.

Landslide susceptibility models are classified into empirical, traditional statistical, and machine learning models. Empirical models depend on expert knowledge and are uncertain. Traditional statistical models are quantitative, using mathematical models that assume independent factors. Machine learning models use large datasets and computer programs for data simulation [11,12]. Advances in information technology have led to the integration of 3S technology with disaster models, enhancing intuitive regional assessments [13].

Hunan Province accounted for 65% of national landslide occurrences with 2561 events. Landslides can also trigger secondary disasters like debris flows and floods. Table 1 presents statistics on landslide disasters in Xupu County townships. With distinct seasonal rainfall patterns, concentrated early summer rainfall, significant terrain relief, and numerous fault structures, Xupu County has multiple factors triggering landslides, making it suitable as the study area for this research [14–16].

The Deterministic Coefficient Model (CF) and Frequency Ratio Model (FR), despite being straightforward traditional statistical models [17], do not suffice for precise landslide prediction. To overcome their limitations and the challenges of high data demands and inefficient calculations, they were each combined with the Logistic Regression Model (LR) to form the CF + LR and FR + LR coupled models [18,19]. These integrations leverage the strengths of the original models and use logistic regression to mitigate the effects of hard-to-quantify data, simplifying the process.

This article presents a statistical analysis of geological disasters and landslides in China from 2011 to 2023, using data from the Natural Resource Statistical Bulletins, National Geological Disaster Bulletins, and National Geological Disaster Prevention and Control Work Reports by the Ministry of Natural Resources of China. In 2022, China experienced 5659 geological disasters, including 3919 landslides, resulting in 90 fatalities, 16 missing persons, and direct economic losses of RMB 1.5 billion.

Table 1. Landslide Disaster Statistics for Xupu County.

Xupu County Townships	Area/km ²	Number of Landslide Points	Proportion of Total Disaster Points	Landslide Density
Beidou Xi Town	172.8	5	0.045	0.029
Dajiangkou Town	236.4	5	0.045	0.021
Luoazhuang Town	205.0	5	0.045	0.024
Gezhi Ping Town	98.9	18	0.162	0.182
Guanyin Pavilion Town	168.7	4	0.036	0.024
Huangmaoyuan Town	114.1	10	0.090	0.088
Junping Town	89.4	2	0.018	0.022
Liangyaping Town	85.0	3	0.027	0.035
Longtan Town	244.5	11	0.099	0.045
Longzhuang Bay	45.2	1	0.009	0.022
Lufeng Town	142.3	2	0.018	0.014
Qiaojing Town	146.9	3	0.027	0.020
Sanjiang Town	287.9	1	0.009	0.003
Shenzi Lake Town	244.1	3	0.027	0.012
Shu Rongxi	59.9	3	0.027	0.050
Shuang Well Town	105.0	2	0.018	0.019
Shuidong Town	145.5	5	0.045	0.034
Sisong Town	85.1	2	0.018	0.023
Taojin Ping Town	68.4	8	0.072	0.117
Tongxi River Town	89.0	4	0.036	0.045
Xiao Henglong	129.7	2	0.018	0.015
Yanxi	139.2	5	0.045	0.036
Youyang	84.0	1	0.009	0.012
Zhongdu	120.6	4	0.036	0.033
Zushi Temple Town	115.9	2	0.018	0.017

This article chooses Xupu County, Hunan Province, as the study area, dividing it into evaluation units [20]. It uses six models—the CF, FR, CF-LR, FR-LR, Support Vector Machine, and Random Forest [21,22]—to assess landslide susceptibility. The accuracy of each model is evaluated and compared using ROC curves to identify the best method for the region.

2. Study Area

Xupu County is located in the western part of Hunan Province and the northeastern part of Huaihua City, on the middle reaches of the Yuan River. It shares borders with Dongkou County and Hongjiang City to the south, Yuanling County and Anhua County to the north, Xinhua County and Longhui County to the east, and Chenxi County and Zhongfang County to the west. Its geographical coordinates range from 110°15'E to 111°01'E and 27°19'N to 28°17'N, covering a total area of 3438 km², as shown in Figure 2. It is in the mid-low mountains and hills of western Hunan, with a subtropical humid monsoon climate—hot, rainy summers; mild, moist winters; and heavy spring-summer rainfall. The region has significant climatic variations, microclimates, and vertical climate differences. Its geology, with frequent structural movements and fault structures, makes it prone to landslides. The nature of strata and rock formations has a profound impact on the structure of slopes and the types of accumulated layers, making it a key factor controlling slope stability. Xupu County has ancient and complex geological conditions, with strata from the Proterozoic, Lower Paleozoic, Upper Paleozoic, Mesozoic, and Cenozoic eras all present. Among them, the stratigraphic units of the Upper Sinian System of the Lower Paleozoic Era, the Doucun Formation, and the Dengying Formation are the most widely exposed in the study area. This study focuses on Xupu's landslide susceptibility, using 111 landslide sites as primary data for model assessment.

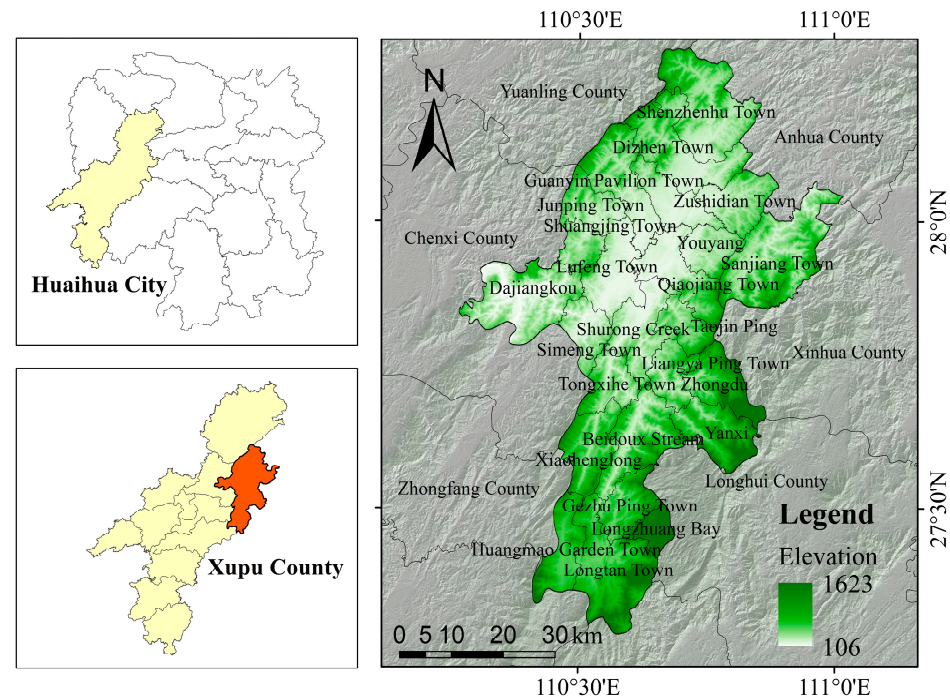


Figure 2. The study area.

3. Methods

Landslides, as random events, are governed by various factors. Modeling has significantly enhanced the study of landslides, boosting research precision and depth. This paper assesses landslide susceptibility in the study area with traditional statistical and machine learning models, producing susceptibility maps for each. The models' performance is then evaluated with ROC curves, with the most accurate model chosen through comparison.

3.1. CF Model

The Deterministic Coefficient Model assumes that the conditions for future landslide occurrences are consistent with the past. Hence, the risk of landslides can be analyzed probabilistically based on the prior probability of past landslide occurrences and the correlation with triggering factors.

$$CF = \begin{cases} \frac{pp_a - pp_s}{pp_a(1 - pp_s)} (pp_a \geq pp_s) \\ \frac{pp_s - pp_a}{pp_s(1 - pp_a)} (pp_a < pp_s) \end{cases} \quad (1)$$

where pp_a represents the ratio of the number of grid cells with geological hazard points present in feature a to the total number of grid cells in feature a ; pp_s represents the ratio of the number of grid cells with landslide geological hazard points in the entire study area to the total number of grid cells in the study area.

3.2. FR Model

The Frequency Ratio method is based on statistical analysis and involves categorizing influencing factors into intervals. The principle is as follows: Assume the landslide area is 'L' and the influencing factor is 'F'. 'F' is divided into different intervals according to specific rules, and then the area of landslides falling within these intervals is calculated. The frequency ratio F_jR for the j interval of the influencing factor 'F' is:

$$F_jR = \frac{P(LF_j)}{P(F_j)} = \frac{A_{LF_j}/A_L}{A_{F_j}/A} = \frac{A_{LF_j}/A_{F_j}}{A_L/A} = \frac{P(L|F_j)}{P(L)} \quad (2)$$

where $P(LF_j)$ denotes the frequency of F_j within the landslide area 'L'; $P(F_j)$ represents the frequency of F_j in the entire study area; A_{LF_j} signifies the area of F_j within 'L'; A_L is the total area of 'L'; A_{F_j} indicates the total area of F_j ; and 'A' stands for the total area of the study region.

3.3. LR Model

The logistic regression model is a predictive model used for binary or multiclass classification [23,24]. Its principle involves calculations through functions such as Sigmoid. The schematic diagram of the Sigmoid function is shown in Figure 3. In this paper, the levels of various landslide-influencing factors are treated as independent variables, while the occurrence or absence of landslides is considered the dependent variable [25]. Let 'P' represent the probability of a landslide occurring, with its value ranging between [0, 1]. The expression is as follows:

$$f(x) = a + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (3)$$

where x_1, x_2, \dots, x_n represent the various landslide influencing factors in the study area; b_1, b_2, \dots, b_n are the logistic regression coefficients corresponding to these factors, and a is a constant.

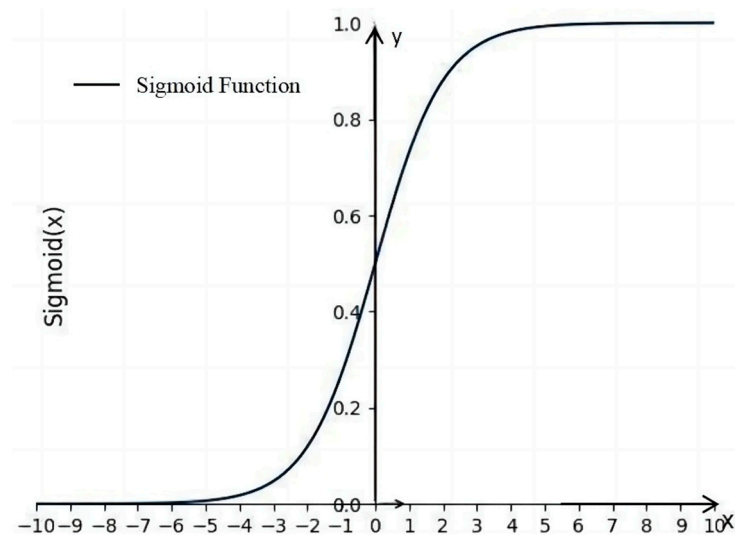


Figure 3. Schematic Diagram of the Sigmoid Function Principle.

3.4. SVM Model

The Support Vector Machine (SVM) is a powerful supervised learning technique utilized for both linear and non-linear classification and regression tasks. Drawing from the concept of Structural Risk Minimization in statistical learning theory, SVM employs kernel methods to project data from an original finite-dimensional space into a more expansive, higher-dimensional space. Within this augmented space, SVM identifies the optimal hyperplane for classification, which efficiently discriminates between various data categories, maximizing the margin between different classes. This strategy is particularly advantageous for problems that are not linearly separable in their original dimensionality. The schematic diagram of the principle of the Support Vector Machine model is shown in Figure 4.

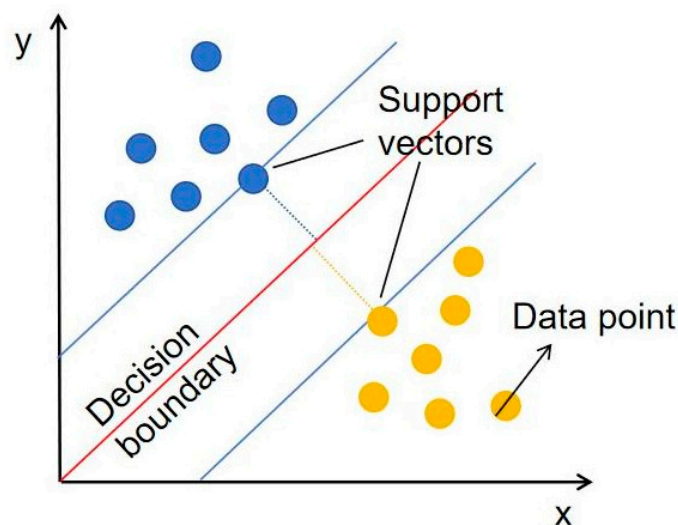


Figure 4. Schematic Diagram of the Principles of the Support Vector Machine Model.

3.5. RF Model

Random Forest (RF) is an ensemble algorithm formed by combining decision trees [26–28]. An ensemble algorithm refers to one that combines multiple weak learners to create a strong learner [29]. Its objective is to use the combination of weak learners to achieve better learning outcomes than what is possible with any single weak learner [30,31]. RF is well-suited for processing high-dimensional data, as it does not require dimensionality reduction or feature selection. It offers fast training speeds, is less prone to overfitting, and is insensitive to missing values. Even with a significant portion of features missing, RF can still maintain its accuracy.

3.6. ROC Curve Accuracy Assessment

The Receiver Operating Characteristic analysis can be used to study binary classification problems in engineering and scientific fields. Essentially, the ROC analysis method is a supervised analysis method, where we need to pre-know the labels of the dataset and the corresponding probability density functions. According to the confusion matrix, the sensitivity and 1-specificity are calculated using the formula, and the ROC curve is plotted with 1-specificity as the horizontal coordinate and sensitivity as the vertical coordinate. The performance of binary classifiers in machine learning is evaluated by the area under the ROC curve (Area Under the Curve, AUC).

3.7. Susceptibility Assessment

Selection of Landslide Influencing Factors

The occurrence of landslides is influenced by various factors, and their selection and analysis are essential for susceptibility assessment. This study examines the topography, landforms, and development types of the area, identifying four factor categories: topographical and morphological, hydrological, lithological, and land-use, totaling twelve evaluation factors. Topographical configuration is critical, as it influences soil moisture and vegetation, affecting landslide occurrence. Lithological factors affect slope material properties and are primary controllers of slope stability. Hydrological factors influence slope morphology and stability through erosion and sediment removal. Human activities, such as road construction, agriculture, and logging, can disturb slope internal forces and potentially trigger landslides.

3.8. Data Sources and Processing

In this research, elevation, geological, and remote sensing data were collected for Xupu County. The elevation data were obtained from the ASTER GDEM digital elevation

model with a 30 M resolution available in the geospatial data cloud; geological data were sourced from the 1:250,000 scale geological maps provided by the Geological Science Data Publication System of the China Geological Data Repository; remote sensing data were acquired from the Landsat 8–9 OLI/TIRS C2 L2 dataset in the geospatial data cloud. Additionally, regional landslide disaster data were derived from the Geological Disaster Survey of Xupu County, Hunan Province, conducted by the Chinese Academy of Sciences' Institute of Resources and Environment. The datasets were processed using ArcGIS 10.8 software: (1) Extracting factors such as rock groups, distance to faults, and distance to roads from the 1:250,000 geological maps; (2) Utilizing elevation data to derive factors like slope, aspect, terrain ruggedness, profile curvature, plan curvature, distance to rivers, and the topographic wetness index; (3) Employing remote sensing data to extract land use types and vegetation cover factors. Among them: the distance from rivers, the distance from faults, and the distance from water systems are measured using the equidistant method, while the other factors are assessed using the natural break point method. The classification results for these different evaluation factors are illustrated in the accompanying Figures 5–7.

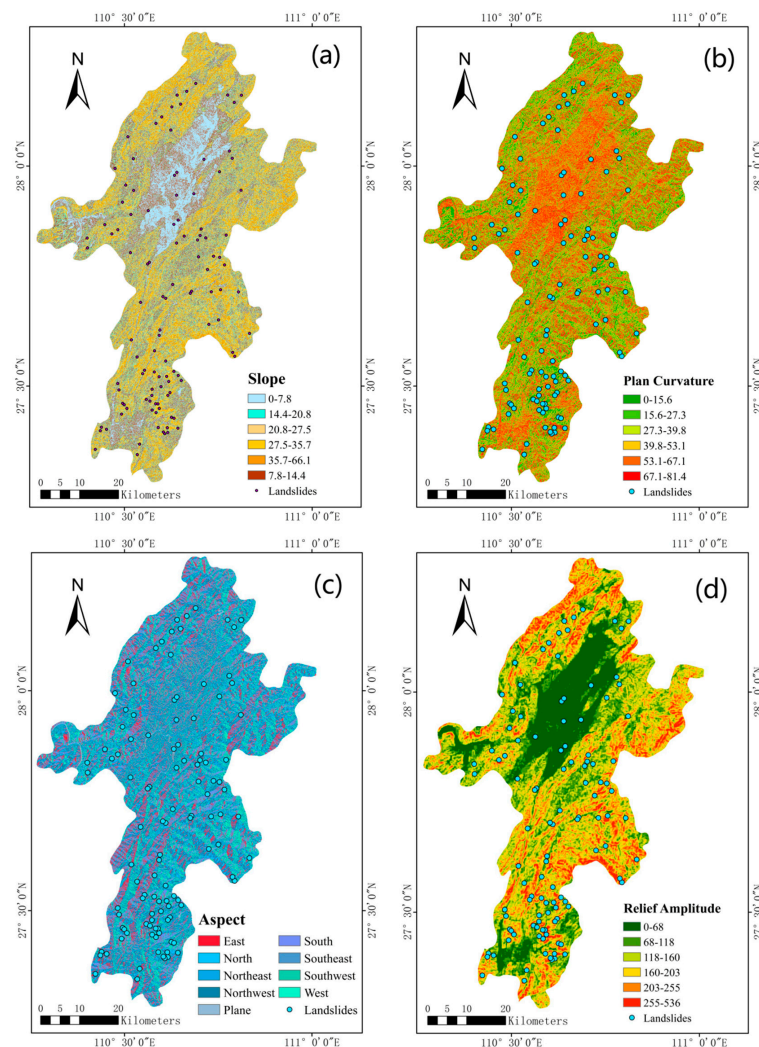


Figure 5. Different Evaluation Factors Grading Results. (a) slope, (b) plan curvature, (c) aspect, (d) relief amplitude.

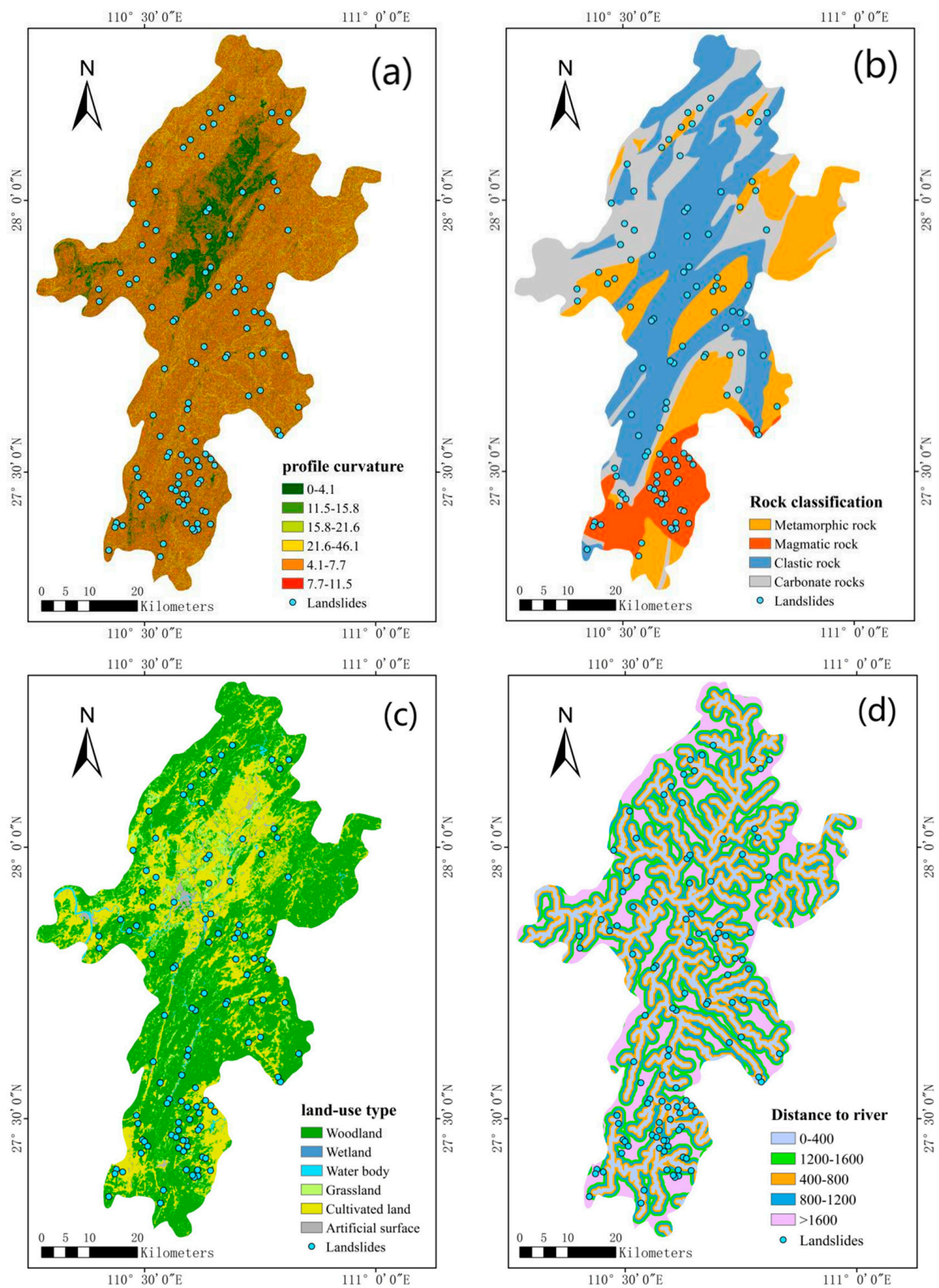


Figure 6. Different Evaluation Factors Grading Results. (a) profile curvature, (b) rock classification, (c) land-use type, (d) distance to a river.

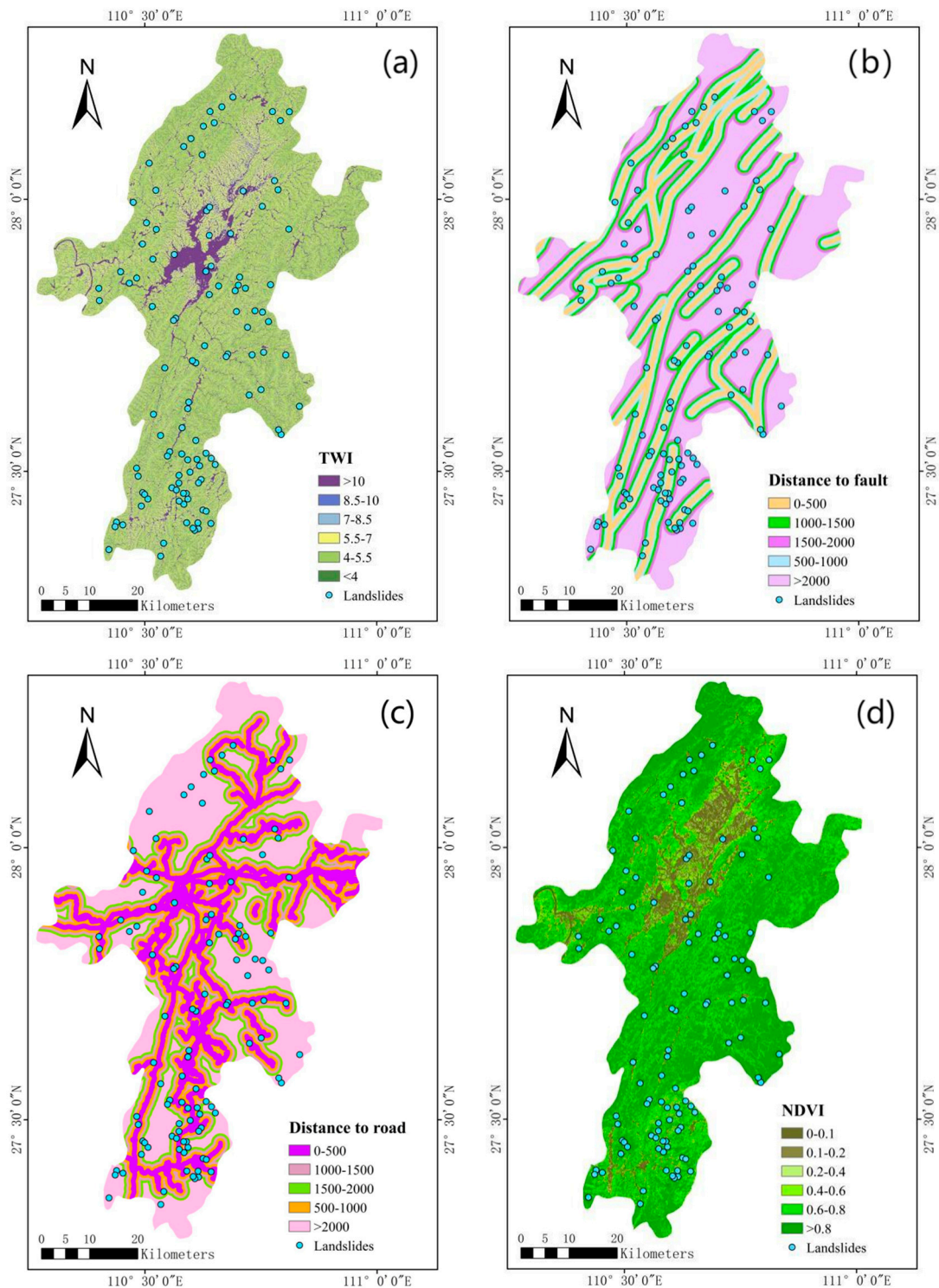


Figure 7. Different Evaluation Factors Grading Results. (a) Twi, (b) distance to fault, (c) distance to road, (d) NDVI.

3.9. Utilization of CF + LR Model and FR + LR Model for Landslide Susceptibility Zoning

The study used CF and FR to quantify factor influences at specific intervals, as detailed in Table 2. SPSS correlation analysis was used to exclude irrelevant variables, resulting in a dataset of 111 non-landslide and 111 landslide points for training. CF and FR values were assigned as independent variables in binary logistic regression, with landslide occurrence as the dependent variable. Regression coefficients (B) indicated the importance of each factor, with statistical significance determined by the Sig value (<0.05). Variables with

Sig > 0.05 were removed, and the remaining factors were re-evaluated, confirming their statistical significance, as shown in Tables 3 and 4.

Table 2. Evaluation Factors FR Values and CF Values.

Factor	Categories	FR	CF	Factor	Categories	FR	CF
DEM /m	106–298	0.888	−0.112	Profile curvature	0–4.1	0.930	−0.070
	298–462	0.991	−0.009		4.1–7.7	1.191	0.161
	462–638	0.787	−0.213		7.7–11.5	0.959	−0.041
	638–846	1.154	0.133		11.5–15.8	0.721	−0.279
	846–1107	1.785	0.440		15.8–21.6	1.190	0.159
	1107–1623	0.788	−0.212		21.6–46.1	0.931	−0.069
Slope /°	0–7.8	0.486	−0.514	Land-use type	Cultivated land	1.896	0.472
	7.8–14.4	1.027	0.027		Woodland	0.589	−0.411
	14.4–20.8	1.290	0.225		Grassland	1.153	0.132
	20.8–27.5	1.466	0.318		Wetland	0.000	−1.000
	27.5–35.7	0.641	−0.359		Waterbody	0.729	−0.271
	35.7–66.1	0.760	−0.240		Artificial surface	1.755	0.430
Aspect	Plane	0.000	−1.000	Distance to road /m	0–500	1.198	0.165
	North	0.795	−0.205		500–1000	1.549	0.354
	Northeast	0.668	−0.332		1000–1500	0.550	−0.450
	East	0.750	−0.250		1500–2000	1.030	0.030
	Southeast	0.658	−0.342		>2000	0.762	−0.238
	South	1.538	0.350	Plan curvature	0–15.6	1.517	0.341
	Southwest	0.919	−0.081		15.6–27.3	1.008	0.008
	West	1.348	0.258		27.3–39.8	1.317	0.241
	Northwest	1.308	0.236		39.8–53.1	0.768	−0.232
NDVI	0–0.1	1.709	0.415	Distance to fault /m	53.1–67.1	0.658	−0.342
	0.1–0.2	0.531	−0.469		67.1–81.4	0.697	−0.303
	0.2–0.4	1.710	0.415		0–500	1.424	0.298
	0.4–0.6	1.030	0.029		500–1000	1.183	0.154
	0.6–0.8	1.108	0.098		1000–1500	1.172	0.147
	>0.8	0.705	−0.295		1500–2000	1.130	0.115
Relief amplitude /m	0–68	0.621	−0.379	Rock classification	>2000	0.650	−0.350
	68–118	0.896	−0.104		Carbonate rock	1.082	0.076
	118–160	1.298	0.230		Metamorphic rock	0.645	−0.355
	160–203	0.838	−0.162		Clastic rock	0.727	−0.273
	203–255	1.369	0.269		Magmatic rock	2.750	0.679
	255–536	0.847	−0.153	Twi	<4	0.483	−0.517
Distance to river /m	0–400	1.584	0.369		4–5.5	1.041	0.039
	400–800	0.830	−0.170		5.5–7	1.083	0.076
	800–1200	0.710	−0.290		7–8.5	0.831	−0.169
	1200–1600	0.995	−0.005		8.5–10	1.714	0.416
	>1600	0.724	−0.276		>10	0.619	−0.381

Table 3. Results of Logistic Regression Analysis for CF Models.

Factor	B	Standard Error	Wald Test Statistic	Degrees of Freedom	Significance	Exp(B)
Twi	2.26	0.905	6.24	1	0.012	9.587
Land-use type	1.442	0.411	12.344	1	0	4.231
Rock classification	1.77	0.466	14.457	1	0	5.874
Fault	1.484	0.623	5.672	1	0.017	4.411
River	1.608	0.601	7.165	1	0.007	4.994
Road	2.209	0.635	12.11	1	0.001	9.111
Constant	0.122	0.162	0.569	1	0.451	1.13

Table 4. Results of Logistic Regression Analysis for FR Models.

Factor	B	Standard Error	Wald Test Statistic	Degrees of Freedom	Significance	Exp(B)
Twi	1.891	0.737	6.591	1	0.01	6.627
Land-use type	1.057	0.286	13.654	1	0	2.877
Rock classification	0.906	0.238	14.487	1	0	2.474
Fault	1.411	0.554	6.492	1	0.011	4.1
River	1.115	0.442	6.353	1	0.012	3.049
Road	2.05	0.538	14.528	1	0	7.771
Constant	−8.918	1.445	38.08	1	0	0

3.10. Utilization of RF Model and SVM Model for Landslide Susceptibility Zoning

This research approaches landslide occurrence as a binary classification issue. A dataset was compiled by identifying 111 landslide occurrences and establishing 50 m buffer zones around these points, which served as the control areas for generating 800 additional landslide points (classified as 1). Across the entire study region, 10,000 random non-landslide points (classified as 0) were created, with those within the 50 m buffers excluded, leading to a total of 10,722 sample points. To assess model efficacy, the dataset was split into training and testing subsets. A random 70% of the sample points were allocated to the training set, while the remaining 30% formed the test set, ensuring the models' robust generalization and predictive capabilities across diverse datasets.

In this study, grid search was employed for hyperparameter tuning of the basic parameters, with the rest set to their default values. Grid search is a method of parameter optimization that involves iterating over all possible parameter combinations to select the best-performing combination.

The hyperparameters of Support Vector Machines (SVM) encompass *C*, *kernel*, and *gamma*. The regularization parameter *C* signifies the tolerance for data error, governing the degree of model complexity. The *kernel* function is responsible for transforming feature spaces to facilitate data separation, and an appropriate kernel can enhance model efficacy. *gamma* is a hyperparameter specific to the Radial Basis Function (RBF), influencing the scope of impact that individual training samples have on the decision boundary. The tuning process for SVM model parameters is illustrated in Figure 8, with detailed parameter values presented in Table 5.

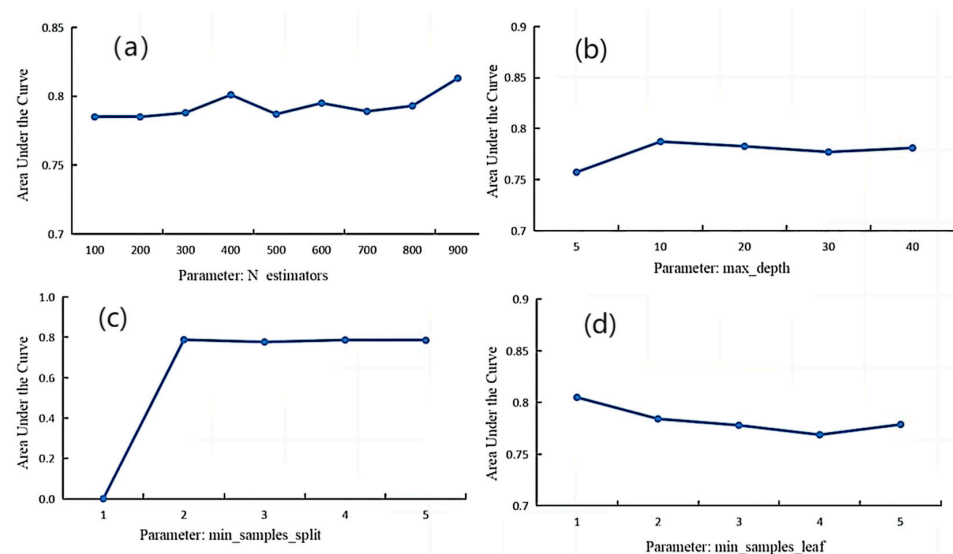
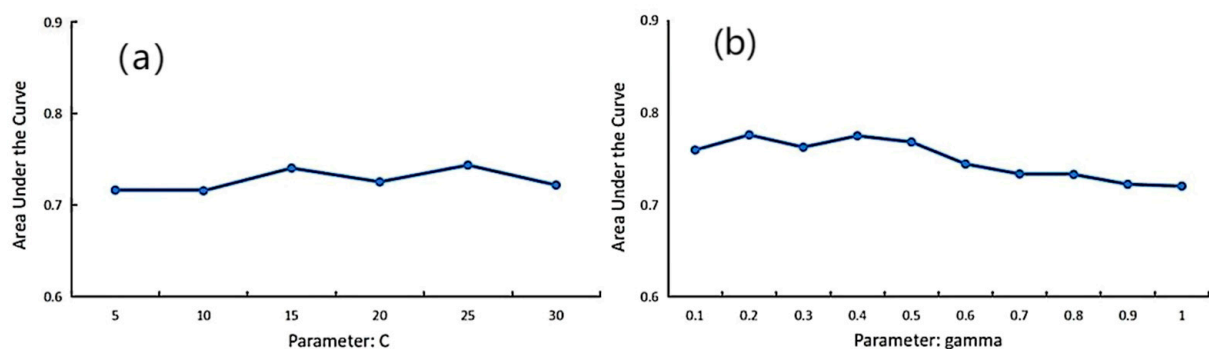


Figure 8. Tuning process for Random Forest model parameters, (a) N_estimators, (b) max_depth, (c) min_sample_split, (d) min_sample_leaf.

Table 5. Hyperparameters of SVM.

Type	Parameter
C	3
kernel	RBF
gamma	0.4

The hyperparameters of Random Forest (RF) include *n_estimators* (the number of decision trees), *max_depth* (the depth of each decision tree), *min_samples_split* (the minimum number of samples required to split an internal node), and *min_samples_leaf* (the minimum number of samples required to be at a leaf node). The meticulous adjustment of these parameters aims to achieve an optimal balance between model performance and computational efficiency. The optimization of these parameters is crucial for ensuring the accuracy and generalization capability of the Random Forest model in landslide susceptibility assessments. The tuning process for RF model parameters is depicted in Figure 9, and a comprehensive list of parameters is provided in Table 6.

**Figure 9.** Tuning process for Support Vector Machine model parameters, (a) C, (b) gamma.**Table 6.** Hyperparameters of RF.

Type	Parameter
n_estimators	900
max_depth	10
min_samples_split	2
min_samples_leaf	1

4. Results

4.1. Landslide Susceptibility Zoning Results

GIS spatial analysis tools were employed to derive susceptibility evaluation results for the CF + LR and FR + LR models by overlaying layers of deterministic coefficient and frequency ratio values, weighted according to their respective weights. Similarly, for the RF and SVM models, landslide susceptibility evaluation results were produced through a weighted overlay of probabilities calculated by machine learning models. The predicted probabilities of landslides for all models span from 0 to 1. The prediction results of the six models were classified into five levels using the natural break method: very low susceptibility (0–0.3), low susceptibility (0.3–0.4), moderate susceptibility (0.4–0.5), high susceptibility (0.5–0.65), and very high susceptibility (0.65–1), as depicted in Figure 10. The concentration of existing landslide disaster points in the very high and high susceptibility zones as determined by the models suggests the validity of the susceptibility zoning outcomes of these models.

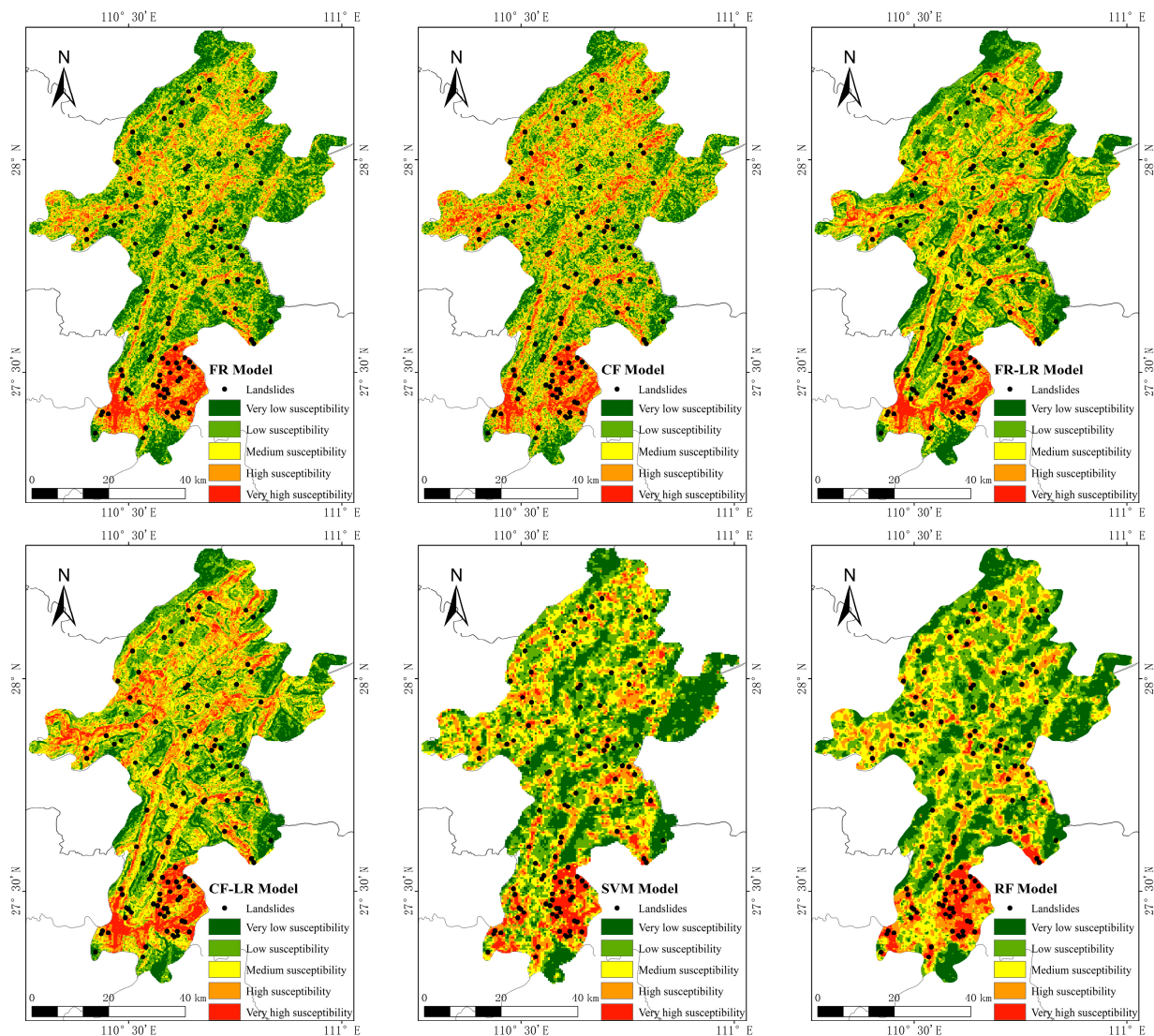


Figure 10. Results of landslide susceptibility prediction models.

As indicated by the graphs, the regions with very high landslide susceptibility are primarily concentrated in areas where igneous rocks are the surface layers and along the banks of rivers, accounting for about 6.5% of the total area of the county. The high susceptibility areas are also predominantly located in the river buffer zones, representing approximately 16% of the county's total area. The areas of moderate, low, and very low susceptibility are mainly distributed in agricultural and forested regions.

4.2. Model Accuracy Assessment

The predictive models' ROC curves are depicted in Figure 11. From the ROC curves, the following Area Under the Curve (AUC) values can be calculated: CF model AUC = 0.756, FR model AUC = 0.764, CF-LR model AUC = 0.776, FR-LR model AUC = 0.781, SVM model AUC = 0.814, and RF model AUC = 0.912. The accuracy of both the RF and CF models slightly improved after coupling with the LR model, indicating that the multi-model coupling has a better predictive effect than single models. With all four models' AUC values falling within the 0.7–1 range, this suggests good overall accuracy for the models, making them reliable for providing references in landslide susceptibility zoning of the study area. Among them, the Random Forest model exhibits the highest accuracy in predicting landslide susceptibility zoning in the study area.

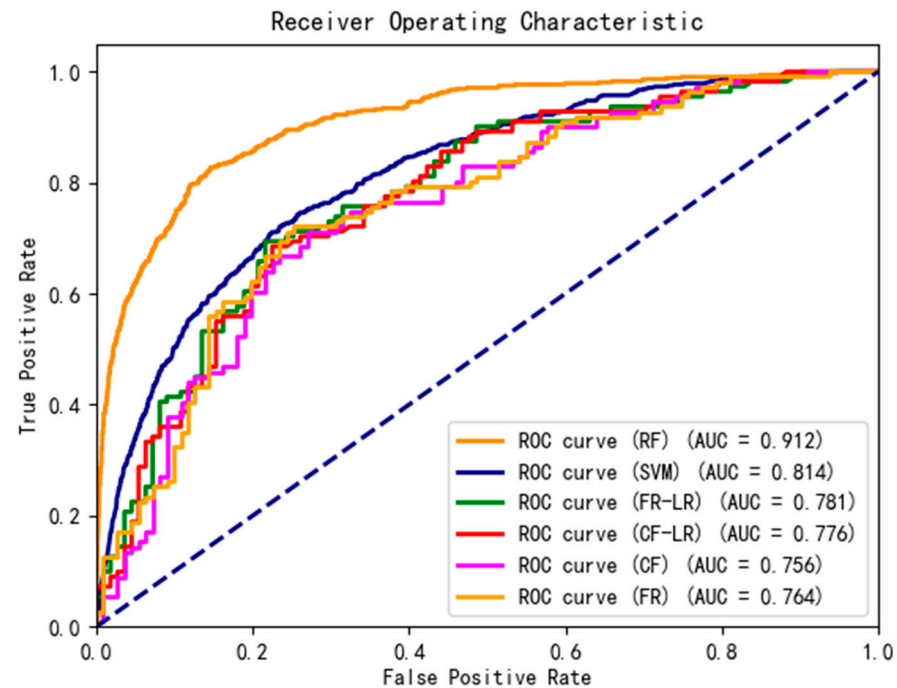


Figure 11. ROC curve and AUC value of the susceptibility assessment model.

4.3. Model Adaptability Analysis

Model adaptability analysis assesses a model's fitness for particular contexts and applications by examining its output. Due to Xupu County's complex geology and heavy rainfall, it is susceptible to landslides. This study focuses on the extremely high susceptibility zones from six zoning maps to gauge model adaptability. The performance of these models in predicting landslides within these zones is analyzed for practical utility. The identified extremely high susceptibility areas are depicted in Figure 12, and the corresponding landslide occurrence counts are summarized in Table 7.

Table 7. Number of Landslide Occurrences.

Model	Very High Susceptibility	High Susceptibility	Medium Susceptibility	Low Susceptibility	Very Low Susceptibility
FR Model	27	46	15	16	7
CF Model	31	48	13	16	3
FR-LR Model	22	43	25	16	5
CF-LR Model	26	44	25	12	4
SVM Model	51	28	26	5	1
RF Model	81	22	8	0	0

Table 7 reveals substantial variations in landslide frequency among extremely high susceptibility zones defined by various models. The RF model had the most landslides, 81, and the greatest density in these zones, while the FR-LR model had the fewest, only 22. These findings, along with prior AUC values from ROC curves, confirm the RF model's superior adaptability in Xupu County across different conditions. These data provide us with an intuitive understanding of the model's performance.

Figure 12 illustrates that areas with very high susceptibility are predominantly concentrated in the southern part of Xupu County, predominantly in regions where the exposed rocks are igneous. This study highlights this area separately, as shown in Figure 13. Using the multi-point extraction tool, a comparative analysis of the number of landslide occurrences in each model area was conducted, and the results are shown in Table 8.

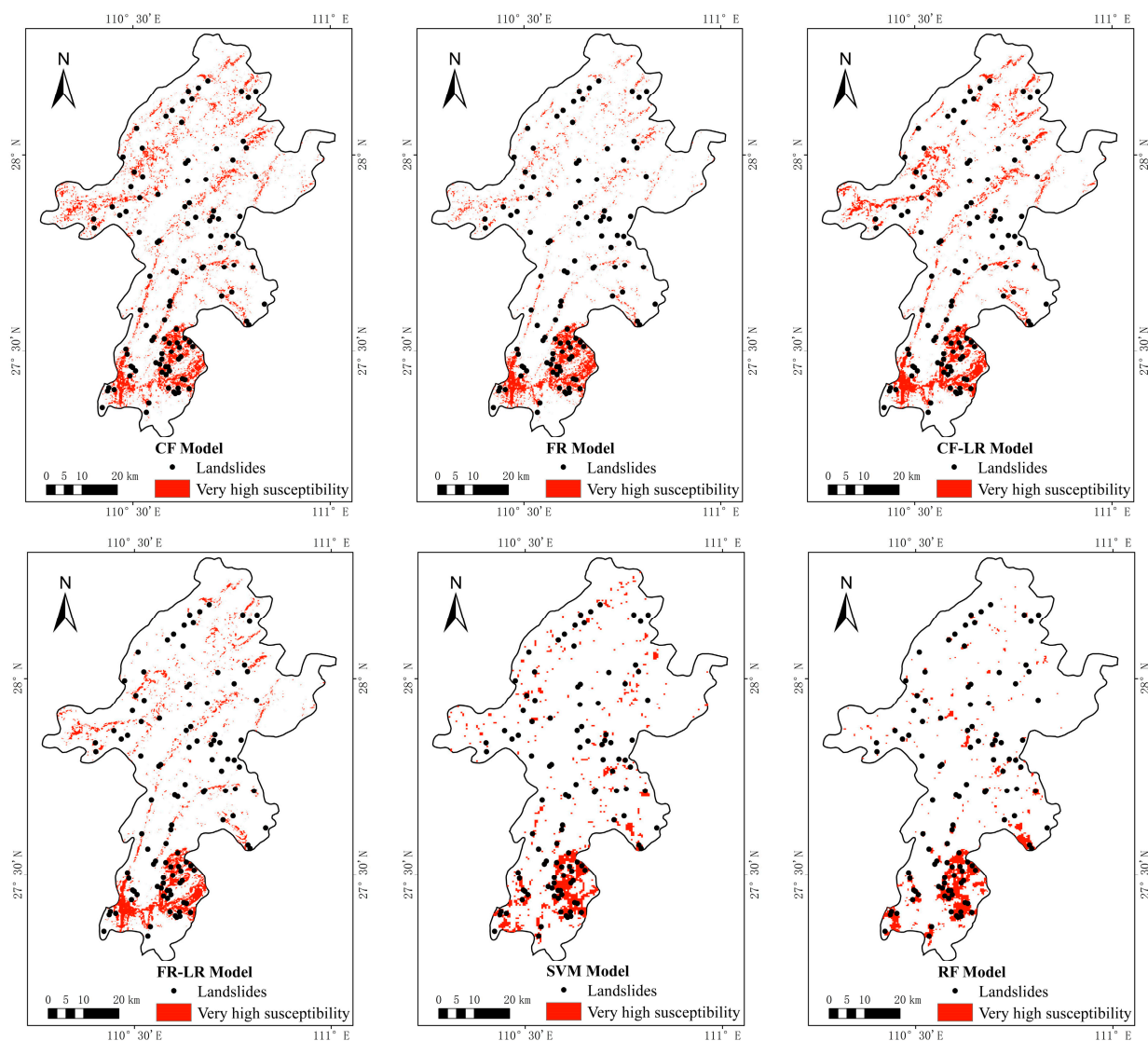


Figure 12. Six Models' High Landslide Susceptibility Zoning Maps.

Table 8. Number of Landslides in the Very High Susceptibility Zone.

Model	Number of Landslide	Occurrences Area/km ²	Landslide Density
FR Model	18	223.5	0.08
CF Model	18	321.9	0.06
FR-LR Model	16	245.9	0.07
CF-LR Model	15	306.2	0.05
SVM Model	28	243.6	0.11
RF Model	31	221.1	0.14

It can be observed that the RF model has the highest number of landslide occurrences within its very high susceptibility zone, totaling 31 events. Considering the AUC values obtained from the ROC curves in the previous section as well as the very high susceptibility zoning across the entire Xupu County, it is evident that the RF model demonstrates the best adaptability within the Xupu County region under both general conditions and specific circumstances.

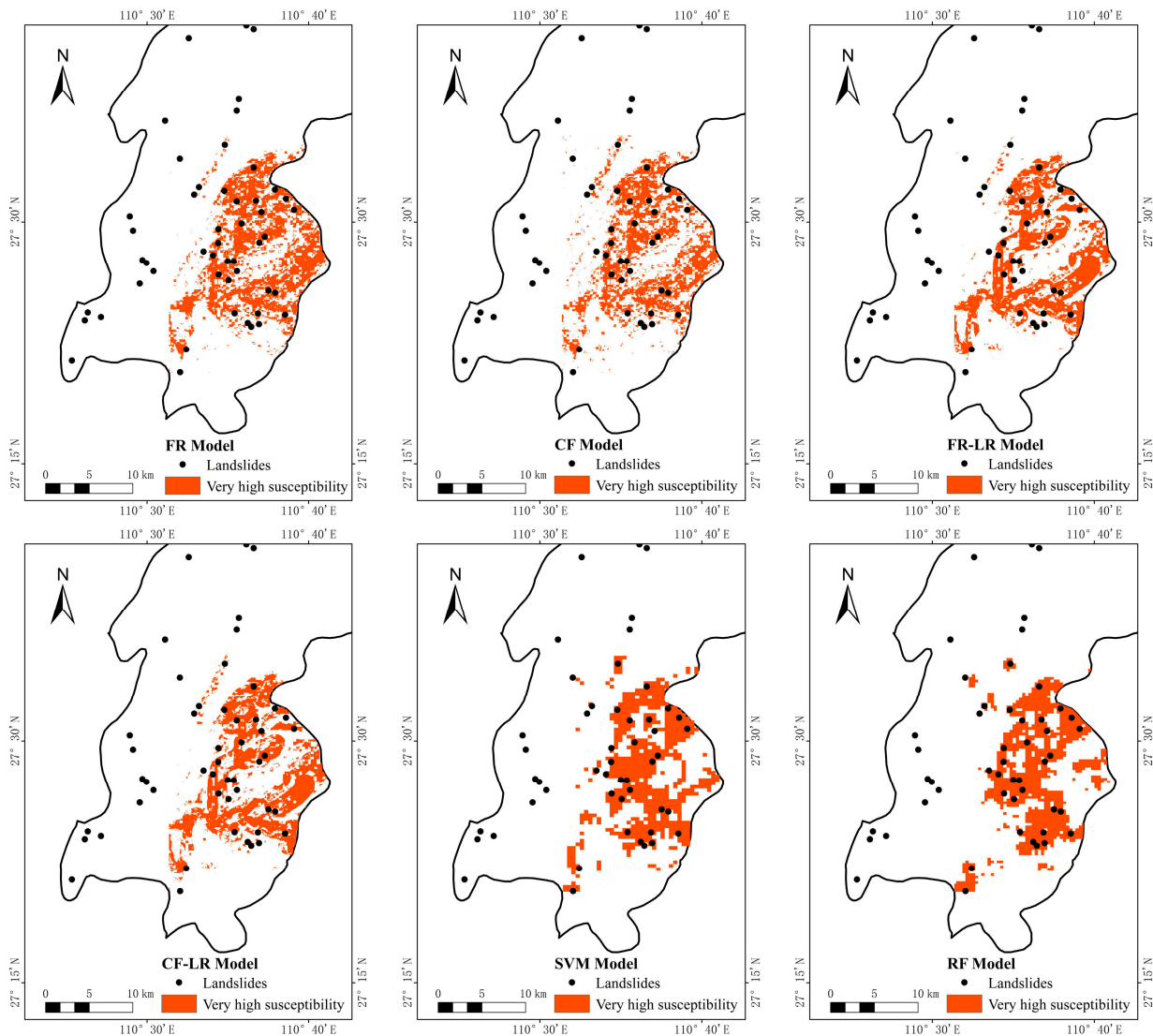


Figure 13. Regional Very High Susceptibility Zonation Map.

5. Discussion

5.1. Limitations in Landslide Research

The reliance on long-term regional landslide data for susceptibility assessment is undeniable, yet it brings forth challenges related to data completeness and accuracy. This study's use of data from 2001 to 2020 highlights the inherent gaps in certain years and regions, particularly in high-altitude and jungle areas. The intermittent nature of data collection, often relying on visual inspections conducted annually or biennially, further exacerbates these challenges.

The potential omission of minor landslides due to infrequent inspections raises concerns about the comprehensiveness of the dataset and its ability to capture the full spectrum of landslide occurrences. This gap in data could lead to underestimations of landslide susceptibility in certain areas, potentially compromising the effectiveness of mitigation strategies.

Exploring alternative data sources and methodologies is crucial to address these limitations. Remote sensing imagery, with its ability to provide frequent and consistent observations of the landscape, offers a valuable opportunity to supplement traditional data collection methods. Additionally, crowd-sourced data platforms, leveraging the power of citizen science, can contribute to more comprehensive and up-to-date landslide information.

The integration of these alternative data sources requires careful consideration of data quality, consistency, and compatibility. Developing robust data fusion techniques to combine information from various sources will be essential for creating a more accurate and reliable picture of landslide susceptibility in the study area.

5.2. Selection of Landslide Influencing Factors

The selection of landslide influencing factors plays a crucial role in the accuracy and reliability of susceptibility assessments. This study initially identified twelve potential factors, encompassing topographic, hydrological, lithological, and land use characteristics. However, the high correlation observed between several topographic factors and elevation raises important questions about their redundancy and the need for further refinement.

The decision to exclude elevation based on its strong correlation with other topographic variables necessitates a deeper exploration of the underlying relationships, even though elevation is undoubtedly an important driver of landslide susceptibility. It would be beneficial to consider the possibility of incorporating elevation indirectly through its impact on other factors.

Furthermore, the significance test employed in the study highlights the importance of carefully evaluating the relevance of each factor. The exclusion of certain factors based on statistical significance may not always reflect their true influence on landslide susceptibility. Exploring alternative methods for assessing factor importance, such as feature importance analysis in machine learning models, could provide a more comprehensive understanding of the contributing factors.

The differential approach taken for traditional statistical models (CF-LR and FR-LR) and machine learning models (SVM and RF) in terms of factor selection raises interesting questions about model complexity and computational efficiency. The inclusion of additional factors in the machine learning models may lead to improved predictive performance but could also increase the risk of overfitting. A more nuanced analysis of the trade-off between model complexity and accuracy would be valuable in determining the optimal approach for landslide susceptibility assessment in different contexts.

6. Conclusions

Landslide disasters pose significant hazards due to their sudden, regional, and explosive nature, coupled with the difficulty of timely prediction. They are a key focus in the field of disaster prevention and mitigation. This paper, by integrating factors such as topography, hydrology, lithology, and land use, establishes six landslide prediction models. Through the analysis of model accuracy, the most precise model suitable for the region is selected. This outcome can alert relevant departments to formulate preventive measures early on and serve as a reference for the disaster prevention and mitigation efforts of the Xupu County government.

Author Contributions: The ideas presented in this paper were provided by J.S., with Z.Z. responsible for writing it. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on the Geospatial Data Cloud.

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

References

1. Gariano, L.S.; Guzzetti, F. Landslides in a changing climate. *Earth-Sci. Rev.* **2016**, *162*, 227–252. [[CrossRef](#)]
2. Alessandro, N.; Catherine, P.; Kathryn, L.; Sophie, T.; Itahisa, G.; Emma, M.; Christian, A.; Annie, W. Mapping landslides from space: A review. *Landslides* **2024**, *21*, 1041–1052.
3. Sun, J. Hard particle force in a soft fracture. *Sci. Rep.* **2019**, *9*, 3065. [[CrossRef](#)] [[PubMed](#)]
4. Sun, J.; Wang, G. Transport model of underground sediment in soils. *Sci. World J.* **2013**, *2013*, 367918.

5. Sun, J.; Wang, G. Research on underground water pollution caused by geological fault through radioactive stratum. *J. Radioanal. Nucl. Chem.* **2013**, *297*, 27–32.
6. Sciarra, M.; Coco, L.; Urbano, T. Assessment and validation of GIS-based landslide susceptibility maps: A case study from Feltrino stream basin (Central Italy). *Bull. Eng. Geol. Environ.* **2017**, *76*, 437–456. [\[CrossRef\]](#)
7. Jiménez-Perálvarez, D.J.; Irigaray, C.; Hamdouni, E.R.; Chacón, J. Building models for automatic landslide-susceptibility analysis, mapping, and validation in ArcGIS. *Nat. Hazards* **2009**, *50*, 571–590. [\[CrossRef\]](#)
8. Huang, Y.; Zhao, L. Review on landslide susceptibility mapping using support vector machines. *Catena* **2018**, *165*, 520–529. [\[CrossRef\]](#)
9. Kadavi, R.P.; Lee, C.; Lee, S. Application of Ensemble-Based Machine Learning Models to Landslide Susceptibility Mapping. *Remote Sens.* **2018**, *10*, 1252. [\[CrossRef\]](#)
10. Guzzetti, F.; Cardinali, M.; Reichenbach, P.; Cipolla, F.; Sebastiani, C.; Galli, M.; Salavati, P. Landslides triggered by the 23 November 2000 rainfall event in the Imperia Province, Western Liguria, Italy. *Eng. Geol.* **2004**, *73*, 229–245. [\[CrossRef\]](#)
11. Schlögel, R.; Marchesini, I.; Alvioli, M.; Reichenbach, P.; Rossi, M. Optimizing landslide susceptibility zonation: Effects of DEM spatial resolution and slope unit delineation on logistic regression models. *Geomorphology* **2018**, *301*, 10–20. [\[CrossRef\]](#)
12. Wang, Z.; Wang, D.; Guo, Q.; Wang, D. Regional landslide hazard assessment through integrating susceptibility index and rainfall process. *Nat. Hazards* **2020**, *104*, 2153–2173. [\[CrossRef\]](#)
13. Yu, C.; Chen, J. Application of a GIS-based slope unit method for landslide susceptibility mapping in Helong City: Comparative assessment of ICM, AHP, and RF model. *Symmetry* **2020**, *12*, 1848. [\[CrossRef\]](#)
14. Wang, Z.; Ma, C.; Qiu, Y.; Xiong, H.; Li, M. Refined zoning of landslide susceptibility: A case study in Enshi County, Hubei, China. *Int. J. Environ. Res. Public Health* **2022**, *19*, 9412. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Ba, Q.; Chen, Y.; Deng, S.; Yang, J.; Li, H. A comparison of slope units and grid cells as mapping units for landslide susceptibility assessment. *Earth Sci. Inform.* **2018**, *11*, 373–388. [\[CrossRef\]](#)
16. Mansouri, R.M.R. Landslide susceptibility zonation using analytical hierarchy process and GIS for the Bojnurd region, northeast of Iran. *Landslides* **2014**, *11*, 1079–1091. [\[CrossRef\]](#)
17. Silalahi, S.E.F.; Pamela Arifianti, Y.; Hidayat, F. Landslide susceptibility assessment using frequency ratio model in Bogor, West Java, Indonesia. *Geosci. Lett.* **2019**, *6*, 10. [\[CrossRef\]](#)
18. Pourghasemi, R.H.; Moradi, R.H.; Aghda, F.M.S. Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. *Nat. Hazards* **2013**, *69*, 749–779. [\[CrossRef\]](#)
19. Choi, J.; Oh, H.; Lee, H.; Lee, C.; Lee, S. Combining landslide susceptibility maps obtained from frequency ratio, logistic regression, and artificial neural network models using ASTER images and GIS. *Eng. Geol.* **2011**, *124*, 12–23. [\[CrossRef\]](#)
20. Van Den Eeckhaut, M.; Reichenbach, P.; Guzzetti, F.; Rossi, M.; Poesen, J. Combined landslide inventory and susceptibility assessment based on different mapping units: An example from the Flemish Ardennes, Belgium. *Nat. Hazards Earth Syst. Sci.* **2009**, *9*, 507–521. [\[CrossRef\]](#)
21. Deng, H.; Wu, X.; Zhang, W.; Liu, Y.; Li, W.; Li, X.; Zhou, P.; Zhuo, W. Slope-unit scale landslide susceptibility mapping based on the random forest model in deep valley areas. *Remote Sens.* **2022**, *14*, 4245. [\[CrossRef\]](#)
22. Yeon, Y.; Han, J.; Ryu, H.K. Landslide susceptibility mapping in Injae, Korea, using a decision tree. *Eng. Geol.* **2010**, *116*, 274–283. [\[CrossRef\]](#)
23. Wang, L.J.; Guo, M.; Sawada, K.; Lin, J.; Zhang, J. Landslide susceptibility mapping in Mizunami City, Japan: A comparison between logistic regression, bivariate statistical analysis and multivariate adaptive regression spline models. *Catena* **2015**, *135*, 271–282. [\[CrossRef\]](#)
24. Shi-biao, B.; Ping, L.; Jian, W. Landslide Susceptibility Assessment of the Youfang Catchment using Logistic Regression. *J. Mt. Sci.* **2015**, *12*, 816–827.
25. Xing, X.; Wu, C.; Li, J.; Zhang, L.; He, R. Susceptibility assessment for rainfall-induced landslides using a revised logistic regression method. *Nat. Hazards* **2021**, *106*, 97–117. [\[CrossRef\]](#)
26. Park, S.-J.; Lee, C.-W.; Lee, S.; Lee, M.-J. Landslide susceptibility mapping and comparison using decision tree models: A Case Study of Jumunjin Area, Korea. *Remote Sens.* **2018**, *10*, 1545. [\[CrossRef\]](#)
27. Wu, X.; Song, Y.; Chen, W.; Kang, G.; Qu, R.; Wang, Z.; Wang, J.; Lv, P.; Chen, H. Analysis of Geological Hazard Susceptibility of Landslides in Muli County Based on Random Forest Algorithm. *Sustainability* **2023**, *15*, 4328. [\[CrossRef\]](#)
28. Van Den, E.M.; Hervás, J.; Jaedicke, C.; Malet, J.; Montanarella, L.; Nadim, F. Statistical modelling of Europe-wide landslide susceptibility using limited landslide inventory data. *Landslides* **2012**, *9*, 357–369. [\[CrossRef\]](#)
29. He, Q.; Jiang, Z.; Wang, M.; Liu, K. Landslide and wildfire susceptibility assessment in Southeast Asia using ensemble machine learning methods. *Remote Sens.* **2021**, *13*, 1572. [\[CrossRef\]](#)

30. Kim, J.C.; Lee, S.; Jung, H.S.; Lee, S. Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. *Geocarto Int.* **2018**, *33*, 1000–1015. [[CrossRef](#)]
31. Saito, H.; Nakayama, D.; Matsuyama, H. Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: The Akaishi Mountains, Japan. *Geomorphology* **2009**, *109*, 108–121. [[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.