remote sensing

MDPI

Article

# Landslide Susceptibility Zoning in Yunnan Province Based on SBAS-InSAR Technology and a Random Forest Model

**Meiyu Liu [1,2], Bing Xu [1,2,*], Zhiwei Li [1,2], Wenxiang Mao [1,2], Yan Zhu [1,2], Jingxin Hou [1,2] and Weizheng Liu [3,4]**

[1] School of Geosciences and Info-Physics, Central South University, Changsha 410083, China
[2] Radar Remote Sensing and Image Geodesy, Central South University, Changsha 410083, China
[3] National Engineering Research Center of High-Speed Railway Construction Technology, Changsha 410075, China
[4] School of Civil Engineering, Central South University, Changsha 410075, China
* Correspondence: xubing@csu.edu.cn; Tel.: +86-133-3251-4642

**Abstract:** Yunnan Province, China, has complex topography and geomorphology, many ravines and valleys, and frequent landslide geological disasters and is of great significance in the assessment of regional landslide geological disasters in Yunnan Province for disaster prevention and mitigation. In this study, Yunnan Province was selected as the research area, and the average annual deformation rate of radar line-of-sight in Yunnan Province over four years from 2018 to 2021 was obtained with SBAS-InSAR technology, which was used as one of the index factors for the susceptibility evaluation of Yunnan Province. The deformation rate reflects the slow movement of the land surface. In addition, elevation, slope, aspect, lithological classification, geological structure, rainfall, distance from roads, distance from rivers, topographic undulation, and NDVI were selected as evaluation index factors and combined with the annual mean deformation rate. A random forest model was used to evaluate and accurately analyze landslide geological disasters in Yunnan Province. The results showed that as an important index factor, the annual mean deformation rate of Yunnan Province can be added to the random forest model to improve the prediction accuracy. The area with high susceptibility accounted for 10% of the entire province, and the number of landslides in the region accounted for 68% of the province. Additionally, the results for prone zoning were highly correlated with the landslide distribution. The accuracy of the random forest model prediction was 0.80, and the AUC value was 0.87, indicating that the random forest model was a highly accurate and reliable evaluation method for studying landslide geological disasters. It is very suitable for the evaluation of landslide susceptibility in Yunnan Province.

**Keywords:** SBAS-InSAR; random forest; Yunnan Province; susceptibility assessment



**Citation:** Liu, M.; Xu, B.; Li, Z.; Mao, W.; Zhu, Y.; Hou, J.; Liu, W. Landslide Susceptibility Zoning in Yunnan Province Based on SBAS-InSAR Technology and a Random Forest Model. *Remote Sens.* **2023**, *15*, 2864. https://doi.org/10.3390/rs15112864

Academic Editors: Marco Polcari, Letizia Anderlini and Antonio Montuori

Received: 25 April 2023
Revised: 28 May 2023
Accepted: 29 May 2023
Published: 31 May 2023



## 1. Introduction

China's Yunnan Province has a complex environment, many ravines and valleys, and large terrain fluctuations, leading to a high incidence of geological disasters in the country. Landslide disasters are the most frequent geological disasters in the region and occur frequently in Yunnan Province, representing a great threat to people's lives and property [1].

As one of the most common disasters in Yunnan Province, China, landslides have been studied many times. Research on landslide susceptibility mainly focused on two aspects: landslide susceptibility evaluation systems and landslide susceptibility methods. At present, landslide disaster susceptibility methods mainly include empirical models (fuzzy logic, analytic hierarchy, etc.) [2,3], statistical models (information quantity, deterministic coefficients, etc.) [4,5], and machine learning models (neural networks, decision trees, support vector machines, etc.) [6,7]. Among them, when analyzing and predicting large areas at the provincial and municipal levels, the accuracy and applicability of empirical models and statistical analysis models were found to be low. Machine learning models have the problems of weak explanatory value and overfitting of prediction results. With the rapid development of machine learning, the random forest model is now commonly used by scholars because of its unique advantages. A random forest model can not only meet the problem of nonlinear features of landslides adapted to other machine learning models but also introduce randomness, avoid overfitting the model, and improve prediction accuracy [8]. Many studies have proved that the random forest model has a good prediction effect and applicability in the study of landslide geological hazard susceptibility assessment [8–11]. Therefore, in this study, the random forest model was used to evaluate the vulnerability of landslide geological disasters in Yunnan Province, China, and obtain more accurate landslide susceptibility zoning in Yunnan Province.

Regarding the study of landslide susceptibility evaluation factors, in 2018, Reichenbach et al. [12] found that 596 landslide susceptibility index factors were identified in the previous research literature. However, topography, landform, hydroclimate, and other landslide development conditions are different in different study areas, so many researchers select only some of these conditions as index factors for their research according to the actual situation in the study area [13–15]. For example, in 2022, for the investigation and zoning of geological disasters in Helong City, Wang X.D. et al. [16] selected 13 disaster factors including elevation, slope, aspect, curvature, lithology, distance from faults, rainfall, distance from water systems, NDVI, soil texture, hydraulic erosion, population density, and distance from roads. They evaluated the vulnerability of the study area based on the investigation and zoning of geological hazards combined with the distribution law and influencing factors of geological disasters and comprehensively considered five factors: topography, geology, meteor hydrology, soil vegetation, and human engineering activities. However, the impact of land surface deformation information on landslide disasters is rarely considered in the above index factors. The land surface deformation rate can reflect the development stage of the landslide. The surface deformation rate can also detect hidden landslide dangers that cannot be found using optical remote sensing images and can evaluate the vulnerability of landslides more comprehensively, so it is very useful to add the annual mean deformation rate as an index factor to the vulnerability evaluation.

Minor surface deformation information will be observable before landslides occur, and the land surface deformation information can be obtained with temporal InSAR (Interferometric Synthetic Aperture Radar) technology. InSAR technology has the ability to obtain a large range of small deformation information and has a good monitoring effect on the area undergoing deformation. This method has achieved many fruitful applications in landslide monitoring [17–20]. InSAR technology is not affected by weather and has a wide coverage, which is very suitable for long-term observations and geological disaster investigations over a large area [20]. Yunnan Province is an alpine and hilly area. SBAS-InSAR (Small Baseline Subset–Interferometric Synthetic Aperture Radar) technology uses a small baseline set to obtain the deformation rate, which can effectively avoid the problem of temporal and spatial coherence, and it is an effective method to monitor the surface deformation in mountainous areas. To date, few scholars have used annual mean deformation rate results obtained with InSAR technology as landslide susceptibility factors, and no researchers have applied annual average deformation rate results as influencing factors to random forest models for susceptibility evaluation studies. To fill these gaps, in the present study, results for the annual mean deformation rate obtained using time-series InSAR technology are introduced into a random forest model as the influencing factors to evaluate the susceptibility of landslide geological disasters in Yunnan Province.

## 2. Methods

### *2.1. SBAS-InSAR Method and Deformation Result Acquisition*

#### 2.1.1. SBAS-InSAR Method

The SBAS-InSAR method, proposed by Berardino et al., divides SAR (Synthetic Aperture Radar) images obtained in the same range that meet selected conditions into several small baseline sets by selecting short time and space baselines for differential interference processing. The small spatial and temporal baselines in the SAR images from each small set that meet the relevant conditions can improve the coherence of image pairs and increase the number of differential interferograms. Then, we calculate the phase changes in each baseline set using the least squares method to obtain the desired surface deformation information. Since the SBAS method introduces multiple images as the main image in the process of realizing differential interference, matrix B is more prone to rank loss. Therefore, it is necessary to solve the minimum norm of the matrix with the SVD (Singular Value Decomposition) method. Then, by differentiating the deformation velocity values in different time periods, the shape variables generated in different time periods can be obtained [21–24].

#### 2.1.2. Deformation Result Acquisition

In this paper, SBAS-InSAR technology is used to obtain the annual average deformation rate results in Yunnan province, and the results are used as an index factor to evaluate the susceptibility. Sentine1-1A ascending track data for four years, from January 2018 to December 2021, were reviewed and selected according to the regional size of Yunnan Province. Sentine1-1A data is freely available. With a short period and C-band, it can obtain small deformation results on a wide range of surfaces. Due to the large area of Yunnan Province, the Sentine1-1A data contained four tracks, and the image data on each track basically covered the area of Yunnan Province. Table 1 includes the number of tracks, the number of images, and the number of baselines selected for data processing.

**Table 1.** Image data from Sentine1-1A in Yunnan Province.

| Track Number | Number of Images | Interfere Pairs/Piece |
|---|---|---|
| 26 | 120 | 237 |
| 99 | 116 | 229 |
| 128 | 121 | 239 |
| 172 | 117 | 231 |

The main process for Sentine1-1A data processing in Yunnan Province was as follows: (1) The first scene image of each track is selected as the main image, and the other images are used as auxiliary images. Baseline selection is carried out using the triangulation locking method because there are a large number of images in each track. If selection is carried out using the time and space baseline threshold, unsatisfied baselines need to be eliminated for areas with poor coherence, which is very difficult to complete. Finally, M image pairs are obtained for each track. (2) The auxiliary image on each track is registered with the corresponding main image; that is, all the auxiliary image data are resampled to the main image space, and then the M image pairs are interfered according to the baseline map to obtain the required interference map. Next, an 8:2 multiple view ratio is used to suppress the speckle noise from the interferogram and improve the pixel signal-to-noise ratio. The terrain phase in Yunnan Province is simulated using the external SRTM (Shuttle Radar Topography Mission) DEM (Digital Elevation Model) with a resolution of 30 m. Different processing is carried out with the interferogram, and the obtained differential interferogram is used to remove the terrain phase. A Gaussian filter is used to filter the differential interferogram to further reduce the signal noise. After setting a certain coherence threshold, the pixels whose coherence is higher than the threshold are selected for phase disentanglement processing. (3) We select a control point with high coherence, estimate the residual phase in the initial disentanglement phase according to the phase information on the selected control point, and remove the residual terrain phase. (4) The singular value decomposition method is used to transform the differential interference results into the same reference target, and then high-pass filtering in the time domain and low-pass filtering in the space domain are used to remove the influence of the atmosphere.

From these processes, the final results for the annual average deformation rate for the ascending orbit in Yunnan Province are obtained [21–24].

The results for the annual average deformation rate of the ascending orbit in Yunnan Province are shown in Figure 1. Here, the positive and negative maximum annual average deformation rates in Yunnan province are −25.6 and 19.0 cm/y, respectively.

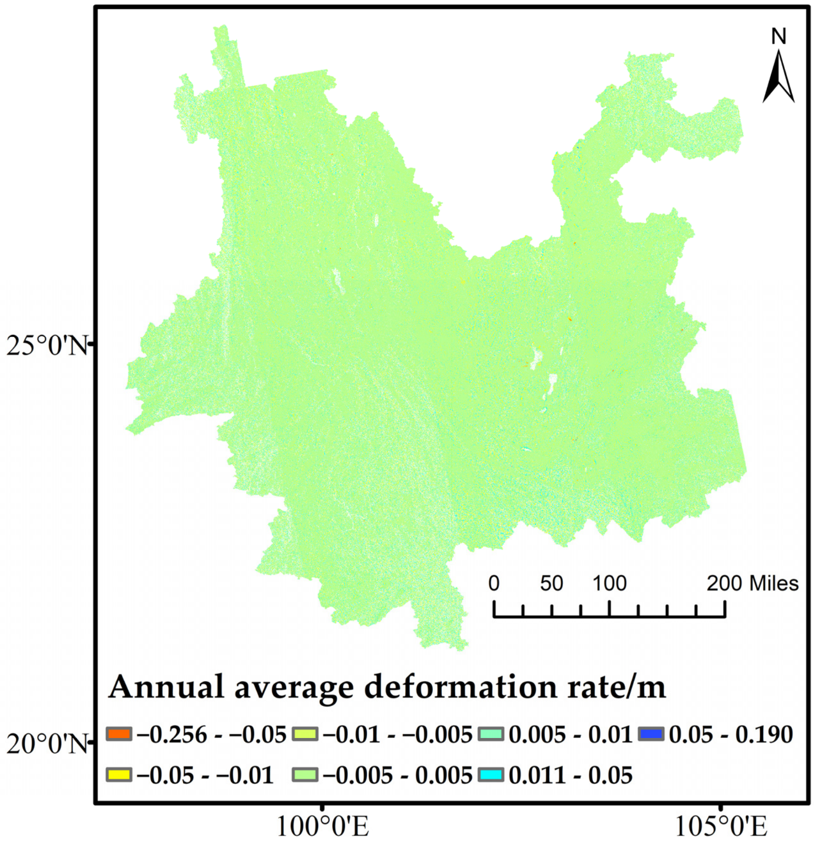


**Figure 1.** Classification of annual average deformation rate results in Yunnan Province.

According to the deformation rate results for Yunnan Province and the analysis of the deformation rate for hidden landslide dangers, the deformation rate results are manually divided into the following seven levels according to a certain level in the ArcGIS software: −0.256~−0.050, −0.050~−0.010, −0.010~−0.005, −0.005~0.005, 0.005~0.010, 0.010~0.050, and 0.050~0.180 m.

## 2.2. Random Forest Model

### 2.2.1. Random Forest Model Method

In 2001, Breiman built the random forest (RF) model by assembling many classified trees [25]. The principle of the random forest model is to extract k samples from the total sample as training samples. The number of decision trees to be trained is then determined using the number of training samples drawn, thereby generating k decision trees to form a random forest model. Each decision tree acts as a classifier to produce a classification result, and the final classification prediction is made by voting [12]. For random forest models, the number of training samples selected is generally two-thirds the total number of samples, and the other one-third is used as test samples to test model performance. In recent years, sufficient experimental and theoretical studies have verified that this model is superior to other susceptibility evaluation models because the random forest model not only improves the accuracy of prediction but also has good tolerance for errors such as noise; thus, RF is the best susceptibility evaluation model that can be selected [13]. The prediction flow in the random forest model is shown in Figure 2.

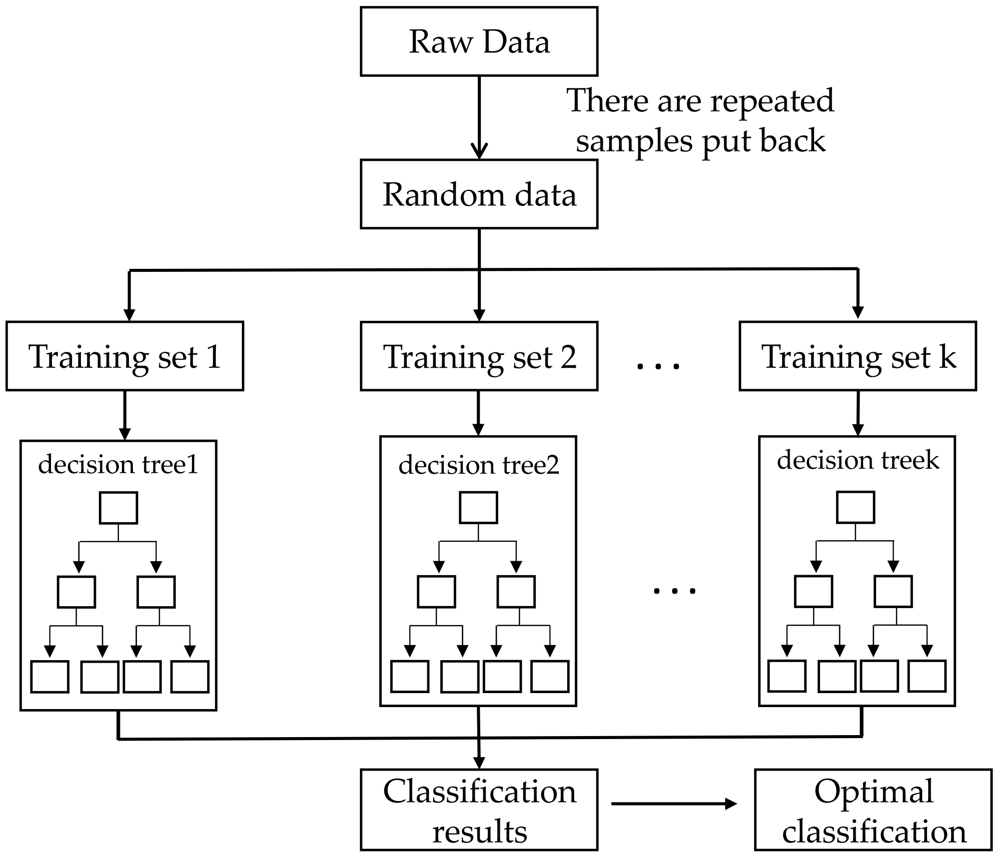


**Figure 2.** Flowchart showing predictions under the random forest model.

2.2.2. Model Training

In this paper, the landslide interpretation and identification of Yunnan Province are carried out using visual interpretation, and 1385 hidden landslide danger points with obvious deformation characteristics are delineated. These points are used here as positive data sets, and the same number of random points are randomly created with the ArcGIS software as negative data sets. A total of 2770 points are used in the susceptibility assessment. Thirty percent of the data for the entire sample is selected as the model training set, and the rest of the data are selected as the model test sample. The training set for the random forest model measures the fitting ability of the model, while the test set reflects the generalization ability of the model. Thus, the performance of the model should be evaluated using both the training set and the test set.

## 3. Study Area and Data Sources

### *3.1. Study Area*

Yunnan Province is located in the hilly area of the Qinghai–Tibet Plateau, located between 21°8′32″–29°15′8″ northern latitude and 97°31′39″–106°11′47″ eastern longitude. The elevation gradually decreases from north to south, where the southeast region is lower than the northwest region, and the highest point is distributed on the main peak of Meili Snow Mountain. The province mainly contains basins, mountains, and plateaus, of which basins occupy a minimal area, while mountains and plateaus account for 94% of the area in the province, including canyons and deep mountains. Yunnan Province features more ravines with a large height difference, and the terrain is more undulating than that in other regions. Yunnan Province is located at the junction of the Eurasian plate and the Indian plate, and the new and old structural geology is quite complex, mainly based on faults. Fault zones are distributed throughout the province, providing essential materials and energy for the development of geological disasters. River systems passing through the province mainly include the Nan Pan River, the Nu River, the Irrawaddy River, the Red River, the Lancing River, and the Jinsha River. Most rivers are representative of mountainous rivers, which are characterized by rapid water flow and large drops. The rivers are mainly distributed along the valley such that the valley slopes on both sides are steep, thereby providing the main breeding area for mountain disasters [1]. Figure 3 shows the distribution map of historical landslide points in Yunnan Province.

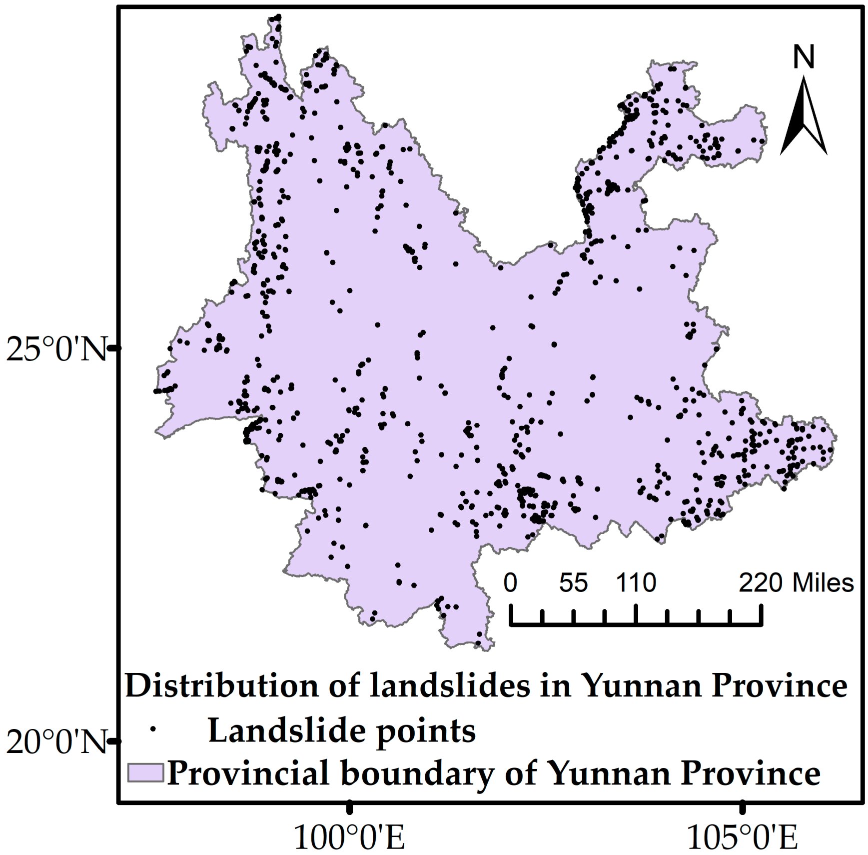


**Figure 3.** Distribution map showing administrative boundaries and landslide hazard points in Yunnan Province.

*3.2. Data Sources*

The occurrence of landslide disasters is closely related to topography, hydrometeorology, human activities, and other factors and is affected by the external environment of the landslide and the slope's own conditions. Additionally, the formation mechanism is complex [26]. On the basis of previous research [27,28], combined with the development characteristics and geological environmental conditions of landslide geological disasters in Yunnan Province, as well as the difficulty and operability of obtaining influencing factors, the following influencing factors were selected for the risk assessment of landslide geological disasters in Yunnan Province: slope, aspect, elevation, and topographic undulation factors were selected from the topography and geomorphology; fault and lithology factors were selected from the geological structure; river and rainfall factors were selected from meteorological hydrology; the road factor was selected from human activity; the NDVI factor was selected from the vegetation cover degree; and the annual average deformation rate factor. An index system for evaluating the risk of landslide geological disasters in Yunnan Province was then established. The landslide disaster data used in this paper are composed of geological disaster data with obvious landslide signs on optical images, and other index factor data sources are shown in Table 2.

**Table 2.** Data sources.

| Data Source | Data Name | Type | Scales |
|---|---|---|---|
| Geo-monitoring cloud platform | DEM | Raster | 30 m |
| | Slope | Raster | 30 m |
| | Aspect | Raster | 30 m |
| | Rainfall | Vector | 1000 m |
| Resource Science Data Center, Chinese Academy of Sciences | Road | Vector | 1:100,000 |
| | River | Vector | 1:100,000 |
| | Fault | Vector | 1:100,000 |
| | Lithology | Vector | 1:10,000 |
| | NDVI | Raster | 30 m |
| | Terrain undulation | Raster | 90 m |
| Optical image sketching | Landslide disaster points | Vector | |
| Annual average deformation rate | Sentinel-1 data | | |

## 4. Indicator Factor Data Processing

This paper uses a variety of classification and grading methods to study the classification and grading of index factors. According to the characteristics of the selected index factors, such as elevation, slope, annual average rainfall, annual mean deformation rate, and NDVI five factors, we used a manual method to divide the grades, making the landslide distribution more concentrated and reasonable. The aspect indicator was graded by direction. For the fault, road, and river index factors, the GIS software buffer was used for grading with a buffer step size of 400 m. We divided the different types of rocks according to the qualitative division table of the hardness of the rocks [29]. According to the degree of topographic undulation in the study area, the terrain undulation degree was also divided into 7 common types.

### *4.1. Topography and Landform*

#### 4.1.1. DEM

Landslide disasters in Yunnan Province are affected by altitude and other factors, and elevation is closely related to these factors. There are great differences in altitude among various areas in Yunnan Province, and the difference between the north and the south reaches 6000 m. Altitude affects the potential energy of the slope body, which indirectly affects the disaster degree of the landslide. Therefore, elevation is regarded as one of the most important evaluation factors in the process of landslide susceptibility evaluation [30].

The overall elevation of Yunnan Province shows that the northwest region is higher than the southeast region. Here, the manual grading method was used to classify the 30 m resolution NASA DEM, and the elevation was statistically analyzed according to five levels: 0–1000, 1000–2000, 2000–3000, 3000–4000, and 4000–6700 m. Table 3 shows the number and proportion of landslides at all levels of elevation in Yunnan Province. Figure 4 shows the distribution of landslides in Yunnan Province. As shown in Table 3, landslide disasters mainly occur at elevations of less than 3000 m.

**Table 3.** The number and proportion of landslides distributed at each grade of elevation.

| DEM Grading (m) | Number of Landslide Disasters | Proportion of Total Disasters (%) |
|---|---|---|
| 0–1000 | 213 | 15 |
| 1000–2000 | 846 | 61 |
| 2000–3000 | 229 | 17 |
| 3000–4000 | 70 | 5 |
| 4000–6700 | 27 | 2 |

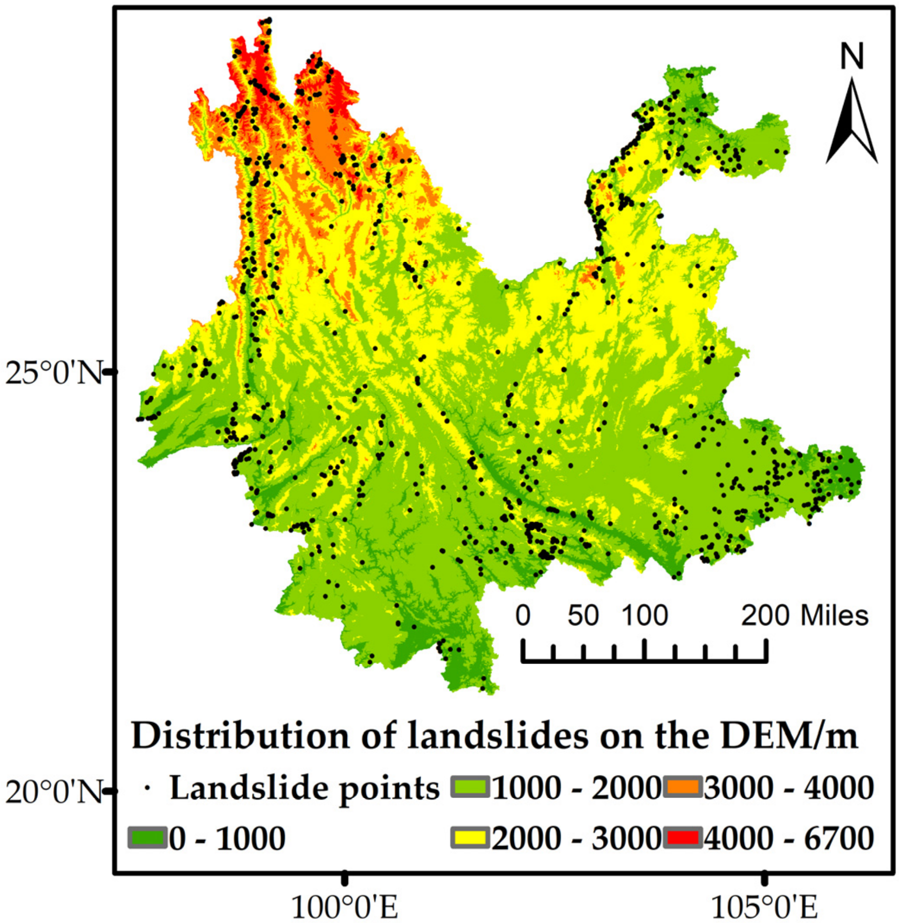


**Figure 4.** Distribution of landslides based on the DEM in Yunnan Province.

#### 4.1.2. Slope

Slope refers to the steepness of the slope body, which has a direct impact on the stability of the slope body. When a landslide occurs, it requires a certain effective landing surface, and there is also a large correlation between the slope and landing surface [31].

Using ArcGIS software to analyze the slope of DEM with a resolution of 30 m in Yunnan Province, the slope distribution range in Yunnan Province was obtained. The statistics found that the probability of landslides occurring on relatively gentle slopes less than 10°, and greater than 50° was very low, so the slopes were divided into six intervals based on a grade of 10°: <10°, 10–20°, 20–30°, 30–40°, 40–50°, and >50°. Table 4 shows the number and proportion of landslides for all slope grades in Yunnan Province. Figure 5 shows the distribution of landslides in Yunnan Province. According to Table 4, landslides mainly occur on slope bodies with slopes less than 50°.

**Table 4.** The number and proportion of landslides distributed under each grade of slope.

| Slope Grading (°) | Number of Landslide Disasters | Proportion of Total Disasters (%) |
|---|---|---|
| <10 | 69 | 5 |
| 10–20 | 282 | 21 |
| 20–30 | 438 | 32 |
| 30–40 | 411 | 30 |
| 40–50 | 131 | 10 |
| >50 | 23 | 2 |

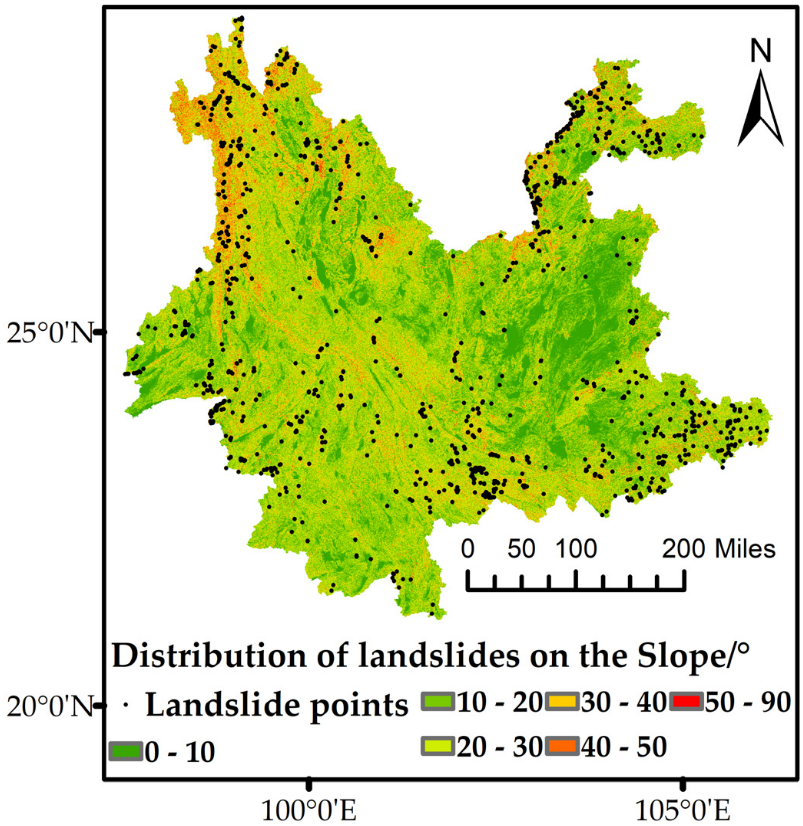


**Figure 5.** Distribution of landslides in Yunnan Province based on slope.

4.1.3. Aspect

Aspect is the basic element of terrain composition, and different aspects will cause differences in sunshine, vegetation growth, surface water flow, and surface temperature. This is indirectly related to the safety of the slope. Therefore, aspect is regarded as one of the most important evaluation factors in the process of landslide susceptibility evaluation.

In this paper, ArcGIS software is used to analyze the aspects of the DEM in Yunnan Province, with the aspects graded based on the direction to obtain the aspect distribution range in Yunnan Province. The slope direction in the study area was divided into nine grades: plane (−1), north (337.5–22.5), northeast (22.5–67.5), east (67.5–112.5), southeast (112.5–157.5), south (157.5–202.5), southwest (202.5–247.5), west (247.5–292.5), and northwest (292.5–337.5). Table 5 presents the number and proportion of landslides under each grade of slope in Yunnan Province, while Figure 6 shows the distribution of landslides in Yunnan Province. As shown in Table 5, landslides are not likely to occur on a flat surface, and landslides are most commonly distributed in the southeast direction.

**Table 5.** The number and proportion of landslides distributed in each grade of aspect.

| Aspect Grading (°) | Number of Landslide Disasters | Proportion of Total Disasters (%) |
|---|---|---|
| −1 | 1 | 0 |
| 337.5–22.5 | 141 | 11 |
| 22.5–67.5 | 123 | 9 |
| 67.5–112.5 | 180 | 13 |
| 112.5–157.5 | 258 | 19 |
| 157.5–202.5 | 190 | 14 |
| 202.5–247.5 | 191 | 14 |
| 247.5–292.5 | 140 | 11 |
| 292.5–337.5 | 125 | 9 |

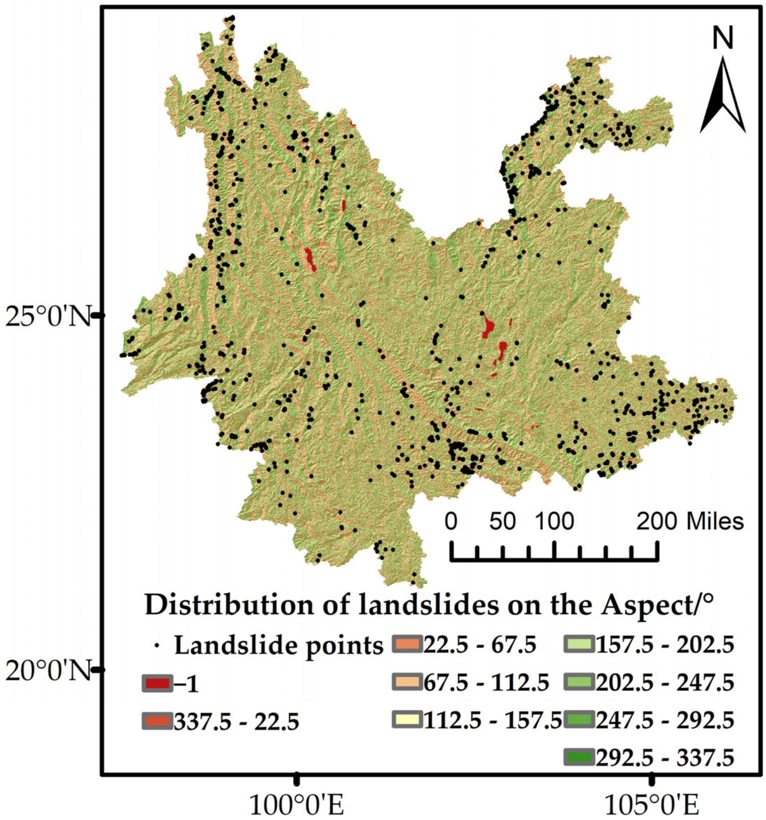


**Figure 6.** Distribution of landslides in Yunnan Province based on aspect.

#### 4.1.4. Terrain Undulation

Terrain undulation is the geomorphological characteristic of the difference between the altitude of the highest point and the lowest point in a certain area. Terrain undulation is related to the stress distribution and size inside the slope body, and as its value increases, the stability of the slope becomes worse. Therefore, terrain undulation is regarded as one of the evaluation factors in the process of landslide susceptibility evaluation.

According to the degree of terrain undulation, we divide the topographic undulations of Yunnan Province into seven categories: plains, terraces, hills, small undulating mountains, medium undulating mountains, large undulating mountains, and extremely large undulating mountains. Topographic undulation can be used as an important reference index for dividing geomorphological morphology. Table 6 presents the number and proportion of landslides for each level of topographic undulation in Yunnan Province, and Figure 7 shows the distribution of landslides in Yunnan Province. As shown in Table 6, the number of landslides is high on medium undulating mountains.

**Table 6.** The number and proportion of landslides for each grade of terrain undulation.

| NDVI Grading | Number of Landslide Disasters | Proportion of Total Disasters (%) |
|---|---|---|
| Plain | 6 | 0 |
| Mesa | 10 | 1 |
| Hilly | 29 | 2 |
| Small undulating mountains | 210 | 15 |
| Medium undulating mountains | 709 | 52 |
| Large undulating mountains | 372 | 27 |
| Extremely large undulating mountains | 36 | 3 |

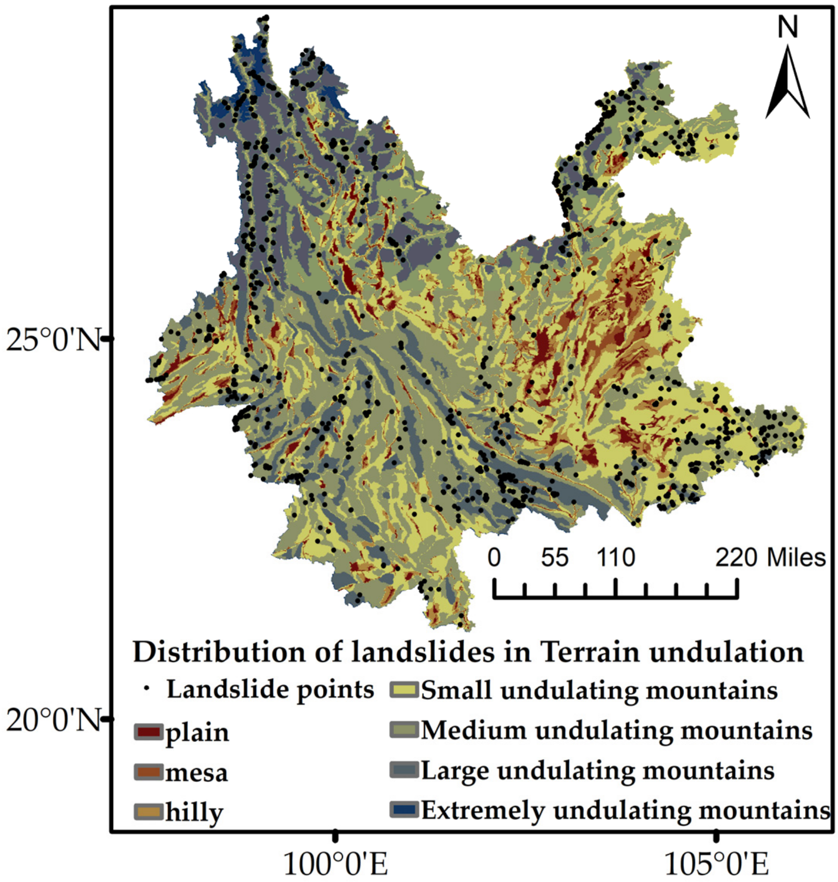


**Figure 7.** Distribution of landslides in Yunnan Province based on terrain undulation.

## *4.2. Geological Structure*

### 4.2.1. Distance from the Fault

Faults have a direct impact on the regional characteristics of landslides. Faults affect surface infiltration, internal structure, and the stress of rock and soil. Therefore, faults are regarded as one of the evaluation factors in the process of landslide susceptibility evaluation.

In this study, we used the buffer analysis tool in ArcGIS software to buffer the data according to a distance of 400 m, which was divided into seven grades: 0–400, 400–800, 800–1200, 1200–1600, 1600–2000, 2000–2400, and more than 2400 m. However, the range of Yunnan Province is very large, and most landslides are distributed at a distance greater than 2400 m. Table 7 presents the distribution and proportion of landslides at each distance from the fault in Yunnan Province, and Figure 8 shows the distribution of landslides in Yunnan Province based on fault distances.

**Table 7.** The number and proportion of landslide distribution at each grade based on fault distance.

| Distance from Fault Classification (m) | Number of Landslide Disasters | Proportion of Total Disasters (%) |
|---|---|---|
| 0–400 | 31 | 2 |
| 400–800 | 31 | 2 |
| 800–1200 | 33 | 2 |
| 1200–1600 | 39 | 3 |
| 1600–2000 | 22 | 2 |
| 2000–2400 | 26 | 2 |
| >2400 | 1202 | 87 |

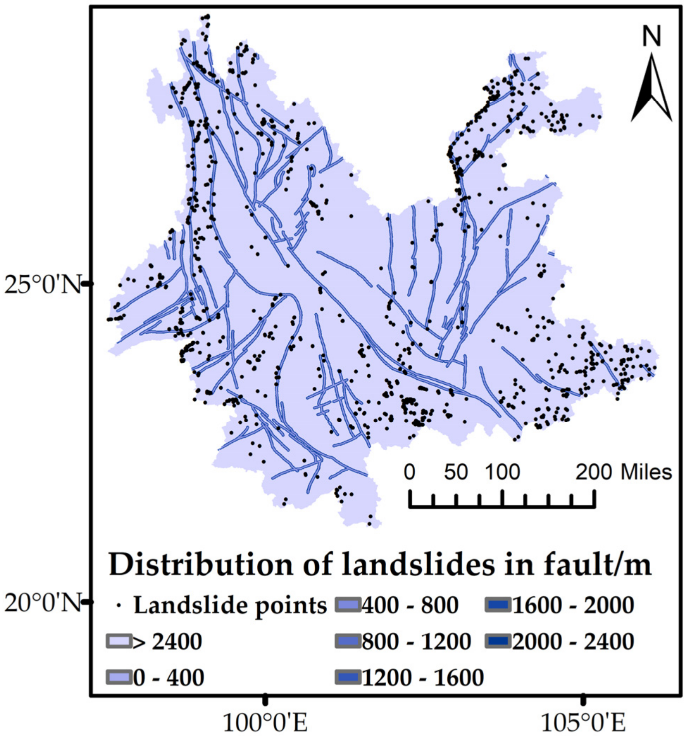


**Figure 8.** Distribution of landslide distances from faults in Yunnan Province.

### 4.2.2. Lithological Classification

The hardness, type, structure, etc., of rocks all determine the stability and integrity of the slope. Lithology plays a decisive role in the resistance of the slope to erosion, weathering, and fragmentation. Therefore, lithology is regarded as one of the evaluation factors in the process of landslide susceptibility evaluation.

Here, we divided the different types of rocks according to the qualitative division table of the hardness of rocks [29]. The geological rock group was divided into ten categories: harder rock sandwiching soft rock, harder rock, harder rock sandwiching softer rock, loose hard rock sandwiching softer rock, hard rock, hard rock sandwiching soft rock, softer rock, loose body, water body, and soft rock. Table 8 shows the number and proportion of landslides for each grade of the annual lithology classification in Yunnan Province, and Figure 9 shows the distribution of landslides based on the lithology classification of Yunnan Province. Table 8 shows that harder rocks sandwiching softer rock are common in areas with the greatest landslide distribution.

**Table 8.** The number and proportion of landslides for each grade of lithological classification.

| Lithological Classification Grading | Number of Landslide Disasters | Proportion of Total Disasters (%) |
|---|---|---|
| Harder rock sandwiching soft rock | 232 | 17 |
| Harder rock | 180 | 13 |
| Harder rock sandwiching softer rock | 360 | 26 |
| Loose hard rock sandwiching softer rock | 19 | 1 |
| Hard rock | 214 | 16 |
| Hard rock sandwiching soft rock | 158 | 12 |
| Softer rock | 27 | 2 |
| Loose body | 11 | 1 |
| Water body | 0 | 0 |
| Soft rock | 158 | 12 |

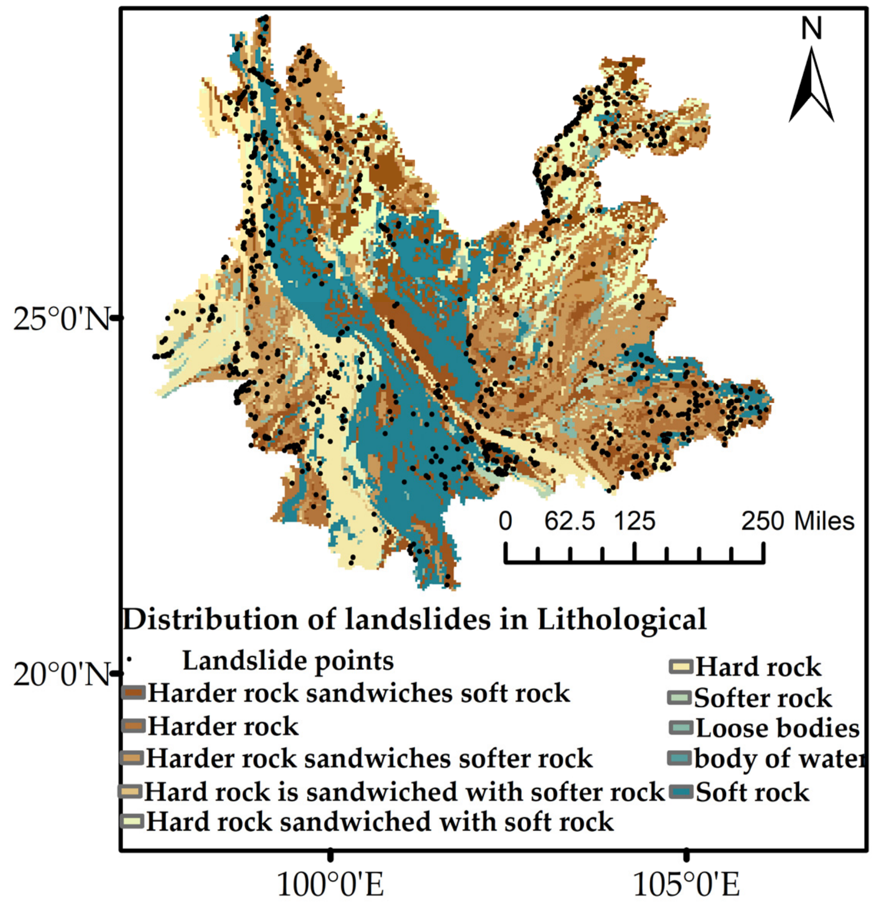


**Figure 9.** Distribution of landslides in Yunnan Province based on lithological classification.

### *4.3. Meteorological Hydrology*

#### 4.3.1. Distance from Rivers

The river cutting depth and river erosion in Yunnan Province has a certain role in promoting the formation of landslides. The erosion of the river can increase the effective empty surface of the slope, which in turn reduces the stability of the rock and soil mass. Therefore, rivers are regarded as one of the evaluation factors in the process of landslide susceptibility evaluation.

Using the buffer analysis tool in ArcGIS software, the distance between the study area and a river was buffered according to a distance of 400 m, which was divided into seven grades: 0–400, 400–800, 800–1200, 1200–1600, 1600–2000, 2000–2400, and more than 2400 m. However, the range of Yunnan Province is very large, and most landslides are distributed at a distance greater than 2400 m. Table 9 presents the distribution and proportion of landslides at each level of distance from rivers in Yunnan Province, and Figure 10 shows the distribution of landslides in Yunnan Province based on river distances.

**Table 9.** The number and proportion of landslides distributed at each grade based on river distance.

| Distance from River Classification (m) | Number of Landslide Disasters | Proportion of Total Disasters (%) |
|---|---|---|
| 0–400 | 94 | 4 |
| 400–800 | 63 | 3 |
| 800–1200 | 54 | 2 |
| 1200–1600 | 43 | 3 |
| 1600–2000 | 46 | 2 |
| 2000–2400 | 47 | 2 |
| >2400 | 1037 | 86 |

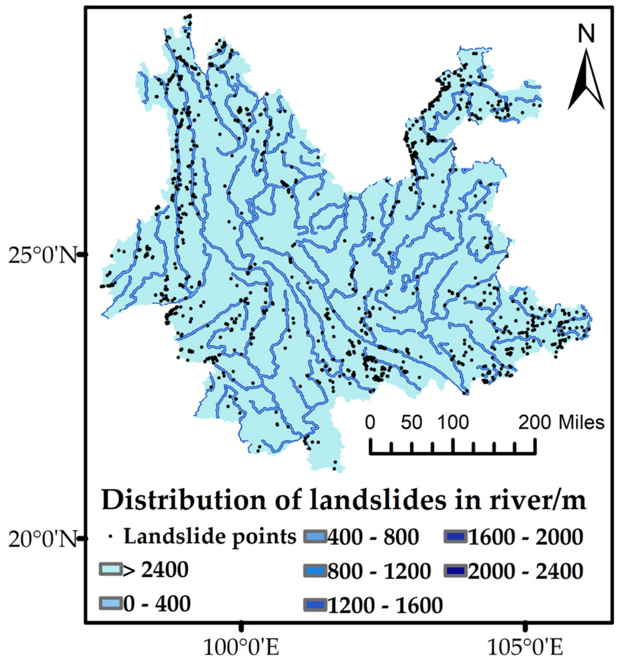


**Figure 10.** Distribution of landslide distances from rivers in Yunnan Province.

### 4.3.2. Average Annual Rainfall

An increase in rainfall increases the weight of a slope itself, weakens its shear resistance, and is the trigger factor that causes rock and soil mass to lose balance. Therefore, rainfall is regarded as one of the important evaluation factors in the process of landslide susceptibility evaluation.

In this paper, the average annual rainfall in the study area was graded using the manual grading method and divided into the following six levels: 1.3–2.0, 2.0–3.0, 3.0–4.0, 4.0–5.0, and >5.0 m. Table 8 presents the distribution and proportion of landslides for each grade of average annual rainfall in Yunnan Province, and Figure 11 shows the distribution of landslides in Yunnan Province. As shown in Table 10, landslides are mainly distributed in areas with an average annual rainfall of 2–5 m.

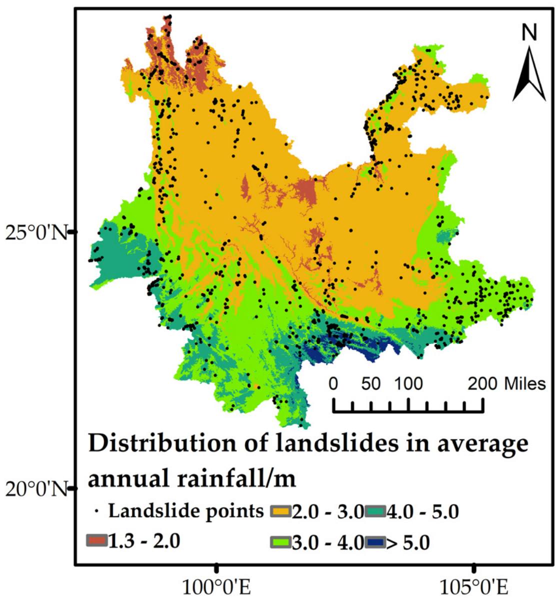


**Figure 11.** Distribution of landslides in Yunnan Province based on average annual rainfall.

**Table 10.** The number and proportion of landslides for each grade of average annual rainfall.

| Average Annual Rainfall Grading (m) | Number of Landslide Disasters | Proportion of Total Disasters (%) |
|---|---|---|
| 1.3–2.0 | 112 | 8 |
| 2.0–3.0 | 565 | 41 |
| 3.0–4.0 | 385 | 28 |
| 4.0–5.0 | 279 | 21 |
| >5.0 | 29 | 2 |

## *4.4. Human Activity*

### Distance from Roads

The construction of a road directly affects the balance of a slope, destroys the stability of the slope body, and makes the effective empty surface steeper. Therefore, roads are regarded as one of the evaluation factors in the process of landslide susceptibility evaluation.

Using the buffer analysis tool in ArcGIS software, the roads in the study area were analyzed according to a distance of 400 m, which was divided into seven grades: 0–400, 400–800, 800–1200, 1200–1600, 1600–2000, 2000–2400, and more than 2400 m. However, the range of Yunnan Province is very large, and most landslides are distributed at a distance greater than 2400 m. Table 11 presents the number and proportion of landslides at each level of distance from roads in Yunnan Province, and Figure 12 shows the distribution of landslides in Yunnan Province based on road distance.

**Table 11.** The number and proportion of landslide distribution at each grade based on road distance.

| Distance from River Classification (m) | Number of Landslide Disasters | Proportion of Total Disasters (%) |
|---|---|---|
| 0–400 | 56 | 4 |
| 400–800 | 41 | 3 |
| 800–1200 | 29 | 2 |
| 1200–1600 | 26 | 3 |
| 1600–2000 | 22 | 2 |
| 2000–2400 | 25 | 2 |
| >2400 | 1185 | 86 |

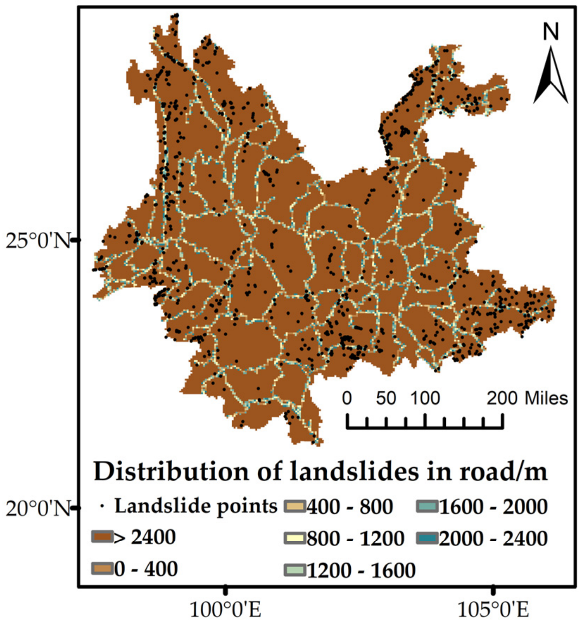


**Figure 12.** Distribution of landslide distances from roads in Yunnan Province.

*4.5. The Degree of Vegetation Cover*

The root system of vegetation is directly related to the ability of rock and soil to retain water. The recharge and transpiration of vegetation groundwater play an important role in affecting the shear resistance of slopes. Therefore, NDVI is regarded as one of the evaluation factors in the process of landslide susceptibility evaluation.

According to the characteristics of vegetation cover, the NDVI in the study area was divided into the following seven categories: −0.06–0.3, 0.3–0.4, 0.4–0.5, 0.5–0.6, 0.6–0.7, 0.7–0.8, and 0.8–0.9. Table 10 presents the number and proportion of landslides at each grade of NDVI in Yunnan Province, and Figure 13 shows the distribution of landslides in Yunnan Province based on the NDVI. As shown in Table 12, the main distribution range of landslides is 0.5–0.8.

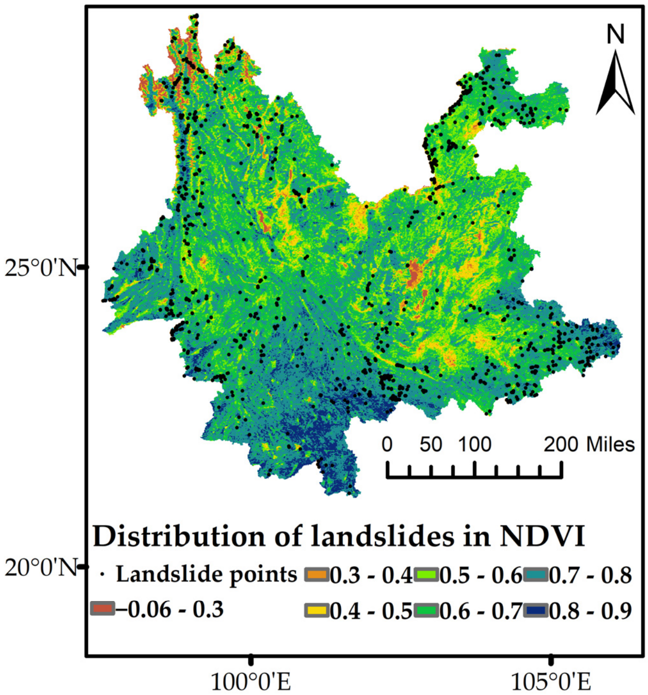


**Figure 13.** Distribution of landslides in Yunnan Province based on NDVI.

**Table 12.** The number and proportion of landslides for each grade of NDVI.

| NDVI Grade | Number of Landslide Disasters | Proportion of Total Disasters (%) |
|---|---|---|
| −0.06–0.3 | 6 | 0 |
| 0.3–0.4 | 35 | 3 |
| 0.4–0.5 | 73 | 5 |
| 0.5–0.6 | 193 | 14 |
| 0.6–0.7 | 489 | 36 |
| 0.7–0.8 | 536 | 39 |
| 0.8–0.9 | 37 | 3 |

## 5. Results and Discussion

*5.1. Model Parameter Settings and Accuracy Verification*

In this paper, 70% of the training samples were used for model establishment and training, and the accuracy in the remaining sample size classification results was verified. Compared to other prediction models, the random forest model has relatively few parameters. In this article, we mainly adjusted the following parameters: n_estimators, max_features, max_depth, min_samples_split, and min_samples_leaf. GridSearchCV was used for searching, and the parameter setting process is shown in Table 13.

**Table 13.** Model parameter adjustment.

| Parameter Settings | Mean Square Error | Accuracy | Area under Curve |
|---|---|---|---|
| Default parameters | 0.23 | 0.77 | 0.85 |
| n_estimators = 86, max_features = 3, max_depth = 10 | 0.20 | 0.80 | 0.87 |
| n_estimators = 86, max_features = 2, max_depth = 10, min_samples_split = 20, min_samples_leaf = 2 | 0.21 | 0.79 | 0.86 |

Table 13 shows that when only the grid search method is used to set the n_estimators, max_features, and max_depth parameters, the accuracy indicators are better, so the following article parameters are used: n_estimators = 86, max_features = 3, and max_depth = 10.

To validate the random forest model, this paper uses the out-of-bag score to verify the model. Because data sampling uses put-back sampling, about 30% of the samples are not selected. We sought to use the data that were not sampled (OOB) as a validation set to verify the correctness of the training set. Here, the oob_score = 0.783, and the acc = 0.90 for the training set, indicating that the random forest model offers good verification accuracy, which is suitable for a vulnerability evaluation of the study area in this paper.

For the performance metrics of the random forest model, we select a confusion matrix to evaluate the model performance and the MSE (mean square error), acc (accuracy), and AUC (area under curve) indicators to determine the evaluation accuracy. The first two parameters are based on statistical indicators used to measure the fit degree of the model and the proportion of samples that are accurately classified, respectively. MSE as a statistical parameter is the mean of the sum of squares for the corresponding point errors from the predicted data and the original data and represents a good or bad degree of data fitting. The value of MSE = 0.2 indicates that the model fits well. Accuracy is the simplest and most intuitive evaluation index for classification problems and can reflect the correct proportion. Here, the acc of the training set and the test set are 0.90 and 0.80, respectively, and accuracy is 0.8 and over, indicating that the model is reasonable and accurate for the evaluation of vulnerability in Yunnan Province. The AUC parameter represents the area under the ROC (receiver operating characteristic) curve, which is calculated using the specificity and sensitivity under a series of different assumptions. The ROC curve, which is the receiver working characteristic curve, is an effective method to evaluate the effect of the binary classification algorithm, reflecting the relationship between the predicted value and the sample value. The AUC value is between 0.5 and 1, where the closer the AUC value is to 1, the higher the prediction accuracy of the model becomes. Here, the value of AUC = 0.87 indicates that the random forest model offers the best prediction accuracy for landslides in Yunnan Province. The above three parameters are commonly used as representative indicators. Figure 14 and Table 14 show the area under the ROC curve and the confusion matrix, respectively. The test results indicate that the random forest model offers high accuracy in the spatial prediction of landslide geological hazard susceptibility in Yunnan Province, good prediction performance, and no overfitting problems.

**Table 14.** Confusion matrix.

| RF Prediction | True Value | | Recall |
|---|---|---|---|
| | Landslide | Non-Landslide | |
| Landslide | 343 | 72 | 0.83 |
| Non-landslide | 95 | 321 | 0.77 |
| Precision | 0.78 | 0.82 | |

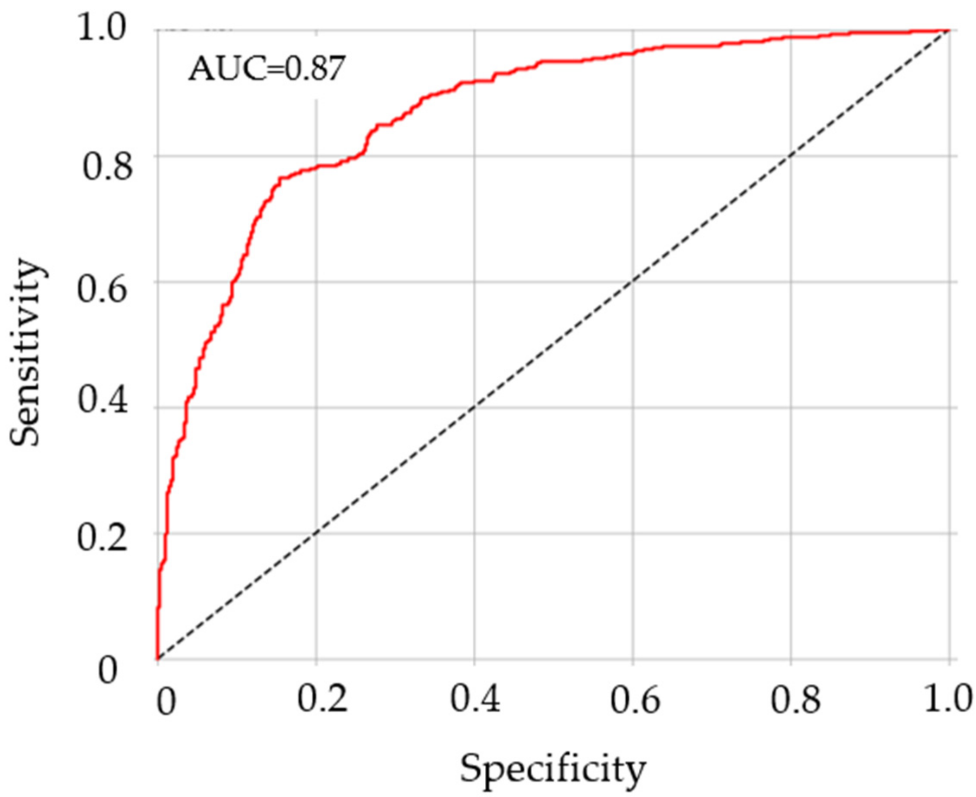


**Figure 14.** Random forest ROC curve.

*5.2. Susceptibility Assessment Results*

In this paper, the random forest model was used to predict the spatial susceptibility of landslide disasters in the study area according to the characteristics of landslide disasters and geological environment data in Yunnan Province. According to previous research experience [32,33], we used the natural discontinuity method to divide landslide susceptibility in Yunnan Province. The natural discontinuity method is a classification and grading method in GIS software based on the principle of numerical statistics, which can maximize the difference between each category. This method is often used to divide susceptibility into the following five levels: low-susceptibility areas (0–0.162), lower-susceptibility areas (0.162–0.32), medium-susceptibility areas (0.32–0.498), higher-susceptibility areas (0.498–0.691), and high-susceptibility areas (0.691–0.984). The landslide susceptibility grade and landslide susceptibility zone map for Yunnan Province are presented, respectively, in Table 15 and Figure 15. The natural discontinuity point classification method can decrease the number of landslides in low- and lower-susceptibility areas and increase the area. Here, most of the landslides fall into higher- and high-susceptibility areas. The density of disaster points corresponds to the ratio of the number of disaster points in each interval to the area of the interval. The density method can intuitively reflect the difference in the number and area of landslides based on the vulnerability level of the model. Based on statistical data, with an increase in susceptibility within the prone zones divided by the model, the density of landslide hazard points in each grade increases, and the density of landslide hazard points in high-susceptibility areas reaches the maximum. The random forest model is densely distributed in areas with high susceptibility. Ultimately, the model results are more consistent with the actual distribution characteristics of disaster points and offer higher accuracy.

**Table 15.** Landslide susceptibility zoning table for Yunnan Province.

| Degree of Susceptibility | Number of Landslides/Pcs | Proportion of Landslides/(%) | Zoning Area/km$^2$ | Proportion of Area/(%) | Hazard Points Density/ Place/10,000 km$^2$ |
|---|---|---|---|---|---|
| Low-susceptibility areas | 13 | 1 | 130,066 | 34 | 1 |
| Lower-susceptibility areas | 36 | 3 | 90,291 | 23 | 4 |
| Medium-susceptibility areas | 131 | 9 | 76,607 | 20 | 16 |
| Higher-susceptibility areas | 264 | 19 | 49,536 | 13 | 53 |
| High-susceptibility areas | 941 | 68 | 37,317 | 10 | 235 |

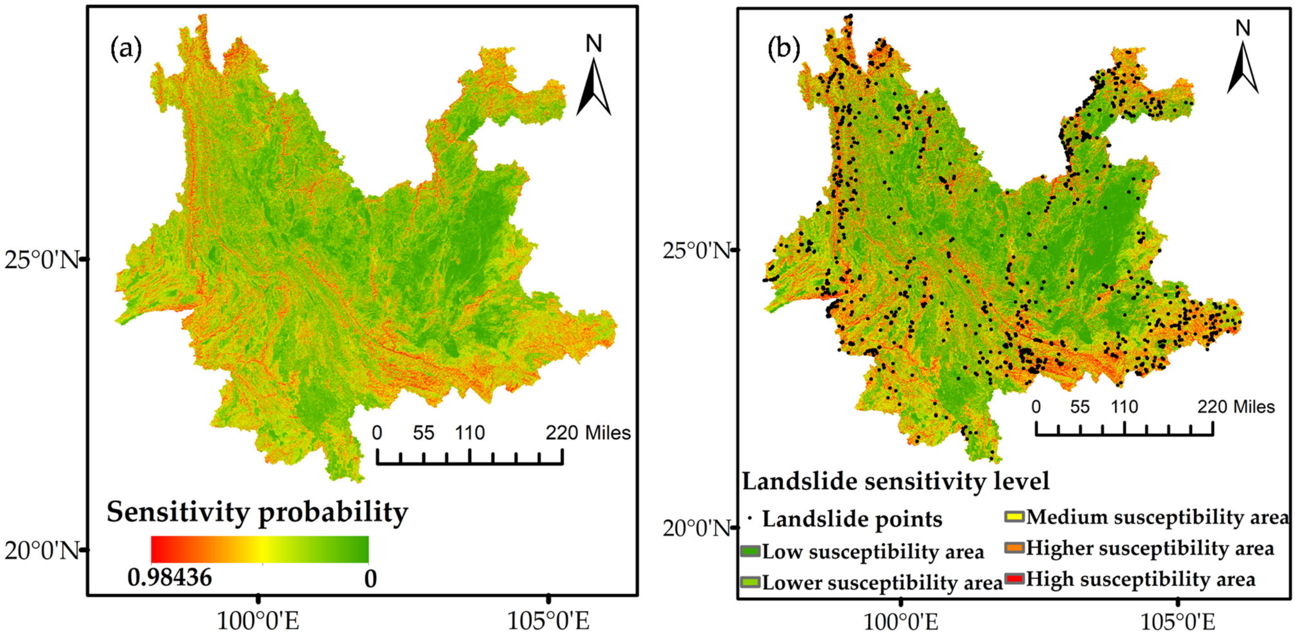


**Figure 15.** Susceptibility map of Yunnan Province: (**a**) sensitivity probability and (**b**) landslide susceptibility level.

It can be seen from the zoning map that the high- and higher-susceptibility areas of landslide disasters are mainly distributed in linear clusters on both sides of the valley and near mountainous areas, on both sides of rivers and both sides of roads, and in other areas with large slope differences. The terrain in the high- and higher-susceptibility areas is complex, with complex geomorphological conditions and longitudinal and horizontal ravines. Additionally, the erosion capacity of the rivers is strong, and the vegetation coverage is low, resulting in serious soil erosion. Steep river valleys also provide an effective surface for landslide development. The moderate- and lower-susceptibility areas of landslides are mainly distributed in areas of thick vegetation cover and valley-to-valley junctions with a gentle regional transition. The low-susceptibility areas for landslides are mainly located in the middle region of Yunnan Province, which belongs to the alpine plain. Here, the slope is small, and the overall distribution is blocks. Additionally, the area is large, which is not suitable for the development and occurrence of landslides and conforms to the characteristics of landslide development.

### *5.3. Impact Factor Importance Analysis*

The importance of landslide susceptibility factors covered in this paper is ranked based on Gini impurity, as shown in Figure 16. It can be seen from Figure 16 that the slope direction, lithology, slope, and annual mean deformation rate of the key variables leading to landslide occurrence are the most important elements affecting the development of landslides. This result also shows that the annual mean deformation rate index factor added in this paper has an important impact on the prediction of landslide susceptibility and that the addition of this index factor can improve the prediction accuracy of landslide susceptibility in Yunnan Province to a certain extent.

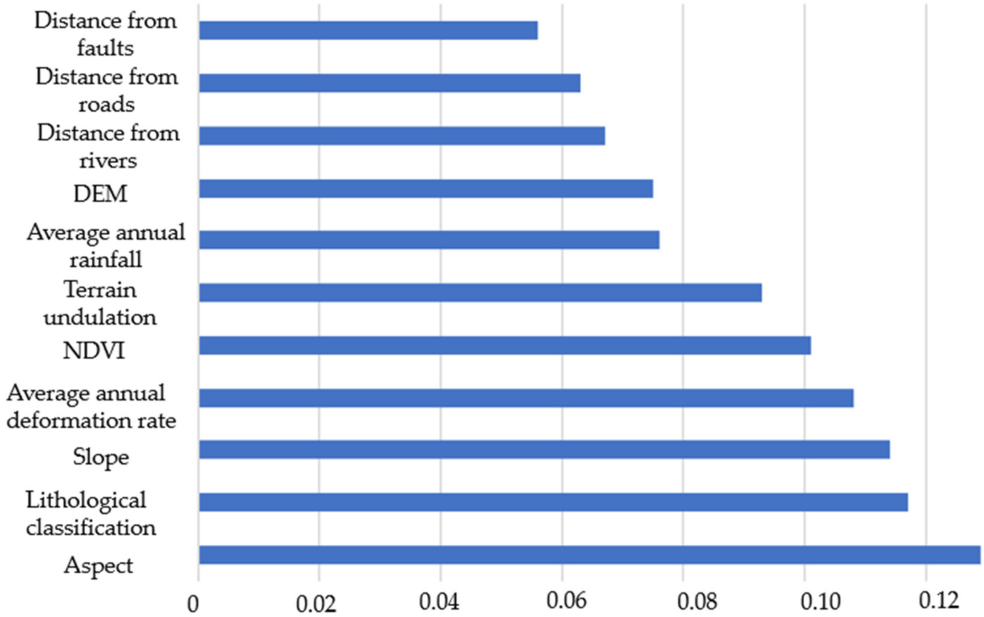


**Figure 16.** Ranking of impact factor importance.

## 6. Conclusions

In this paper, the land surface deformation in the entire Yunnan Province was obtained using SBAS-InSAR technology. The annual mean deformation rate was added as an index factor to the landslide susceptibility evaluation study in Yunnan Province, and its importance as an index factor was explored. We found that the annual mean deformation rate was the most important index factor in addition to aspect, lithology, and slope, indicating that it had an important indicative effect on the evaluation of landslide susceptibility. Thus, the addition of the annual mean deformation rate index factor improved the prediction accuracy of landslide susceptibility in Yunnan Province to a certain extent.

When studying the application of the random forest model to landslide susceptibility evaluation in Yunnan Province, we calculated MSE = 0.20, acc = 0.80, and AUC = 0.87. The obtained landslide susceptibility results also coincided with the actual landslide distribution, which further demonstrates the reliability of the results in this paper. Therefore, the landslide susceptibility evaluation results obtained in this paper can provide a basis for disaster prevention and mitigation in Yunnan Province.

In the study area, the area proportion of the low- and lower-susceptibility areas was 57%, and the landslide proportion was 4%. The area proportion of the higher- and high-susceptibility areas was 23%, and the landslide proportion was 87%. The statistical results showed that most of the landslide hazard points were located in higher- and high-susceptibility areas, which is in line with the characteristics for the concentrated distribution of hidden landslide dangers in high- and higher-susceptibility areas.

**Author Contributions:** Conceptualization, M.L., Z.L. and B.X.; methodology, M.L.; software, M.L.; validation, B.X., Z.L. and W.M.; formal analysis, Y.Z. and J.H.; investigation, B.X.; resources, M.L.; data curation, Z.L., W.M. and Y.Z.; writing—original draft preparation, M.L.; writing—review and editing, M.L., Z.L., W.M. and B.X.; visualization, Z.L., B.X., W.M., Y.Z., W.L. and J.H.; supervision, B.X., W.L., W.M. and J.H.; project administration, B.X., W.L. and Z.L.; funding acquisition, B.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was partly funded by the National Science Fund for Distinguished Young Scholars (No. 41925016) and the National Natural Science Foundation of China (No. 41804008) and Science and Technology Research and Development Program Project of China railway group limited (Major Special Project, No. 2021-Special-08).

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** The authors thank the following institutions for providing materials for this study: European Space Agency (ESA) for Sentinel-2 images; the Geo-monitoring cloud platform; and the Resource Science Data Center, Chinese Academy of Sciences.

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

## References

1. He, F.; Tan, S.; Liu, H. Mechanism of rainfall induced landslides in Yunnan Province using multi-scale spatiotemporal analysis and remote sensing interpretation. *Microprocess. Microsyst.* **2022**, *90*, 104502. [CrossRef]
2. Suvam, H.; Sarkar, S.; Kanungo, D. GIS-based landslide susceptibility zonation mapping using the analytic hierarchy process (AHP) method in parts of Kalimpong Region of Darjeeling Himalaya. *Environ. Monit. Assess.* **2022**, *194*, 234.
3. Panchal, S.; Shrivastava, A.K. Landslide hazard assessment using analytic hierarchy process (AHP): A case study of National Highway 5 in India. *Ain Shams Eng. J.* **2022**, *13*, 101626. [CrossRef]
4. Li, Y.; Deng, X.; Ji, P.; Yang, Y.; Jiang, W.; Zhao, Z. Evaluation of Landslide Susceptibility Based on CF-SVM in Nujiang Prefecture. *Int. J. Environ. Res. Public Health* **2022**, *19*, 14248. [CrossRef] [PubMed]
5. Li, Y.Y.; Mei, H.B.; Ren, X.J.; Hu, X.; Li, M. Geological disaster susceptibility evaluation based on certainty factor and support vector machine. *J. Geo-Inf. Sci.* **2018**, *20*, 1699–1709. [CrossRef]
6. Van Dao, D.; Jaafari, A.; Bayat, M.; Mafi-Gholami, D.; Qi, C.; Moayedi, H.; Van Phong, T.; Ly, H.-B.; Le, T.-T.; Trinh, P.T.; et al. A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. *Catena* **2020**, *188*, 104451. [CrossRef]
7. Wu, X.; Ren, F.; Niu, R. Landslide susceptibility assessment using object mapping units, decision tree, and support vector machine models in the Three Gorges of China. *Environ. Earth Sci.* **2013**, *71*, 4725–4738. [CrossRef]
8. Sun, D.; Wen, H.; Wang, D.; Xu, J. A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. *Geomorphology* **2020**, *362*, 107201. [CrossRef]
9. Arabameri, A.; Saha, S.; Roy, J.; Chen, W.; Blaschke, T.; Bui, D.T. Landslide Susceptibility Evaluation and Management Using Different Machine Learning Methods in The Gallicash River Watershed, Iran. *Remote Sens.* **2020**, *12*, 475. [CrossRef]
10. He, Q.; Wang, M.; Liu, K. Rapidly assessing earthquake-induced landslide susceptibility on a global scale using random forest. *Geomorphology* **2021**, *391*, 107889. [CrossRef]
11. Wang, X.Q.; Ding, M.T.; He, Y.F. Assessment of debris flow vulnerability in Xiangcheng County, Ganzi Tibetan Autonomous Prefecture based on RF model. *J. Catastrophol.* **2023**, *38*, P642.
12. Reichenbach, P.; Rossi, M.; Malamud, B.D.; Mihir, M.; Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth Sci. Rev.* **2018**, *180*, 60–91. [CrossRef]
13. Devara, M.; Tiwari, A.; Dwivedi, R. Landslide susceptibility mapping using MT-InSAR and AHP enabled GIS-based multi-criteria decision analysis. *Geomat. Nat. Hazards Risk* **2021**, *12*, 675–693. [CrossRef]
14. Yang, S.; Li, D.Y.; Yan, L.X.; Huang, Y.; Wang, M.Z. Landslide Susceptibility Assessment in High and Steep Bank Slopes along Wujiang River Based on Random Forest Model. *Saf. Environ. Eng.* **2021**, *28*, 131–138. [CrossRef]
15. Huang, L.X.; Hao, M.J.; Li, W.P.; Zhaoye, Z.; Peiqian, J. Evaluation of Landslide Susceptibility Based on RBF Neural Network-Information Coupling Model: A Case Study of Min County, Gansu Province. *Chin. J. Geol. Hazard Control* **2021**, *32*, 116–126. [CrossRef]
16. Wang, X.D.; Zhang, C.B.; Wang, C. Geological Disaster Susceptibility in Helong City Based on Logistic Regression and Random Forest. *J. Jilin Univ. (Earth Sci. Ed.)* **2022**, *52*, 1957–1970. [CrossRef]
17. Li, M.; Zhang, L.; Yang, M.; Liao, M. Complex surface displacements of the Nanyu landslide in Zhouqu, China revealed by multi-platform InSAR observations. *Eng. Geol. Eng. Geol.* **2023**, *317*, 107069. [CrossRef]
18. Kouhartsiouk, D.; Perdikou, S. The application of DInSAR and Bayesian statistics for the assessment of landslide susceptibility. *Nat. Hazards* **2021**, *105*, 2957–2985. [CrossRef]
19. Ran, P.; Li, S.; Zhuo, G.; Wang, X.; Meng, M.; Liu, L.; Chen, Y.; Huang, H.; Ye, Y.; Lei, X. Early Identification and Influencing Factors Analysis of Active Landslides in Mountainous Areas of Southwest China Using SBAS−InSAR. *Sustainability* **2023**, *15*, 4366. [CrossRef]
20. Wang, Y.; Dong, J.; Zhang, L.; Deng, S.; Zhang, G.; Liao, M.; Gong, J. Automatic detection and update of landslide inventory before and after impoundments at the Lianghekou reservoir using Sentinel-1 InSAR. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *118*, 103224. [CrossRef]

21. Lauknes, T.R.; Zebker, H.A.; Larsen, Y. InSAR Deformation Time Series Using an $L_1$-Norm Small-Baseline Approach. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 536–546. [CrossRef]
22. Schmidt, D.A.; Bürgmann, R. Time-dependent land uplift and subsidence in the Santa Clara valley, California, from a large interferometric synthetic aperture radar data set: Santa Clara Valley INSAR Time Series. *J. Geophys. Res. Solid Earth* **2003**, *108*, B002267. [CrossRef]
23. Lauknes, T.; Shanker, A.P.; Dehls, J.; Zebker, H.; Henderson, I.; Larsen, Y. Detailed rockslide mapping in northern Norway with small baseline and persistent scatterer interferometric SAR time series methods. *Remote Sens. Environ.* **2010**, *114*, 2097–2109. [CrossRef]
24. Berardino, P.; Fornaro, G.; Lanari, R.; Sansosti, E. A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 2375–2383. [CrossRef]
25. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
26. 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]
27. Sangeeta; Singh, S.K. Influence of anthropogenic activities on landslide susceptibility: A case study in Solan district, Himachal Pradesh, India. *J. Mt. Sci.* **2023**, *20*, 429–447. [CrossRef]
28. Addis, A. GIS-Based Landslide Susceptibility Mapping Using Frequency Ratio and Shannon Entropy Models in Dejen District, Northwestern Ethiopia. *J. Eng.* **2023**, *2023*, 1–14. [CrossRef]
29. Mao, Y.P.; Wu, B.Q.; He, W.Y.; Yu, W.X.; Yang, J.W. The neotectonic movement and the chronological feature of active fault in yunnan. *J. Seismol. Res.* **1997**, *20*, 98–106.
30. Huang, H. *Research on Landslide Susceptibility Evaluation Based on Time Series InSAR Technology*; Liaoning University of Science and Technology: Anshan, China, 2021. [CrossRef]
31. Tian, C.Y.; Zhang, W.; Zhang, G. Evaluation of Geological Disaster Vulnerability in Xifeng County Based on Information Content Method. *J. Cap. Norm. Univ. (Nat. Sci. Ed.)* **2020**, *41*, 32–40.
32. Xu, J.H.; Sun, D.L.; Wang, Y.; Wen, H.; Wang, D. Landslide Susceptibility Mapping of Fegjie County Based on GIS and Improved Analytic Hierarchy Process. *J. Chongqing Norm. Univ. (Nat. Sci.)* **2020**, *37*, 36–44. [CrossRef]
33. Chen, X.Y.; Shi, Y.; Wen, Y.X.; Shi, R.Y.; Mi, X.M. Landslide susceptibility evaluation based on certainty factor support vector machines. *Sci. Technol. Eng.* **2023**, *23*, 518–527.