


Technical Note

Landslide Susceptibility Research Combining Qualitative Analysis and Quantitative Evaluation: A Case Study of Yunyang County in Chongqing, China

Wengang Zhang ¹ , Songlin Liu ¹, Luqi Wang ^{1,*}, Pijush Samui ², Marcin Chwała ³ and Yuwei He ¹

¹ School of Civil Engineering, Chongqing University, Chongqing 400045, China; zhangwg@cqu.edu.cn (W.Z.); songlinl@cqu.edu.cn (S.L.); 20164985@cqu.edu.cn (Y.H.)

² Department of Civil Engineering, National Institute of Technology Patna, Patna 800005, Bihar, India; pijush@nitp.ac.in

³ Department of Geotechnics and Hydrotechnics, Faculty of Civil Engineering, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland; marcin.chwala@pwr.edu.pl

* Correspondence: wlq93@cqu.edu.cn

Abstract: Machine learning-based methods are commonly used for landslide susceptibility mapping. Most of the recent publications focused on quantitative analysis, i.e., improving data processing methods, comparing and perfecting the data-driven model itself, but rarely taking the qualitative aspects of the local landslide occurrences into consideration and the further analysis of the key features was always lacking. This study aims to combine qualitative and quantitative analysis and examine its effect on mapping accuracy; based on the feature importance ranks and the related literature, the key features for identifying landslide/non-landslide points of different sub-zones were further analyzed. Before modeling, the study area Yunyang County, Chongqing City, China, was manually divided into four sub-zones based on the information from geological hazards exploration in Chongqing, including the mechanism of landslide formation and sliding failure and geomorphic unit characteristics. Upon the qualitative analysis basis, five grid searches tuned random forest models (one for the whole region and four for the sub-zones independently) were established by 1654 data points and 20 conditioning features. Compared with the conventional data-driven method, the integrated quantitative evaluation based on the qualitative analysis results showed higher reliability, which not only improved the mapping accuracy but also increased the AUC values of all four sub-models, which were 8.8%, 2.3%, 1.9% and 9.1% higher than that of the parent model. Moreover, the quantitative evaluation based on the qualitative analysis revealed the key factors affecting local landslide formation. Therefore, qualitative analysis is recommended in future landslide susceptibility modeling with the additional combination of data-driven methods.

Keywords: landslide susceptibility mapping; random forest model; qualitative analysis; quantitative evaluation; Yunyang County



Citation: Zhang, W.; Liu, S.; Wang, L.; Samui, P.; Chwała, M.; He, Y. Landslide Susceptibility Research Combining Qualitative Analysis and Quantitative Evaluation: A Case Study of Yunyang County in Chongqing, China. *Forests* **2022**, *13*, 1055. <https://doi.org/10.3390/f13071055>

Academic Editors: Chong Xu, Haijia Wen, Weile Li and Hiromu Daimaru

Received: 11 May 2022

Accepted: 28 June 2022

Published: 4 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

Landslides are one of the most destructive geological hazards, which not only cause enormous damage to houses and infrastructure, such as bridges and roads, but also lead to loss of life [1]. According to the World Health Organization, approximately 4.8 million people were affected, and more than 18,000 deaths were caused by landslides between 1998 and 2017. Specifically, as one of the countries with a high incidence of landslides, China suffered severe loss of life [2,3]. The China Statistical Yearbook indicates that during 2000 to 2015, 373,630 landslides occurred in this country, killing 10,996 people, which is approximately 690 landslide-related deaths per year [4]. To mitigate the serious social impact caused by landslides, constructive and productive activities should be avoided in areas with high susceptibility to landslides. Therefore, developing an efficient method

to distinguish landslide-prone zones is an essential need for both local governments and research institutes [5]. Landslide susceptibility describes the likelihood that a landslide will occur in a certain area based on local terrain conditions [6]. Landslide susceptibility mapping (LSM) is one of the most widely used assessment methods, it visualizes the spatial distribution of zones with different probabilities of occurrence of landslides in a certain area.

Various methods such as probabilistic analysis, statistical analysis, analytic process, and weighted overlay were widely applied to LSM by researchers in the early stages. With the development of Artificial Intelligence (AI) and Geographic Information System (GIS), machine learning-based methods, with the capability of solving complex nonlinear problems, are becoming increasingly popular compared to opinion-driven models and statistical learning, making the accuracy and precision of susceptibility models evolve rapidly [7,8]. Huang et al. [9] adopted logic regression (LR), support vector machine (SVM), and random forest (RF) on LSM for model comparison. He et al. [10] used RF in the global assessment of earthquake-induced landslide susceptibility. Sun et al. [11] applied the Bayes algorithm to optimize the hyper-parameters of the RF model for LSM. Smith et al. [12] compared the effect of landslide inventories assembled by different methods on the performance of RF and LF for LSM. Lim et al. [13] applied the RF model to estimate the probability of a landslide. Nhu et al. [14] investigated and compared the Logistic Model Tree, LR, NBTree, Artificial Neural Network (ANN), and SVM in the shallow landslide susceptibility mapping for Bijar City in Kurdistan City. Zhang et al. [15] used the predictive performance of RF, XGBoost, SVM, and LR on landslide susceptibility mapping in Yunyang County. Hu et al. [16] compared the effect of different non-landslide sampling methods on the performance of SVM and NB for LSM. Zhou et al. [17] applied GeoDetector and RFE for factor optimization and then used the selected factors as inputs to train an RF model to obtain the LSM of Wuxi County. Sun et al. [18] proposed a hybrid landslide warning model based on RF susceptibility zoning and precipitation. Zhou et al. [19] constructed an interpretable model for the susceptibility to rainfall-induced shallow landslides based on SHAP and XGBoost.

Among those methods, RF is the most commonly used method in large-scale mapping and classification [20–22] due to its characteristics of low computational cost, low data requirement, convenience of hyper-parameters tuning, and robustness in solving complex nonlinear problems [23]. Previous work usually focused on quantitative analysis, such as the selection and improvement of models and input features, but rarely took into account the qualitative analysis of landslide areas. Actually, as one of the major geological hazards, landslides are highly area dependent, the mechanism of landslide formation and its corresponding triggering factors are undoubtedly different in distinct areas. The frequent fluctuation of reservoir water seriously reduces the stability of the slopes in the reservoir area, making them prone to landslides [24]. For mountainous areas, however, rainfall is the major triggering factor for the occurrence of landslides [25]. With increased population, human activities have become the major issue that accelerates landslide formation in areas with high population density. Therefore, manually dividing a relatively large region into different sub-zones according to the qualitative analysis of the landslide formation and geomorphic unit characteristics will theoretically improve mapping accuracy. This paper aims to use the 827 historical landslide data points in Yunyang County and the 20 conditioning factors to build 5 RF models, including an RF model (referred to as the parent model below) for the whole region and four RF models for the divided four sub-zones (referred to as sub-model one to sub-model four below). Then, the feature importance and the performances of the parent model and the four sub-models are analyzed and compared to verify the effectiveness of applying experience-based zonation before modeling.

2. Study Area

Chongqing City is located in the mountainous area around the eastern Sichuan basin and the slope area of the basin margin. It spans two tectonic units, namely the Yangtze

quasi-platform and the Qinling fold system. The landscape of Chongqing City is mainly mountains and hills, which make up 92% of its total area. There are many adverse geological conditions accelerating the formation of landslides, dangerous rock collapse, ground collapse, debris flow, and other geological disasters, including developed surface water networks, strongly cut terrain, complex rock and soil structure, and geological structure, making it one of the cities with the highest geological disaster frequency in the country.

The spatial distribution of geological hazards in Chongqing City shows a certain degree of concentration and can be concluded as a striped distribution and vertical zonal distribution; moreover, its temporal distribution presents a seasonal cluster pattern. According to the statistics, there are currently 14,926 geological hazard-prone points in Chongqing City, of which 5776 (38.7%) are located in the 7 districts and counties of northeast Chongqing City (Wanzhou, Kaizhou, Chengkou, Wuxi, Wushan, Fengjie, Yunyang), 1864 (12.49%) are in the 5 districts and counties of southeastern Chongqing City (Wulong, Youyang, Qianjiang, Pengshui, Xiushan), and 1320 (8.84%) are in the 11 districts of the main city. Therefore, northeast Chongqing City is the key area with a high probability of potential geological disasters.

As one of the seven districts and counties in the northeast of Chongqing City (Figure 1), Yunyang County (spans $108^{\circ}24'37''$ – $109^{\circ}14'47''$ E and $30^{\circ}34'59''$ – $31^{\circ}26'28''$ N) is located in the middle of the Three Gorges Reservoir Project area, being the important hub of the ecological and economic zone along the Yangtze River. According to the announcement of the Chongqing Forest Bureau, while the forest area of Chongqing city reaches 54.5%, that of Yunyang County exceeds 58.5%, making it one of the greenest counties in China. Based on the Seventh National Census of China, there were 929,034 long-term residents (48% of them are urban residents) in this area in the year 2020. Yunyang County is crossed by twelve major folds, namely Changdianfang Syncline (1), Macaoba Anticline (2), Qvmahe Syncline (3), Tiefengshan Anticline (4), Yangliuwan Syncline (5), Dongcun Anticline (6), Xinchang Anticline (7), Huangpoxi Syncline (8), Guling Syncline (9), Fangdoushan Anticline (10), Ganchang Syncline (11), and Longjukan Syncline (12). Under the subtropical monsoon climate, Yunyang County has an average annual rainfall of 1123.7 to 1264.8 mm and an average annual temperature from 10.2 to 18.5 °C.

Mountainous areas are generally susceptible to mass movements due to preparatory and triggering causal factors [26]; not only the weathering effects but anthropogenic activities in the region also commonly accelerate the formation of unstable areas on both the earth material and on hill slopes [27]. As a part of Chongqing City, Yunyang County has always been a significant hotspot for landslide occurrences. There are a total of 836 historical landslides recorded in the dataset; 827 data points are left after data cleaning. A total of 28.2% of them are small landslides, 51.8% are medium landslides, and 20% are large landslides. Among them, trust-load-caused landslides accounted for 53.7%, and loosen-caused landslides and multi-caused landslides accounted for 14.5% and 31.8%, respectively. To build sub-models, we manually divided the study area into four different sub-zones (Figure 2) based on the information from the exploration of geological hazards in Chongqing City, such as the mechanism of landslide formation and sliding failure and the geomorphic unit characteristics. Among the four sub-zones, sub-zone II contains all the strip-distributing landslides along the mainstream of the Yangtze River, so it can also be called the Yangtze River mainstream zone. From a larger scope, a part of Yunyang County belongs to the low-hills section that crosses Yunyang, Fengjie, and Kaizhou; this area is classified as sub-zone IV. Sub-zone I (south of sub-zone II) is crossed by the main highway called S305, and the main area of sub-zone III (between sub-zone II and IV) is crossed by the S103 and S305. Similarly, the density of the road network is also at a high stage in the other two parts of sub-zone III. The landslides that occurred in these two sub-zones are found to be mainly along the roads (Figure 3). After zonation, 89 landslides are located in sub-zone I, 285 of them are in sub-zone II, sub-zone III contains 44 landslides, and with the largest area, 408 of the historical landslides occurred in sub-zone IV.

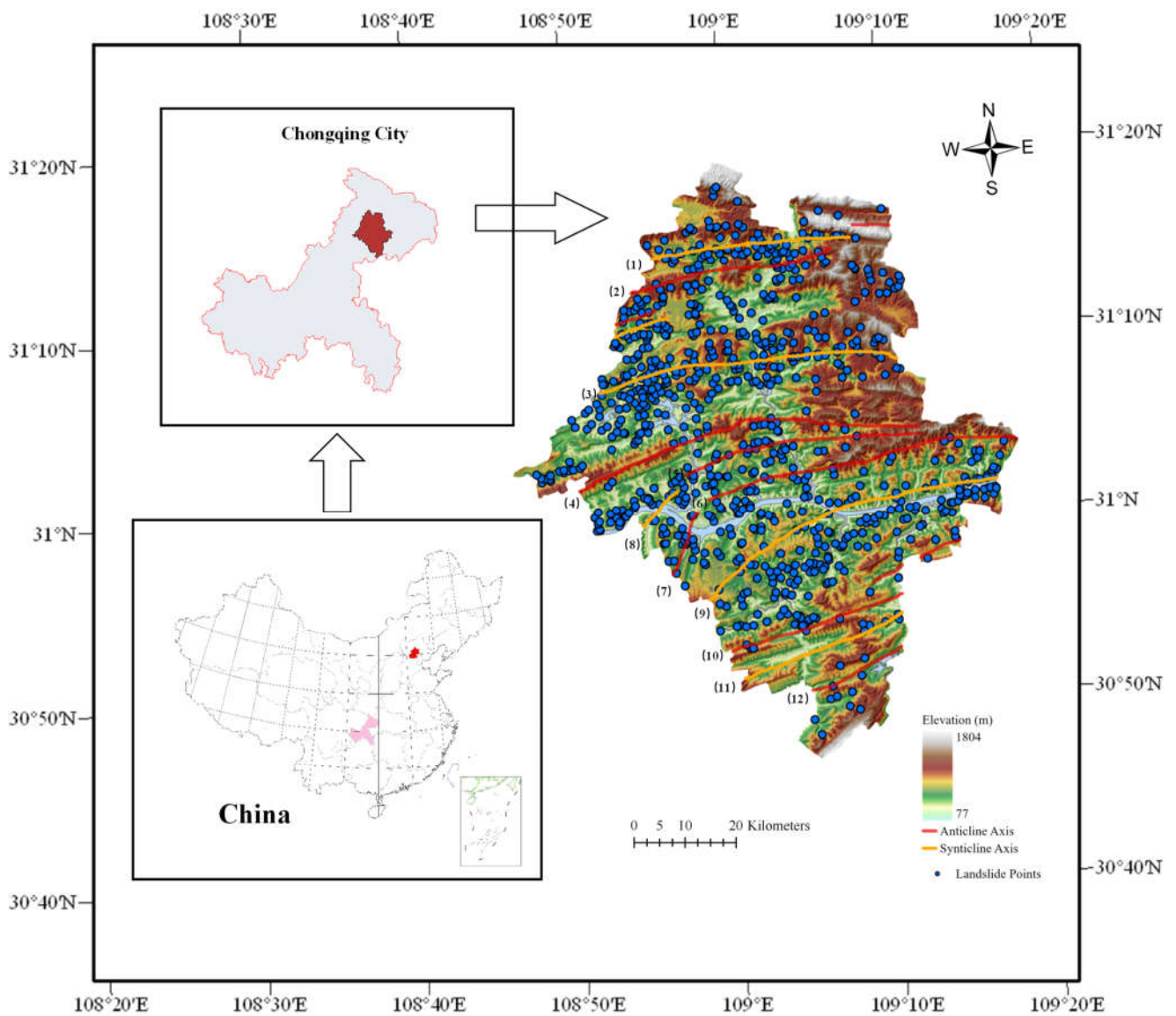


Figure 1. Location, landslide distribution and tectonic map of the study area. (1) Changdianfang Syncline, (2) Macaoba Anticline, (3) Qymahe Syncline, (4) Tiefengshan Anticline, (5) Yangliuwan Syncline, (6) Dongcun Anticline, (7) Xinchang Anticline, (8) Huangpoxi Syncline, (9) Guling Syncline, (10) Fangdoushan Anticline, (11) Ganchang Syncline, and (12) Longjukan Syncline.

As one of the typical landslides in the Three Gorges Reservoir area, the Jiuxianping landslide (in sub-zone II) is located on the left bank of the Yangtze River (Figure 2b). After the Three Gorges Reservoir project, the fluctuation of the Three Gorges Reservoir water level restarted the displacement and deformation of the ancient landslide, making this area more prone to geological hazards. A subsidence of about one meter occurred on a roadway in the middle of the landslide body after heavy rain in 2003 and 2004, causing the roadway to be abandoned. With the impact of continuous heavy rain, landslides occurred in the back accumulation of Jiuxianping on 19 and 22 June 2007, causing the houses of the villagers to collapse, and the mountain body cracked. On 9 June 2009, the back-accumulation of Jiuxianping deformed again under the impact of heavy rain, causing cracks on both the accumulation body and the houses of the villagers. Recently, under the continuous effect of the Three Gorges Reservoir, this area has been in the overall creep deformation stage for years, especially the cliffs near the river, which often suffer from local collapse and damage.

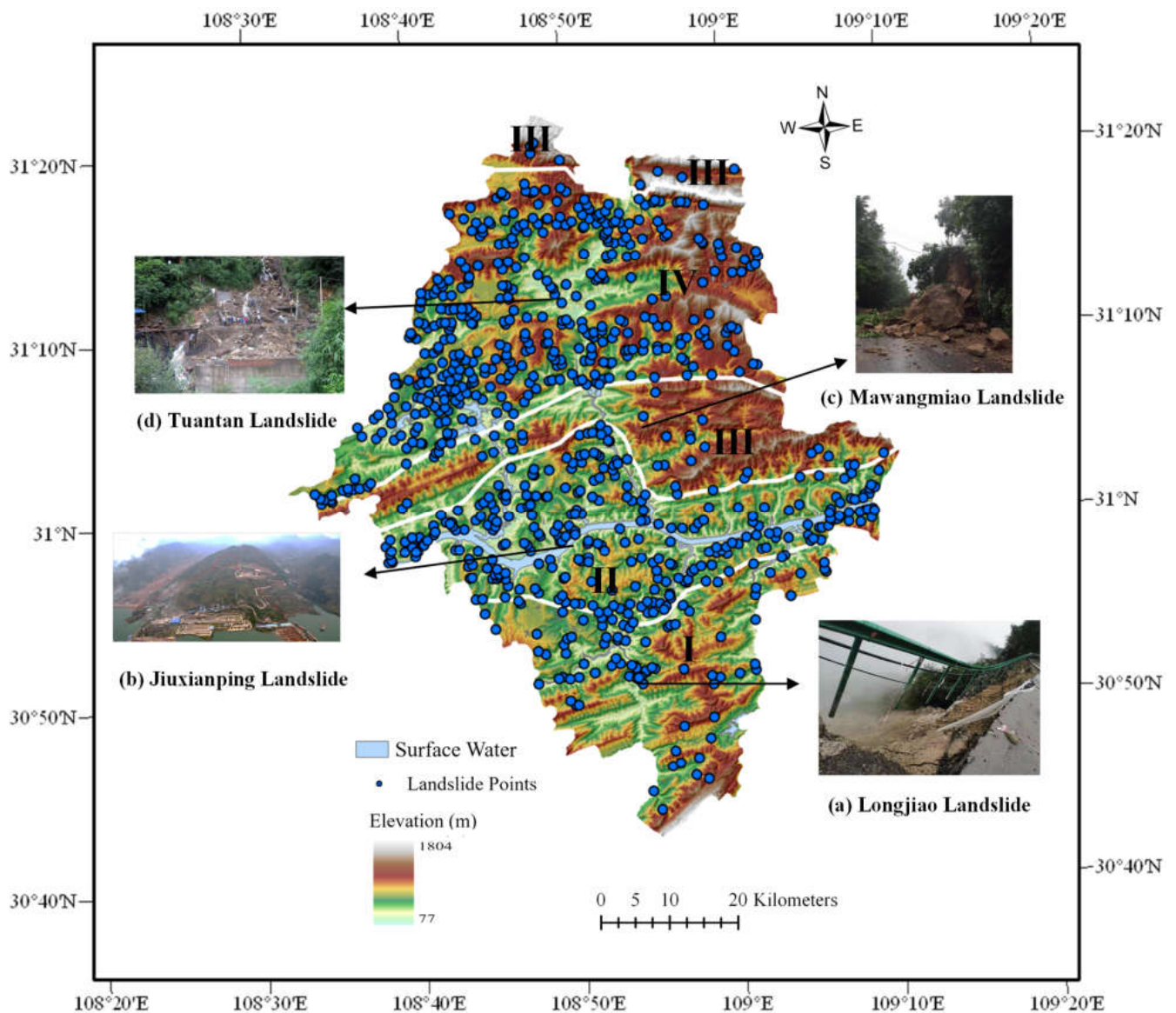


Figure 2. Zonation and typical landslides: (a) Longjiao Landslide, (b) Jiuxianping Landslide, (c) Mawangmiao Landslide, (d) Tuantan Landslide.

The continuous heavy rains from 30 August to 1 September 2014 made the accumulated rainfall in Jiangkou Town more than 300 mm. The day after that, the Tuantan landslide (Figure 2d) occurred on the back mountain and on the left side of the Yongfa Coal Mine staff dormitory in Tuantan village, Jiangkou Town, Yunyang County (in sub-zone IV). Although the employees were notified to evacuate from the area subjected to the massive landslide, twelve of them were buried on the spot. Unfortunately, only one of the twelve was saved.

Typically, in sub-zone I, under the impact of rainfall, a landslide occurred on the S202 Highway in the direction from Longjiao to Rucao (Figure 2a) on 13 July 2021. Similarly, there was a 10,000 cubic-meter landslide triggered by heavy rainfall in the area of Mawang Temple (Figure 2c), which trapped two four-wheel cars and a motorcycle, and blocked the highway section for five days.

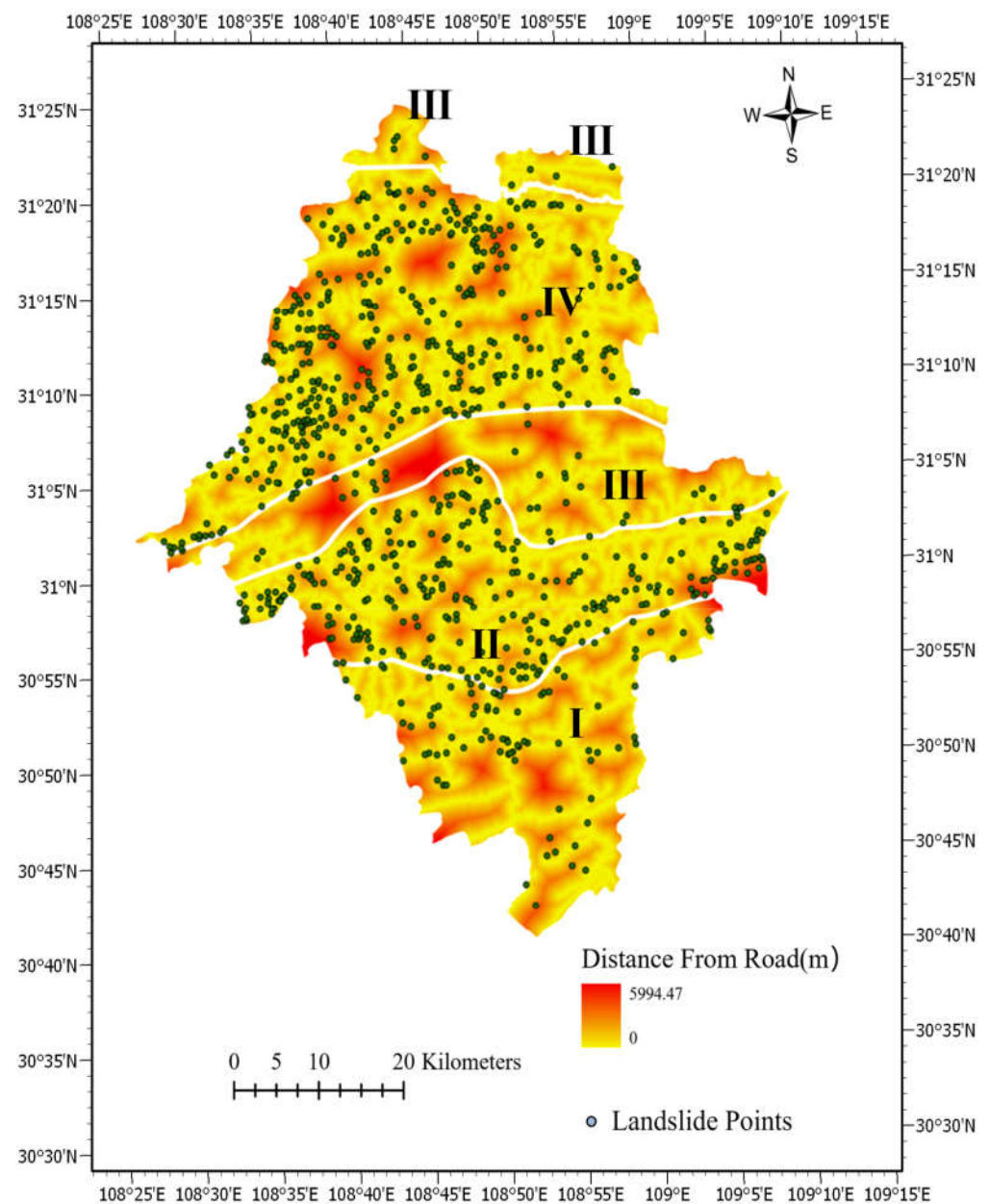


Figure 3. Landslides distribution versus distance from road in different sub-zones.

3. Method Explanation

3.1. Random Forest

As one of the most popular classification methods, RF was first proposed by Breiman [28] and Cutler [29]. During its training process, different data subsets obtained by random sampling are used to train multiple decision trees as independent estimators; each of them is only allowed to fit the data based on the part of the input features, and the final output of the RF model is based on the voting results of the constructed estimators. With the double randomness, those estimators will be trained as distinct ones, and such a special structure makes RF less sensitive to noise and outliers, less likely to overfit and has a lower dependency on feature selection, but more robust and can provide more accurate predictions compared with the other machine learning models.

The RF training process can be concluded as the flowing three simple steps (Figure 4):

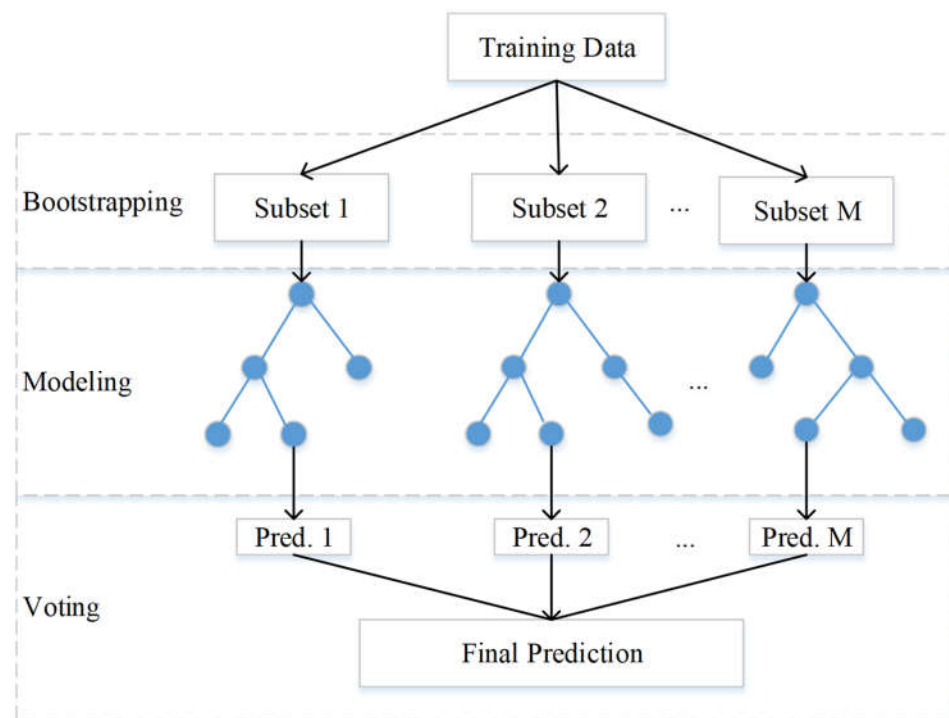


Figure 4. Random forest flowchart.

Bootstrapping: To build M decision trees (estimators), M subsets will be generated as training sets from the original dataset by sampling with replacement.

Modeling: Each subset obtained by bagging is used to train the corresponding estimator. Similarly, for the purpose of getting estimators as distinct as possible, the available features for each estimator are also randomly chosen from the entire feature list.

Voting: Each estimator will output a result independently, then the final result of the model will be generated based on the voting results (Equation (1)).

$$f(x) = \frac{1}{M} \sum_{m=1}^M f_m(x) \quad (1)$$

where $f(x)$ represents the final output, and $f_m(x)$ means the output of the m th tree.

3.2. Grid Search

As pre-set parameters, hyper-parameters play a crucial role in model performance, and there are rare algorithms that are hyper-parameter-free [30]. Therefore, proper hyper-parameter tuning is essential for improving model performance. Grid search, as one of the conventional automatic hyper-parameter tuning methods, is widely used because of its simple operation. Its basic idea is to choose the best hyper-parameter combination by enumerating and iterating over all possible combinations. Although it is computationally expensive because of the exhaustive search process, grid search suits random forest very well, as random forest is a hyper-parameter tuning friendly model; only two hyper-parameters are to be tuned in our case.

3.3. Performance Measure

For classification problems, accurate-oriented modeling sometimes ends up with “rabid” models that tend to classify all samples as a certain type, especially for unbalanced datasets. Therefore, the confusion matrix is introduced in this case (Table 1).

Table 1. Confusion matrix.

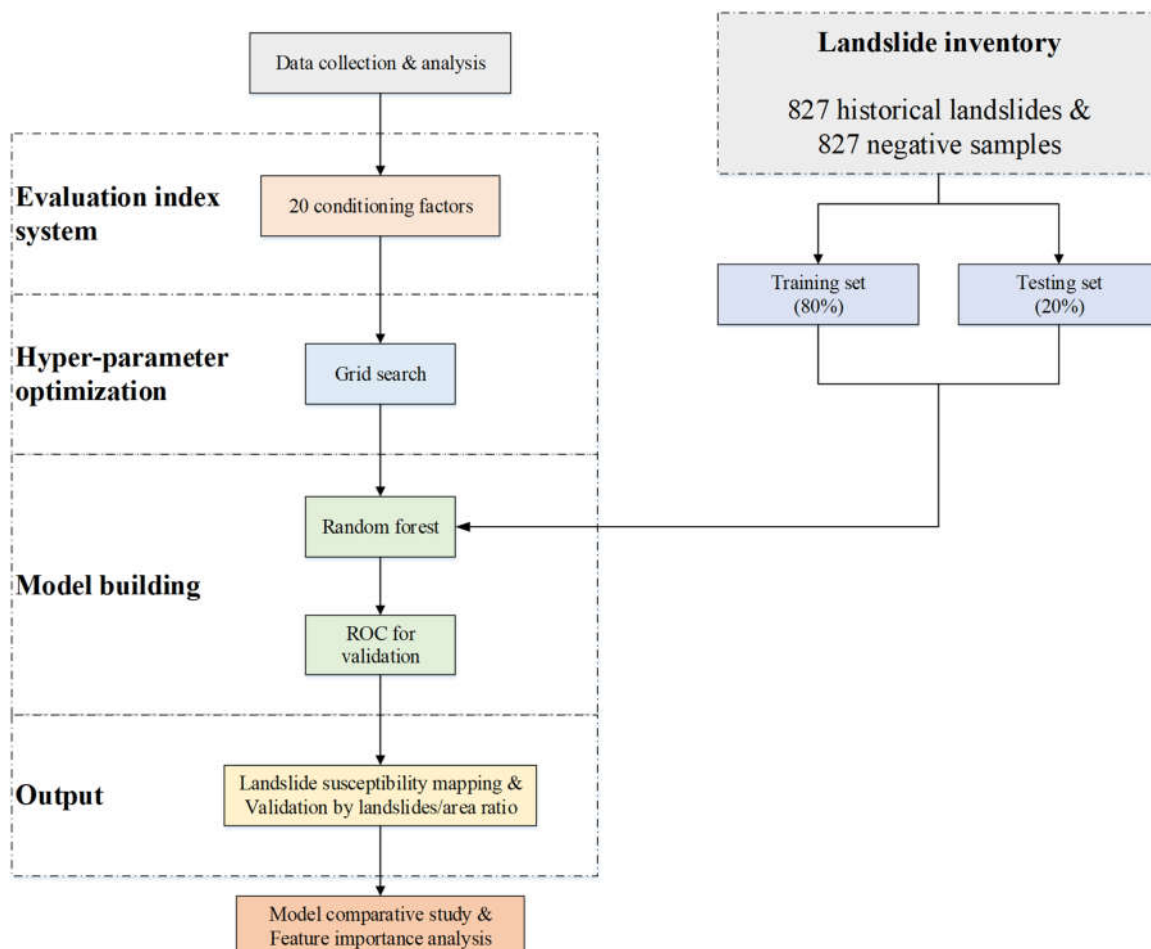
Predicted Values	Actual Values	
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

According to Table 1, the true positive rate, also called the recall, is defined as $TPR = TP / (TP + FN)$, and the false positive rate is defined as $FPR = FP / (FP + TN)$. After model establishment, its corresponding TPR versus FPR at different cutoff values can be plotted, which is called as receiver operation characteristic (ROC) curve and can be used to represent the predictive ability of the model. The area under the ROC curve (AUC) is commonly used as an index that represents the true classification accuracy of the model.

4. Methodology

4.1. Data Collection and Preparation

Data collection and data cleaning are of the top importance in machine learning applications; high-quality data is the foundation of accurate prediction models. Model perfection also plays a vital role; choosing appropriate hyper-parameters will help models to extract useful information from input data more effectively and precisely. In this section, we will introduce the procedure for data preparation and model establishment. The analysis of the performance and the factor importance of different models will be presented in the following sections. The flowchart for this study is shown in Figure 5.

**Figure 5.** Flowchart of this study.

Triggered by the joint effects of natural factors and human activities, the mechanism of the occurrence of landslides is very complex [5]. Based on the assumption that future landslides will occur under the same conditions as past landslides [31], evaluating these factors and analyzing their relationships with historically recorded landslides can contribute to forming the basis for the prediction of future landslides in an area [32,33]. However, after reviewing the studies related to the evaluation of landslide sensitivity from 1983 to 2016, Reichenbach et al. [34] found that there was a total of 596 different factors that affected landslide formation, considering all of them will be laborious and time-consuming. According to previous articles, topography, hydrology, geology, land cover, and natural and human-related factors are generally used for landslide susceptibility analysis [14,17]. Therefore, in this study, topographical factors (aspect, elevation, plane curvature, profile curvature, relief amplitude, slope), hydrological factors (aridity, distance from rivers, index of moisture (IM), and topographic wetness index (TWI)), geological factors (lithology, distance from anticline axis, distance from syncline axis), factors related to land cover (namely land use and normalized difference vegetation index (NDVI)), human factors (such as human activity intensity of land surface (HAILS), distance from road, and population), and natural factors (average annual temperature, average annual rainfall) are all considered (Figure 6).

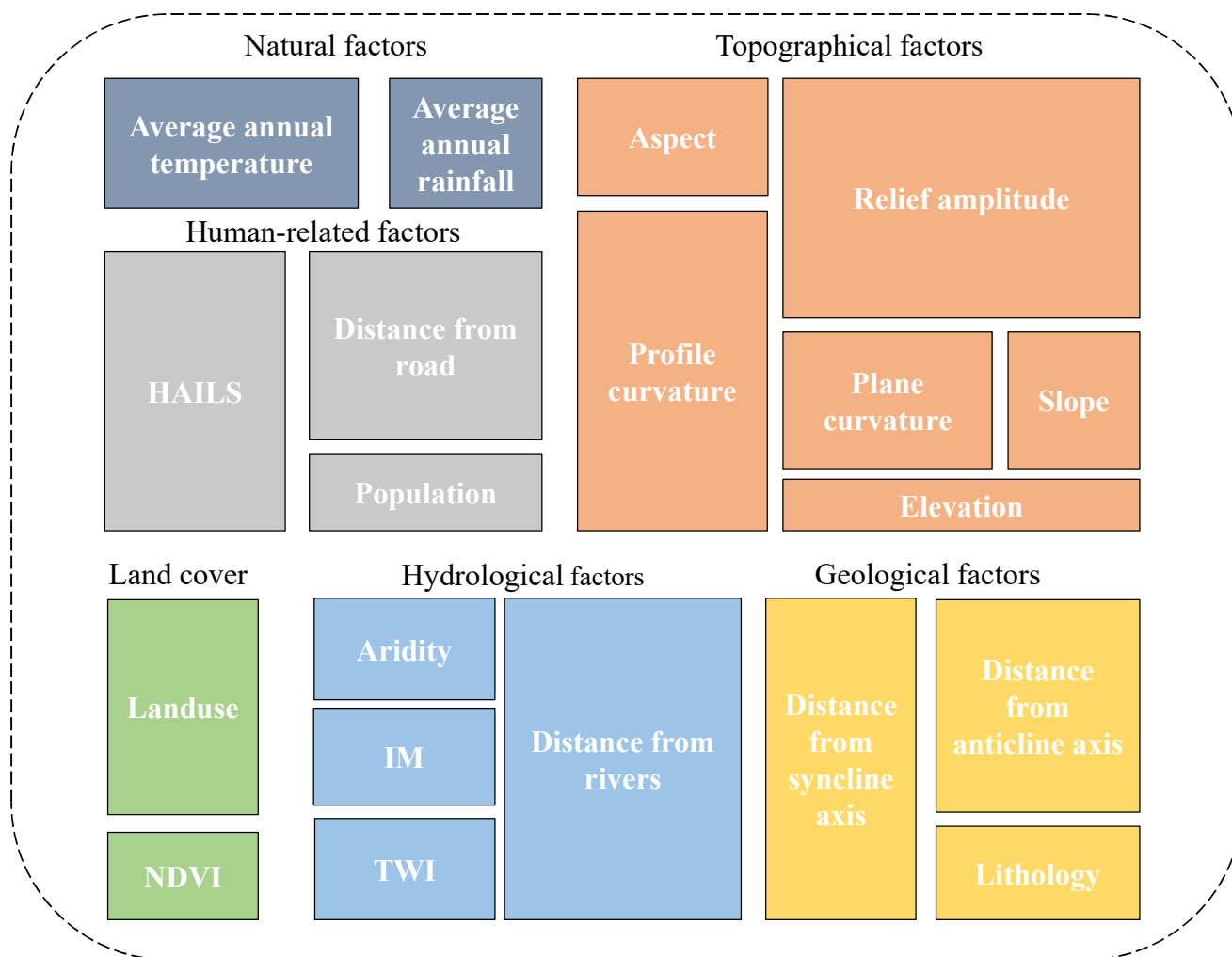


Figure 6. Conditioning factor types.

As the grid data layers cannot be directly used for model training, the Yunyang County fishnet with a cell size of 25×25 m was created for the purpose of data extraction for model training and measuring the distances from specific structures/natural sources, the

total number of cells is 5,831,382. The secondary data includes aspect, plane curvature, profile curvature, relief amplitude, slope, distance from rivers, TWI, distance from anticline axis, distance from syncline axis, and distance from road. Among them, TWI, aspect, plane curvature, profile curvature, relief amplitude, and slope were obtained by ArcGIS processing of DEM. The characteristics associated with distances (i.e., distance from rivers, distance from anticline axis, and distance from syncline axis) were obtained by applying the near function of ArcGIS to the corresponding primary data (i.e., the river network of Chongqing City, the road network of Chongqing City, and the geological structure of Yunyang County) and the created fishnet of Yunyang County. The ranges and categories of the 20 characteristics are summarized in Table 2.

Table 2. (a) Categories of conditioning factors. (b) Ranges of conditioning factors.

Feature ID	Feature Name	# of Classes	Class Name
(a)			
1	lithology	37	1. T ₁ d ₃ ; 2. T ₁ d ₄ ; 3. J ₂ s ₂ ; 4. Ss; 5. T ₁ J ₄ ; 6. T ₁ j ₁ ; 7. T ₁ j ₂ ; 8. J ₂ s ₁ ; 9. J ₂ x; 10. J ₁ z ₁ da; 11. J ₁ z ₁ m; 12. J ₁ z ₁ d; 13. J ₁ z; 14. T ₁ j ₃ ; 15. T ₂ b ₂ ; 16. T ₂ b ₃ ; 17. T ₂ b ₂ ; 18. T ₂ b ₁ ; 19. T ₃ xj; 20. J ₁ zlm; 21. J ₁ zld; 22. J ₃ sn; 23. J ₃ p ₂ ; 24. J ₃ p; 25. T ₃ xj ₃ ; 26. T ₃ xj ₂ ; 27. T ₃ xj ₁ ; 28. J ₁ zlda; 29. J ₁ z ₁ ; 30. J ₂ sn; 31. T ₁ d ₂ ; 32. T ₁ d; 33. P ₃ w ₂ ; 34. P ₃ d; 35. P ₂ m – g; 36. P ₃ wl; 37. P ₂ q
2	landuse	15	1. Paddy filed; 2. Dry land; 3. Forestland; 4. Shrub land; 5. Sparse woods; 6. Other woods; 7. Grassland with high coverage; 8. Grassland with medium coverage; 9. Grassland with low coverage; 10. Cannel; 11. Reservoir; 12. Urban land use; 13. Rural settlements; 14. Other types of building; 15. Marshland
(b)			
Feature ID	Feature Name	Value Range	
3	Slope (°)	0–67.0356	
4	Average Annual Temperature (°C)	10.2–18.5	
5	Average Annual Rainfall (mm)	1123.7–1264.8	
6	Aridity	0.698–0.952	
7	IM	14.5–53.47	
8	Population	88.8744–695.201	
9	HAILS	0–78.2	
10	NDVI	–0.1372–0.8837	
11	Distance from Syncline Axis (m)	0–19,026.1	
12	Distance from Anticline Axis (m)	0–16,958.8	
13	Distance from Road (m)	0–5994.47	
14	Distance from Rivers (m)	0–17,236.8	
15	Elevation (m)	76.0774–1794.29	
16	Profile Curvature	0–44.2457	
17	Aspect (°)	0–360	
18	Plane Curvature	0–360	
19	TWI	–0.82–30.1	
20	Relief Amplitude	0–148.865	

All 20 factors were visualized by ArcGIS with a resolution of 25 m (Figure 7). Then the corresponding classes and rating values of the 20 factors were assigned to each cell of the prepared fishnet.

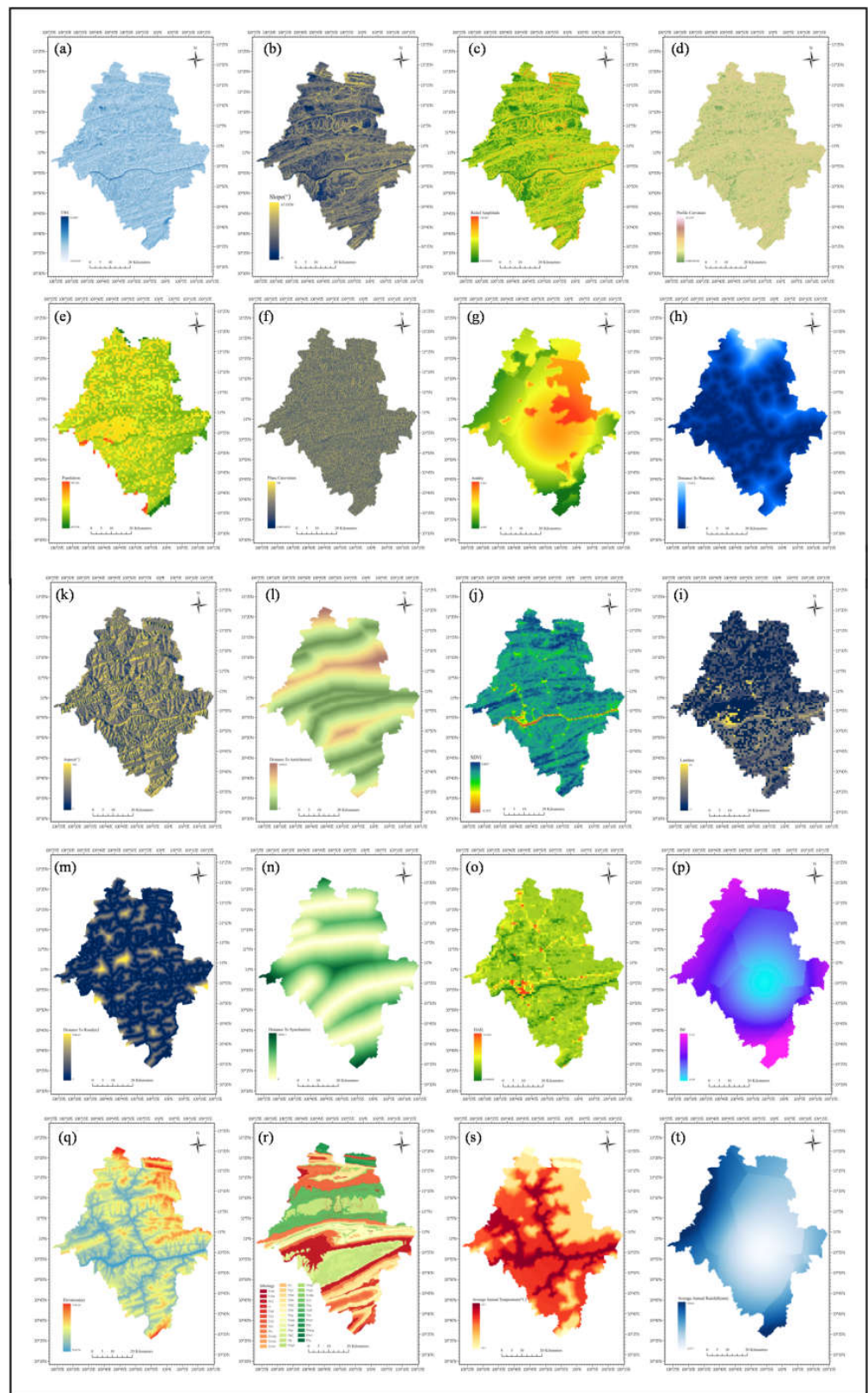


Figure 7. Conditioning factors of landslides susceptibility: (a) TWI; (b) Slope; (c) Relief amplitude; (d) Profile curvature; (e) Population; (f) Plane curvature; (g) Aridity; (h) Distance from rivers; (i) Land use; (j) NDVI; (k) Aspect; (l) Distance from anticline axis; (m) Distance from road; (n) Distance from syncline axis; (o) HAILS; (p) IM; (q) Elevation; (r) Lithology; (s) Average annual temperature; (t) Average annual rainfall.

After obtaining the 827 historical landslide cells in the study area as positive samples (1), the same number of non-landslide cells were randomly extracted from the landslide-free areas as negative samples (0) [35,36]. To make sure that the sub-models can be constructed with the same data as the parent model, the number of negative samples selected from each sub-zone was determined by the number of the positive samples in the sub-zone (i.e., 89 for sub-zone I, 286 for sub-zone II, 44 for sub-zone III, and 408 for sub-zone IV). Thus the five sample datasets were formed (Table 3). Based on previous research [14,37], 80% of each dataset was randomly selected to train the corresponding statistical model, and the remaining 20% was used for validation purposes.

Table 3. The number of data points in different zonation and datasets.

Sub-Zone	# of Positive Samples	# of Negative Samples	Total	Percentage of Training Set	# of Training Samples
I	89	89	178	80%	142
II	286	286	572	80%	457
III	44	44	88	80%	70
IV	408	408	816	80%	652
Whole Region	827	827	1654	80%	1323

4.2. Model Development and Application

Five independent random forest-based models were established and validated based on the sample datasets constructed before. The hyper-parameters of each model were tuned by the grid search method. Being considered as the most straightforward optimization method [38], the grid search method needs to iterate over the entire interval of each hyper-parameter. Therefore, the number and the iterating interval of the tuned hyper-parameters have an obvious impact on its search efficiency. The number of estimators determines the stability of a random forest model, and adding more estimators will lower its mean squared prediction delta (MSPD) and hence improve model stability [39], but will increase computational cost [40]. The maximum depth of a tree controls the stability of a random forest model in a different way. The stability will decrease as depth increases since increasing the depth will make the model tend to just memorize the training data, but if the forest is too shallow, the model will underfit, resulting in low AUC [39]. In this case, the number of estimators and the maximum depth of a tree are the two hyper-parameters to be tuned; the results are shown in Table 4.

Table 4. Hyper-parameters of models.

Model	Number of Estimators	Maximum Depth of a Tree
Parent Model	175	21
Sub-Model One	10	6
Sub-Model Two	95	12
Sub-Model Three	10	4
Sub-Model Four	100	12

After obtaining well-trained models, we used all 5,831,382 cells prepared in Section 4.1 as input to generate the landslide susceptibility maps for the study area.

5. Results

The landslide susceptibility maps generated by the parent model and sub-models are displayed in Figure 8, where Figure 8a represents the map outputted by the parent model, and Figure 8b is the map produced by sub-models, Figure 8c,d are the detailed scopes for the part of both maps. The entire region is divided and classified into five zones of susceptibility to different levels of landslides (very low, low, moderate, high, and very

high) by the method of natural breaks method. From the result of the parent model in Table 5, 14.6% of the whole area is classified as a very low landslide-prone zone, 22.8% as a low landslide-prone zone, 25.23% as a moderate landslide-prone zone, 21.47% as a high landslide-prone zone, and 15.90% as a very high landslide-prone zone. From the result of the sub-models, their ratios are 16.42%, 20.81%, 22.56%, 21.54%, and 18.67%, respectively.

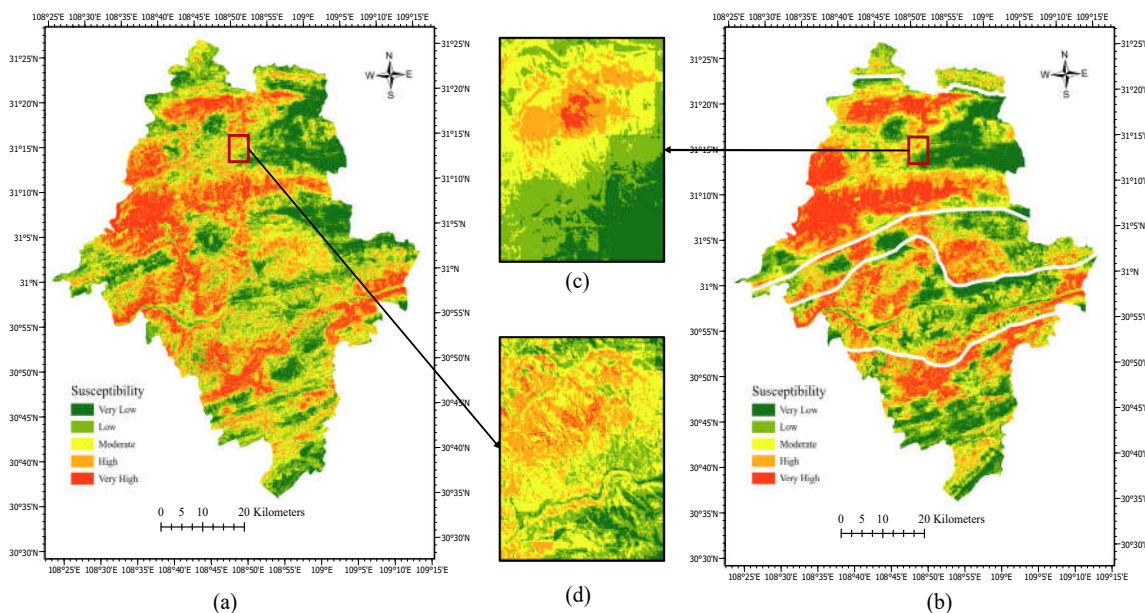


Figure 8. Comparison between the landslide susceptibility mapping of the parent model and sub-models: (a) Landslide susceptibility map generated by the parent model, (b) Landslide susceptibility map generated by the sub-models, (c) Detailed view of landslide susceptibility map generated by the sub-models, (d) Detailed view of landslide susceptibility map generated by the parent model.

Table 5. Statistic results of landslide susceptibility in different levels of different models.

Model	Landslides Susceptibility Level	Landslides Susceptibility Threshold	Landslides Ratio	Area Ratio	Landslides/Area Ratio
Parent Model	Very Low	0.000–0.239	0.24%	14.60%	0.016
	Low	0.239–0.408	2.18%	22.80%	0.096
	Moderate	0.408–0.573	6.05%	25.23%	0.248
	High	0.573–0.741	19.22%	21.47%	0.895
	Very High	0.741–1.000	72.31%	15.90%	4.548
Sub-Models	Very Low	-	0.12%	16.42%	0.007
	Low	-	1.21%	20.81%	0.058
	Moderate	-	4.48%	22.56%	0.199
	High	-	18.55%	21.54%	0.861
	Very High	-	75.64%	18.67%	4.051
Sub-Model One	Very Low	0.000–0.235	0.00%	20.22%	0.000
	Low	0.235–0.435	0.00%	23.88%	0.000
	Moderate	0.435–0.631	8.99%	22.36%	0.402
	High	0.631–0.831	25.84%	18.64%	1.386
	Very High	0.831–1.000	65.17%	14.90%	4.374
Sub-Model Two	Very Low	0.000–0.278	0.35%	15.14%	0.023
	Low	0.278–0.435	2.46%	22.03%	0.112
	Moderate	0.435–0.588	4.56%	24.99%	0.182
	High	0.588–0.749	13.33%	22.45%	0.594
	Very High	0.749–1.000	79.30%	15.39%	5.153

Table 5. Cont.

Model	Landslides Susceptibility Level	Landslides Susceptibility Threshold	Landslides Ratio	Area Ratio	Landslides/Area Ratio
Sub-Model Three	Very Low	0.000–0.303	0.00%	13.68%	0.000
	Low	0.303–0.421	0.00%	24.81%	0.000
	Moderate	0.421–0.536	9.10%	25.78%	0.353
	High	0.536–0.662	31.81%	22.06%	1.442
	Very High	0.662–1.000	59.09%	13.67%	4.323
Sub-Model Four	Very Low	0.000–0.227	0.00%	16.59%	0.000
	Low	0.227–0.427	0.74%	16.20%	0.0457
	Moderate	0.427–0.612	2.94%	19.38%	0.152
	High	0.612–0.792	19.12%	22.26%	0.859
	Very High	0.792–1.000	77.20%	25.57%	3.020

Logically, the landslides/area ratio should increase from a very low landslide-prone zone to a very high landslide-prone zone, which is exactly what our models indicate. According to the results of the parent model, the landslides/area ratio increases from 0.016 to 4.548, from a very low landslide-prone zone to a very high landslide-prone zone, and that also increases from 0.007 to 4.051 from the results of the sub-models. Figure 9 displays such a tendency, and it can be seen that the outputs of the sub-models have more obvious gaps between the very low landslide-prone zone and the very high landslide-prone-zone, which reflects another merit of the sub-models compared with the parent model.

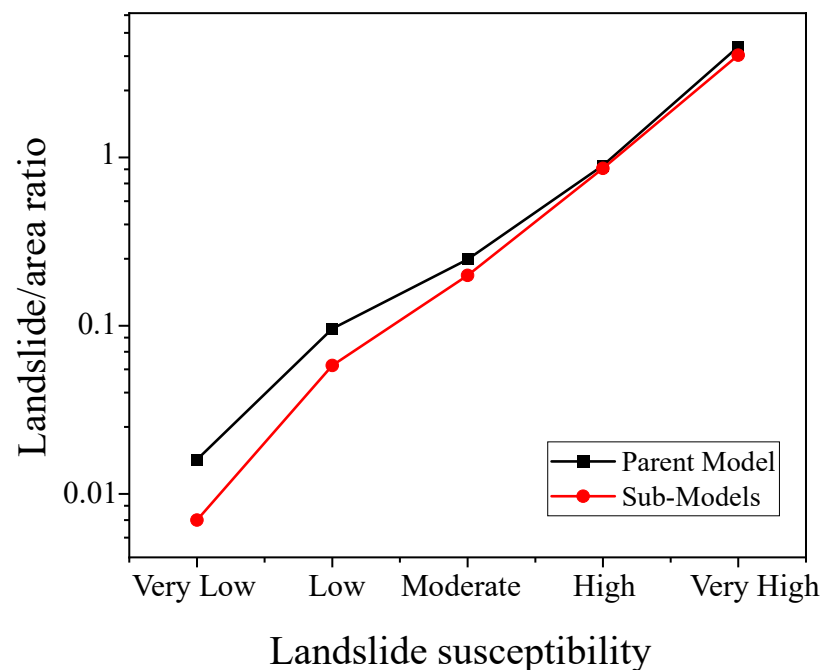


Figure 9. Landslides/area ratio.

The validation AUC values of the five models are shown in Figure 10; the AUC value of the parent model is 0.872, while that of sub-model one to sub-model four are 0.949, 0.892, 0.889, and 0.951, respectively. All of the sub-models outperformed the parent model, which proved the aforementioned hypothesis. However, the increases are not at the same level; sub-model four achieved the highest improvement (9.1%), while the lowest improvement is 1.9%, which was obtained by sub-model two.

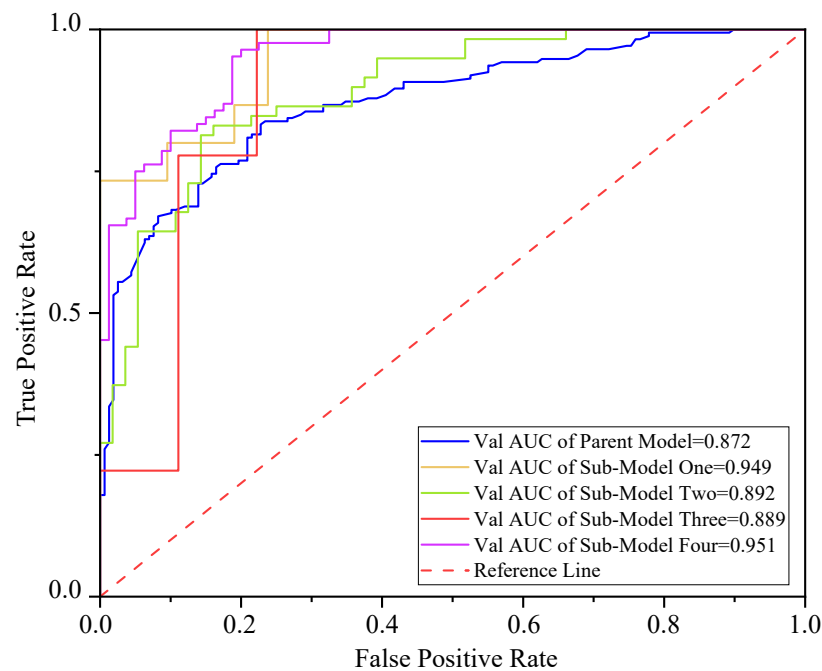


Figure 10. Validation AUC values of different models.

The importance of features generally represents how much a specific feature contributes to the decision-making process of a model. In this case, the most important features can be the key factors in identifying landslide/non-landslide points. As shown in Table 6 and Figure 11, the distance from syncline axis, aridity, elevation, distance from rivers, and average annual temperature are of the highest importance for the parent model. NDVI, average annual rainfall, distance from road, and elevation are the main features that are associated with the formation of landslides in sub-zone I. For sub-model two, elevation is the most important feature, which is followed by average annual rainfall, distance from rivers, distance from anticline axis, and average annual temperature. Distance from road, HAILS, elevation, plane curvature, and average annual temperature are the top features for sub-model three. Last but not least, distance from syncline axis, average annual temperature, elevation, aridity, and average annual rainfall play important roles during the predicting process of sub-model four.

Table 6. Top five important factors in different sub-zones.

Zone	Feature Importance Rank				
	1st	2nd	3rd	4th	5th
Whole Region	Distance from syncline axis (m)	Aridity	Elevation (m)	Distance from rivers (m)	Average Annual Temperature (°C)
Sub-Zone I	Distance from syncline axis (m)	NDVI	Average Annual Rainfall (mm)	Distance from Road (m)	Elevation (m)
Sub-Zone II	Elevation (m)	Average Annual Rainfall (mm)	Distance from rivers (m)	Distance from anticline axis (m)	Average Annual Temperature (°C)
Sub-Zone III	Distance From Road (m)	HAILS	Elevation (m)	Plane Curvature	Average Annual Temperature (°C)
Sub-Zone IV	Distance from syncline axis (m)	Average Annual Temperature (°C)	Elevation (m)	Aridity	Average Annual Rainfall (mm)

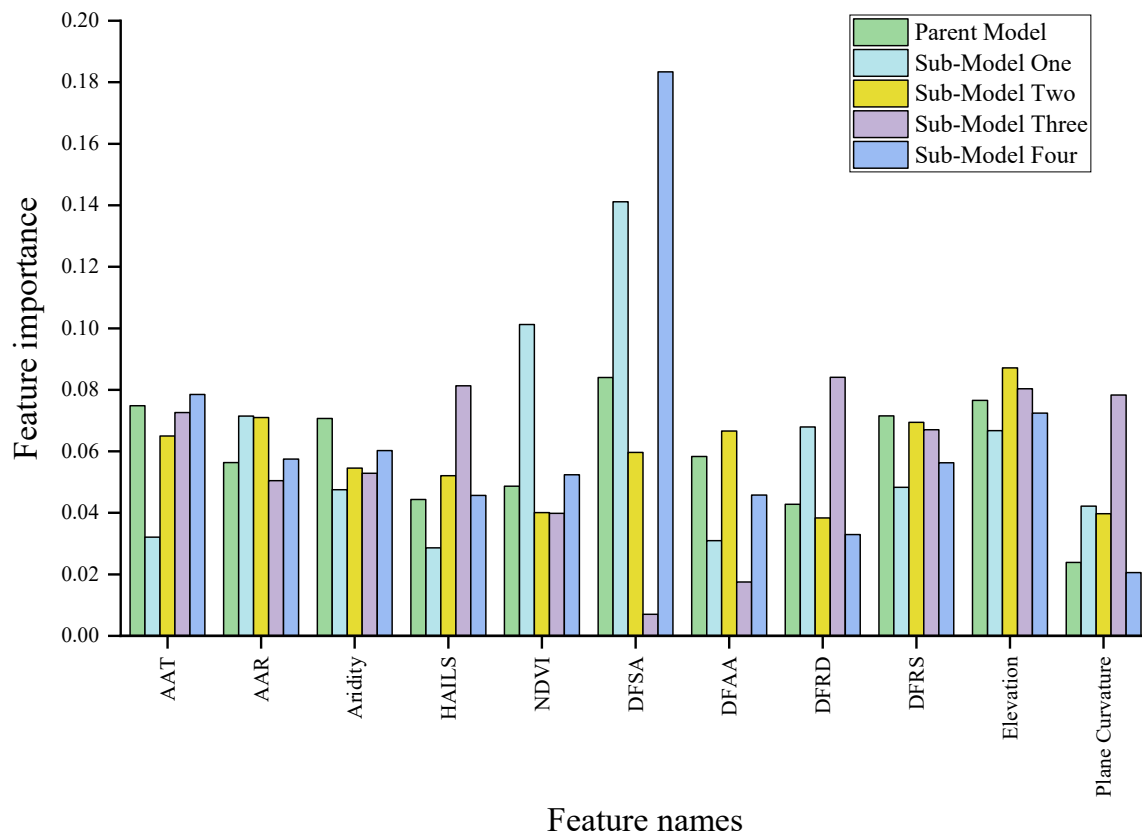


Figure 11. Top five important factors of different models. AAT: Average annual temperature; AAR: Average annual rainfall; DFSA: Distance from syncline axis; DFAA: Distance from anticline axis; DFRD: Distance from road; DFRS: Distance from rivers.

6. Discussion

6.1. Feature Importance Analysis

As indicated in Table 6, the distance from syncline axis is at the dominant place among the twenty factors for the parent model, sub-model one, and sub-model four. Factors associated with water, such as average annual rainfall and the distance from rivers, are also important places. The triggering factors of shallow landslides are highly dependent on the rainfall water infiltration and its further redistribution [41]. Due to its softening effect and long-term erosion effect, the distance from rivers has a significant influence on the development of landslides [42]. Temperature has a remarkable effect on landslide formation; experimental results indicated that the shear strength of slip surface soils reduces with decreasing temperature, which will negatively affect slope instability [43]. Therefore, obtained by the accumulated temperature and rainfall, aridity is also of great importance. Although different plants generally have a different contribution of rainfall to soil water [44], which will affect slope stability differently, on average, the effect of water uptake from the plant cover makes the vegetated slopes averagely 12.84% drier, and matric suctions three times higher than the fallow slope, which contributes to slope stability [45]; so NDVI is taken as one of the main considerations. Unlike vegetation cover, human-made land covers usually have negative influences on slope stability. The development of new building areas can potentially increase susceptibility to landslides [46]. Road networks not only directly destabilize existing slopes by disturbing their original structures during construction processes, but transport activities also have negative effects on slope stability. Similarly, HAILS is another influential factor. For topological factors, the two most important factors, in this case, are elevation and plane curvature.

6.2. Model Comparison

6.2.1. Parent Model

When dealing with the whole region, the model tends to capture information that is more general, and thus its feature importance has a higher universality. In our case, six major synclines are lying in the whole domain, and their generally dominant effects on identifying landslide/non-landslide points are successfully captured by the parent model. Aridity, distance from rivers, and average annual temperature affect the formation of landslides in meteorological and hydrological aspects. Although ranked fourth as the most important one, elevation is the most important topological factor in the research area.

6.2.2. Sub-Models

After dividing the whole region into different sub-zones, the models will have the chance to learn the knowledge that is specifically right for each sub-zone. Thus their performances will be improved.

Sub-model One. Located in the south of Yunyang County, sub-zone I occupies 741.7 square kilometers, which is only 20% of the entire research area. Nevertheless, it is crossed by two of the six major synclines, which makes the syncline effect more obvious, and, therefore, the distance from the syncline axis is still the most important factor for sub-zone I. Similarly, the landslide susceptibility in sub-zone I is also affected by both hydrological and topological factors, namely average annual rainfall and elevation. However, specifically, as it is crossed by the major highway S305, land cover (NDVI) and human activities (distance from road) are the other two main influential factors.

Sub-model Two. As we discussed before, sub-zone II is the most special sub-zone since it is crossed by the Three Gorges Reservoir area, most of the landslides in Yunyang County were impacted by water occurred here. Therefore, theoretically, the landslide formation in sub-zone II is sensitive to hydrological factors. In this case, the average annual rainfall and the distance from rivers are ranked as the second and the third-most important factors. The reason why elevation is of the greatest importance is that lower places are usually prone to the influence of the periodic variation of reservoir water level. Statistically, 76% of the landslides impacted by water in sub-zone II occurred below the elevation of 249 m. The tectonic action and natural factors also contribute a lot to the landslide formation in sub-zone II; the distance from anticline axis and the average annual temperature are ranked in fourth and fifth place, respectively.

Sub-model Three. For sub-zone III, specifically, its data points are relatively far away from the major synclines, which mitigates the importance of the feature. Instead, as an area crossed by major highways and dense road networks, human activities, including distance from road and HAILS, are the two most important factors that have the main contribution to the landslide formation of the sub-zone. Topographical factors, namely elevation and plane curvature, are ranked in third and fourth place, respectively, and the average annual rainfall is evaluated as the fifth-most important factor.

Sub-model Four. Crossed by three major synclines, the landslide identification in sub-zone IV is also significantly dependent on the distance from the syncline axis. Furthermore, since it is part of the low hills section crossing Yunyang, Fengjie, and Kaizhou, the formation of landslides in this area is primarily influenced by the factors associated with the stability of mountain bodies, such as the average annual temperature, elevation, aridity, and the average annual rainfall.

7. Conclusions

In this study, Yunyang County is manually zoned into four parts based on the qualitative analysis of geological hazards exploration in Chongqing City, including the mechanism of landslide formation and sliding failure and geomorphic unit characteristics. Based on the qualitative analysis result, five random forest landslide susceptibility models are constructed using historical landslides data points and twenty relating factors for the following quantity analysis. These models, including a parent model and four sub-models,

are optimized by the grid search method individually. A comparison between the parent model and the combination of the sub-models is conducted. The following conclusions are drawn:

The AUC value of the parent model achieves 0.872, which shows that the traditional RF with the hyper-parameters tuned by the grid search method has a reliable performance on landslide susceptibility mapping. In this study, synclines have the most important effects on the formation of landslides in Yunyang County, followed by aridity, elevation, distance from rivers, and average annual temperature.

However, more general information extracted from “mainstream” landslides would usually cover that of the “minority” landslides when treating a large region equally, resulting in low information utility and the inability to identify potential landslides under special geological conditions. With enough data points, experience-based zoning before modeling is proved to be an effective solution to the issue; the qualitative analysis serves the purpose of pre-classification based on the information from geological hazards exploration, which groups the landslides that occurred under similar geological conditions, and thus enables the models to obtain the specific knowledge under each condition. Therefore, in our case, while the traditional RF obtained the general prediction skill for the entire region of Yunyang County, all the sub-models have become “experts” in their respective sub-areas. The test AUC values of sub-model one to four are 8.8%, 2.3%, 1.9%, and 9.1% higher than those of the parent model. Furthermore, the proposed method also contributes to further revealing the key factors that include local landslide instability under specific geological conditions, which can be used by planners and policymakers for a more specific and accurate landslide control in certain areas, thus further improving the safety of life and public property.

For sub-zone I, the top five conditioning factors are distance from syncline axis, NDVI, average annual rainfall, distance from road, and elevation. For sub-zone III, without the influence of major synclines, its top factors are distance from road, HAILS, elevation, plane curvature, and average annual temperature. For sub-zone IV, the distance from syncline axis becomes the most important factor again, and it is followed by average annual temperature, elevation, aridity, and average annual rainfall.

Sub-zone II is crossed by the Three Gorges Reservoir area. Suffered by periodic variation in reservoir water level and the impacts of other factors related to the reservoir band, the modified method based on general conditioning factors has relatively less effect on improving the accuracy of the mapping. The effect of more specific factors on the formation of landslides on the banks of the reservoir will be analyzed in further research. In the case of this paper, the results of sub-model two point out that elevation, average annual rainfall, distance from rivers, distance from anticline axis, and average annual temperature are the top five conditioning factors among the existing twenty factors for sub-zone II.

Author Contributions: Methodology, S.L.; validation, L.W.; software, Y.H.; resources, P.S.; formal analysis, M.C.; supervision, W.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China (2019YFC1509605), the China Postdoctoral Science Foundation funded project (2021M700608), the Natural Science Foundation of Chongqing, China (cstc2021jcyj-bsh0047), High-end Foreign Expert Introduction program (G20200022005 and DL2021165001L), and Science and Technology Research Program of Chongqing Municipal Education Commission (HZ2021001).

Data Availability Statement: Restrictions apply to the availability of these data. Data was obtained from National Earth System Science Data Center, Resource and Environment Science and Data Center and Chongqing Geological monitoring station and are available at <http://www.geodata.cn/?jsessionid=4FDC4730785878003B1ED71B62212634> (accessed on 1 May 2022), and <https://www.resdc.cn/Default.aspx> (accessed on 1 May 2022) with the permission of National Earth System Science Data Center and Resource and Environment Science and Data Center.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Huang, F.; Huang, J.; Jiang, S.; Zhou, C. Landslide displacement prediction based on multivariable chaotic model and extreme learning machine. *Eng. Geol.* **2017**, *218*, 173–186. [\[CrossRef\]](#)
- Petley, D. Global patterns of loss of life from landslides. *Geology* **2012**, *40*, 927–930. [\[CrossRef\]](#)
- Lin, L.; Lin, Q.; Wang, Y. Landslide susceptibility mapping on a global scale using the method of logistic regression. *Nat. Hazards Earth Syst. Sci.* **2017**, *17*, 1411–1424. [\[CrossRef\]](#)
- Sheng, L.; Wang, W.; Zhu, W. *China Statistical Yearbook 2016*; China Statistics Press: Beijing, China, 2016. (In Chinese)
- Sun, D.; Xu, J.; Wen, H.; Wang, D. Assessment of landslide susceptibility mapping based on Bayesian hyperparameter optimization: A comparison between logistic regression and random forest. *Eng. Geol.* **2021**, *281*, 105972. [\[CrossRef\]](#)
- Brabb, E. Innovative approaches to landslide hazard mapping. In Proceedings of the 4th International Symposium on Landslides, Toronto, ON, Canada, 16–21 September 1984; Volume 1, pp. 307–324.
- Wu, C.; Hong, L.; Wang, L.; Zhang, R.; Pijush, S.; Zhang, W. Prediction of wall deflection induced by braced excavation in spatially variable soils via convolutional neural network. *Gondwana Res.* **2022**. [\[CrossRef\]](#)
- Wang, Y.; Tang, H.; Huang, J.; Wen, T.; Ma, J.; Zhang, J. A comparative study of different machine learning methods for reservoir landslide displacement prediction. *Eng. Geol.* **2022**, *298*, 106544. [\[CrossRef\]](#)
- Huang, F.; Chen, J.; Liu, W.; Huang, J.; Hong, H.; Chen, W. Regional rainfall-induced landslide hazard warning based on landslide susceptibility mapping and a critical rainfall threshold. *Geomorphology* **2022**, *408*, 108236. [\[CrossRef\]](#)
- He, Q.; Wang, M.; Liu, K. Rapidly assessing earthquake-induced landslide susceptibility on a global scale using random forest. *Geomorphology* **2021**, *391*, 107889. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Smith, H.G.; Spiekermann, R.; Betts, H.; Neverman, A.J. Comparing methods of landslide data acquisition and susceptibility modelling: Examples from New Zealand. *Geomorphology* **2021**, *381*, 107660. [\[CrossRef\]](#)
- Lim, C.H.; Kim, H.J. Can Forest-Related Adaptive Capacity Reduce Landslide Risk Attributable to Climate Change?—Case of Republic of Korea. *Forests* **2022**, *13*, 49. [\[CrossRef\]](#)
- Nhu, V.-H.; Shirzadi, A.; Shahabi, H.; Singh, S.K.; Al-Ansari, N.; Clague, J.J.; Jaafari, A.; Chen, W.; Miraki, S.; Dou, J.; et al. Shallow landslide susceptibility mapping: A comparison between logistic model tree, logistic regression, naïve Bayes tree, artificial neural network, and support vector machine algorithms. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2749. [\[CrossRef\]](#)
- Zhang, W.; Li, H.; Han, L.; Chen, L.; Wang, L. Slope stability prediction using ensemble learning techniques: A case study in Yunyang County, Chongqing, China. *J. Rock Mech. Geotech. Eng.* **2022**, *in press*. [\[CrossRef\]](#)
- Hu, Q.; Zhou, Y.; Wang, S.; Wang, F. Machine learning and fractal theory models for landslide susceptibility mapping: Case study from the Jinsha River Basin. *Geomorphology* **2020**, *351*, 106975. [\[CrossRef\]](#)
- Zhou, X.; Wen, H.; Zhang, Y.; Xu, J.; Zhang, W. Landslide susceptibility mapping using hybrid random forest with GeoDetector and RFE for factor optimization. *Geosci. Front.* **2021**, *12*, 101211. [\[CrossRef\]](#)
- Sun, D.; Gu, Q.; Wen, H.; Shi, S.; Mi, C.; Zhang, F. A Hybrid Landslide Warning Model Coupling Susceptibility Zoning and Precipitation. *Forests* **2022**, *13*, 827. [\[CrossRef\]](#)
- Zhou, X.; Wen, H.; Li, Z.; Zhang, H.; Zhang, W. An interpretable model for the susceptibility of rainfall-induced shallow landslides based on SHAP and XGBoost. *Geocarto Int. TGEI* **2022**, *1–27*, (online). [\[CrossRef\]](#)
- Akar, Ö.; Güngör, O. Integrating multiple texture methods and NDVI to the Random Forest classification algorithm to detect tea and hazelnut plantation areas in northeast Turkey. *Int. J. Remote Sens.* **2015**, *36*, 442–464. [\[CrossRef\]](#)
- Hengl, T.; Heuvelink, G.B.M.; Kempen, B.; Leenaars, J.G.B.; Walsh, M.G.; Shepherd, K.D.; Sila, A.; MacMillan, R.A.; Mendes De Jesus, J.; Tamene, L.; et al. Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. *PLoS ONE* **2015**, *10*, e0125814. [\[CrossRef\]](#)
- Feng, Q.; Liu, J.; Gong, J. Urban flood mapping based on unmanned aerial vehicle remote sensing and random forest classifier—A case of Yuyao, China. *Water* **2015**, *7*, 1437–1455. [\[CrossRef\]](#)
- Taalab, K.; Cheng, T.; Zhang, Y. Mapping landslide susceptibility and types using Random Forest. *Big Earth Data* **2018**, *2*, 159–178. [\[CrossRef\]](#)
- Yin, Y.; Wang, L.; Zhang, W.; Dai, Z. Research on the collapse process of a thick-layer dangerous rock on the reservoir bank. *Bull. Eng. Geol. Environ.* **2022**, *81*, 109. [\[CrossRef\]](#)
- Wang, L.; Zhang, Z.; Huang, B.; Hu, M.; Zhang, C. Triggering mechanism and possible evolution process of the ancient Qingshi landslide in the Three Gorges Reservoir. *Geomat. Nat. Hazards Risk* **2021**, *12*, 3160–3174. [\[CrossRef\]](#)
- Nakileza, B.R.; Nedala, S. Topographic influence on landslides characteristics and implication for risk management in upper Manafwa catchment, Mt Elgon Uganda. *Geoenviron. Disasters* **2020**, *7*, 27. [\[CrossRef\]](#)
- Nefeslioglu, H.A.; Gokceoglu, C.; Sonmez, H.; Gorum, T. Medium-scale hazard mapping for shallow landslide initiation: The Buyukoy catchment area (Cayeli, Rize, Turkey). *Landslides* **2011**, *8*, 459–483. [\[CrossRef\]](#)
- Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [\[CrossRef\]](#)

29. Cutler, A.; Stevens, J.R. Random forests for microarrays. *Methods Enzymol.* **2006**, *411*, 422–432. [[CrossRef](#)]
30. Zhang, W.; Wu, C.; Zhong, H.; Li, Y.; Wang, L. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. *Geosci. Front.* **2020**, *12*, 469–477. [[CrossRef](#)]
31. Lee, S.; Pradhan, B. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides* **2007**, *4*, 33–41. [[CrossRef](#)]
32. Raghuvanshi, T.K.; Negassa, L.; Kala, P.M. GIS based grid overlay method versus modeling approach—A comparative study for landslide hazard zonation (LHZ) in Meta Robi District of west showa zone in Ethiopia. *Egypt. J. Remote Sens. Space Sci.* **2015**, *18*, 235–250. [[CrossRef](#)]
33. Girma, F.; Raghuvanshi, T.K.; Ayenew, T.; Hailemariam, T. Landslide hazard zonation in Ada Berga district, central Ethiopia- A GIS based statistical approach. *J. Geomat.* **2015**, *9*, 25–38.
34. 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](#)]
35. Yesilnacar, E.; Topal, T. Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek Region (Turkey). *Eng. Geol.* **2005**, *79*, 251–266. [[CrossRef](#)]
36. Mathew, J.; Jha, V.K.; Rawat, G.S. Landslide susceptibility zonation mapping and its validation in part of Garhwal Lesser Himalaya, India, using binary logistic regression analysis and receiver operating characteristic curve method. *Landslides* **2008**, *6*, 17–26. [[CrossRef](#)]
37. Wang, X.; Li, S.; Liu, H.; Liu, L.; Liu, Y.; Zeng, S.; Tang, Q. Landslide susceptibility assessment in Wenchuan County after the 5.12 magnitude earthquake. *Bull. Eng. Geol. Environ.* **2021**, *80*, 5369–5390. [[CrossRef](#)]
38. LaValle, S.M.; Branicky, M.S. *On the Relationship between Classical Grid Search and Probabilistic Roadmaps*; Springer Tracts in Advanced Robotics: Berlin/Heidelberg, Germany, 2004; pp. 59–75. [[CrossRef](#)]
39. Liu, C.H.; Chamberlain, B.P.; Little, D.A.; Cardoso, Â. Generalising random forest parameter optimisation to include stability and cost. *Mach. Learn. Knowl. Discov. Databases* **2017**, *10536*, 102–113. [[CrossRef](#)]
40. Lujan-Moreno, G.A.; Howard, P.R.; Rojas, O.G.; Montgomery, D.C. Design of experiments and response surface methodology to tune machine learning hyperparameters, with a random forest case-study. *Expert Syst. Appl.* **2018**, *109*, 195–205. [[CrossRef](#)]
41. Ivanov, V.; Arosio, D.; Tresoldi, G.; Hojat, A.; Zanzi, L.; Papini, M.; Longoni, L. Investigation on the role of water for the stability of shallow landslides—insights from experimental tests. *Water* **2020**, *12*, 1203. [[CrossRef](#)]
42. Wang, F.; Xu, P.; Wang, C.; Wang, N.; Jiang, N. Application of a GIS-based slope unit method for landslide susceptibility mapping along the Longzi River, southeastern Tibetan Plateau, China. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 172. [[CrossRef](#)]
43. Shibasaki, T.; Matsuura, S.; Okamoto, T. Experimental evidence for shallow, slow-moving landslides activated by a decrease in ground temperature. *Geophys. Res. Lett.* **2016**, *43*, 6975–6984. [[CrossRef](#)]
44. Zhang, Y.; Zhang, B.; Xu, Q.; Gao, D.; Xu, W.; Ren, R.; Jiang, J.; Wang, S. The Effects of Plant and Soil Characteristics on Partitioning Different Rainfalls to Soil in a Subtropical Chinese Fir Forest Ecosystem. *Forests* **2022**, *13*, 123. [[CrossRef](#)]
45. Gonzalez-Ollauri, A.; Mickovski, S.B. Hydrological effect of vegetation against rainfall-induced landslides. *J. Hydrol.* **2017**, *549*, 374–387. [[CrossRef](#)]
46. Promper, C.; Puissant, A.; Malet, J.-P.; Glade, T. Analysis of land cover changes in the past and the future as contribution to landslide risk scenarios. *Appl. Geogr.* **2014**, *53*, 11–19. [[CrossRef](#)]