remote sensing

MDPI

Article

A Comparative Analysis of Certainty Factor-Based Machine Learning Methods for Collapse and Landslide Susceptibility Mapping in Wenchuan County, China

Xinyue Yuan [1,2], Chao Liu [1,2], Ruihua Nie [1,2], Zhengli Yang [1,2], Weile Li [3], Xiaoai Dai [4], Junying Cheng [4], Junmin Zhang [4], Lei Ma [5], Xiao Fu [6], Min Tang [7], Yina Xu [7] and Heng Lu [1,2,*]

1 State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Chengdu 610065, China; yuanxinyue@stu.scu.edu.cn (X.Y.); liuchao@scu.edu.cn (C.L.); nierh@scu.edu.cn (R.N.); yangzhengli@scu.edu.cn (Z.Y.)
2 College of Hydraulic and Hydroelectric Engineering, Sichuan University, Chengdu 610065, China
3 State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu 610059, China; liweile08@mail.cdut.edu.cn
4 College of Earth Science, Chengdu University of Technology, Chengdu 610059, China; daixiaoa@mail.cdut.edu.cn (X.D.); chunjunying@stu.cdut.edu.cn (J.C.); zhangjunmin@stu.cdut.edu.cn (J.Z.)
5 School of Geography and Ocean Science, Nanjing University, Nanjing 210093, China; maleinju@nju.edu.cn
6 Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China; fuxiao@my.swjtu.edu.cn
7 China Railway Eryuan Engineering Group Co., Ltd., Chengdu 610031, China; tangmin05@ey.crec.cn (M.T.); xuyn@ey.crec.cn (Y.X.)
* Correspondence: luheng@scu.edu.cn

Citation: Yuan, X.; Liu, C.; Nie, R.; Yang, Z.; Li, W.; Dai, X.; Cheng, J.; Zhang, J.; Ma, L.; Fu, X.; et al. A Comparative Analysis of Certainty Factor-Based Machine Learning Methods for Collapse and Landslide Susceptibility Mapping in Wenchuan County, China. *Remote Sens.* **2022**, *14*, 3259. https://doi.org/10.3390/rs14143259

Academic Editor: Francesca Ardizzone

Received: 15 June 2022
Accepted: 4 July 2022
Published: 6 July 2022

Abstract: After the "5·12" Wenchuan earthquake in 2008, collapses and landslides have occurred continuously, resulting in the accumulation of a large quantity of loose sediment on slopes or in gullies, providing rich material source reserves for the occurrence of debris flow and flash flood disasters. Therefore, it is of great significance to build a collapse and landslide susceptibility evaluation model in Wenchuan County for local disaster prevention and mitigation. Taking Wenchuan County as the research object and according to the data of 1081 historical collapse and landslide disaster points, as well as the natural environment, this paper first selects six categories of environmental factors (13 environmental factors in total) including topography (slope, aspect, curvature, terrain relief, TWI), geological structure (lithology, soil type, distance to fault), meteorology and hydrology (rainfall, distance to river), seismic impact (PGA), ecological impact (NDVI), and impact of human activity (land use). It then builds three single models (LR, SVM, RF) and three CF-based hybrid models (CF-LR, CF-SVM, CF-RF), and makes a comparative analysis of the accuracy and reliability of the models, thereby obtaining the optimal model in the research area. Finally, this study discusses the contribution of environmental factors to the collapse and the landslide susceptibility prediction of the optimal model. The research results show that (1) the areas prone to extremely high collapse and landslide predicted by the six models (LR, CF-LR, SVM, CF-SVM, RF and CF-RF) have an area of 730.595 km^2, 377.521 km^2, 361.772 km^2, 372.979 km^2, 318.631 km^2, and 306.51 km^2, respectively, and the frequency ratio precision of collapses and landslides is 0.916, 0.938, 0.955, 0.956, 0.972, and 0.984, respectively; (2) the ranking of the comprehensive index based on the confusion matrix is CF-RF>RF>CF-SVM>CF-LR>SVM>LR and the ranking of the AUC value is CF-RF>RF>CF-SVM>CF-LR>SVM>LR. To a certain extent, the coupling models can improve precision more over the single models. The CF-RF model ranks the highest in all indexes, with a POA value of 257.046 and an AUC value of 0.946; (3) rainfall, soil type, and distance to river are the three most important environmental factors, accounting for 24.216%, 22.309%, and 11.41%, respectively. Therefore, it is necessary to strengthen the monitoring of mountains and rock masses close to rivers in case of rainstorms in Wenchuan county and other similar areas prone to post-earthquake landslides.

Keywords: collapse and landslide disaster; susceptibility evaluation; machine learning model; certainty factor; Wenchuan county

1. Introduction

Collapse is a sudden sharp inclination and falling movement of rock-soil mass on a steep hillside under the action of gravity, while landslide is a process in which the rock-soil mass on the side slope slides along the weak surface as a whole or dispersedly, under the action of gravity under the influence of surface water infiltration, river erosion, seismic activity, human activities, and other factors [1–3]. Unstable slope refers to a slope in a critical state that is about to lose stability, that is, a slope with the potential for collapse and landslide. With the development of society, the exploitation of the natural environment by human activities has continuously intensified, and the frequency of occurrence of collapse and landslide disasters has become higher and higher. In 2019, the total number of collapse and landslide disasters in China was 4220, accounting for more than 88% of the total number of geological disasters, and secondary disasters caused by these also block rivers, trigger floods, and form debris flows, posing a serious threat to human life and property, infrastructure, and natural resources [4–7]. Collapses and landslides often occur together under the same or similar conditions. There are many factors affecting the occurrence of collapses and landslides, mainly divided into four categories: topography, geological conditions, endogenic and exogenic geological processes, and human activities [8]. Therefore, it is of great significance for the early prediction, prevention, and mitigation of collapse and landslide disasters to analyze the impact factors that cause the occurrence of collapse and landslide disasters, build a regional collapse and landslide disaster susceptibility evaluation model, and evaluate the susceptibility level of collapse and landslide disasters [9–12].

With the continuous development of remote sensing, geographic information systems, global positioning systems, and other spatial information technologies and computer hardware equipment, the susceptibility prediction model of collapse and landslide disasters has developed from a qualitative model to a quantitative model [13,14]. A qualitative model is mainly driven by knowledge and the quality of the evaluation results is closely related to the evaluator's own experience, as in the fuzzy comprehensive evaluation method, analytical hierarchy process, and so on [14,15]. Driven by data, quantitative models are widely used evaluation models at present, and their evaluation results are more objective, as with logistic regression, weight of evidence, frequency ratio, certainty factor, information value, etc. [16–20]. As the volume of data increases, the complexity of terrain, geology, and other elements cannot be completely solved by simulation analysis through traditional mathematical methods. Therefore, some scholars have introduced machine learning methods to establish collapse and landslide disaster susceptibility prediction models, which can automatically analyze the input data and connect the nonlinear relationship between targets and factors, as in neural networks, support vector machines, random forests, and maxent; there also have high computational efficiency for high-dimensional data [21–28]. In order to improve prediction accuracy, deep learning methods are widely used, such as convolutional neural networks (CNN) and deep neural networks (DNN) [22,29]. The coupling model combines two or more models, integrates the collapse and landslide sample selection, feature selection, and information extraction for collapse and landslide disaster prediction, and synthesizes the advantages of each model so as to effectively improve the prediction precision of the models [30–33].

On 12 May 2008, Wenchuan county in the Ngawa Tibetan and Qiang Autonomous Prefecture of Sichuan Province was struck by an 8.0-magnitude earthquake, which was the most destructive earthquake in modern China. The earthquake caused a large number of potential geological disasters, induced about 50,000 collapses and landslides covering an area of 750 km^2, and formed 5.25 billion m^3 of loose sediments. Under the inducement of heavy rainfall, disasters such as flash floods and debris flow can easily occur [34–36].

Research has shown that the geological disasters after earthquakes show a vibration attenuation trend, with a peak period of 4–5a in 20–25a, and finally recover to the pre-earthquake level [37]. Therefore, it is of great significance to carry out research on the susceptibility assessment and prediction of collapse and landslide disasters in Wenchuan County for the early warning, prevention, and mitigation of collapse and landslide disasters. Previous studies on landslide susceptibility in Wenchuan County have mostly adopted a single machine learning model [30–33], there have been few studies on the coupling of statistical methods and machine learning methods, and there are few opinions on landslide disaster prevention in Wenchuan County. Taking Wenchuan County as the research object and based on data of historical collapse and landslide disasters, as well as the natural environment, this paper selects a total of 13 environmental factors and couples them with certainty factors with three machine learning methods (namely logistic regression, support vector machine, and random forest) to build six collapse and landslide susceptibility prediction models to evaluate the susceptibility of collapse and landslide disasters in Wenchuan County. It then obtains the laws of the impact of each environmental factor on the development of collapse and landslide in its attribute intervals.

2. Materials and Methods

2.1. Overview of the Research Area and Collapse & Landslide Information

Wenchuan, a county in the Ngawa Tibetan and Qiang Autonomous Prefecture of Sichuan Province, China, is located at the northwest edge of the Sichuan basin. It has an east–west width of 84 km and a north–south length of 105 km, and also a total area of 4084 km^2, with a spatial range between 30°45′–31°43′north latitude and 102°51′–103°44′east longitude (Figure 1). In terms of topography, with the Longmen mountains in the northeast and the Qionglai mountain systems in the southwest, its terrain is mainly high and middle mountains, with an altitude of 745–5927 m, and its topography inclines from northwest to southeast. In Wenchuan County, the river system is very developed. There are many rivers and nearly 200 tributaries, which include the Minjiang river, Zagunao river, Shoujiang river, Caopo river, etc. Wenchuan County has an average annual air temperature of 13.5 °C and an annual rainfall of 500 mm, and belongs to the sub-humid climate region of the Qinghai Tibet Plateau, with the climate rising with the topography from southeast to northwest on the whole, and the rainfall gradually decreasing from south to north. The research area lies in the Longmen mountain structural belt on the eastern edge of the Qinghai Tibet Plateau, and there are two major fault zones, that is, Maoxian-Wenchuan fault zone and Beichuan-Yingxiu fault zone. In terms of stratum lithology, the stratum types are well developed. The rocks were formed in the Cenozoic Quaternary Period, Mesozoic Jurassic Period, Cretaceous Period, and Paleozoic Period. The lithology mainly includes magmatic rocks, granite, diorite, and gabbro.

According to the data of disaster points over the years provided by the Sichuan Geological Survey, the earliest disaster point was recorded on 1 July 1958, while the latest disaster point was recorded on 26 June 2020. There are 1081 unstable slopes, collapses and landslides in the research area, including 454 collapses, 360 landslides, and 267 unstable slopes, accounting for 42.00%, 33.30%, and 24.70%, respectively.

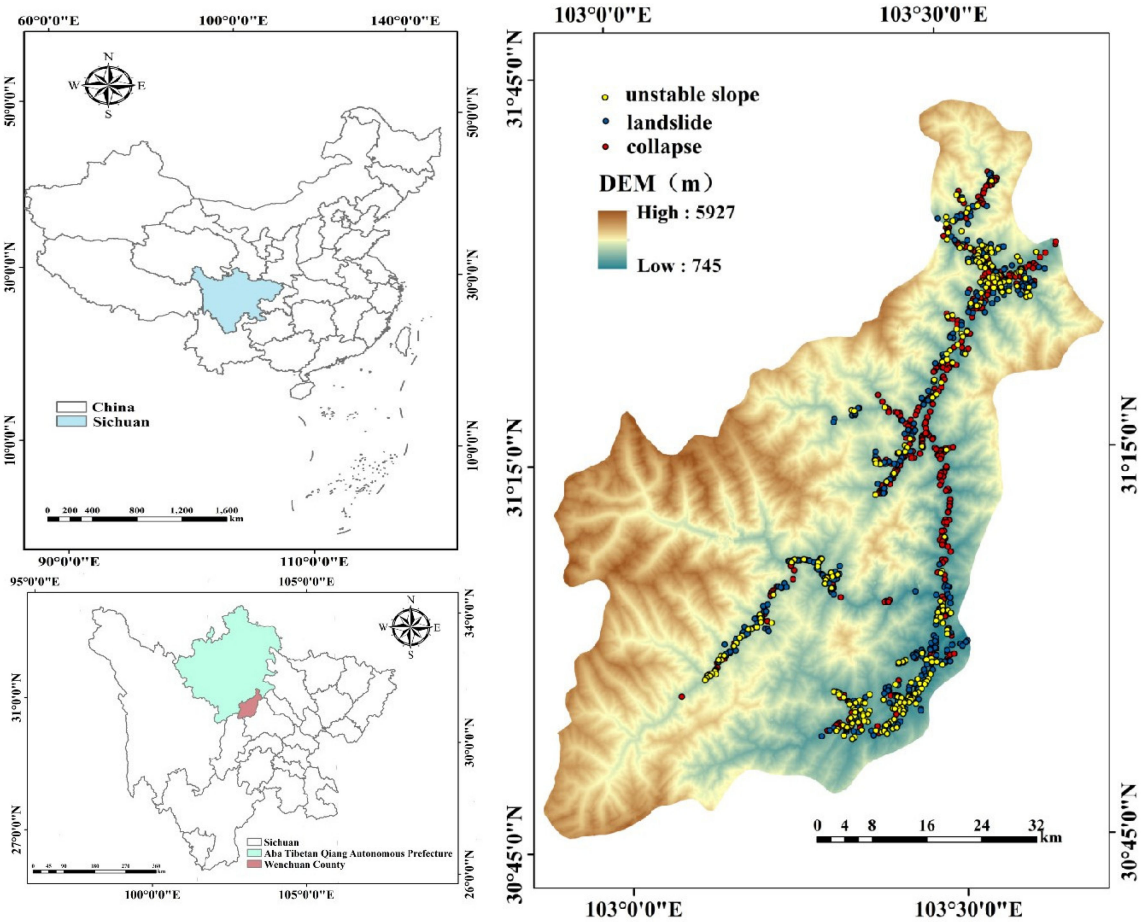


Figure 1. Geographical location of the research area and distribution of historical collapses and landslides.

2.2. Data Sources

The basic geographic data and environmental data used in this research are as follows: (1) the distribution data of geological disaster points are from the Sichuan Geological Survey and are mainly used to divide the training set and the validation set; (2) the data of digital elevation model (DEM) is ASTER GDEM 30m resolution digital elevation data, which is sourced from the NASA official website (https://search.asf.alaska.edu/#/, accessed on 20 March 2022) and used to obtain slope, aspect, curvature, terrain relief, and topographic wetness index; (3) the river data comes from the thematic map of the river system in China from 91 satellite map assistant software, and is used to obtain the distance to river; (4) the fault data is from the 1:500,000 geological map of 91 satellite map assistant software, and used to obtain the distance to fault; (5) the rainfall data refers to the spatial interpolation data set of annual rainfall in China since 1980, and is from the Resource and Environment Science and Data Center of Chinese Academy of Sciences (http://www.resdc.cn/, accessed on 18 March 2022); (6) the remote sensing image data is the Landsat 8 OLI image on 9 April 2018, which is from the geospatial data cloud network (http://www.gscloud.cn/, accessed on 1 March 2022) and is used to obtain the normalized difference vegetation index; (7) the lithology data comes from the Sichuan Geological Survey; (8) the soil data comes from the geospatial data cloud (http://www.gscloud.cn/, accessed on 1 March 2022); (9) data on land use comes from the geospatial data cloud (http://www.gscloud.cn/, accessed on 2 March 2022); (10) the seismic peak ground acceleration is from the United States Geological Survey (USGS) (https://earthquake.usgs.gov/, accessed on 2 March 2022). The details of the data sources are shown in Table 1.

Table 1. Data sources.

Data Type	Data Sources	Usage	Spatial Resolution
The distribution data of geological disaster points	Sichuan Geological Survey	Divide the training set and the validation set	Vector data
The digital elevation model (DEM)	NASA official website (https://search.asf.alaska.edu/#/)	Obtain slope, aspect, curvature, terrain relief, and topographic wetness index	30 m × 30 m
The river data	The thematic map of the river system in China from 91 satellite map assistant software	Obtain the distance to river	1:500,000
The fault data	Geological map from 91 satellite map assistant software	Obtain the distance to fault	1:500,000
The rainfall data	The Resource and Environment Science and Data Center of Chinese Academy of Sciences (http://www.resdc.cn/)	Obtain average annual rainfall	1000 m × 1000 m
The Landsat 8 OLI image on 9 April 2018	The geospatial data cloud network (http://www.gscloud.cn/)	Obtain the normalized difference vegetation index	30 m × 30 m
The lithology data	Sichuan Geological Survey	Obtain the lithology	30 m × 30 m
The soil data	The geospatial data cloud network (http://www.gscloud.cn/)	Obtain the soil	30 m × 30 m
Land use	Geospatial data cloud (http://www.gscloud.cn/)	Obtain the data of land use	30 m × 30 m
The seismic peak ground acceleration	The United States Geological Survey (USGS) (https://earthquake.usgs.gov/)	Obtain seismic peak ground acceleration	Vector data

2.3. Data Description of Environmental Factors

Collapse and landslide disaster is a natural phenomenon of earth activities on the earth's surface. The main environmental factors that induce such disasters include terrain, geology, land cover, ecology, hydrology, meteorology, earthquake, and human engineering activity. Therefore, the effective selection of environmental factors is the basis for establishing a susceptibility evaluation system for collapse and landslide disasters, which has a great impact on the reliability and accuracy of evaluation results. In combination with the field survey data and the occurrence of historical geological disasters in the research area, this paper selects six categories of environmental factors (13 environmental factors in total) including topography, geological structure, hydrology, seismic impact, ecology, and human activity as the susceptibility evaluation index of collapse and landslide disasters; the reasons for the selection of environmental factors are shown in Table 2. Grid data with a resolution of 30 m × 30 m and a projection of WGS1984 and UTM-Zone48 are converted in ArcGIS 10.5 software, as shown in Figure 2.

Table 2. Reasons for the selection of environmental factors.

Data Type	Factors	Reason for Selecting the Parameters
Topographic	Slope	Slope affects water flow direction and soil development, which is one of the important reasons for slope instability [20]. The more the slope increases, the more concentrated the shear stress in the slope is, and the greater the possibility of occurrence of collapse and landslide disasters will be [23].
	Aspect	The influence of aspect on collapse and landslide is the regular difference of microclimate and water heat ratio of hillside. The sunshine duration, solar radiation intensity, and daily temperature difference are different on slopes with different aspects [38].
	Curvature	Curvature is defined as the change rate of the slope and the shape of the earth's surface, which has a great impact on the transportation of collapse and landslide materials [39]. The greater the concave–convex degree of the slope is, the more unstable the slope is, and the more likely it is that collapse and landslide will occur [40]. Negative curvature, zero curvature, and positive curvature represent concave surfaces, plane surfaces, and convex surfaces, respectively.
	Terrain relief	The terrain relief reflects the difference between the highest point and the lowest point of altitude in a specific area, and controls the gravitational potential energy that can cause collapse and landslide disasters [41]. The greater the terrain relief is, the more fractured the terrain is, the higher the instability of the surface soil layer and slope is, and the more likely it is that collapse and landslide disasters will occur.
	Topographic wetness index (TWI)	The topographic wetness index refers to the influence of the scale and terrain of the saturated runoff zone on the region, and is used to quantify the control of terrain on hydrological processes. By comprehensively considering the impact of terrain and soil characteristics on soil moisture distribution, Beven and Kirkby proposed [42] the calculation formula $\boldsymbol{TWI} = \mathbf{ln}\frac{A_S}{tan\beta}$, where A_S represents the drainage area and β represents the slope angle.
Geological	Lithology	The rock-soil type and structural characteristics control the stress distribution, strength, and deformation and failure characteristics [43] of the rock-soil mass of the slope. Slopes with different lithology have different shear strength and stability, and also have different probability of occurrence of collapse and landslide disasters.
	Soil type	Different soil types have different shear strength and hydraulic conductivity, which affect the stability of slopes [31].
	Distance to fault	The rock mass is broken, the rock has poor erosion and weathering resistance, and the slope has poor stability near the fault zone [44].
Hydrological	Rainfall	Rainfall infiltration not only softens the rock-soil mass of the slope, but also increases the seepage pressure. The formed surface runoff will scour and erode the slope, resulting in the instability of the slope. The average annual rainfall affects the slope and its ecological environment, thus affecting the occurrence of collapse and landslide disasters [24].
	Distance to river	The softening, scouring, and erosion caused by river erosion have a serious impact on the stability of the slope. Slopes located in the coastal area of a river are eroded by the river and infiltrated by water, which leads to changes in internal stress and a greater probability of occurrence of collapses and landslides [45].
Seismic	Peak ground acceleration (PGA)	As an important dynamic factor to measure the impact of earthquakes on collapse and landslide, seismic peak ground acceleration reflects the overall vibration intensity of the earth's surface after an earthquake. The intense activity of the earth's surface reduces the stability of the rock-soil mass and increases the possibility of occurrence of collapse and landslide disasters [46].
Ecological	Normalized difference vegetation index (NDVI)	As an important index that can reflect the growth status and coverage of vegetation, NDVI can inhibit the occurrence of collapses and landslides to a certain extent [47]. The calculation formula is $\boldsymbol{NDVI} = \frac{IR-R}{IR+R}$, where $\boldsymbol{IR}$ represents the reflectance in near-infrared wavelength and $\boldsymbol{R}$ represents the reflectance in red light wavelength.
Human activity	Land use	The type of land use not only affects soil moisture and surface runoff, but also indirectly affects the development of landslides and collapses [48].

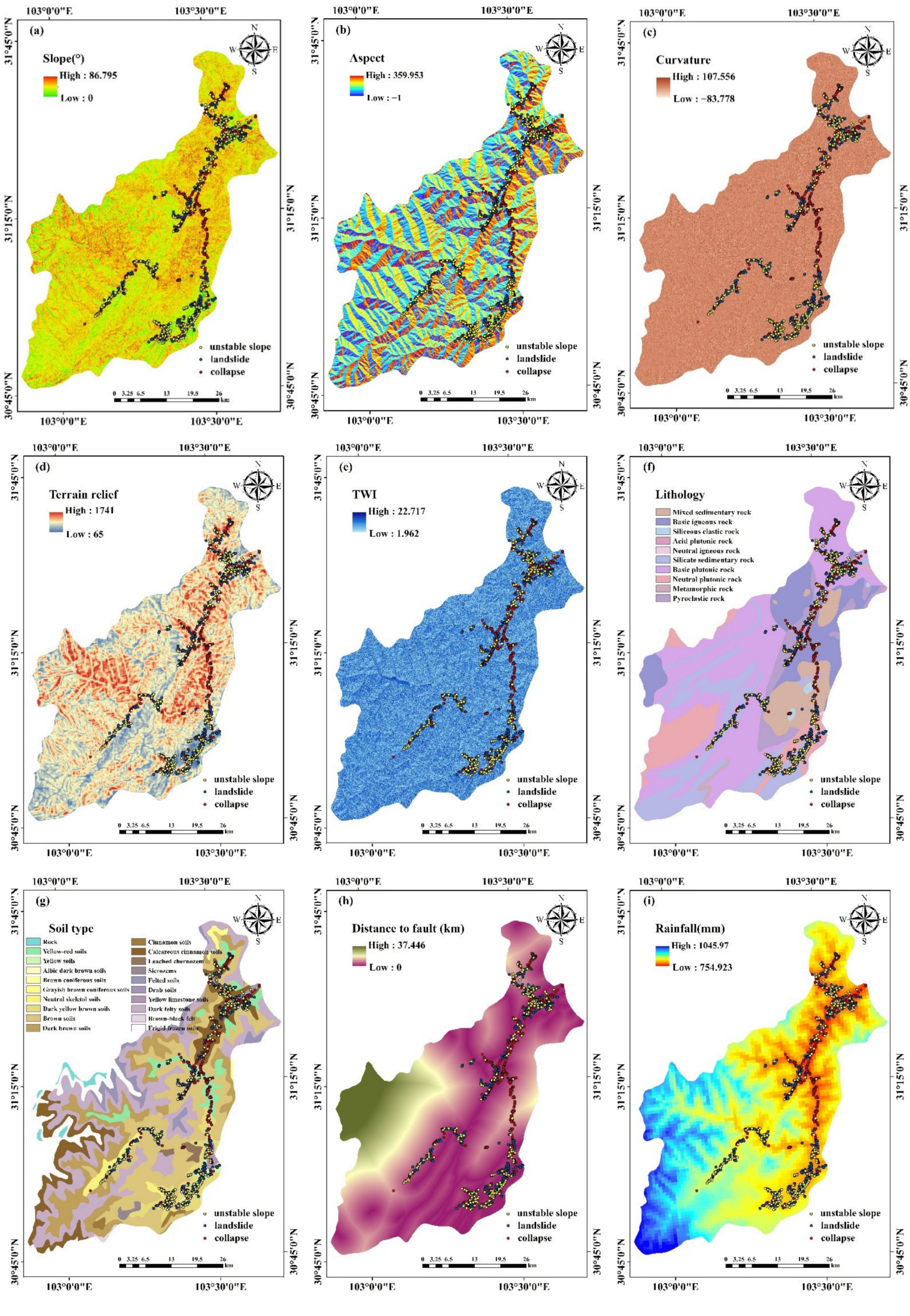


Figure 2. *Cont.*

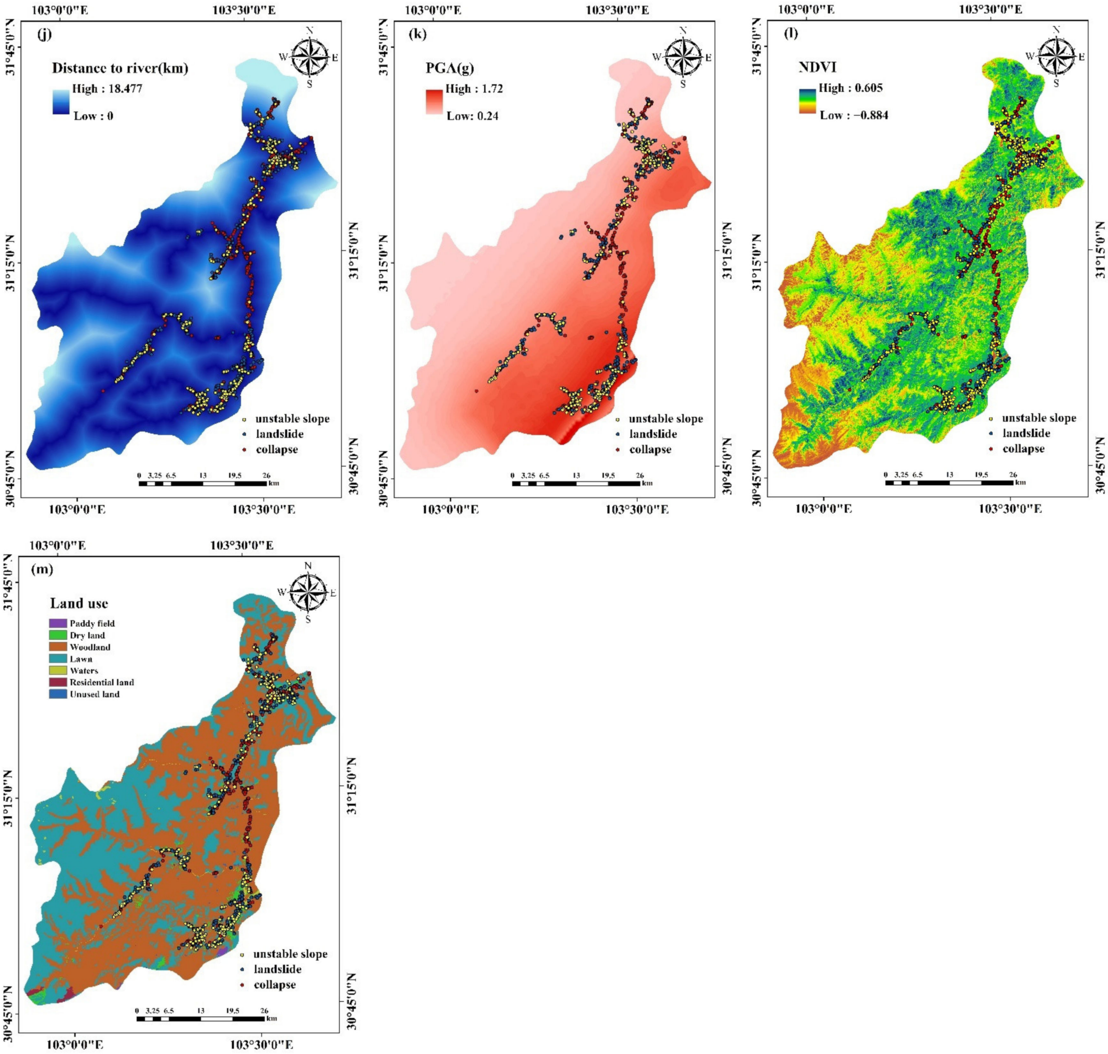


Figure 2. Schematic diagram of environmental factors: (**a**) slope, (**b**) aspect, (**c**) curvature, (**d**) terrain relief, (**e**) TWI, (**f**) lithology, (**g**) soil type, (**h**) distance to fault, (**i**) rainfall, (**j**) distance to river, (**k**) PGA, (**l**) NDVI, (**m**) land use.

2.4. Research Methods

2.4.1. Research Technical Routes

The idea of this research is to couple the certainty factor with three machine learning methods (namely logistic regression, support vector machine, and random forest) to build six collapse and landslide susceptibility prediction models, compare and analyze the performance of the single model and the coupling model to obtain the optimal model, and finally discuss the contribution of each environmental factor to the collapse and landslide susceptibility prediction of the optimal model. The main technical flow of this paper is shown in Figure 3 and includes the following steps:

Step 1 is to collect data related to collapse and landslide disasters in the research area, including data of historical collapse and landslide disaster points and environmental impact factors;

Step 2 is to carry out an independence test of environmental impact factors through Pearson correlation coefficient and multi-collinearity diagnostics;

Step 3 is to obtain the certainty factor value of each environmental factor by the certainty factor methods, and obtain the laws of the impact of each environmental factor on the development of collapse and landslide in its attribute intervals;

Step 4 is to obtain a training set and validation set by dividing historical collapse and landslide disaster points and randomly selecting non-collapse and landslide points at a ratio of 7:3, build six collapse and landslide susceptibility prediction models (LR, CF-LR, SVM, CF-SVM, RF, and CF-RF), and draw the collapse and landslide susceptibility mapping based on GIS;

Step 5 is to use the validation set to evaluate the models based on the confusion matrix, ROC curve, and AUC value, and compare and analyze the model performance to obtain the optimal model;

Step 6 is to discuss the importance of each environmental factor based on the optimal model, rank the contribution of environmental factors to the model, and obtain the important trigger factors of collapse and landslide disasters in the research area.

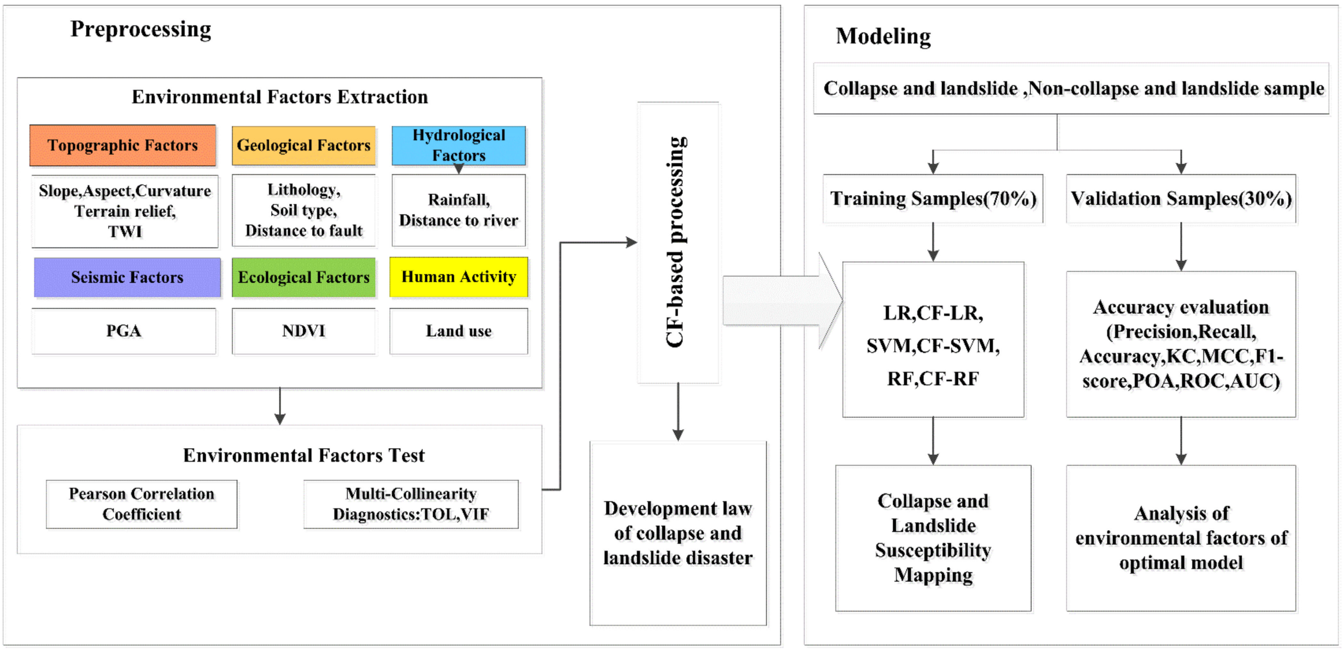


Figure 3. Technical flow.

2.4.2. Screening of Environmental Factors

The environmental factors that affect the occurrence of collapse and landslide disasters are diverse and complex, and each factor has a certain correlation. The high correlation between factors leads to complexity of the model. Therefore, it is very important to perform an independence test of each factor and eliminate factors with high correlation for the subsequent modeling [30]. For this reason, the Pearson correlation coefficient (PCC), variance inflation factor (VIF), and tolerance (TOL) are adopted for independence test in this research.

(1) Correlation analysis of the factors:

PCC can measure the similarity between collapse and landslide susceptibility environmental factors, with a value range of −1 to 1. The closer the absolute value is to 1, the more similar the samples are; the closer the absolute value is to 0, the less similar the samples are. When the correlation coefficient is within the range of 0.8–1, it indicates that the factors have extremely high correlation; 0.6–0.8 indicates high correlation, 0.4–0.6 indicates moderate correlation, 0.2–0.4 indicates weak correlation, and 0.0–0.2 indicates no correlation [49]. The calculation formula is:

$$PCC = \frac{\sum_{i=1}^{n}(x_i - \overline{x})\sum_{j=1}^{n}\left(y_j - \overline{y}\right)}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2\sum_{j=1}^{n}\left(y_j - \overline{y}\right)^2}} \tag{1}$$

where $\boldsymbol{PCC}$ represents the correlation coefficient between samples $\boldsymbol{x_i}$ and $\boldsymbol{y_j}$, $\boldsymbol{x_i}$ and $\boldsymbol{y_j}$ represent the variable values of $\boldsymbol{X_i}$ and $\boldsymbol{Y_j}$ respectively, $\overline{\boldsymbol{x}}$ and $\overline{\boldsymbol{y}}$ represent the average values of $\boldsymbol{X_i}$ and $\boldsymbol{Y_j}$, respectively.

(2) Multi-collinearity test

Multi-collinearity means that there is a high correlation between two or more predictive variables in a multiple regression model. Tolerance (TOL) and variance inflation factor (VIF) are commonly used in collinearity diagnostics. When the TOL value is less than 0.1 or the VIF value is greater than 10, it indicates that there is serious collinearity among the factors. When the TOL value is less than 0.2 or the VIF value is greater than 5, it indicates that there is strong collinearity [50] among the factors. The calculation formula is:

$$\boldsymbol{VIF_i} = \frac{\mathbf{1}}{\mathbf{1} - \boldsymbol{R_i^2}} = \frac{\mathbf{1}}{\boldsymbol{TOL}}(\boldsymbol{i} = \mathbf{1}, \mathbf{2}, \mathbf{3} \ldots\ldots \boldsymbol{k}) \tag{2}$$

where $\boldsymbol{R_i^2}$ represents the certainty factor between the ith independent variable X_i and other $k - 1$ independent variables.

2.4.3. Processing of the CF-Based Environmental Factors

The certainty factor (CF) is a piecewise probability function, which was first proposed by E.H. Shortliffe and B.G. Buchanan [51] and later improved by Heckerman [52]. It is an index used to analyze the susceptibility of various factors that can affect the occurrence of collapses and landslides, and its calculation formula is shown in Formula (3). It can also establish the quantitative relationship between landslide activities and control factors. At present, the certainty factor model has been used in many studies for evaluating the susceptibility of regional collapse and landslide disasters [53–56]. After the CF-coupled machine learning model is used for collapse and landslide susceptibility modeling, this paper divides the selected basic environmental factors into eight attribute intervals (of which lithology, soil type, and land use type are divided according to natural attributes) by natural discontinuity method, and obtains the certainty factor values of each environmental factor in the attribute intervals. The value of the certainty factor reflects the probability of occurrence of collapse and landslide disasters for environmental factors in this attribute interval.

$$\boldsymbol{CF_\alpha^i} = \begin{cases} \frac{\boldsymbol{PP_\alpha^i} - \boldsymbol{PP_s}}{\boldsymbol{PP_\alpha^i}(\mathbf{1} - \boldsymbol{PP_s})}, \boldsymbol{PP_\alpha^i} < \boldsymbol{PP_s} \\ \frac{\boldsymbol{PP_\alpha^i} - \boldsymbol{PP_s}}{\boldsymbol{PP_s}\left(\mathbf{1} - \boldsymbol{PP_\alpha^i}\right)}, \boldsymbol{PP_\alpha^i} \geq \boldsymbol{PP_s} \end{cases} \tag{3}$$

where $\boldsymbol{CF_\alpha^i}$ is the certainty factor of the influence factor i at the jth level; $\boldsymbol{PP_\alpha^i}$ is the conditional probability of occurrence of collapse and landslide disaster of the influence factor i at the jth level. The number of collapse and landslide disaster points of the influence factor i at the jth level is used to replace the ratio of the number of grids of the influence factor i at the jth level in the research area. $\boldsymbol{PP_s}$ is the prior probability of occurrence of collapse and landslide disasters in the research area. The CF has a value range of −1 to 1. If it is greater than 0, it means that the probability of occurrence of collapse and landslide disaster is high in this factor interval; if it is less than 0, it means that the probability of occurrence of collapse and landslide disaster is low in this factor interval; if it is equal to 0, it means that the probability of occurrence of collapse and landslide disaster is uncertain in this factor interval.

2.4.4. Machine Learning Model

(1) Logistic regression

A logistic regression (LR) model is a regression model that describes binary-classification dependent variables and a series of independent variables [32]. In collapse and landslide disaster susceptibility modeling, LR model is used to find the optimal fitting function to describe the relationship between the occurrence of collapses and landslides and a group of independent indexes such as slope, lithology, etc. The independent variables in the model are the influence factors for the occurrence of collapses and landslides, while the binary-classification dependent variables represent the occurrence (represented as 1 in the model) or non-occurrence (represented as 0 in the model) of collapses and landslides. LR represents the relationship between the occurrence probability of collapses and landslides and the independent variables, as shown in Formula (4).

$$\boldsymbol{P} = \frac{\boldsymbol{e^Y}}{\boldsymbol{1 + e^Y}} \tag{4}$$

where $\boldsymbol{P}$ represents the occurrence probability of collapses and landslides. $\boldsymbol{Y}$ represents the fitting function of multiple factors, which is expressed in the following Formula (5).

$$\boldsymbol{Y = B + A_1X_1 + A_2X_2 + \ldots + A_nX_n} \tag{5}$$

where $\boldsymbol{B}$ represents the constant term obtained by logistic regression, $\boldsymbol{A_i}$ represents the logistic regression coefficient of each independent variable, and $\boldsymbol{X_i}$ represents the influence factors for the occurrence of collapses and landslides.

(2) Support vector machine

As a supervised learning method based on the principle of mathematical statistics and structural risk minimization method, SVM is used for classification and regression, and was first proposed by Vapnik [57]. Its working principle is to maximize the distance between the nearest sample points on both sides by constructing an optimal separating hyperplane. It has the advantages of high accuracy, strong popularization, and good generalization ability in processing high-dimensional invisible data [58]. In this research, the training set is set as $\boldsymbol{T = \{(x_i, y_i)\}_{i=1}^{M}}$; $\boldsymbol{x}$ is the input vectors, including slope, aspect, curvature, terrain relief, TWI, lithology, soil type, distance to fault, rainfall, distance to river, PGA, NDVI, and land use. In the formula $\boldsymbol{y_i \in \{0, 1\}}$, 1 and 0 represent collapse and non-collapse, respectively. SVM classification aims to find an optimal separating hyperplane that can distinguish between collapse and non-collapse from the above training set. The prediction accuracy of support vector machine depends on the choice of kernel function. There are four commonly used kernel functions, that is, linear, polynomial, radial basis function (RBF), and sigmoid, of which RBF is widely used in the susceptibility prediction of collapses and landslides due to its advantages of few parameters, strong flexibility, and good performance. Therefore, the RBF kernel function is used in this research to build the SVM model, as shown in Formula (6). For the RBF kernel function, the regularization parameter $\boldsymbol{\vartheta}$ and the gamma parameter γ are parameters that need to be determined. The greater the regularization parameter is, the less error is allowed, because once error occurs, it is easy to have overfitting, otherwise, it is easy to have under-fitting. The gamma parameter γ controls the degree of nonlinearity of the model.

$$\boldsymbol{k(x_i, x_j) = \exp\left(-\gamma \parallel x_i - x_j \parallel^2\right)} \tag{6}$$

where $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ are the input vectors and γ is the gamma parameter.

(3) Random forest

As an integrated classification and regression model composed of multiple decision trees, RF obtains the optimal classification results according to the voting results of each decision tree. It was first proposed by Breiman [59] and its working process is divided into three steps. The first step is to draw K samples from the original training set in a manner of sampling with replacement by bootstrap sampling, and the characteristic number of each sample is the same as that of the original training set. The second step is to build decision tree models for each K sample, randomly select the mtry characteristics at each node of the decision tree as the splitting characteristics, and calculate the optimal node partition according to Gini standard (Formula (7)) to generate child nodes. Then, the whole K tree forms a random forest model. The third step is to vote according to the K classification results to determine the final classification. The randomness of the RF model is reflected in the randomness of the training set and the optimal attribute of node splitting, which can avoid model over-fitting and enhance its stability. The main characteristic of random forest is that it can provide the Gini index of the corresponding input variables, that is, the importance ranking of each input variable.

$$Gini = 1 - \sum_{i=1}^{2} p_i^2 \tag{7}$$

where p_i represents the probability that the observed sample falls in category i.

2.4.5. Performance Evaluation of the Models

(1) Confusion matrix

Model validation and performance evaluation are important steps in the process of collapse and landslide susceptibility evaluation. A confusion matrix is often used for performance evaluation of the binary-classification models. A confusion matrix includes the following four parameters (see Table 3). True positive (TP) is the number of collapse and landslide points predicted by the model that are actual collapse and landslide points. False negative (FN) is the number of non-collapse and non-landslide points predicted by the model that are actually collapse and landslide points. False positive (FP) is the number of collapse and landslide points predicted by the model that are actually non-collapse and non-landslide points. True negative (TN) is the number of non-collapse and non-landslide points predicted by the model that are actually non-collapse and non-landslide points. On this basis, the performance evaluation of each model is carried out in this research with 7 statistical indexes, including precision, recall, accuracy, kappa coefficient (KC), MCC, F1-score, and performance overall (POA) [60]. The description of each index is as shown in Table 4.

Table 3. Confusion matrix of prediction results.

Prediction Situation	**Actual Situation**	
	Positive Sample	**Negative Sample**
Positive sample	True positive (TP)	False positive (FP)
Negative sample	False negative (FN)	True negative (TN)

Table 4. Description of characteristics of statistical indexes based on confusion matrix.

Index	**Statistical Definition**	**Usage**
Precision	$\frac{TP}{TP+FP}$	Evaluating the proportion of the TP sample in all predicted positive samples
Recall	$\frac{TP}{TP+FN}$	Quantifying the proportion of the TP sample in all true positive samples
Accuracy	$\frac{TP+TN}{TP+FP+FN+TN}$	Quantifying the proportion of correctly predicted samples
KC	$\frac{P_0-P_e}{1-P_e}$ $P_0 = \frac{TP+TN}{TP+FN+FP+TN}$ $P_e = \frac{(TP+FN)(TP+FP)+(TN+FN)(FP+TN)}{(TP+FN+FP+TN)^2}$	Checking consistency and measuring classification precision
MCC	$\frac{TP\times TN-FP\times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$	Describing the correlation coefficient between the actual classification and the predicted classification, with a value range of -1 to 1. When the value is 1, it indicates the perfect prediction of the receiver; when the value is 0, it indicates that the predicted result is not as good as the randomly predicted result; when the value is -1, it indicates that the predicted classification is completely inconsistent with the actual classification.
F1-score	$\frac{2\times Precision\times Recall}{Precision+Recall}$	Representing the harmonic mean of accuracy and recall, with a value range of -1 to 1.
POA	$Accuracy + MCC + F1 - score$	Representing the sum of the accuracy, the Matthews correlation coefficient and the harmonic mean; the comprehensive performance index can quantify the overall performance of the model.

(2) ROC curve and AUC value

The receiver operating characteristic (ROC) curve is a useful technology to verify the performance of the probability model and is also a common method to verify collapse and landslide susceptibility models [21]. In the ROC curve, the specificity is taken as the abscissa (that is, the percentage of the number of collapse and landslide points predicted by the model that are actually non-collapse and non-landslide points to the number of all non-collapse and non-landslide points actually determined by the model), the susceptibility is taken as the ordinate, and the integral (namely the area enclosed by the curve and the x-axis) of the curve in the value range of 0 to 1 is the AUC value (area under curve). The closer the ROC curve is to the upper left corner, the greater the AUC value is, which indicates that the occurrence of collapse and landslide disasters will be predicted more successfully and the accuracy of the model will be higher.

3. Experimental Results

3.1. Independent Test of Environmental Factors

3.1.1. Correlation Analysis of the Factors

The correlation coefficient matrix among 13 environmental factors is obtained by using the "Band Collection Statistics" tool of the ArcGIS toolbox, and R4.1.3 is used for visualization to obtain Figure 4. In the figure, red indicates positive correlation and blue indicates negative correlation. The redder the color is, the stronger the correlation between the two factors. The size of the circle directly reflects the magnitude of correlation coefficients. From the figure, it can be observed that the correlation coefficient of all factors is less than 0.6, and the correlation degree is weak, so the degree of interaction between all factors is small.

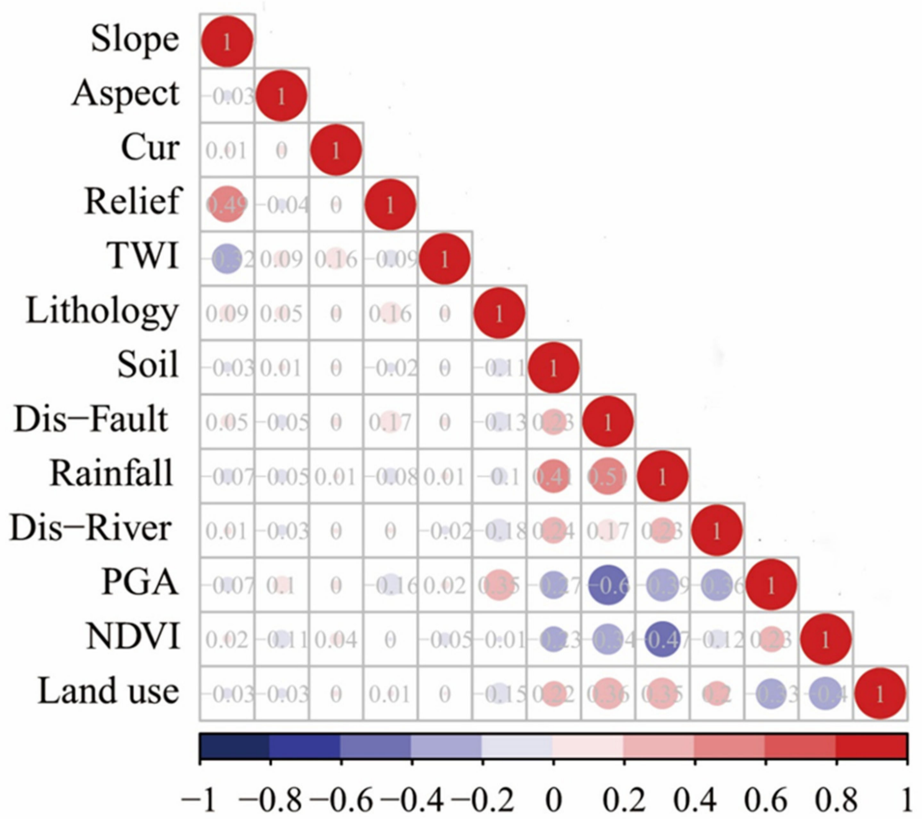


Figure 4. Pearson correlation values between factors.

3.1.2. Multi-Collinearity Test

The TOL and VIF values of each collapse and the landslide susceptibility evaluation factor are obtained through collinearity analysis by SPSS software, as shown in Table 5. The tolerance (TOL) of all factors is above 0.4, much higher than the threshold value of 0.1; the

VIF value is less than 2.2, significantly lower than the threshold value of 5 or 10, indicating that there is no multi-collinearity among the selected factors and verifying the rationality of the evaluation index again.

Table 5. Collinearity diagnostic results of influence factors.

Factors	TOL	VIF
Slope	0.630	1.586
Aspect	0.951	1.052
Curvature	0.952	1.051
Terrain relief	0.606	1.651
TWI	0.829	1.206
Lithology	0.824	1.214
Soil type	0.753	1.328
Distance to fault	0.494	2.026
Rainfall	0.554	1.805
Distance to river	0.691	1.448
PGA	0.459	2.179
NDVI	0.783	1.278
Land use	0.833	1.201

3.2. Attribute Interval Classification and Certainty Coefficient Calculation of Environmental Factors

The ratio of the total number of collapse and landslide disaster points (1081) to the total number of grids (4,565,153) in the research area is used as a replacement in this research. The attribute interval classification and certainty coefficient value of environmental factors are shown in Table 6. The certainty coefficient values of each environmental factor in its attribute intervals are visualized by Origin, as shown in Figure 5. The law of the occurrence possibility of collapse and landslide disasters in the research area based on various factors is as follows. (1) Topography factor: when the slope is 0–10°, the probability of occurrence of collapses and landslides is the highest, and the CF value is 0.739; the southeast and northwest aspects are most favorable for the occurrence of collapses and landslides. When the curvature is −2–1, the probability of occurrence of collapses and landslides is the highest; with the increase of terrain relief, the probability of occurrence of collapses and landslides decreases gradually. When the terrain relief is 65–380, the probability of occurrence of collapses and landslides is the highest and the CF value is 0.764. When the TWI is 10.9–13.2, it is not conducive to the development of collapses and landslides, while when the TWI is 14.5–22.8, the probability of occurrence of collapses and landslides is the highest and the CF value is greater than 0.8. (2) Geological structure factor: the areas with neutral igneous rock and basic plutonic rock lithology are most conducive to the occurrence of collapses and landslides and the CF value is greater than 0.5. The areas with soil types of yellow-red soils, yellow soils, neutral skeletal soils, dark yellow brown soils, calcareous cinnamon soils, drag soils, and yellow limestone soils are conducive to the occurrence of collapses and landslides and the CF value is greater than 0.5. In the areas with soil type of drag soils, the probability of occurrence of collapses and landslides is the highest and the CF value is 0.931. With the increase of distance to fault, the probability of occurrence of collapses and landslides decreases gradually. When the distance to fault is less than 2 km, the probability of occurrence of collapses and landslides is the highest and the CF value is greater than 0.5. (3) Hydrological factor: when the rainfall is 750–820 mm, the probability of occurrence of collapses and landslides is higher and the CF value is greater than or equal to 0.5. With the increase of distance to river, the probability of occurrence of collapses and landslides gradually decreases. When the distance to river is less than 1 km, the probability of occurrence of collapses and landslides is higher and the CF value is greater than or equal to 0.7. (4) Seismic factor: the probability of occurrence of collapses and landslides is the highest when the PGA is 1.5–1.8, and the CF value is 0.813. (5) Ecological factor: with the increase of NDVI, the probability of occurrence of collapses and landslides will increase first and then decrease. The probability of occurrence of collapses and landslides is higher in areas with NDVI value of −0.04–0.14, and the CF value is greater than 0.5. (6) Human activity factor: the areas with land use types of paddy field, dry land, waters, and residential land are conducive to the occurrence of collapses and landslides and the CF value is greater than or equal to 0.8.

Table 6. Attribute interval classification and certainty factor value of environmental factors.

Factors	Classes	Number of Collapse and Landslide Points	Number of Grids in the Interval Area	PPa ($\times 10^4$)	PPs ($\times 10^4$)	CF
Slope	0–10	99	108,985	9.084	2.368	0.739
	10–20	203	407,289	4.984	2.368	0.525
	20–30	293	1,087,132	2.695	2.368	0.121
	30–35	145	851,185	1.704	2.368	−0.281
	35–40	132	859,690	1.535	2.368	−0.352
	40–50	169	977,535	1.729	2.368	−0.27
	50–60	36	236,575	1.522	2.368	−0.357
	60–90	4	36,765	1.088	2.368	−0.541
Aspect	Flat (−1)	0	207	0	2.368	−1
	North (0–22.5, 337.5–360)	85	551,252	1.542	2.368	−0.349
	Northeast (22.5–67.5)	112	509,589	2.198	2.368	−0.072
	East (67.5–112.5)	177	643,946	2.749	2.368	0.139
	Southeast (112.5–157.5)	209	665,656	3.140	2.368	0.246
	South (157.5–202.5)	97	552,841	1.755	2.368	−0.259
	Southwest (202.5–247.5)	105	559,647	1.876	2.368	−0.208
	West (247.5–292.5)	120	521,380	2.302	2.368	−0.028
	Northwest (292.5–337.5)	176	560,635	3.139	2.368	0.246
Curvature	−84–−5	6	42,253	1.420	2.368	−0.4
	−5–−2	86	457,015	1.882	2.368	−0.205
	−2–−1	207	696,140	2.974	2.368	0.204
	−1–0	340	1,177,275	2.888	2.368	0.18
	0–1	292	1,106,230	2.640	2.368	0.103
	1–3	130	895,326	1.452	2.368	−0.387
	3–6	15	169,212	0.886	2.368	−0.626
	6–108	5	21,702	2.304	2.368	−0.027
Terrain relief	65–380	320	319,281	10.023	2.368	0.764
	380–490	344	786,325	4.375	2.368	0.459
	490–585	199	1,008,688	1.973	2.368	−0.167
	585–670	120	945,998	1.269	2.368	−0.464
	670–770	64	823,541	0.777	2.368	−0.672
	770–895	29	496,861	0.584	2.368	−0.754
	895–1280	5	183,238	0.273	2.368	−0.885
	1280–1745	0	1221	0	2.368	−1
TWI	1.9–5	255	1,234,257	2.066	2.368	−0.128
	5–7.8	357	1,497,445	2.384	2.368	0.007
	7.8–10.9	317	1,227,863	2.582	2.368	0.083
	10.9–13.2	70	470,993	1.486	2.368	−0.372
	13.2–14.5	33	107,219	3.078	2.368	0.231
	14.5–16.2	18	12,781	14.083	2.368	0.832
	16.2–18.6	20	10,820	18.484	2.368	0.872
	18.6–22.8	11	3775	29.139	2.368	0.919

Table 6. *Cont.*

Factors	Classes	Number of Collapse and Landslide Points	Number of Grids in the Interval Area	PPa ($\times 10^4$)	PPs ($\times 10^4$)	CF
Lithology	Mixed sedimentary rock	373	1,842,775	2.024	2.368	−0.145
	Basic igneous rock	0	23,850	0	2.368	−1
	Siliceous clastic rock	169	515,300	3.280	2.368	0.278
	Acid plutonic rock	99	641,152	1.544	2.368	−0.348
	Neutral igneous rock	176	300,428	5.858	2.368	0.569
	Silicate sedimentary rock	205	817,660	2.507	2.368	0.056
	Basic plutonic rock	14	22,126	6.327	2.368	0.626
	Neutral plutonic rock	0	3137	0	2.368	−1
	Metamorphic rock	44	395,842	1.112	2.368	−0.531
	Pyroclastic rock	1	2883	3.469	2.368	0.317
Soil type	Rock	0	40,521	0	2.368	−1
	Yellow-red soils	117	217,479	5.380	2.368	0.56
	Yellow soils	127	57,096	22.243	2.368	0.894
	Albic dark brown soils	0	7	0	2.368	−1
	Brown coniferous soils	0	6806	0	2.368	−1
	Grayish brown coniferous soils	0	25,873	0	2.368	−1
	Neutral skeletal soils	56	84,700	6.612	2.368	0.642
	Dark yellow brown soils	171	225,969	7.567	2.368	0.687
	Brown soils	205	1,139,431	1.799	2.368	−0.24
	Dark brown soils	0	657,341	0	2.368	−1
	Cinnamon soils	0	26,458	0	2.368	−1
	Calcareous cinnamon soils	252	504,598	4.994	2.368	0.526
	Leached chernozem	1	35,339	0.283	2.368	−0.881
	Sierozems	0	20,783	0	2.368	−1
	Felted soils	0	153,662	0	2.368	−1
	Drab soils	134	39,080	34.289	2.368	0.931
	Yellow limestone soils	18	22,031	8.170	2.368	0.71
	Dark felty soils	0	1,056,976	0	2.368	−1
	Brown-black felt	0	8942	0	2.368	−1
	Frigid frozen soils	0	242,061	0	2.368	−1
Distance to fault	0–2	751	1,511,153	4.970	2.368	0.524
	2–5	237	1,050,165	2.257	2.368	−0.047
	5–8	75	742,206	1.011	2.368	−0.573
	8–13	18	519,422	0.347	2.368	−0.854
	13–17	0	227,833	0	2.368	−1
	17–23	0	218,938	0	2.368	−1
	23–29	0	174,581	0	2.368	−1
	29–38	0	120,855	0	2.368	−1
Rainfall	750–790	535	353,745	15.124	2.368	0.844
	790–820	343	731,798	4.687	2.368	0.495
	820–845	149	821,064	1.815	2.368	−0.234
	845–870	53	834,612	0.635	2.368	−0.732
	870–900	1	641,016	0.016	2.368	−0.993
	900–930	0	527,244	0	2.368	−1
	930–970	0	461,923	0	2.368	−1
	970–1050	0	193,751	0	2.368	−1

Table 6. *Cont.*

Factors	Classes	Number of Collapse and Landslide Points	Number of Grids in the Interval Area	PPa ($\times 10^4$)	PPs ($\times 10^4$)	CF
Distance to river	0–1	756	854,540	8.847	2.368	0.733
	1–2	167	767,860	2.175	2.368	−0.082
	2–4	77	1,197,286	0.643	2.368	−0.728
	4–6	46	718,322	0.640	2.368	−0.73
	6–8	19	478,968	0.397	2.368	−0.833
	8–10	10	299,504	0.334	2.368	−0.859
	10–13	6	162,894	0.368	2.368	−0.844
	13–19	0	85,779	0.000	2.368	−1
PGA	0.2–0.4	71	1,252,569	0.567	2.368	−0.761
	0.4–0.6	72	830,025	0.867	2.368	−0.634
	0.6–0.8	322	584,544	5.509	2.368	0.570
	0.8–1	306	816,726	3.747	2.368	0.368
	1–1.1	36	342,609	1.051	2.368	−0.556
	1.1–1.3	57	426,004	1.338	2.368	−0.435
	1.3–1.5	171	276,284	6.189	2.368	0.618
	1.5–1.8	46	36,392	12.640	2.368	0.813
NDVI	−0.89–0.33	1	12,760	0.784	2.368	−0.669
	−0.33–0.16	6	313,938	0.191	2.368	−0.919
	−0.16–0.04	54	351,964	1.534	2.368	−0.352
	−0.04–0.05	222	602,824	3.683	2.368	0.357
	0.05–0.14	421	1,046,104	4.024	2.368	0.412
	0.14–0.23	209	853,744	2.448	2.368	0.033
	0.23–0.34	140	868,752	1.612	2.368	−0.319
	0.34–0.61	28	515,067	0.544	2.368	−0.770
Land use	Paddy field	11	9514	11.562	2.368	0.795
	Dry land	240	125,190	19.171	2.368	0.877
	Woodland	387	2,639,503	1.466	2.368	−0.381
	Lawn	367	1,736,151	2.114	2.368	−0.107
	Waters	50	30,997	16.131	2.368	0.853
	Residential land	24	19,453	12.337	2.368	0.808
	Unused land	2	4345	4.603	2.368	0.486

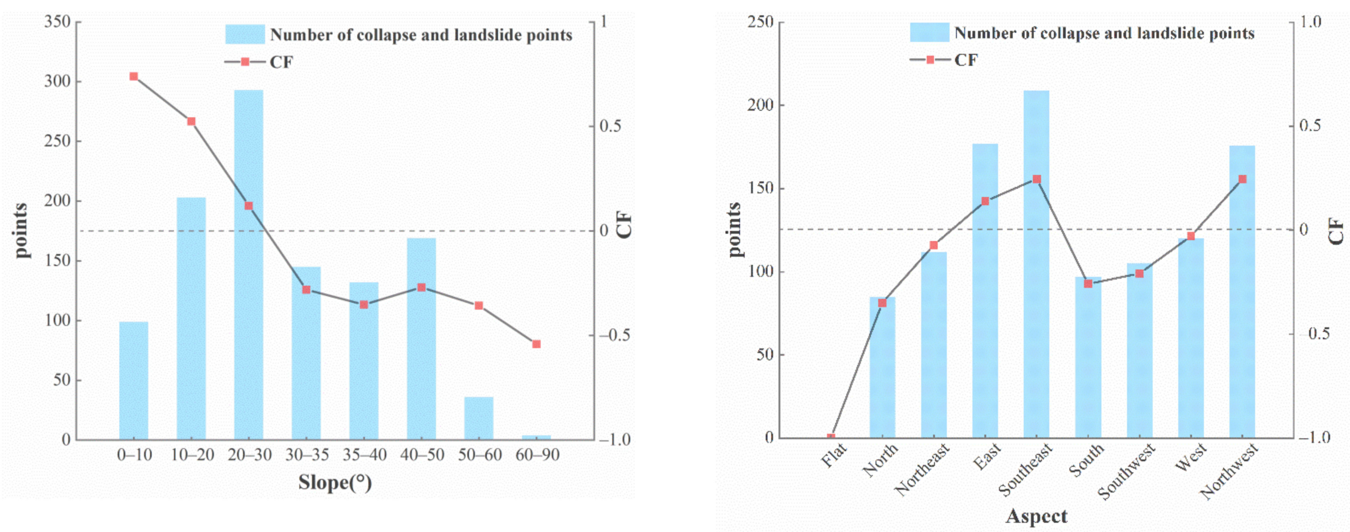


Figure 5. *Cont.*

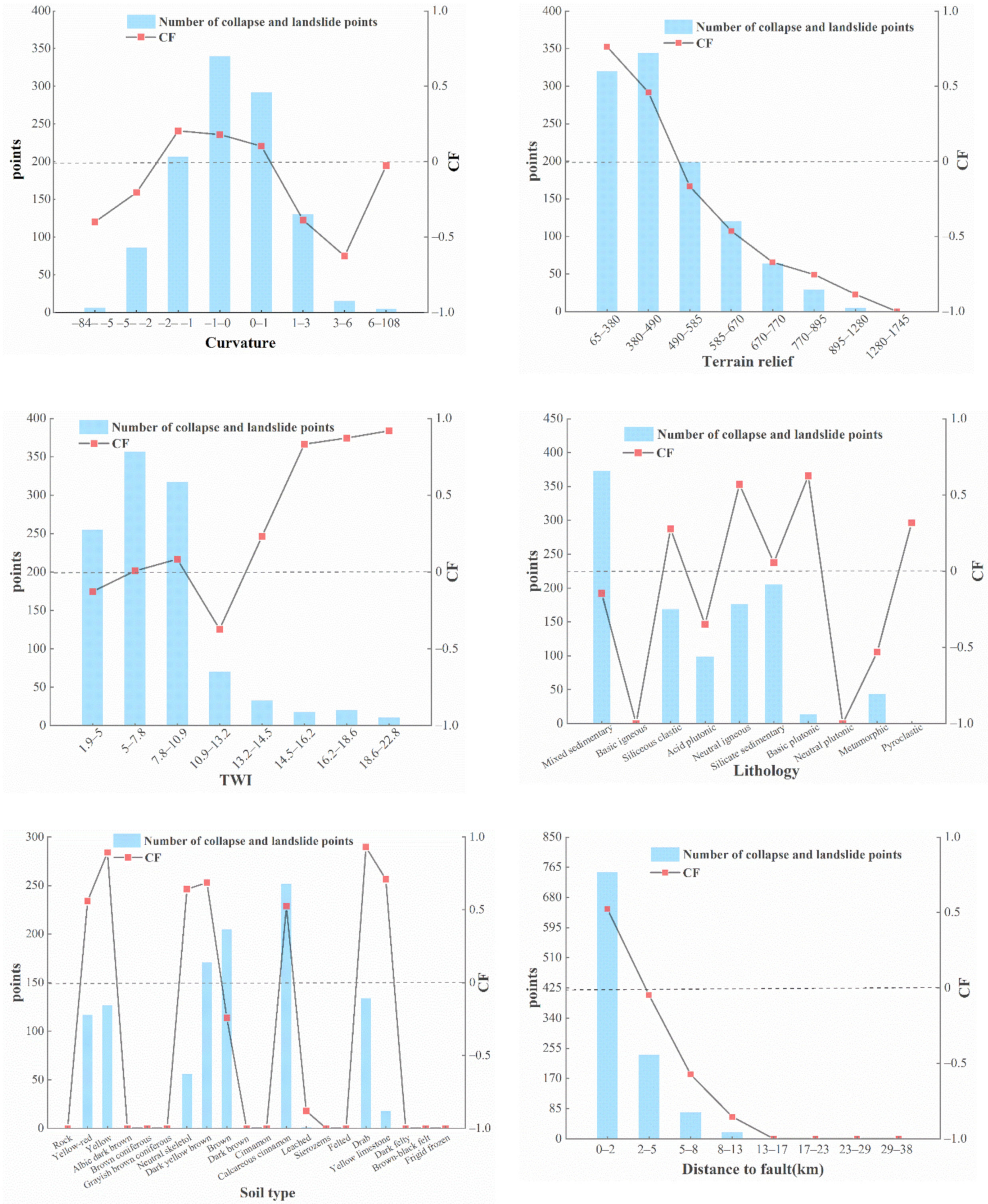


Figure 5. *Cont.*

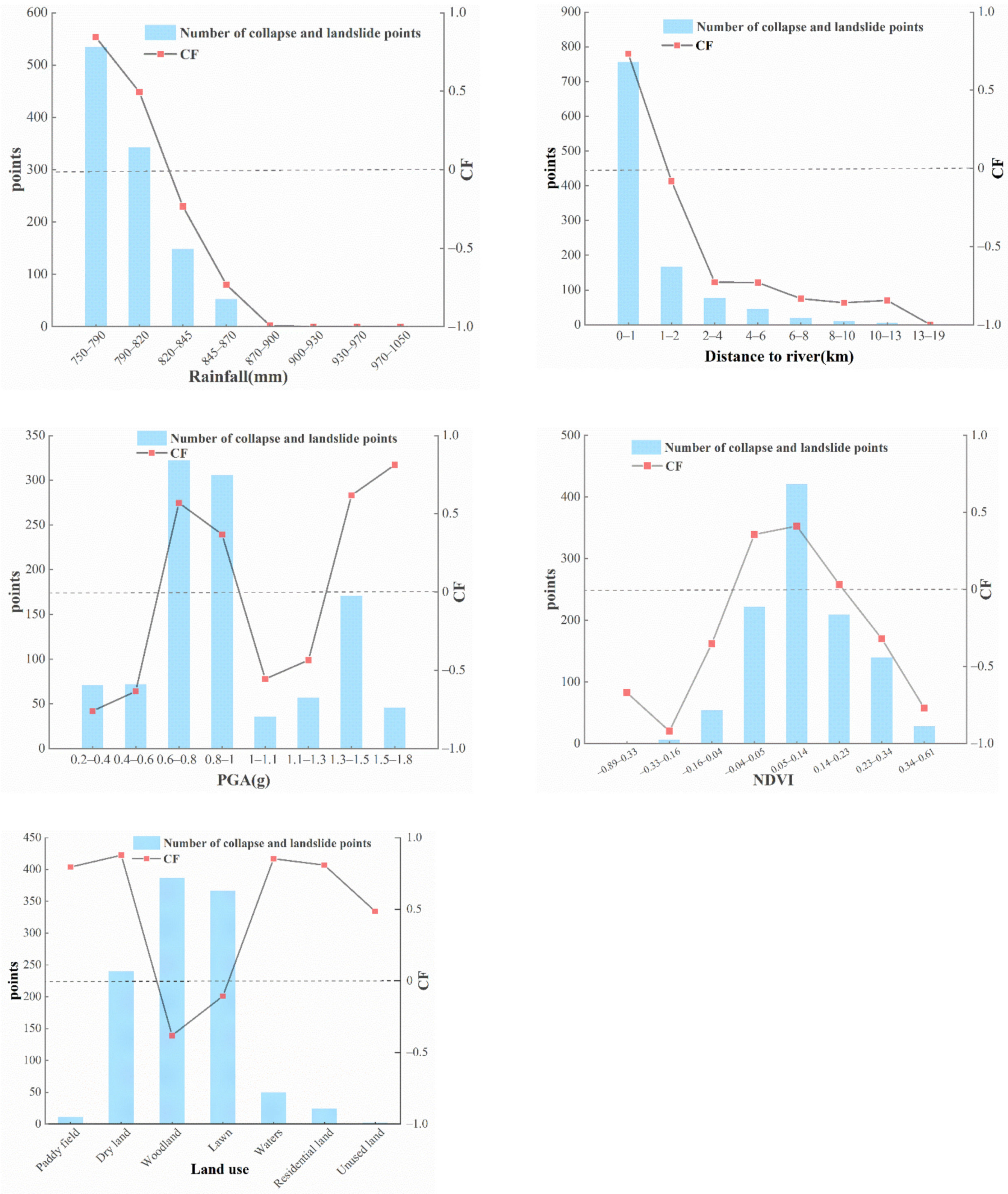


Figure 5. Distribution of the number of collapse and landslide points and the CF value in the attribute interval of environmental factors.

3.3. Modeling Results

3.3.1. LR and Coupling Models

In this research, the original value and certainty factor value (CF) of each environmental factor of the training sample were input into SPSS 25 software for binary logistic regression calculation to obtain the regression coefficient and constant term of each environmental factor, as shown in Table 7. The R^2 of the CF-LR model is 0.760, and the R^2 of the LR model is 0.672, so the fitting degree of the CF-LR model is better than that of single the LR model. Then, the obtained regression coefficient and constant term are put into Formula (4) and calculated by using the grid calculator of ArcGIS 10.5 to predict the probability of occurrence of collapses and landslides for each grid unit.

Table 7. Coefficients and constant terms for LR and CF-LR.

Environmental Factor	LR	CF-LR
Slope	0.113	−0.209
Aspect	0	1.482
Curvature	0.367	0.286
Terrain relief	0.531	1.408
TWI	0.167	1.11
Lithology	0.104	0.585
Soil type	0.269	1.505
Distance to fault	0	1.171
Rainfall	0.460	1.661
Distance to river	1.209	0.928
PGA	0.482	0.277
NDVI	0.669	0.994
Land use	0.140	0.634
Constant	−6.188	−0.416

3.3.2. SVM and Coupling Models

In this research, the training set and test set data were imported into R4.1.3, the "e1071" and "caret" packages were installed, and the tune function was called to adjust the regularization parameter ϑ and the gamma parameter value of the RBF kernel function by grid search method and five-fold cross-validation method. The regularization parameter ϑ of the single SVM model is 20 and the gamma parameter is 0.1, while the regularization parameter ϑ of the CF-SVM model is 0.4, and the gamma parameter is 0.1. The regularization parameter ϑ value of the SVM model is higher than that of the CF-SVM model, indicating that the error tolerance of the single SVM model is less than that of the CF-SVM model and that the CF-SVM model has better generalization capacity. Then, the trained models were used for collapse and landslide susceptibility prediction of 4,565,153 point objects in the whole research area, and the susceptibility values of all points were imported into ArcGIS 10.5 and converted to 30×30 m grid units.

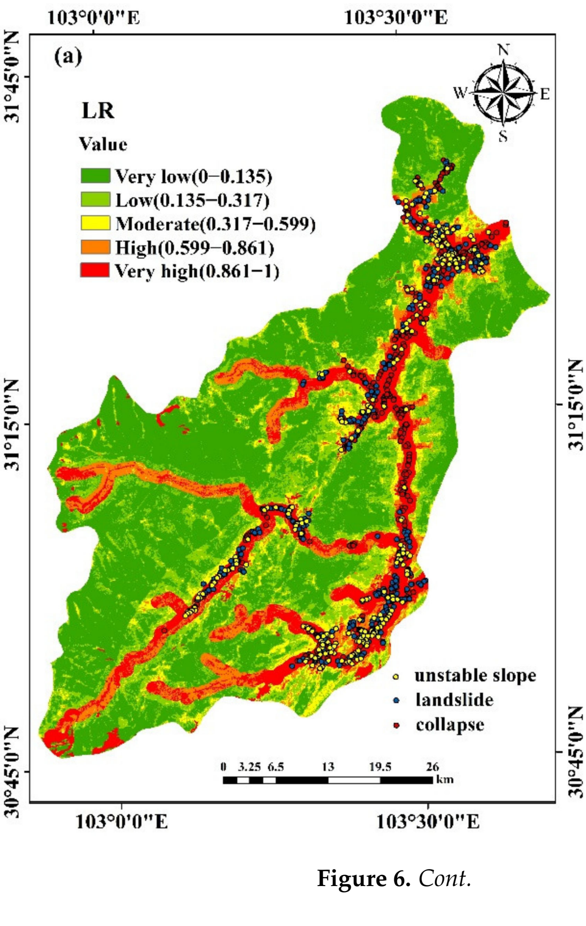
3.3.3. RF and Coupling Models

In this research, the training set and test set data were imported into R4.1.3 and the "randomForest" and "caret" packages were installed. The two parameters ntree and mtry in random forest have an important impact on the accuracy of the model. Ntree is the number of decision trees. The prediction performance of the RF model increases with the increase of ntree, but the amount of calculations in the model gradually increases and the modeling time becomes longer and longer. Research has showed that when ntree increases to 300 [61], the prediction performance of the RF is stable. Therefore, ntree was selected as 300 to establish the RF model for collapse and landslide disaster prediction in this research. Mtry is the number of characteristic nodes of each tree. When Mtry is small, the correlation between decision trees decreases and the classifier fitting is poor. When Mtry is large, the running speed of the model will slow down. On the basis that the value of ntree is equal to 300, this research takes the highest accuracy as the standard and ivtaubs the optimal Mtry value by grid search method and five-fold cross-validation method. The Mtry value of both RF model and CF-SVM model is 5. Then, the trained models were used for collapse and landslide susceptibility prediction of 4,565,153 point objects in the whole research area and the susceptibility values of all points were imported into ArcGIS 10.5 and converted to 30 $\times$ 30 m grid units.

3.4. Collapse and Landslide Susceptibility Prediction Mapping

According to the probability values of collapse and landslide disasters predicted by the six models, the research area was divided into five intervals—very low, low, moderate, high, and very high—by natural discontinuity method (see Figure 6). The statistical results of susceptibility evaluation are shown in Table 8. It can be seen from Figure 6 that the areas with high and very high probability of disaster occurrence are mainly distributed in the middle, northwest, and southeast of Wenchuan County, and extend from north to south in strips, while the areas with low and very low probability of disaster occurrence are mainly distributed around the west and southwest of Wenchuan County. Table 7 shows that the areas with very low, low, moderate, high, and very high probability of collapse and landslide disaster occurrence predicted by the LR model have areas of 1943.382 km^2, 807.659 km^2, 298.507 km^2, 328.496 km^2, and 730.595 km^2, accounting for 47.3%, 19.658%, 7.265%, 7.995%, and 17.782%, respectively. The areas with very low, low, moderate, high, and very high probability of collapse and landslide disaster occurrence predicted by the CF-LR model have areas of 2577.819 km^2, 583.316 km^2, 320.097 km^2, 249.885 km^2, and 377.521 km^2, accounting for 62.741%, 14.197%, 7.791%, 6.082%, and 9.188%, respectively. The areas with very low, low, moderate, high, and very high probability of collapse and landslide disaster occurrence predicted by the SVM model have areas of 2047.961 km^2, 904.71 km^2, 469.662 km^2, 324.533 km^2, and 361.772 km^2, accounting for 49.845%, 22.020%, 11.431%, 7.899%, and 8.805%, respectively. The areas with very low, low, moderate, high, and very high probability of collapse and landslide disaster occurrence predicted by the CF-SVM model have areas of 2887.869 km^2, 454.71 km^2, 206.651 km^2, 186.43 km^2, and 372.979 km^2, accounting for 70.288%, 11.067%, 5.030%, 4.538%, and 9.078%, respectively. The areas with very low, low, moderate, high, and very high probability of collapse and landslide disaster occurrence predicted by the RF model have areas of 2712.996 km^2, 447.601 km^2, 328.298 km^2, 301.112 km^2, and 318.631 km^2, accounting for 66.032%, 10.894%, 7.990%, 7.329%, and 7.755%, respectively. The areas with very low, low, moderate, high, and very high probability of collapse and landslide disaster occurrence predicted by the CF-RF model have areas of 2790.878 km^2, 443.463 km^2, 314.755 km^2, 253.031 km^2, and 306.51 km^2, accounting for 67.927%, 10.793%, 7.661%, 6.159%, and 7.460%, respectively.

In order to validate the reliability and accuracy of the susceptibility mapping level, the frequency ratio FR (that is, the ratio of the percentage of the number of collapse and landslide points at each susceptibility level to the percentage of the area at each level) was calculated. The frequency ratios of the areas with very low, low, moderate, high, and very high probability of collapse and landslide disaster occurrence predicted by the LR model are 0.01, 0.118, 0.395, 0.694, and 4.994, respectively. The frequency ratios of the areas with very low, low, moderate, high, and very high probability of collapse and landslide disaster occurrence predicted by the CF-LR model are 0.006, 0.078, 0.641, 2.282, and 8.668, respectively. The frequency ratios of the areas with very low, low, moderate, high, and very high probability of collapse and landslide disaster occurrence predicted by the SVM model are 0.011, 0.126, 0.364, 0.878, and 9.718, respectively. The frequency ratios of the areas with very low, low, moderate, high, and very high probability of collapse and landslide disaster occurrence predicted by the CF-SVM model are 0.009, 0.134, 0.386, 1.733, and 9.701, respectively. The frequency ratios of the areas with very low, low, moderate, high, and very high probability of collapse and landslide disaster occurrence predicted by the RF model are 0.004, 0.051, 0.313, 1.931, and 10.640, respectively. The frequency ratios of the areas with very low, low, moderate, high, and very high probability of collapse and landslide disaster occurrence predicted by the CF-RF model are 0.001, 0.043, 0169, 1.277, and 12.103, respectively. In all collapse and landslide susceptibility mapping, the frequency ratio FR ranges from 0.01 to 12.103. Most collapses and landslides occur in areas with very high probability of disaster occurrence, while only a small number of collapses and landslides occur in areas with very low and low probability of disaster occurrence. With the increase of collapse and landslide susceptibility level, the frequency ratio in each level gradually increases, which verifies the accuracy of the model. The frequency ratio precision of the collapse and landslide susceptibility results can be obtained by dividing the frequency ratios of the areas with high and very high probability of disaster occurrence by the sum of all frequency ratios. The frequency ratio precisions of the results predicted by the LR, CF-LR, SVM, CF-SVM, RF, and CF-RF models are 0.916, 0.938, 0.955, 0.956, 0.972, and 0.984 respectively, indicating that the prediction accuracy of each model is high and that they can predict the occurrence of collapse and landslide disasters. The frequency ratio precision of the RF model is higher than that of the LR and SVM models, which indicates that the collapse and landslide susceptibility predicted by the RF model better reflects the spatial aggregation characteristics and distribution rules of regional collapses and landslides. In addition, the frequency ratio precision of the coupling model is higher than that of the single model, which indicates that the CF-based coupling model can improve prediction accuracy. In particular, the CF-RF model predicts that the area with high and very high probability of disaster occurrence has the lowest area ratio, but the frequency ratio precision reaches 0.984, indicating that the CF-RF model is the optimal model.



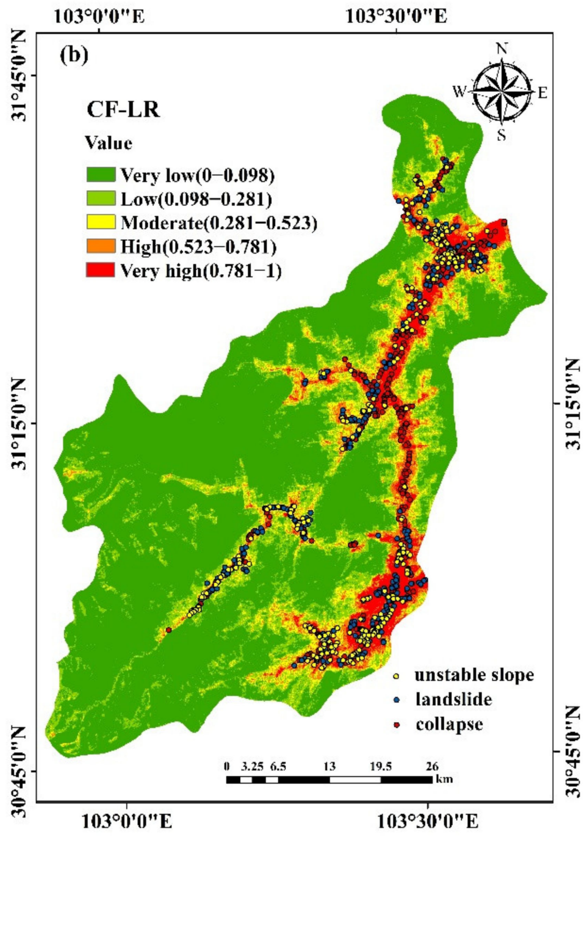


Figure 6. *Cont.*

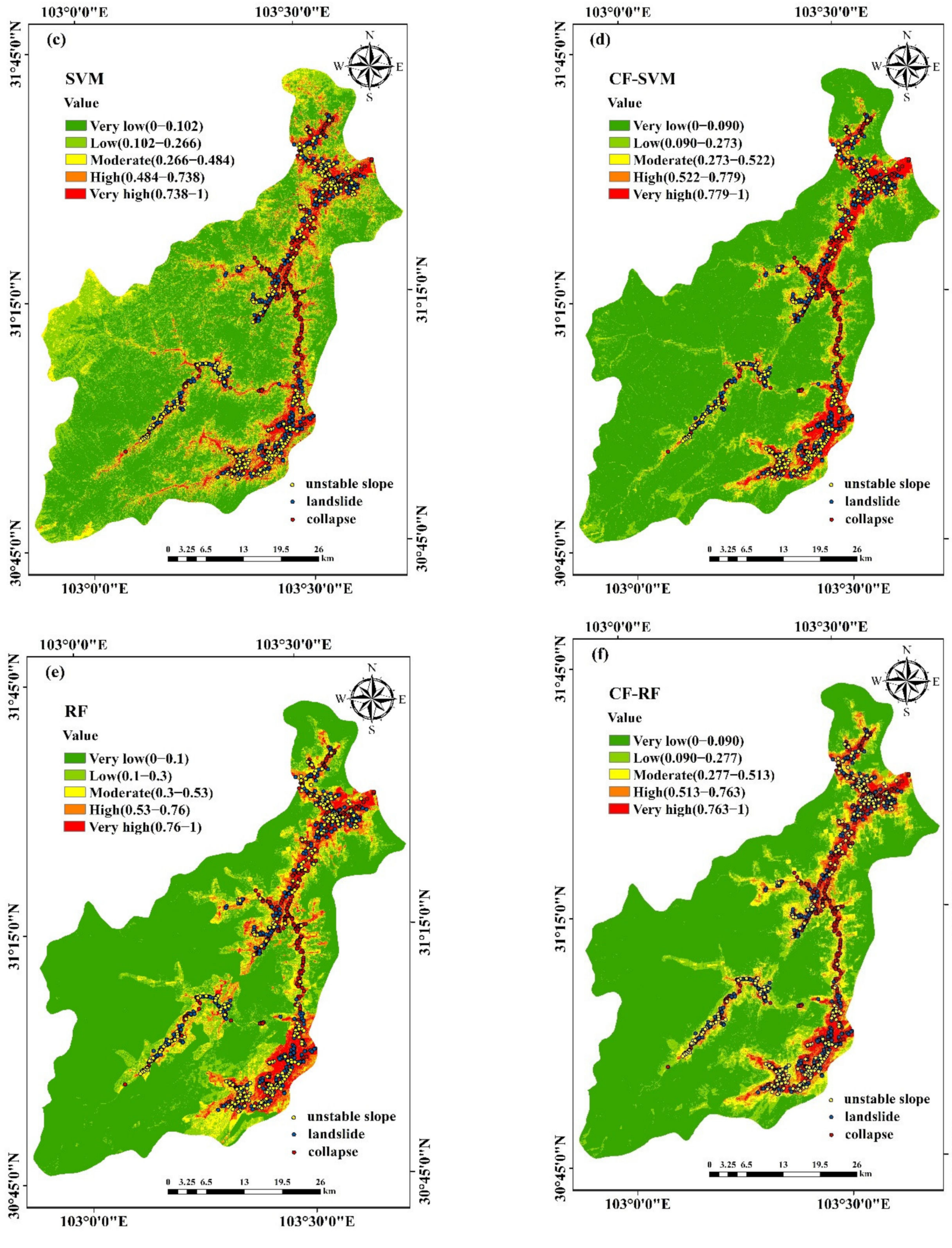


Figure 6. Collapse and landslide susceptibility mapping of different models: (**a**) LR, (**b**) CF-LR, (**c**) SVM, (**d**) CF-SVM, (**e**) RF, (**f**) CF-RF.

Table 8. Distribution of collapses and landslides at all susceptibility levels with different models.

Model	Geohazard Level	Area (km^2)	Area Percentage (%)	Number of Collapse and Landslide Points	Ratio of Collapse and Landslide (%)	Frequency Ratio (FR)
LR	Very low	1943.382	47.300	5	0.463	0.010
	Low	807.659	19.658	25	2.313	0.118
	Moderate	298.507	7.265	31	2.868	0.395
	High	328.496	7.995	60	5.550	0.694
	Very high	730.595	17.782	960	88.807	4.994
CF-LR	Very low	2577.819	62.741	4	0.370	0.006
	Low	583.316	14.197	12	1.110	0.078
	Moderate	320.097	7.791	54	4.995	0.641
	High	249.885	6.082	150	13.876	2.282
	Very high	377.521	9.188	861	79.648	8.668
SVM	Very low	2047.961	49.845	6	0.555	0.011
	Low	904.710	22.020	30	2.775	0.126
	Moderate	469.662	11.431	45	4.163	0.364
	High	324.533	7.899	75	6.938	0.878
	Very high	361.772	8.805	925	85.569	9.718
CF-SVM	Very low	2887.869	70.288	7	0.648	0.009
	Low	454.710	11.067	16	1.480	0.134
	Moderate	206.651	5.030	21	1.943	0.386
	High	186.430	4.538	85	7.863	1.733
	Very high	372.979	9.078	952	88.067	9.701
RF	Very low	2712.996	66.032	3	0.278	0.004
	Low	447.601	10.894	6	0.555	0.051
	Moderate	328.298	7.990	27	2.498	0.313
	High	301.112	7.329	153	14.154	1.931
	Very high	318.631	7.755	892	82.516	10.640
CF-RF	Very low	2790.878	67.927	1	0.093	0.001
	Low	443.463	10.793	5	0.463	0.043
	Moderate	314.755	7.661	14	1.295	0.169
	High	253.031	6.159	85	7.863	1.277
	Very high	306.510	7.460	976	90.287	12.103

3.5. Precision Evaluation of the Models

3.5.1. Evaluation of Precision Validation Parameters

See Table 9 for the results of confusion matrix and statistical indexes; each index is shown in Figure 7. In terms of precision, as seen from CF-RF>RF>CF-LR>CF-SVM>SVM>LR, the precision of the CF-RF model is highest, indicating that the CF-RF model has the strongest ability to distinguish negative samples. In terms of recall, as seen from LR>CF-SVM>CF-RF>RF>CF-LR>SVM, the recall of the LR model is highest, indicating that the CF-RF model has the strongest ability to distinguish positive samples. As precision and recall are contradictory measurements (when precision is high, recall is often low; when recall is high, precision is often low), the F1-score index is introduced, which takes both precision and recall into account. The F1-score value of each model is greater than 0.8; the order of each model according to the F1-score value is CF-RF>RF>CF-SVM>CF-LR>SVM>LR, indicating that all models can reflect the collapse and landslide susceptibility in the research area, and the performance of the CF-RF model is relatively high. In terms of accuracy, as seen from CF-RF>RF>CF-SVM>CF-LR>SVM>LR, the accuracy of the CF-RF model is highest, which indicates that the RF model can better predict the occurrence of collapse and landslide disasters than the SVM and LR models, and the coupling model can improve the prediction accuracy of the model. In terms of KC, the KC values of the six models for susceptibility evaluation are greater than 0.6, indicating that the models have high consistency, i.e., the difference between the prediction results and the actual classification results of the model is small and the classification accuracy is high. As seen from CF-RF>CF-SVM>CF-LR>RF>SVM>LR, the coupling model can improve the classification accuracy. In terms of MCC, the MCC value in all models is greater than 0.6 and the order is CF-RF>RF>CF-LR>CF-SVM>SVM>LR, showing that the models can predict the occurrence of collapse and landslide disasters and the coupling model can improve the classification accuracy. Some models have both good and bad indexes. A single index cannot measure all the advantages and disadvantages of a model. Therefore, the POA index is introduced. It is a comprehensive performance index to quantify the overall performance of a model. The model with the highest POA has the highest overall performance. The order of each model according to the POA value is CF-RF>RF>CF-SVM>CF-LR>SVM>LR. The POA value of the CF-RF model is highest (2.570), followed by the POA value of the RF model (2.552), showing that in the research of collapse and landslide susceptibility in Wenchuan County, the RF model has higher prediction precision than the SVM and LR models; the SVM model ranks second and the LR model has the lowest prediction precision. The coupling model can improve the precision of the model over the single model. The top ranking of the CF-RF model in all indexes shows that it has the highest accuracy and reliability in this research, and is the optimal model.

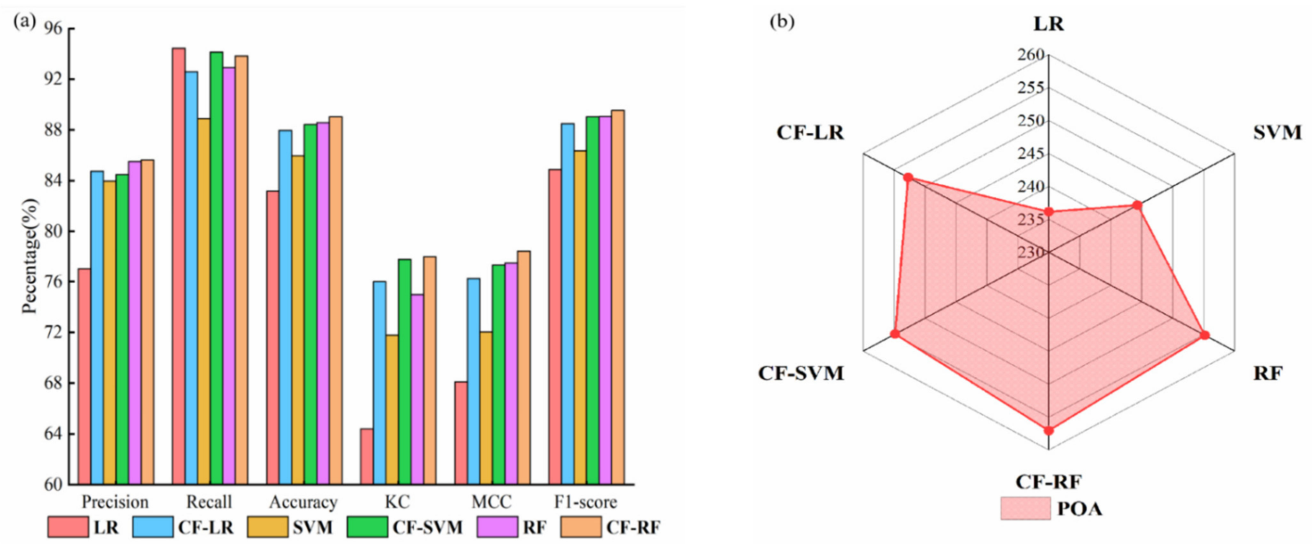


Figure 7. Precision comparison of the models (**a**) precision, recall, accuracy, KC, MCC, and F1-score; (**b**) POA of the models.

Table 9. Analysis of prediction ability of different models by validation samples.

	LR	CF-LR	SVM	CF-SVM	RF	CF-RF
TP	306	300	288	305	301	304
TN	233	270	269	268	273	273
FP	91	54	55	56	51	51
FN	18	24	36	19	23	20
Precision (%)	77.078	84.746	83.965	84.488	85.511	85.634
Recall (%)	94.444	92.593	88.889	94.136	92.901	93.827
Accuracy (%)	83.179	87.963	85.957	88.426	88.580	89.043
KC (%)	64.400	76.000	71.800	77.800	75.000	78.000
MCC (%)	68.109	76.254	72.038	77.358	77.516	78.446
F1-score (%)	84.882	88.496	86.357	89.051	89.053	89.543
POA (%)	236.170	252.712	244.338	254.835	255.216	257.046

3.5.2. Comparison of ROC and AUC Results

The ROC curve of the prediction probability of the validation sample is drawn by SPSS software, and the ROC curve and AUC value of the six modes for susceptibility evaluation are shown in Figure 8. The AUC values of LR, SVM, and RF models are 0.905, 0.918, and 0.935, respectively, while the AUC values of CF-LR, CF-SVM, and CF-RF models are 0.929, 0.933, and 0.946, respectively. The AUC value of the CF-RF model is the highest. It can be seen that the precision of all models is high and that among all single models, the RF model has the highest precision, followed by the SVM model; the LR model has the lowest precision. The AUC value of the CF-based models is greater than that of the single models, which further indicates that the coupling model is helpful in improving the ability to predict collapse and landslide disaster. The CF-RF model in this research has the best performance, which is consistent with the precision test conclusion based on the confusion matrix.

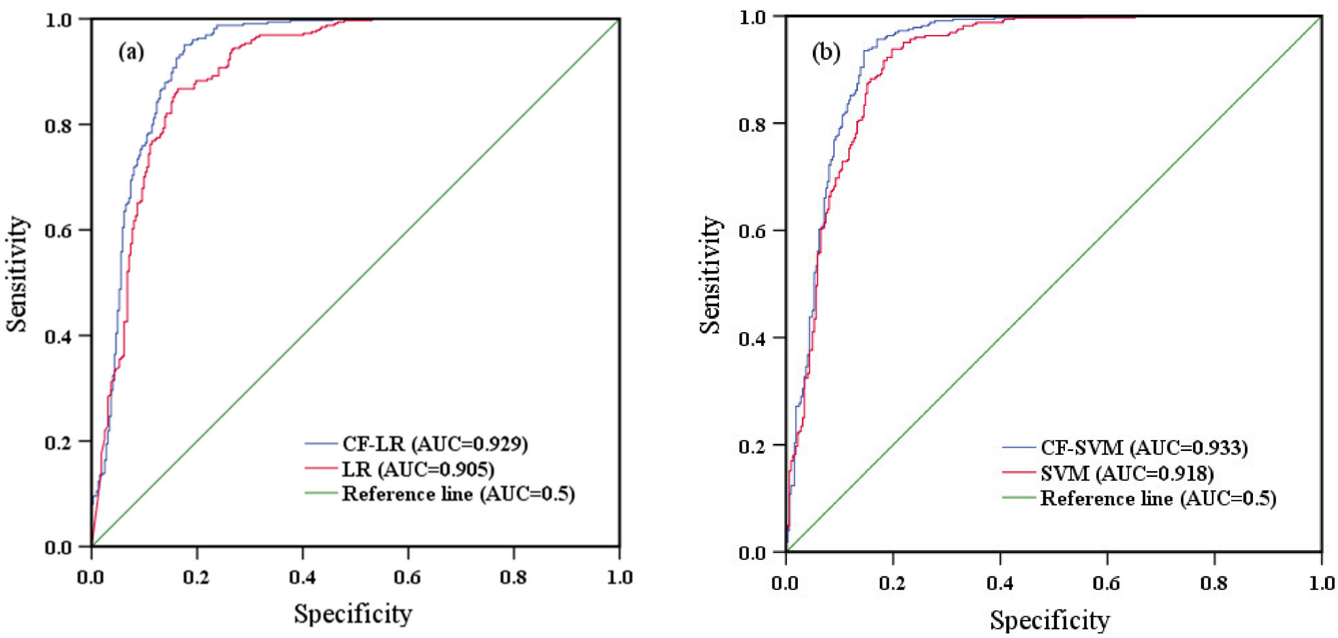


Figure 8. *Cont.*

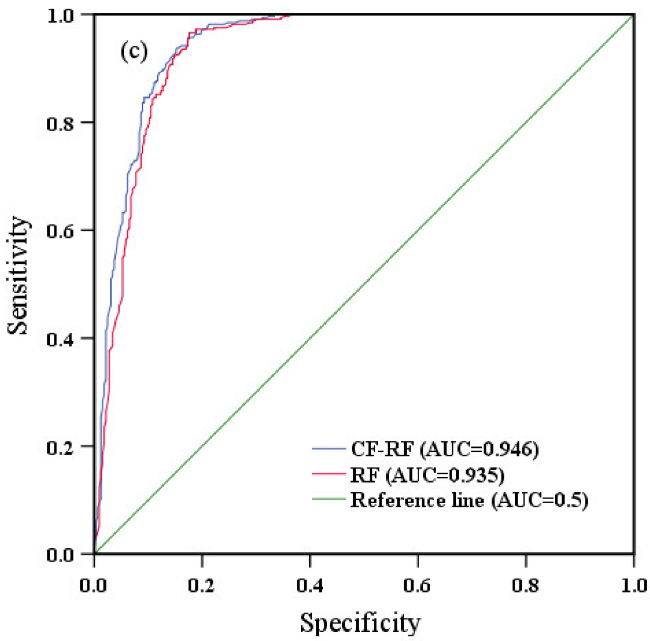


Figure 8. ROC curves with associated AUC values versus validation set: (**a**) CF-LR and LR; (**b**) CF-SVM and SVM; (**c**) CF-RF and RF.

4. Discussion

In this study, we built three single models (LR, SVM and RF) and three CF-based hybrid models (CF-LR, CF-SVM and CF-RF) to generate six collapse and landslide susceptibility maps in Wenchuan County, and compared the prediction accuracy of the six models. The results show that the machine learning models based on the certainty factor have higher prediction accuracy than the single models. Among them, CF-RF model has the highest performance, which is consistent with the research results of Xiao Wang et al. [2]. Previous studies on landslide susceptibility mapping in Wenchuan County mostly used a single model. Yulin Su et al. built a DNN model for earthquake-geological disaster chain study, which is discussed with the support vector machine (SVM) model, logistic regression (LR) model, and random forest (RF) model [30]. Shuai Li et al. studied the change of geological hazard sensitivity and its driving mechanism ten years after the Wenchuan earthquake by using a random forest model [31]. Juan Cao et al. compared the prediction accuracy of logistic regression (LR) and random forest (RF) models in sensitivity mapping of Wenchuan and Lushan earthquake landslides [32]. Xie, P. et al. made earthquake landslide susceptibility maps in Wenchuan County by using a neural network model and a logistic regression model [33]. In recent years, in order to improve the prediction accuracy of landslide susceptibility mapping, deep learning methods have been widely used. Zhang, S. et al. compared the capabilities of advanced convolutional neural networks (CNN) and traditional machine learning methods [26]; Zheng, H.Y. et al. combined this with a deep neural network (DNN) to build a spatial prediction model of landslide disasters [27]. Compared with the single model, machine learning models based on a certainty factor have higher prediction accuracy and are simpler to build than deep learning models. They play an important role in predicting potential landslides in the future and can provide a decision-making basis for the early warning and prevention of landslides in Wenchuan County.

4.1. Importance Ranking of Environmental Factors

In Section 3.2, through the preliminary analysis of the correlation between each environmental factor and the occurrence of collapses and landslides, the intervals of each factor relating to collapses and landslides are obtained, but the contribution of each factor to the collapse and landslide susceptibility prediction are not reflected. From the Section 3.5, it can be seen that the CF-RF is the optimal model in the research area, so the importance of environmental factors is discussed based on the CF-RF model. In the RF tree, optimal segmentation is measured with impurity, and the importance of basic environmental factors is calculated by the reduced value of Gini index of the environmental factor when the node is divided. In this research, the importance of environmental factors is measured by the percentage of the average Gini index decrease to the sum of average Gini index decrease of all environmental factors. The 13 environmental factors in the CF-RF model are analyzed by RStudio4.1.3 and the importance ranking of each factor is obtained by origin2018 software (Figure 9). It can be seen from Figure 9 that the ranking of the 13 environmental factors according to their importance proportion, is rainfall>soil type>distance to river>terrain relief>PGA>land use>NDVI>distance to fault>lithology>aspect>slope>TWI>curvature. As the three most important environmental factors among the 13 environmental factors, rainfall, soil type and distance to river have importance proportions of 24.216%, 22.309%, and 11.41%, respectively, and make the highest contributions to the model, showing that these three environmental factors are important trigger factors of collapse and landslide disasters in the research area, and cannot be ignored in the susceptibility evaluation of collapses and landslides. However, slope, TWI, and curvature account for the lowest importance proportions—3.159%, 3.02%, and 2.813%, respectively—indicating that these three environmental factors have little impact on the susceptibility evaluation of collapses and landslides in the research area.

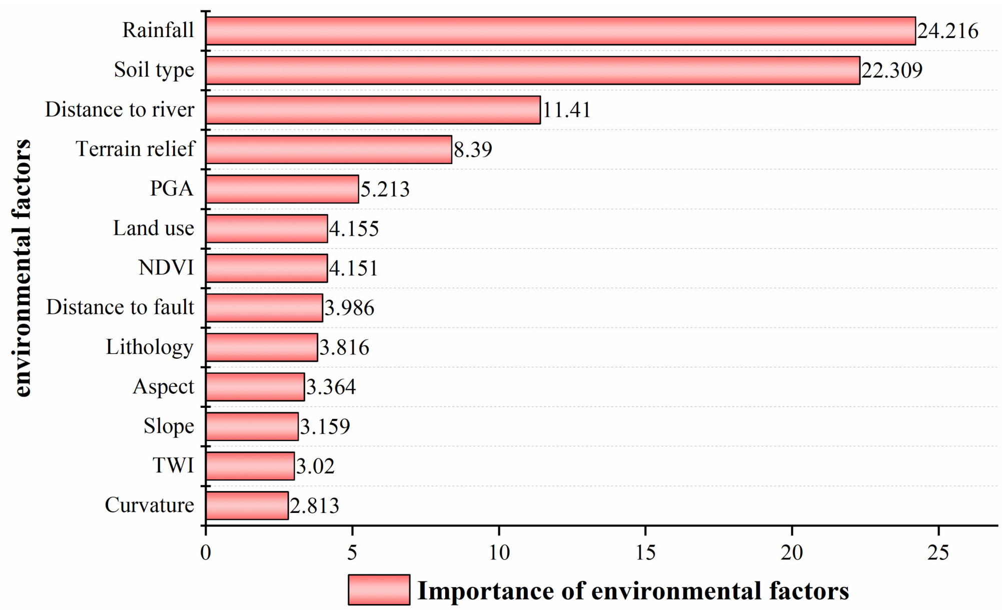


Figure 9. Importance Ranking of Environmental Factors of the CF-RF Model.

4.2. Division of Evaluation Units

The accuracy of collapse and landslide susceptibility evaluation is closely related to the evaluation unit. Common evaluation units are the grid unit, terrain unit and slope unit [62–64]. After the evaluation unit is determined, the value of each environmental factor can be allocated to each unit. A grid unit divides the research area into regular squares of predefined size for storage and calculation; this is widely used in collapse and landslide susceptibility mapping, but cannot fully reflect the terrain relief and the geological and hydrogeological elements of the research area [65]. With the morphology of the earth's surface based on DEM, a terrain unit takes the concave–convex earth's surface as the boundary to divide the areas; the curvature is the key to extract the concave–convex boundary [23]. A slope unit is a watershed area delimited by the drainage line (valley line) and the water boundary (ridge line), as well as the basic terrain and landform unit of geological disasters [66]. The evaluation unit used in this research is the 30 m × 30 m grid element, but in future research, terrain units and slope units can be used to analyze the collapse and landslide susceptibility, and the similarities and differences between terrain units, slope units, and grid units can be analyzed and compared.

4.3. Uncertainty of Hybrid Models

The hybrid model of machine learning and the statistical method are widely used in the research of collapse and landslide susceptibility, and can effectively improve the prediction precision of models. These statistical methods are an important link between the collapse and landslide susceptibility index and its environmental factors; their connection performance is very important to the precision of machine learning models. At present, commonly used statistical methods include certainty factor (CF), weight of evidence (WOE), information value (IV), index of entropy (IOE), and frequency ratio (FR) [32,67–70]. There is no specific evaluation of which statistical methods can improve the precision of machine learning models, and different statistical methods bring great uncertainty to the prediction of susceptibility to collapses and landslides by machine learning models. In this research, the certainty factor is coupled with three machine learning methods: logistic regression, support vector machine, and random forest. In future research, other statistical methods and machine learning methods can be mixed to build collapse and landslide susceptibility models, allowing the exploration of the uncertainty law of collapse and landslide susceptibility prediction by machine learning models based on different statistical methods.

5. Conclusions

This paper takes the historical collapse and landslide disaster points in Wenchuan County as the data source, selects appropriate environmental factors, builds three single models (LR, SVM and RF) and three CF-based hybrid models (CF-LR, CF-SVM and CF-RF), completes the susceptibility mapping of collapse and landslide disasters in Wenchuan County, evaluates the accuracy and reliability of the models, obtains the laws of the impact of each environmental factor on the development of collapse and landslide in its attribute intervals, and explores the contribution of environmental factors to the collapse and landslide susceptibility prediction of the optimal model. The research shows that:

(1) The six models LR, CF-LR, SVM, CF-SVM, RF, and CF-RF can evaluate the susceptibility of collapse and landslide disasters in Wenchuan County. The areas with high and very high probability of disaster occurrence are mainly distributed in the middle, northwest, and southeast of Wenchuan County, and extend from north to south in strips, while the areas with low and very low probability of disaster occurrence are mainly distributed around the west and southwest of Wenchuan County. The areas with very high probability of collapse and landslide disaster occurrence predicted by the models have an area of 730.595 km^2, 377.521 km^2, 361.772 km^2, 372.979 km^2, 318.631 km^2, and 306.51 km^2, accounting for 17.782%, 9.188%, 8.805%, 9.078%, 7.755%, and 7.460%, respectively. The frequency ratio precision of collapses and landslides is 0.916, 0.938, 0.955, 0.956, 0.972, and 0.984, respectively, which validates the accuracy of the models. The frequency ratio precision of the RF model is higher than that of the LR and SVM models, and the coupling models have higher frequency ratio precision than the single models;

(2) The precision of each model is evaluated based on the validation samples. The ranking of the comprehensive POA index based on the confusion matrix is CF-RF>RF>CF-SVM>CF-LR>SVM>LR, while the ranking of the AUC value is CF-RF>RF>CF-SVM>CF-LR>SVM>LR. The RF model has the highest precision. The coupling model can improve the precision of the models over the single models. The highest ranking of the CF-RF model in all indexes shows that it has the highest accuracy and reliability in this research, and is therefore the optimal model.

(3) The importance of environmental factors is explored based on the CF-RF model; the ranking of the 13 environmental factors according to their proportion of importance is rainfall>soil type>distance to river>terrain relief>PGA>land use>NDVI>Distance to fault>lithology>aspect>slope>TWI>curvature. As the three most important environmental factors among the 13 environmental factors, rainfall, soil type and distance to river have

importance proportions of 24.216%, 22.309%. and 11.41%, respectively. Rainfall is the most important trigger factor for collapse and landslide disasters in the research area, while the importance of curvature accounts for 2.813% and contributes the least to the model. Therefore, during disaster prevention and mitigation in Wenchuan region, it is necessary to strengthen the monitoring of mountains and rock masses close to rivers under rainstorm conditions.

Author Contributions: X.Y. and H.L. drafted the manuscript and were responsible for the research design, experiment, and analysis. C.L., R.N., W.L. and X.D. reviewed and edited the manuscript. Z.Y., J.C., J.Z., L.M., X.F., M.T. and Y.X. supported the data preparation and the interpretation of the results. All of the authors contributed to editing and reviewing the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Key R&D Program of China (2019YFC1510700), the National Key R&D Program of China (2021YFC3000401), the National Natural Science Foundation of China (41701499), the funding provided by the Alexander von Humboldt-Stiftung, the Sichuan Science and Technology Program (2018GZ0265), the Geomatics Technology and Application Key Laboratory of Qinghai Province, China (QHDX-2018-07), the Major Scientific and Technological Special Program of Sichuan Province, China (2018SZDZX0027), and the Key Research and Development Program of Sichuan Province, China (2018SZ027, 2019-YF09-00081-SN).

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

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

References

1. Martínez-Moreno, F.J.; Galindo-Zaldívar, J.; González-Castillo, L.; Azañón, J.M. Collapse susceptibility map in abandoned mining areas by microgravity survey: A case study in Candado hill (Málaga, southern Spain). *J. Appl. Geophys.* **2016**, *130*, 101–119. [CrossRef]
2. 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]
3. Carrión-Mero, P.; Montalván-Burbano, N.; Morante-Carballo, F.; Quesada-Román, A.; Apolo-Masache, B. Worldwide Research Trends in Landslide Science. *Int. J. Environ. Res. Public Health* **2021**, *18*, 9445. [CrossRef] [PubMed]
4. Ye, P.; Yu, B.; Chen, W.; Liu, K.; Ye, L. Rainfall-induced landslide susceptibility mapping using machine learning algorithms and comparison of their performance in Hilly area of Fujian Province, China. *Nat. Hazards* **2022**, 1–31. [CrossRef]
5. Youssef, A.M.; Pourghasemi, H.R. Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia. *Di Xue Qian Yuan* **2021**, *12*, 639–655. [CrossRef]
6. Demir, G.; Aytekin, M.; Akgun, A. Landslide susceptibility mapping by frequency ratio and logistic regression methods: An example from Niksar–Resadiye (Tokat, Turkey). *Arab. J. Geosci.* **2015**, *8*, 1801–1812. [CrossRef]
7. Assilzadeh, H.; Levy, J.K.; Wang, X. Landslide Catastrophes and Disaster Risk Reduction: A GIS Framework for Landslide Prevention and Management. *Remote Sens.* **2010**, *2*, 2259–2273. [CrossRef]
8. Azarafza, M.; Ghazifard, A.; Akgün, H.; Asghari-Kaljahi, E. Landslide susceptibility assessment of South Pars Special Zone, southwest Iran. *Environ. Earth Sci.* **2018**, *77*, 805. [CrossRef]
9. Thi Ngo, P.T.; Panahi, M.; Khosravi, K.; Ghorbanzadeh, O.; Kariminejad, N.; Cerda, A.; Lee, S. Evaluation of deep learning algorithms for national scale landslide susceptibility mapping of Iran. *Geosci. Front.* **2021**, *12*, 505–519. [CrossRef]
10. Chen, W.; Pourghasemi, H.R.; Zhao, Z. A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural network models for landslide susceptibility mapping. *Geocarto Int.* **2017**, *32*, 367–385. [CrossRef]
11. Muñoz-Torrero Manchado, A.; Allen, S.; Ballesteros-Cánovas, J.A.; Dhakal, A.; Dhital, M.R.; Stoffel, M. Three decades of landslide activity in western Nepal: New insights into trends and climate drivers. *Landslides* **2021**, *18*, 2001–2015. [CrossRef]
12. Quesada-Román, A. Landslide risk index map at the municipal scale for Costa Rica. *Int. J. Disaster Risk Reduct.* **2021**, *56*, 102144. [CrossRef]
13. Ali, S.A.; Parvin, F.; Vojteková, J.; Costache, R.; Linh, N.T.T.; Pham, Q.B.; Vojtek, M.; Gigović, L.; Ahmad, A.; Ghorbani, M.A. GIS-based landslide susceptibility modeling: A comparison between fuzzy multi-criteria and machine learning algorithms. *Geosci. Front.* **2021**, *12*, 857–876. [CrossRef]
14. Quesada-Román, A.; Fallas-López, B.; Hernández-Espinoza, K.; Stoffel, M.; Ballesteros-Cánovas, J.A. Relationships between earthquakes, hurricanes, and landslides in Costa Rica. *Landslides* **2019**, *16*, 1539–1550. [CrossRef]
15. Bahrami, Y.; Hassani, H.; Maghsoudi, A. Landslide susceptibility mapping using AHP and fuzzy methods in the Gilan province, Iran. *GeoJournal* **2021**, *86*, 1797–1816. [CrossRef]

16. Tao, Y.; Xue, Y.; Zhang, Q.; Yang, W.; Li, B.; Zhang, L.; Qu, C.; Zhang, K. Risk Assessment of Unstable Rock Masses on High-Steep Slopes: An Attribute Recognition Model. *Soil Mech. Found. Eng.* **2021**, *58*, 175–182. [CrossRef]
17. Zhao, X.; Chen, W. Optimization of Computational Intelligence Models for Landslide Susceptibility Evaluation. *Remote Sens.* **2020**, *12*, 2180. [CrossRef]
18. Chen, W.; Li, H.; Hou, E.; Wang, S.; Wang, G.; Panahi, M.; Li, T.; Peng, T.; Guo, C.; Niu, C.; et al. GIS-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models. *Sci. Total Environ.* **2018**, *634*, 853–867. [CrossRef]
19. Hong, H.; Chen, W.; Xu, C.; Youssef, A.M.; Pradhan, B.; Tien Bui, D. Rainfall-induced landslide susceptibility assessment at the Chongren area (China) using frequency ratio, certainty factor, and index of entropy. *Geocarto Int.* **2017**, *32*, 139–154. [CrossRef]
20. Aditian, A.; Kubota, T.; Shinohara, Y. Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. *Geomorphology* **2018**, *318*, 101–111. [CrossRef]
21. Polat, A. An innovative, fast method for landslide susceptibility mapping using GIS-based LSAT toolbox. *Environ. Earth Sci.* **2021**, *80*, 217. [CrossRef]
22. Zheng, H.; Liu, B.; Han, S.; Fan, X.; Zou, T.; Zhou, Z.; Gong, H. Research on landslide hazard spatial prediction models based on deep neural networks: A case study of northwest Sichuan, China. *Environ. Earth Sci.* **2022**, *81*, 258. [CrossRef]
23. Zêzere, J.L.; Pereira, S.; Melo, R.; Oliveira, S.C.; Garcia, R.A.C. Mapping landslide susceptibility using data-driven methods. *Sci. Total Environ.* **2017**, *589*, 250–267. [CrossRef]
24. 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]
25. Arabameri, A.; Chen, W.; Loche, M.; Zhao, X.; Li, Y.; Lombardo, L.; Cerda, A.; Pradhan, B.; Bui, D.T. Comparison of machine learning models for gully erosion susceptibility mapping. *Geosci. Front.* **2020**, *11*, 1609–1620. [CrossRef]
26. Lin, J.; He, P.; Yang, L.; He, X.; Lu, S.; Liu, D. Predicting future urban waterlogging-prone areas by coupling the maximum entropy and FLUS model. *Sustain. Cities Soc.* **2022**, *80*, 103812. [CrossRef]
27. Javidan, N.; Kavian, A.; Pourghasemi, H.R.; Conoscenti, C.; Jafarian, Z.; Rodrigo-Comino, J. Evaluation of multi-hazard map produced using MaxEnt machine learning technique. *Sci. Rep.* **2021**, *11*, 6496. [CrossRef]
28. Rahmati, O.; Golkarian, A.; Biggs, T.; Keesstra, S.; Mohammadi, F.; Daliakopoulos, I.N. Land subsidence hazard modeling: Machine learning to identify predictors and the role of human activities. *J. Environ. Manag.* **2019**, *236*, 466–480. [CrossRef]
29. Zhang, S.; Bai, L.; Li, Y.; Li, W.; Xie, M. Comparing Convolutional Neural Network and Machine Learning Models in Landslide Susceptibility Mapping: A Case Study in Wenchuan County. *Front. Environ. Sci.* **2022**, *10*, 496. [CrossRef]
30. Chen, W.; Peng, J.; Hong, H.; Shahabi, H.; Pradhan, B.; Liu, J.; Zhu, A.; Pei, X.; Duan, Z. Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. *Sci. Total Environ.* **2018**, *626*, 1121–1135. [CrossRef]
31. Chen, W.; Li, Y. GIS-based evaluation of landslide susceptibility using hybrid computational intelligence models. *Catena* **2020**, *195*, 104777. [CrossRef]
32. Trinh, T.; Luu, B.T.; Le, T.H.T.; Nguyen, D.H.; Van Tran, T.; Van Nguyen, T.H.; Nguyen, K.Q.; Nguyen, L.T. A comparative analysis of weight-based machine learning methods for landslide susceptibility mapping in Ha Giang area. *Big Earth Data* **2022**, 1–30, *ahead of print*. [CrossRef]
33. Zhou, X.; Wu, W.; Qin, Y.; Fu, X. Geoinformation-based landslide susceptibility mapping in subtropical area. *Sci. Rep.* **2021**, *11*, 24325. [CrossRef] [PubMed]
34. Qiu, C.; Su, L.; Zou, Q.; Geng, X. A hybrid machine-learning model to map glacier-related debris flow susceptibility along Gyirong Zangbo watershed under the changing climate. *Sci. Total Environ.* **2022**, *818*, 151752. [CrossRef]
35. Xu, C.; Xu, X.; Zhou, B.; Yu, G. Revisions of the M 8.0 Wenchuan earthquake seismic intensity map based on co-seismic landslide abundance. *Nat. Hazards* **2013**, *69*, 1459–1476. [CrossRef]
36. Zhu, J.; Ding, J.; Liang, J. Influences of the Wenchuan Earthquake on sediment supply of debris flows. *J. Mt. Sci.* **2011**, *8*, 270–277. [CrossRef]
37. Huang, R.; Li, W. Post-earthquake landsliding and long-term impacts in the Wenchuan earthquake area, China. *Eng. Geol.* **2014**, *182*, 111–120. [CrossRef]
38. Evik, E.; Topal, T. GIS-based landslide susceptibility mapping for a problematic segment of the natural gas pipeline, Hendek (Turkey). *Environ. Geol.* **2003**, *44*, 949–962.
39. Gnyawali, K.R.; Zhang, Y.; Wang, G.; Miao, L.; Pradhan, A.M.S.; Adhikari, B.R.; Xiao, L. Mapping the susceptibility of rainfall and earthquake triggered landslides along China–Nepal highways. *Bull. Eng. Geol. Environ.* **2020**, *79*, 587–601. [CrossRef]
40. Mersha, T.; Meten, M. GIS-based landslide susceptibility mapping and assessment using bivariate statistical methods in Simada area, northwestern Ethiopia. *Geoenviron. Disasters* **2020**, *7*, 20. [CrossRef]
41. Sajinkumar, K.S.; Anbazhagan, S. Geomorphic appraisal of landslides on the windward slope of Western Ghats, southern India. *Nat. Hazards* **2015**, *75*, 953–973. [CrossRef]
42. Beven, K.J.; Kirkby, M.J.; Schofield, N.; Tagg, A.F. Testing a physically-based flood forecasting model (TOPMODEL) for three U.K. catchments. *J. Hydrol.* **1984**, *69*, 119–143. [CrossRef]
43. Pachuau, L. Zonation of Landslide Susceptibility and Risk Assessment in Serchhip town, Mizoram. *J. Indian Soc. Remote* **2019**, *47*, 1587–1597. [CrossRef]

44. Bucci, F.; Santangelo, M.; Cardinali, M.; Fiorucci, F.; Guzzetti, F. Landslide distribution and size in response to Quaternary fault activity: The Peloritani Range, NE Sicily, Italy. *Earth Surf. Proc. Land.* **2016**, *41*, 711–720. [CrossRef]
45. Steger, S.; Brenning, A.; Bell, R.; Petschko, H.; Glade, T. Exploring discrepancies between quantitative validation results and the geomorphic plausibility of statistical landslide susceptibility maps. *Geomorphology* **2016**, *262*, 8–23. [CrossRef]
46. Yi, Y.; Zhang, Z.; Zhang, W.; Xu, Q.; Deng, C.; Li, Q. GIS-based earthquake-triggered-landslide susceptibility mapping with an integrated weighted index model in Jiuzhaigou region of Sichuan Province, China. Nat. Hazard. *Earth Sys.* **2019**, *19*, 1973–1988. [CrossRef]
47. Peduzzi, P. Landslides and vegetation cover in the 2005 north Pakistan earthquake; a GIS and statistical quantitative approach. *Nat. Hazard. Earth Sys.* **2010**, *10*, 623–640. [CrossRef]
48. Pham, V.D.; Nguyen, Q.; Nguyen, H.; Pham, V.; Vu, V.M.; Bui, Q. Convolutional Neural Network—Optimized Moth Flame Algorithm for Shallow Landslide Susceptible Analysis. *IEEE Access* **2020**, *8*, 32727–32736. [CrossRef]
49. Cheng, J.; Dai, X.; Wang, Z.; Li, J.; Qu, G.; Li, W.; She, J.; Wang, Y. Landslide Susceptibility Assessment Model Construction Using Typical Machine Learning for the Three Gorges Reservoir Area in China. *Remote Sens.* **2022**, *14*, 2257. [CrossRef]
50. Shahzad, N.; Ding, X.; Abbas, S. A Comparative Assessment of Machine Learning Models for Landslide Susceptibility Mapping in the Rugged Terrain of Northern Pakistan. *Appl. Sci.* **2022**, *12*, 2280. [CrossRef]
51. Shortliffe, E.H. A Model of Inexact Reasoning in Medicine. *Math. Biosci.* **1975**, *23*, 351–379. [CrossRef]
52. Heckerman, D.E.; Shortliffe, E.H. From certainty factors to belief networks. *Artif. Intell. Med.* **1992**, *4*, 35–52. [CrossRef]
53. Chen, W.; Li, W.; Chai, H.; Hou, E.; Li, X.; Ding, X. GIS-based landslide susceptibility mapping using analytical hierarchy process (AHP) and certainty factor (CF) models for the Baozhong region of Baoji City, China. *Environ. Earth Sci.* **2016**, *75*, 63. [CrossRef]
54. Dou, J.; Tien Bui, D.; Yunus, A.P.; Jia, K.; Song, X.; Revhaug, I.; Xia, H.; Zhu, Z. Optimization of Causative Factors for Landslide Susceptibility Evaluation Using Remote Sensing and GIS Data in Parts of Niigata, Japan. *PLoS ONE* **2015**, *10*, e133262. [CrossRef]
55. Paryani, S.; Neshat, A.; Javadi, S.; Pradhan, B. Comparative performance of new hybrid ANFIS models in landslide susceptibility mapping. *Nat. Hazards* **2020**, *103*, 1961–1988. [CrossRef]
56. Dou, J.; Yamagishi, H.; Pourghasemi, H.R.; Yunus, A.P.; Song, X.; Xu, Y.; Zhu, Z. An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan. *Nat. Hazards* **2015**, *78*, 1749–1776. [CrossRef]
57. Vapnik, V.N. The support vector method. In Proceedings of the 7th International Conference on Artificial Neural Networks, Lausanne, Switzerland, 8–10 October 1997; pp. 263–271.
58. Cortes, C.; Vapnik, V. Cortes-Vapnik1995_Article_Support-vectorNetworks. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
59. Breiman, L. Breiman2001_Article_RandomForests. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
60. Wang, H.; Zhang, L.; Yin, K.; Luo, H.; Li, J. Landslide identification using machine learning. *Geosci. Front.* **2021**, *12*, 351–364. [CrossRef]
61. Zhou, X.; Wu, W.; Lin, Z.; Zhang, G.; Chen, R.; Song, Y.; Wang, Z.; Lang, T.; Qin, Y.; Ou, P.; et al. Zonation of Landslide Susceptibility in Ruijin, Jiangxi, China. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5906. [CrossRef]
62. Kreuzer, T.M.; Wilde, M.; Terhorst, B.; Damm, B. A landslide inventory system as a base for automated process and risk analyses. *Earth Sci. Inform.* **2017**, *10*, 507–515. [CrossRef]
63. Zhang, T.; Fu, Q.; Quevedo, R.P.; Chen, T.; Luo, D.; Liu, F.; Kong, H. Landslide Susceptibility Mapping Using Novel Hybrid Model Based on Different Mapping Units. *KSCE J. Civ. Eng.* **2022**, *26*, 2888–2900. [CrossRef]
64. Ba, Q.; Chen, Y.; Deng, S.; Yang, J.; Li, H. A comparison of slope units and grid cells as mapping units for landslide susceptibility assessment. *Earth Sci. Inform.* **2018**, *11*, 373–388. [CrossRef]
65. Liu, X.; Su, P.; Li, Y.; Zhang, J.; Yang, T. Susceptibility assessment of small, shallow and clustered landslide. *Earth Sci. Inform.* **2021**, *14*, 2347–2356. [CrossRef]
66. 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]
67. Qin, Y.; Yang, G.; Lu, K.; Sun, Q.; Xie, J.; Wu, Y. Performance Evaluation of Five GIS-Based Models for Landslide Susceptibility Prediction and Mapping: A Case Study of Kaiyang County, China. *Sustainability* **2021**, *13*, 6441. [CrossRef]
68. Zhao, B.; Ge, Y.; Chen, H. Landslide susceptibility assessment for a transmission line in Gansu Province, China by using a hybrid approach of fractal theory, information value, and random forest models. *Environ. Earth Sci.* **2021**, *80*, 441. [CrossRef]
69. Xiao, T.; Yin, K.; Yao, T.; Liu, S. Spatial prediction of landslide susceptibility using GIS-based statistical and machine learning models in Wanzhou County, Three Gorges Reservoir, China. *Acta Geochim.* **2019**, *38*, 654–669. [CrossRef]
70. Mohajane, M.; Costache, R.; Karimi, F.; Bao Pham, Q.; Essahlaoui, A.; Nguyen, H.; Laneve, G.; Oudija, F. Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area. *Ecol. Indic.* **2021**, *129*, 107869. [CrossRef]