

Article

Geological Hazard Susceptibility Evaluation using Information Quantity, Deterministic Coefficient, and Logistic Regression Models and Their Comparison at Xuanwei, China

Shaohan Zhang¹, Shucheng Tan^{2*}, Lifeng Liu¹, Duanyu Ding³, Yongqi Sun¹, Jun Li⁴

¹ Institute of International Rivers and Eco-Security, Yunnan University, Kunming 650500, Yunnan, China

² School of Earth Science, Yunnan University, Kunming 650500, Yunnan, China

* Correspondence: shchtan@yun.edu.cn

³ Faculty of Architecture and City Planning, Kunming University of Science and Technology, Kunming 650500, Yunnan, China

⁴ Yunnan Architectural Engineering Design Company Limited, Kunming 650501, Yunnan, China

Abstract: In China, the majority of mountainous regions are characterized by complex topography and a delicate, sensitive geological environment. Coupled with a generally underdeveloped infrastructure and numerous unreasonable human engineering activities, these regions are often highly susceptible to geological disasters. Geological hazards can cause significant damage to human lives and property, impeding the development of mountainous areas. Consequently, researching the assessment of geological hazard vulnerability is crucial for disaster prevention, emergency management, and economic development in these regions. This study focuses on Xuanwei City and selects eight factors for evaluation, including elevation, gradient, slope aspect, normalized vegetation index, stratigraphic lithology, distance from faults, distance from rivers, and distance from roads. These factors are chosen based on a comprehensive analysis of the spatial and temporal distribution of geological hazards and disaster incubation conditions. Two paired models, the deterministic coefficient model + logistic regression model (CF+LR) and the information quantity model + logistic regression model (I+LR), were employed to quantitatively assess the study area. The accuracy of these models was evaluated using ROC curves and AUC values. The results indicate that: (1) The AUC values for the CF+LR and I+LR coupled models are 0.799 and 0.772, respectively, demonstrating that both models can objectively and reliably assess the vulnerability to geological hazards in the study area; (2) Based on the CF+LR model calculations, the geological hazard susceptibility of Xuanwei City can be categorized into four zones: extremely high susceptibility (6.09%), high susceptibility (31.08%), medium susceptibility (32.26%), and low susceptibility (30.57%); (3) The CF+LR model more accurately represents the evaluation results and offers a strong reference value.

Keywords: information model; certainty factor model; logistic regression model; geological hazard; susceptibility; Xuanwei city

1. Introduction

Assessing the susceptibility of geological disasters is crucial in evaluating the risk they pose. This refers to the likelihood of geological disasters occurring due to various internal factors, such as topography, geomorphology, stratigraphic lithology, and geological structure, in combination with external factors such as extreme rainfall and earthquakes. Accurate prediction and classification of vulnerability for geological hazards are essential for regional economic development, ensuring people's well-being and driving high-quality development. In the 1960s, researchers began investigating the assessment of geohazard susceptibility, transitioning from initial qualitative evaluation methods to quantitative approaches. Since the 21st century, advancements in 3S technology and mathematical

and statistical theories have led to an increasing number of studies utilizing quantitative evaluation techniques. Wang CM et al. [1] applied the information quantity method to assess and delineate geological hazard vulnerability in Wen County; Sun et al. [2] integrated the Dempster-Shafer evidence theory to evaluate geohazard susceptibility along the northern route of the Sichuan-Tibet highway; and Kohno et al. [3] employed the AHP-GIS method to examine earthquake-induced slope damage. However, single quantitative evaluation methods are prone to mutual interference among evaluation factors. If there is a high correlation between these factors, the accuracy of the evaluation model may be compromised, causing the assessment results to fall short of expectations. To overcome this, the logistic regression model has been used to test independence and diagnose covariance between evaluation factors, eliminating highly correlated factors and determining weights of evaluation factors more objectively and scientifically.

Owing to the limitations of single evaluation methods and the benefits of logistic regression models, an increasing number of researchers are effectively integrating the two approaches to assess geological hazard vulnerability in their study areas. After conducting field investigations and tests, the results have been found to be consistent with on-site conditions, and the evaluation model demonstrates greater accuracy. Riegel et al. [4] utilized a geographic information system (GIS) and logistic regression model to investigate landslide susceptibility in Novo Hamburgo, with the model effectively determining landslide occurrence probability and exhibiting strong predictive capabilities. Devkota et al. [5] employed GIS-based certainty factors, entropy index, and logistic regression models to evaluate landslide susceptibility, focusing on the Mugling–Narayanghat road section in the Nepal Himalaya; their results indicated that the model demonstrated high accuracy. Fan et al. [6] assessed geological hazard vulnerability in Wenchuan County using an information quantity-logistic regression model and compared the evaluation accuracy of the information quantity model, logistic regression model, and coupled model simultaneously. Luo et al. [7] applied the CF-Logistic model to evaluate landslide susceptibility in the Jiuzhaigou scenic area, with their findings revealing that the coupled evaluation model had higher accuracy than the single evaluation model and that its zoning results were highly reliable. This study focuses on Xuanwei City, combining the information quantity model and the deterministic coefficient model with the logistic model to evaluate geological hazard vulnerability in the study area. This is achieved through a comprehensive and objective analysis of existing spatial and temporal distribution patterns and disaster conditions of geological hazards in the area. Eight influencing factors were selected: elevation, slope, slope aspect, normalized difference vegetation index (NDVI), stratigraphic lithology, distance from faults, distance from rivers, and distance from roads. By comparing various geological hazard susceptibility evaluation models, results, and accuracy within the same study area, this research explores a coupled quantitative evaluation model with high precision and reliability. This model serves as a reference for assessing geological hazard susceptibility in the study area and potentially in other county-level regions. It offers a scientific basis for disaster prevention and mitigation, development planning, and the selection of suitable locations for major projects in Xuanwei.

2. Overview of the Study Area

Xuanwei City, situated in Qujing City within Yunnan Province, is located in the northeastern part of Yunnan Province (Figure 1. The grid and vector data of the study area are derived from the geospatial data cloud [<https://www.gscloud.cn/>] and processed by the ArcGIS platform.). Its geographical coordinates range from 103°35'30"E to 104°40'50"E longitude and 25°53'30"N to 26°44'50"N latitude. The city occupies a strategic position at the Yunnan-Guizhou interprovincial junction, often referred to as the lock and key into Yunnan Province. To the east, Xuanwei shares borders with Panxian and Shuicheng in Guizhou Province; to the west, it is separated from Huize County; to the south, it is connected to Fuyuan County and Zhanyi County; and to the north, it adjoins Weining in Guizhou Province. Due to its high altitude, cool climate, significant diurnal temperature variations, and relatively low rainfall, the region is suitable for the cultivation of cold-resistant and drought-tolerant crops, generally classified as an alpine mountainous area. Xuanwei City is located at the edge of the Yunnan Plateau and is a transition slope to the Guizhou Plateau, with the terrain generally higher in

the northwest and lower in the southeast. The highest peak in the area is the main peak of East Mountain which has an elevation of 2868 meters, and the lowest point is the Lakeng Iron Cable Bridge of the Beipan River, with an elevation of 920 m.

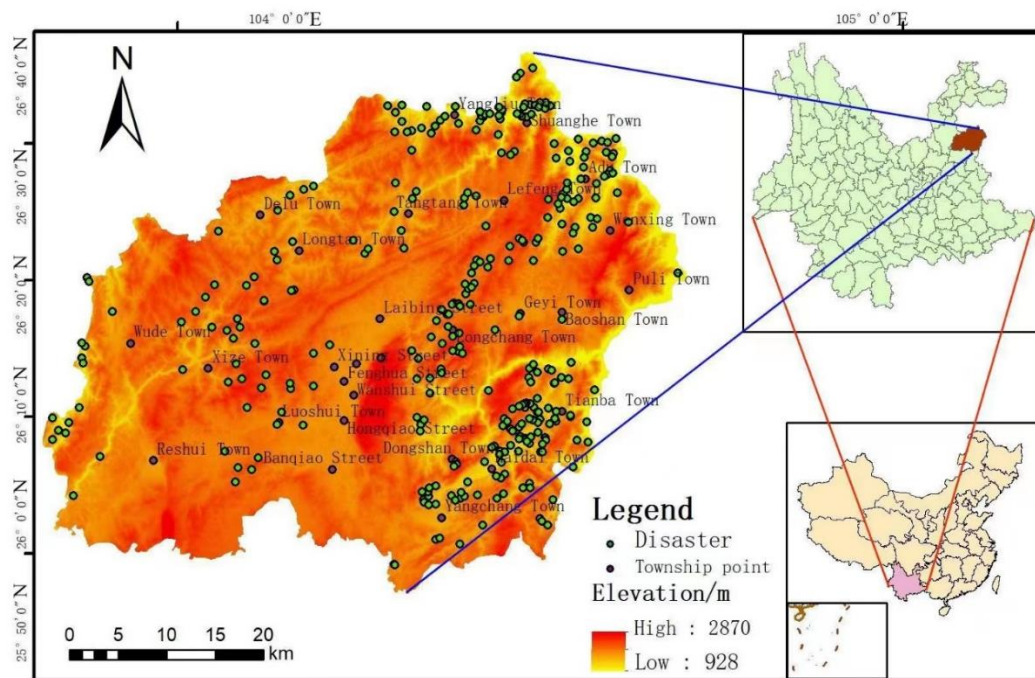


Figure 1. Overview of the study area.

The primary rivers in the region are part of the upper reaches of the Beipan River within the Pearl River system, originating from the northwestern foothills of MaXiong Mountain in Xuanwei City. The area features two major tributaries: the GeXiang River, which flows through Xuanwei City, and the Kedu River, which flows through Kedu. The region's main landforms encompass three types: basin, tectonic erosion and denudation, and karst. The stratigraphic lithology in the area is complex, with the most extensive exposures being the Xuanwei Formation and Emeishan Basalt Formation of the Diametron System. This is followed by the tuffs of the Qixia Maokou Formation, also of the Diametron System, and Carboniferous and Devonian carbonates distributed in the third area. The rock masses throughout the study area exhibit varying degrees of hardness, with generally weak rock quality, low weathering resistance, poor mechanical properties, and suboptimal engineering geological characteristics. The area has a temperate monsoon climate with three-dimensional climatic characteristics, with higher temperatures in the basin and river valley and lower temperatures in the mountainous areas. The annual precipitation ranges from 623 mm to 1348 mm, with an average of 940 mm and concentrated rainfall from May to October.

3. Research Methods

3.1. Certainty Factors Methods (CF)

The Certainty Factor (CF) model is a bivariate statistical analysis probability function that can be used to assess the sensitivity of geohazard occurrence for each causative factor. The model assumes that future geohazard events will occur under the same working conditions as past geohazard events. The calculation formula for the model is:

$$CF = \begin{cases} \frac{PP_a - PP_s}{PP_a(1 - PP_s)} & PP_a \geq PP_s \\ \frac{PP_a - PP_s}{PP_s(1 - PP_a)} & PP_a < PP_s \end{cases} \quad (1)$$

Where ' PP_a ' represents the conditional probability of a geological hazard occurring in the evaluation factor classification level 'a'. It is typically calculated as the ratio of the number of geological hazard points in level 'a' to the ratio of the area of the study area occupied by level 'a'. ' PP_s ' represents the prior probability of the event occurring in all data. It is expressed as the ratio of the number of geohazard sites in the whole study area to the area of the whole study area.

The CF model's values range from -1 to 1, where a positive value closer to 1 indicates a higher predictability of geohazard occurrence, meaning that this type of influence factor significantly contributes to geohazard susceptibility. A negative value closer to -1 means that the certainty of geohazard occurrence is lower, indicating that a geological disaster is less likely to occur under the effect of this type of influence factor. When the result is close to zero, it suggests that the influence of this factor on geohazard susceptibility cannot be determined.

3.2. Information Methods (I)

Information methods for geological hazard susceptibility assessment are based on information theory. The methods were first applied to landslide prediction by Yan Guozhen and later expanded to include other geological hazards [8]. The underlying logic of these methods is to convert the measured values of impact factors reflecting the occurrence of geological hazards into informative values that serve as quantitative indicators for vulnerability zoning. The information value of each evaluation factor is calculated by combining the geological hazard point data with the raster layer of each evaluation factor in the study area. The information quantity values of each layer are superimposed through the raster calculator of the ArcGIS platform to obtain the geological hazard susceptibility evaluation result for the study area. A higher information value indicates a greater susceptibility to geological hazards, and it is calculated using the formula:

$$I(x_i, H) = \ln \frac{N_i/N}{S_i/S} \quad (2)$$

where: ' $I(x_i, H)$ ' represents the information value provided by the evaluation factor ' x_i ' concerning the occurrence of geohazards, which reflects the likelihood of geohazards. A larger I value indicates a higher probability of occurrence. 'S' denotes the total area of the evaluation unit within the study area; ' S_i ' is the area of the cell occupied by the evaluation factor ' x_i ' in the study area; 'N' reeferes to the total number of geological hazard units in the study area; ' N_i ' represents the number of geohazard units containing the evaluation factor ' x_i ' in the study area.

3.3. Logistic Regression Model (LR)

The logistic model is a frequently used statistical analysis model for binomial dependent variables. It is used to describe the relationship between the occurrence of a geological hazard, where 0 represents no geological hazard and 1 represents occurrence, and multiple independent variables ($X_1, X_2, X_3, \dots, X_n$). In the logistic model, the independent variable is the CF value (or information value) for each evaluation factor grading index. This value can be continuous or discrete and does not need to satisfy a normal frequency distribution [7]. The functional expression of the model is:

$$\begin{cases} P(Y = 1|X) = \frac{1}{1+e^{-Z}} \\ Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n \end{cases} \quad (3)$$

Formula (3) describes the logistic model: 'P' represents the probability of geohazard occurrence, ranging from 0 to 1, where 0 indicates that a geohazard cannot occur, while 1 indicates that it will occur. 'n' represents the number of evaluation factors, ' β_i ' is the logistic regression coefficient, and ' x_i ' represents the deterministic coefficient value (or information value) for each evaluation factor.

3.4. Certainty Factors - Logistic Regression Model (CF+LR)

The CF values obtained from equation (1) for different levels of each evaluation factor are utilized as the independent variables in the CF-LR model. The CF values are input into the binary logistic regression analysis model in SPSS software to calculate the logistic regression coefficients of

each evaluation factor. To determine the weights of the evaluation factors, factors with high covariance are eliminated. By substituting the obtained β_i values from the logistic regression analysis into equation (3) using the corresponding deterministic coefficient values, the susceptibility evaluation results of the CF-LR model can be obtained.

3.5. Informativeness-Logistic Regression Model (I+LR)

Similarly, Equation (2) enables the calculation of informativeness values for various levels of each evaluation factor. These computed informativeness values serve as the independent variables for the I-LR model. Utilizing SPSS software, the information value for each factor is substituted, and logistic regression analysis is conducted to obtain logistic regression coefficients. Following a correlation analysis of evaluation factors, highly correlated factors are excluded. The aforementioned procedure is repeated to derive the weights for each evaluation factor. By integrating the information quantity values with the logistic regression coefficients through Equation (3), the I-LR model for geo-hazard susceptibility evaluation can be successfully implemented.

4. Susceptibility Evaluation

4.1. Selection and grading of the evaluation factor

The selection of evaluation factors and models are the two most crucial steps in geological hazard susceptibility evaluation, as they directly influence the scientific accuracy and feasibility of the evaluation results [9]. To select factors for geological hazard susceptibility evaluation, evaluation factors should be chosen in conjunction with field survey information [10]. Informed by prior research and the field investigation data from the study area, this paper primarily examines the natural influencing factors, such as geological conditions and geographical environment in the study area, and selects elevation, slope, aspect, NDVI, stratigraphic lithology, distance from faults, distance from rivers, and distance from roads as evaluation indices for geological disaster susceptibility. Taking into account the spatial and temporal evolution characteristics and disaster-causing mechanisms of geological disasters, the classification results are depicted in Fig. 2.

(1) Elevation: Elevation represents the macroscopic landform within a specific area. Numerous research findings indicate that the occurrence and elevation distribution of geological disasters exhibit significant regional patterns [11]. Simultaneously, elevation largely determines the movement potential energy of the disaster body. The higher the elevation, the greater the dynamic potential energy stored after sliding, the greater the impact on the disaster-bearing body, and the more severe the resulting loss. Considering the spatial distribution characteristics of geological disaster points and elevation in the study area, elevation is divided into five grades: <1000 m, 1000 m~1200 m, 1200 m~1500 m, 1500 m~2000 m, >2000 m (Fig. 2a).

(2) Slope: In the study area, the slope magnitude is intimately connected to the extent of geological disasters. The displacement and impact force of the disaster body largely depend on the slope size. Moreover, the formation mechanism of geological disasters and the critical point of anti-slipping for the sliding body are significantly controlled by the slope factor. The slope of the study area is divided into four grades: <10°, 10°~30°, 30°~50°, >50° (Figure 2b).

(3) Slope aspect: Assuming equal vegetation coverage, the sunny slope contains abundant water and heat, and the rock mass's internal water tends towards saturation. Under water infiltration, the initiation conditions for geological disasters are lower, making disasters more likely to occur [12]. Utilizing the surface analysis function of ArcGIS software, the slope aspect information of the study area was extracted from the DEM data, and it was divided into north, northeast,

east, southeast, south, southwest, west, and northwest (Figure 2c).

(4) Normalized Difference Vegetation Index (NDVI): Vegetation serves the functions of slope protection, stabilization, and soil conservation which contribute to slope stability [13]. Generally, regions with high vegetation coverage exhibit less severe geological disaster development. The coarse roots of vegetation provide significant tension to the slope body and securely anchor to the slope body, making it more resistant to forming a sliding belt under rainwater infiltration. In this study, the normalized vegetation index in the study area is divided into five categories: <0 , $0.0-0.2$, $0.2-0.4$, $0.4-0.6$, >0.6 (Figure 2d).

(5) Stratum lithology: Lithology reflects the physical and chemical properties of the minerals that compose the rock mass. In geological disaster susceptibility evaluation, the chemical properties primarily manifest as the chemical reactions between the minerals in the rock mass and other factors (water, atmospheric rainfall, fertilizer, etc.), which diminish the rock mass's original strength. The physical properties are more evident in the structure, mechanical properties, and engineering geological properties of the mineral itself. The lithology of the study area is divided into four categories: loose soil, soft rock, soft and hard interbedded rock, and medium hard rock (Figure 2e).

(6) Distance from faults: Geological disasters typically occur in areas with more active fault structures, and the two are closely related. Particularly in the cross-composite parts of regional fault structures, the rock is relatively fragmented, often evolving into a structural condition conducive to the formation and development of geological disasters [14]. Based on the 1:50,000 geological map of the study area, fault belt information is extracted using the ArcGIS platform, and a 500 m interval buffer zone is established. The study area's distance from faults is divided into five categories: <500 m, 500 m -1000 m, 1000 m -1500 m, 1500 m -2000 m, >2000 m (Figure 2f).

(7) Distance from rivers: Rivers not only alter surface morphology but also constitute a major cause of geological disasters. Rivers exert an erosive effect on the slopes on both sides. Under the cyclic erosion of hydrodynamics, slopes can easily form an empty face, causing the gravity of the upper rock mass to exceed the critical tension it can withstand and thus triggering geological disasters. Based on the distribution characteristics of the water system and geological disasters in the study area, the distance from rivers is divided into six grades: <200 m, 200 m -400 m, 400 m -600 m, 600 m -800 m, 800 m -1000 m, >1000 m (Figure 2g).

(8) Distance from roads: Roads represent the impact of human engineering activities on rock and soil. During the construction of essential projects, excavating mountains and cutting slopes are inevitable processes. This can cause rock and soil masses to develop gaps due to vibration and disturbance, which facilitates water infiltration and alters the natural stress state of the rock and soil masses, reducing their cohesion and internal friction angle. Consequently, the sliding body is more likely to exceed the equilibrium state. Based on the vector data of the main roads in the study area, this paper establishes a buffer zone with 300 m as a segment in ArcGIS software and divides the distance from roads into six categories of 1500 m (Figure 2h).

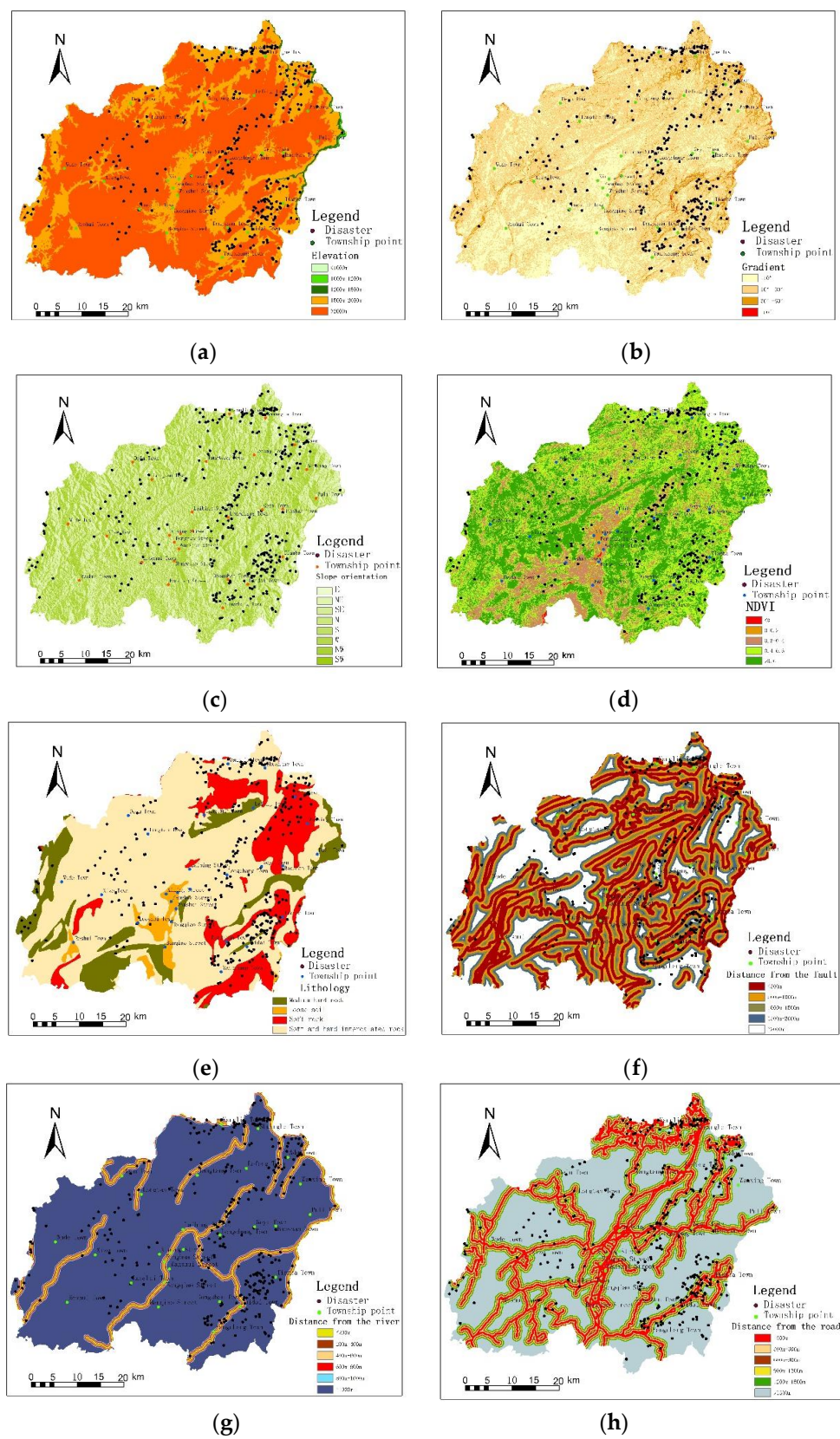


Figure 2. Evaluation factors grading chart.

4.2. Evaluation model

(1) Calculation of CF and I value

The evaluation factor layers were transformed into grid layers and reclassified according to their respective classifications. The number of disaster points and the area of each classification were then

counted to determine the CF and I values for each evaluation factor level using equations (1) and (2), as presented in Table 1. The CF and I values not only demonstrate the relative significance of different grades within the same evaluation factor but also show the contribution rate of each evaluation index in geological disasters.

The CF value characterizes the certainty of geological disasters objectively and fairly. The closer the CF value is to 1, the higher the probability of geological disasters. Based on the results in Table 1, it can be concluded that areas with elevations of <1000 m and 1000 m-1200 m have the lowest certainty of geological disasters, and geological hazards are less likely to occur. A gradient greater than 50° has a high certainty coefficient, indicating a high probability of geological disasters. However, it cannot be accurately determined whether the northeast, east, and south have a driving effect on the occurrence of geological hazards, whereas the certainty of geological disasters in the southwest and west is the lowest. Additionally, the lush the vegetation in the area, the less likely it is for geological disasters to occur. Geological hazards mostly occur in areas close to the fault zone, and the number of geological disaster points decreases with an increase in distance from the fault zone. Furthermore, geological disasters are prone to occur in soft rock mass areas. The CF value is highest in the range of 200 m-400 m from the water system, indicating a high probability of geological disasters in this environment. The influence of distance from the road on geological disasters is inversely proportional to a certain extent, indicating that the farther the distance from the road, the less prone to disasters. However, it is noteworthy that the general law of slope and disaster distribution in this paper is different from previous studies. The reasons for this difference are insufficient accuracy of geological disaster record data, errors in the information extracted by ArcGIS multi-value extraction to point tool during operation, and high-quality vegetation growth in the landform with a slope between 30° and 50°, which has a positive effect on slope reinforcement and slope protection, thereby reducing the probability of geological disasters. When the slope is >50°, the vegetation is sparse, which weakens the ability of the surface cover to fix itself, and the tensile stress of the landslide is increased, making geological hazards more likely to occur.

Table 1. Calculation results for information value and CF value across various evaluation factor classification levels.

Evaluation factor	Level	I	CF	Evaluation factor	Level	I	CF
Elevation	<1000m	0.0000	-	Lithology	Loose	-	-
		0	1.0000		soil	1.1908	0.6960
	1000m-1200m	0.0000	-		Soft	6	5
		0	1.0000		rock	0.5367	0.4153
	1200m-1500m	0.4128	0.3382		Soft and	5	7
		5	5		hard	-	-
Slope	1500m-2000m	0.8313	0.5645	Distance from the fault	inter-layer	0.0490	0.0478
		9	8		rock-	4	6
	>2000m	-	-		mediu	-	-
		0.8433	0.5697		m-hard	0.6720	0.9582
	<10°	-	-		rock	2	3
		0.4066	0.3341		<500m	0.2587	0.2280
	10°-30°	5	4		500m-	9	3
					1000m-		
					1000m-	0.0501	0.0489
					1500m-	7	3
					1500m-		

Gradient	30°-50°	0.2099	0.1894	Distance from the road	2000m	-	-
	>50°	9	1		>2000m	0.3195	0.2735
		-	-			8	6
		0.4416	0.3570			-	-
		7	5			0.1943	0.1766
						6	5
		1.7273	0.8222			-	-
		2	8			0.3959	0.3269
						5	7
		0.1466	0.1364			0.5620	0.4300
		5	1			8	0
		0.0908	0.0868		<200m	0.7943	0.5481
		4	4		200m-	2	4
		-	-		400m	0.6528	0.4794
		0.0026	0.0026		400m-	3	5
		9	9		600m		
NDVI	North	-	-	Distance from the river	600m-	0.7041	0.5055
	Northeast	0.3515	0.2963		800m	5	0
	East	0	8		800m-		
	Southeast	-	-		1000m	0.4635	0.3709
	South	0.0954	0.0910		>1000m	2	5
	Southwest	6	5			-	-
	West	-	-			0.1634	0.9999
	Northwest	0.4036	0.9999			6	4
		9	5				
		0.0431	-			0.4383	0.3549
		2	0.9999			8	4
			2				
		0.3109	-		<300m	0.1015	0.0966
		5	0.9999		300m-	9	1
			0		600m		
		1.7735	0.8303		600m-	0.2265	0.2026
		9	2		900m	0	9
		0.1319	0.1236	Distance from the river	900m-	-	-
		7	4		1200m	0.2022	0.1830
	<0				1200m-	1	8
	0-0.2	0.3764	0.3137		1500m	-	-
	0.2-0.4	6	3		>1500m	0.1629	0.1503
	0.4-0.6					3	5
	>0.6	0.3486	0.2943			-	-
		3	6			0.2070	0.9999
						7	4
		-	-				
		1.5145	0.7800				
		1	9				

The degree of influence of evaluation factors on the susceptibility of geological disasters is reflected by the information quantity of observed values of geological disasters in the study area. The larger the information variable of evaluation factors, the more prone to geological hazards under the action of this factor. Based on the analysis of the calculation results of the information value, it can be concluded that geological hazards often occur in the 1500 m-2000 m altitude range, and when the slope is >50°, the information value is the largest, indicating a high probability of disasters. The northwest slope is prone to geological disasters due to the influence of the natural environment (climate,

sunshine, precipitation, etc.). Vegetation coverage is negatively correlated with the probability of geological disasters. Soft rock masses serve as the foundation and material source for the development of geological disasters. The impact of faults, rivers, and roads on geological disasters demonstrates a certain degree of correlation. Within a specific distance, these three evaluation factors contribute to the promotion of geological hazard occurrences.

(2) Weight calculation

Logistic regression analysis was conducted using SPSS software, where 660 disaster points were used as sample points, with 330 geological disaster points and 330 non-geological disaster points. The occurrence of geological disasters was taken as the dependent variable (1 for occurrence, 0 for non-occurrence), while the CF and I values of each evaluation factor classification level were taken as independent variables. The logistic regression coefficient (β) value and significance (Sig.) value were obtained through binary logistic regression analysis in SPSS software (Table 2), using a stepwise backward calculation method based on the partial likelihood ratio test results to eliminate variables [15]. Typically, if the Sig. value is greater than 0.05, it suggests that there is collinearity between the evaluation factor and other evaluation factors, necessitating collinearity diagnosis. In the collinearity diagnosis table (Table 3), a larger conditional index implies more significant collinearity with other evaluation factors. It is generally considered that multicollinearity exists when the conditional index exceeds 10 [16], requiring a comprehensive assessment of the intrinsic relationship between the evaluation factor and other predisposing factors for disasters, and the elimination of highly correlated evaluation factors. This process is repeated until all Sig. values of the evaluation factors are less than 0.05 (Table 4), indicating that the model is reasonable and accurate, with the corresponding β value representing the weight of the evaluation factor. In the CF model (Table 2), the significance values of distance from the road, distance from the fault, slope direction, and distance from the river were all greater than 0.05, indicating their collinearity with other factors. The collinearity diagnosis results showed that the variance ratio of the seventh dimension from the road distance and the ninth dimension from the river distance were significantly larger. Therefore, the distance factor from the river was removed, considering that roads and other human engineering activities in the study area were mostly built along the river. Furthermore, the aspect factor was eliminated after analysis, considering its minimal contribution to the susceptibility of geological hazards in the study area, and the close relationship of most geological disasters in the study area to the fault zone factor. In the information model (Table 2), the significance values of distance from the river and slope factor were greater than 0.05, indicating their significant collinear effects with other evaluation factors. These two factors were eliminated, and the remaining factors were used as independent variables for binary logistic regression analysis.

Table 2. Results of logistic regression analysis without removing covariate causative factors.

	Factor	β	Standard error	Wald	Variance	Significance	Exp(B)
CF	NDVI	1.689	0.236	51.325	1	0.000	5.415
	Distance from road	-0.032	0.159	0.040	1	0.841	0.969
	Distance from fault	0.796	0.416	3.662	1	0.056	2.217
	Elevation	1.206	0.166	52.516	1	0.000	3.341
	Slope	0.867	0.358	5.865	1	0.015	2.379
	Gradient	-0.073	0.185	0.158	1	0.691	0.929
	Distance from river	0.138	0.153	0.818	1	0.366	1.148
	Lithology	0.701	0.242	8.418	1	0.004	2.016
I	Constant	-0.076	0.172	0.196	1	0.658	0.927
	NDVI	1.130	0.151	55.894	1	0.000	3.096
	Distance from road	-2.031	0.250	66.281	1	0.000	0.131

Distance from fault	1.234	0.397	9.681	1	0.002	3.436
Elevation	1.011	0.127	63.194	1	0.000	2.747
Slope	0.395	0.329	1.443	1	0.230	1.485
Gradient	1.604	0.438	13.398	1	0.000	4.973
Distance from river	0.148	0.305	0.235	1	0.628	1.159
Lithology	0.598	0.260	5.290	1	0.021	1.819
Constant	-0.261	0.102	6.581	1	0.010	0.771

Table 3. The results of evaluation factors covariance diagnosis.

Mode I	Dimen- sion	Eigen- value	Condi- tion in- dex	Variance ratio								
				(Con- stant)	ND VI	Dis- tance from road	Dis- tance from fault	Eleva- tion	Slop e	Gradi- ent	Dis- tance from river	Lithol- ogy
CF	1	2.651	1.000	0.03	0.00	0.04	0.00	0.00	0.00	0.05	0.04	0.01
	2	1.376	1.388	0.01	0.18	0.01	0.13	0.25	0.02	0.01	0.00	0.04
	3	1.146	1.521	0.00	0.09	0.04	0.01	0.00	0.47	0.00	0.00	0.21
	4	0.963	1.659	0.00	0.22	0.00	0.49	0.00	0.04	0.00	0.00	0.22
	5	0.836	1.781	0.00	0.05	0.00	0.23	0.08	0.21	0.00	0.01	0.51
	6	0.728	1.909	0.00	0.41	0.03	0.12	0.41	0.13	0.07	0.00	0.00
	7	0.642	2.032	0.00	0.00	0.58	0.00	0.13	0.12	0.21	0.01	0.00
	8	0.480	2.351	0.02	0.03	0.14	0.01	0.04	0.01	0.49	0.35	0.01
	9	0.179	3.849	0.94	0.02	0.16	0.01	0.08	0.00	0.16	0.58	0.01
I	1	1.520	1.000	0.01	0.08	0.01	0.08	0.20	0.03	0.00	0.13	0.04
	2	1.374	1.052	0.21	0.01	0.28	0.01	0.01	0.06	0.00	0.01	0.05
	3	1.137	1.156	0.13	0.10	0.02	0.08	0.01	0.22	0.19	0.00	0.08
	4	1.073	1.190	0.02	0.03	0.00	0.00	0.00	0.06	0.36	0.18	0.23
	5	0.976	1.248	0.00	0.38	0.01	0.39	0.00	0.10	0.05	0.00	0.04
	6	0.815	1.366	0.01	0.22	0.05	0.01	0.00	0.28	0.05	0.16	0.37
	7	0.796	1.382	0.01	0.00	0.00	0.32	0.00	0.10	0.34	0.25	0.16
	8	0.707	1.466	0.11	0.07	0.13	0.10	0.57	0.00	0.00	0.26	0.01
	9	0.602	1.590	0.49	0.11	0.50	0.00	0.20	0.14	0.00	0.02	0.01

Table 4. Results of logistic regression analysis excluding covariate causative factors.

	Factor	β	Standard error	Wald	Variance	Significance	Exp(B)
CF	NDVI	1.683	0.235	51.358	1	0.000	5.380
	Distance from road	-0.032	0.158	0.041	1	0.840	0.969
	Distance from fault	0.809	0.415	3.808	1	0.048	2.247
	Elevation	1.237	0.163	57.672	1	0.000	3.445
	Slope	0.871	0.357	5.941	1	0.015	2.390
	Lithology	0.690	0.241	8.213	1	0.004	1.994
	Constant	-0.141	0.113	1.578	1	0.209	0.868
I	NDVI	1.113	0.150	54.732	1	0.000	3.044
	Distance from road	-2.054	0.249	68.262	1	0.000	0.128
	Distance from fault	1.243	0.396	9.842	1	0.002	3.464
	Elevation	1.037	0.125	69.372	1	0.000	2.821
	Gradient	1.627	0.436	13.918	1	0.000	5.091
	Lithology	0.656	0.256	6.543	1	0.011	1.926
	Constant	-0.271	0.101	7.184	1	0.007	0.763

(Note: By analyzing the field survey data of geological disasters in the study area, it is found that human engineering activities such as road and bridge construction and unreasonable slope excavation have a great influence on the existing geological disasters, so the distance factor from the road is not removed.)

(3) CF-Logistic model

A functional relationship was established between the CF value and the β value obtained from logistic regression analysis. The model formula can be expressed as:

$$\begin{cases} Y_1 = -0.141 + 1.683X_1 - 0.032X_2 + 0.809X_3 + 1.237X_4 + 0.871X_5 + 0.690X_6 \\ P_1 = \frac{1}{(1+e^{-Y_1})} \end{cases} \quad (4)$$

The formula above presents the probability P_1 of geological disasters, with a value ranging between [0 ~ 1]. $X_1 \sim X_6$ represent the evaluation factors: NDVI, distance from the road, distance from the fault, elevation, slope, and CF value of lithology.

(4) I-Logistic model

A model was developed based on the information value and the β value obtained from logistic regression analysis. The formula for the model is as follows:

$$\begin{cases} Y_2 = -0.271 + 1.113x_1 - 2.054x_2 + 1.243x_3 + 1.037x_4 + 1.627x_5 + 0.654x_6 \\ P_2 = \frac{1}{(1+e^{-Y_2})} \end{cases} \quad (5)$$

In the above formula, P_2 is the probability of geological disasters, and the value range is [0 ~ 1]. $x_1 \sim x_6$ represent the evaluation indexes: NDVI, distance from the road, distance from the fault, elevation, gradient and information value of lithology.

4.3. Evaluation results and verification

The weight and CF value of the CF-Logistic model are employed using formula (4), and grid weighting superposition is conducted through the map algebra function on the ArcGIS platform, ultimately obtaining the susceptibility evaluation results for the study area. The natural breakpoint method is then used to categorize the evaluation results into four grades: extremely high-prone area, high-prone area, medium-prone area, and low-prone area. This categorization helps generate the geological disaster susceptibility zoning map for Xuanwei City based on the CF-Logistic model (Fig. 3a). Similarly, the evaluation factor weight and information value of the CF-I model are utilized, and

the process is repeated using Formula (5) to produce the susceptibility zoning map for the study area based on the I-Logistic model (Fig. 3b).

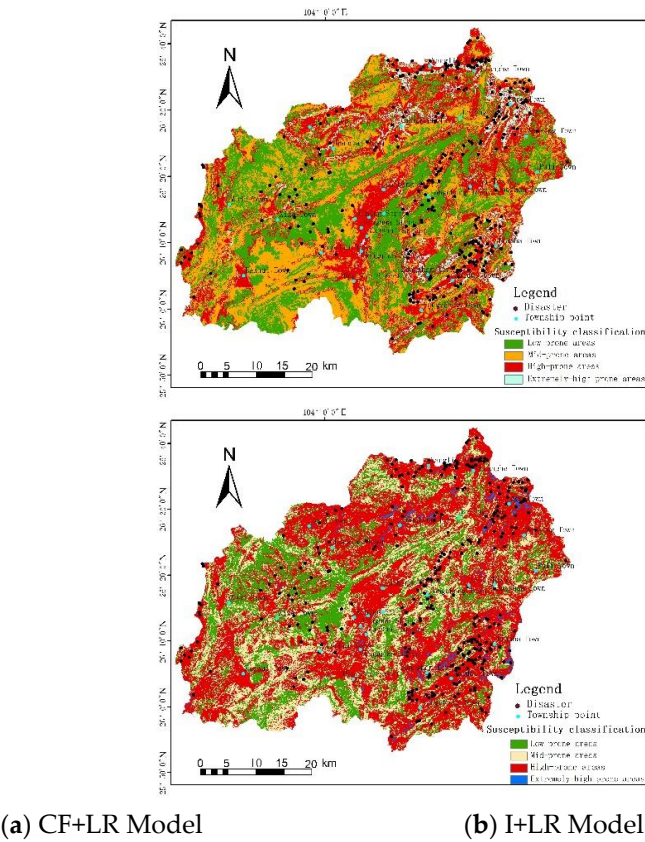


Figure 3. Geological hazard susceptibility zoning map of the study area.

Table 5. Statistical table of the results of geological hazard susceptibility zoning in Xuanwei City.

Susceptibility division	Evaluation model	Area/km ²	Area proportion/% (A)	Disasters/pcs	Proportion of disaster points/% (D)	Ratio (R=D/A)
Extremely-high prone areas	CF+LR	369.46	6.09	92	27.88	4.58
	I+LR	89.96	1.48	11	3.33	2.25
High prone areas	CF+LR	1886.78	31.08	159	48.18	1.55
	I+LR	3137.75	51.69	281	85.15	1.65
Medium susceptible area	CF+LR	1958.34	32.26	69	20.91	0.65
	I+LR	1584.39	26.10	32	9.70	0.37
Low susceptible area	CF+LR	1855.30	30.57	10	3.03	0.10
	I+LR	1257.78	20.72	6	1.82	0.09

By comparing and analyzing the geological disaster susceptibility zoning maps (Fig. 3a, Fig. 3b) obtained from the two evaluation models and the statistical table of the area and the number of disaster points for both models (Table 5), it is possible to verify whether the evaluation results of the two models are reasonable. This comparison also helps to determine whether the model can objectively and fairly represent the actual geological disaster susceptibility in Xuanwei City. From Table 5, it is evident that there is a significant difference in the proportion of disaster points in extremely high-prone and high-prone areas for both the CF-Logistic and I-Logistic models. A thorough analysis of the field survey data in the study area reveals that slope has a considerable impact on geological disasters in the extremely high-prone area. Alongside the analysis of the slope factor classification map (Fig. 2b), it is concluded that when the slope ranges between 30° and 50°, the region's geological disasters are extremely high-prone. The I-Logistic model, due to the high correlation between the

slope factor and other evaluation factors, differs from the CF-Logistic model in the susceptibility evaluation of geological disasters in Xuanwei City, as it lacks the key evaluation factor of slope. By comparing Figure 3a and Figure 3b, it can be seen that the CF-Logistic model and the I-Logistic model are roughly similar in the spatial distribution of the high-risk areas of geological disasters in the study area, and the difference exists in the partition area of the I-Logistic model. The high susceptibility area is characterized by sunnier slopes, resulting in significant physical and chemical impacts, reduced vegetation coverage, elevated surface runoff coefficients, and intensified erosion. These factors have created favorable conditions for geological disasters. However, the CF-Logistic model does not consider aspect as an evaluation parameter, and therefore, its final evaluation outcomes differ from those of the I-Logistic model.

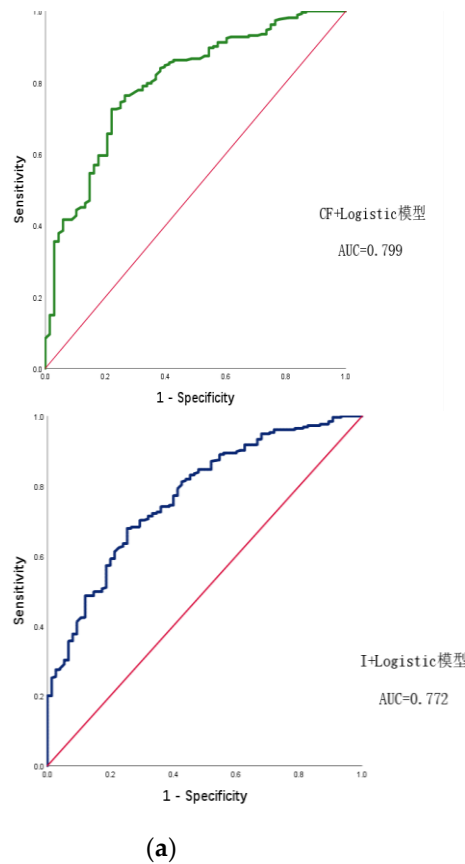


Figure 4. ROC curve of each mode.

The ROC curve is a prevalent method for verifying the accuracy of geological disaster-prone areas. The ROC curve can distinctly, intuitively, and accurately depict the relationship between the specificity and sensitivity of the evaluation model, demonstrating excellent test accuracy. As a result, it has been extensively employed in geological disaster susceptibility evaluations [16]. The AUC value of the ROC curve, representing the area under the curve, serves as a criterion for measuring the model's accuracy. Its value ranges between $[0.5 \sim 1]$, with a value closer to 1 indicating a more convex curve toward the upper left, which signifies higher model accuracy. As shown in Figure 4, the AUC value for the CF-Logistic model is 0.799, while the AUC value for the I-Logistic model is 0.772. This suggests that the CF-Logistic evaluation model exhibits higher evaluation accuracy and more reasonable evaluation results, making it better suited for assessing geological disaster susceptibility in Xuanwei City. The existing geological disasters in the study area are used to verify the accuracy of the CF-Logistic model. It is found that the model can well predict the susceptibility of geological disasters in Xuanwei City, and can be used as a scientific reference for disaster prevention and early warning, major project site selection, and land space planning in Xuanwei City (Figure 5).

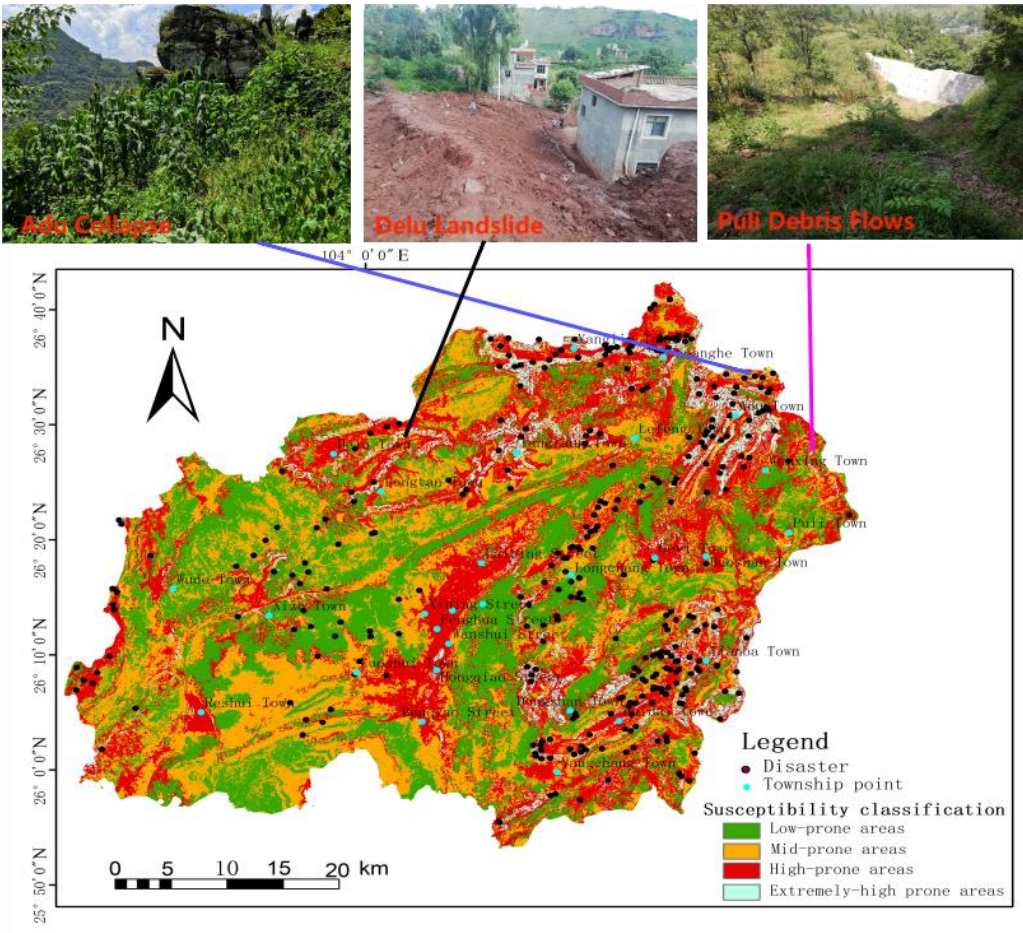


Figure 5. Verification of field disasters.

5. Conclusion

- (1) This study focuses on Xuanwei City as the subject of geological disaster susceptibility evaluation. After examining the spatial distribution and development environment characteristics of existing geological disasters in the region, elevation, slope, aspect, NDVI, stratigraphic lithology, distance from faults, distance from rivers, and distance from roads are chosen as evaluation factors. The CF-Logistic evaluation model and I-Logistic evaluation model are employed to divide geological disaster susceptibility in Xuanwei City using the ArcGIS platform.
- (2) Based on the two coupling models, it is determined that NDVI, elevation, and distance from faults have a significant influence on geological disaster susceptibility in the study area. Particularly when $NDVI < 0$; when the altitude ranges between 1500 m and 2000 m; and when the distance from the fault is less than 500 m, the CF value, I value, and logistic regression coefficients of the three factors are relatively large. This suggests that geological disasters are most likely to occur under these conditions, and areas with such an environment should be given significant attention.
- (3) The extremely high-susceptibility areas for geological disasters are mainly located along roads and densely populated regions. The geological structure in these areas is highly developed, with fragmented rock and significant influence from fault zones. High-prone areas are primarily found near rivers and faults, and the terrain in these regions is highly variable. Medium-prone areas are strongly influenced

by new tectonic movements, possess a relatively developed surface water system, and have complex geological environmental conditions. Low-prone areas are predominantly distributed in the central, north-eastern, and southwestern regions of Xuanwei City. The slope of these areas is mostly between 5° and 10° , hydrogeological conditions are relatively simple, and the disaster environment is not complicated.

(4) The partition results of the two models are roughly similar in spatial distribution. Differences in the partition area arise due to slight variations in evaluation factors among different models. Overall, the evaluation results of the two models align with the distribution of existing geological disasters in the study area and offer valuable reference for geological hazard risk assessments and disaster prevention and emergency work. The AUC values of the CF + Logistic model and the I + Logistic model are 0.799 and 0.772, respectively, indicating that both models meet the requirements for objective and scientific evaluations of geological disasters in Xuanwei City. The CF + Logistic model demonstrates higher evaluation accuracy.

(5) The distribution of hazardous rock mass in the study area is extensive. Engineering treatment would require considerable economic investment. At present, the most effective prevention and control measures are mass observation and preparedness, along with strengthened rainfall monitoring and restrictions on large-scale excavation to mitigate the impact of collapse disasters. Landslide disasters have a significant impact on the safety of people's lives and property in the study area, and collective relocation efforts are challenging and feasible. Based on the geological hazard susceptibility zoning results, appropriate engineering control measures should be selected according to local conditions to manage prone landslides, especially by clarifying prevention and control policies for geological disasters during flood seasons to reduce their impact on landslide disasters. Debris flow disasters mainly occur in deep valleys with steep terrain, fragmented rock, and heavy rainfall, causing severe damage to affected bodies. Planting trees in debris flow formation areas can help reduce soil erosion and control from the root; employing blocking engineering measures in circulation areas to obstruct debris flow; and constructing sedimentation fields or drainage ditches in accumulation areas to prevent debris flow from impacting villages or blocking rivers.

6. Discussion and Prospect

6.1. Discussion

China is among the countries with the most severe geological hazards globally, which constitutes a fundamental national condition. Conducting comprehensive surveys of disaster risks in countries and regions is an inevitable necessity and a requirement for high-quality economic and social development. The selection of evaluation methods, evaluation factors, and weight determination are critical components of the entire evaluation process. In the past, qualitative evaluation methods contained numerous subjective factors, and future-developed quantitative evaluation methods lacked thorough research on the correlation of evaluation factors. These issues can affect the accuracy of the evaluation results. The presented approach of coupling a single quantitative method with a logistic regression model can address both the high collinearity of factors in a single evaluation method and the challenge of the logistic regression model not being able to quantify the classification level of evaluation factors so that the evaluation results are scientific and credible. In this study, the following questions merit in-depth analysis: (1) In the CF-Logistic model, whether the inclusion of the highly correlated factor of distance from the road will directly impact the evaluation results; (2) The two

models in this study reveal a high degree of collinearity between the river and other evaluation indicators. As the river serves as an external force transforming surface morphology, it is scientifically reasonable to eliminate this factor.

6.2. Prospect

Changes in location affect the geological conditions, such as topography, geological structure, and stratigraphic lithology, which in turn influence the disaster mechanism of a region. Therefore, the evaluation model used for each region differs because of its unique characteristics. The accuracy of this evaluation model primarily depends on the quality, correlation, reliability, and representativeness of the input evaluation factor data. The selection of appropriate evaluation factors and models depends on the disaster-causing conditions of geological disasters in a region which can lead to more accurate prediction results that meet ideal requirements. In our future research, we aim to explore a set of quantitative evaluation models for geological disaster susceptibility. With the assistance of machine learning algorithms, the model can autonomously, objectively, rigorously, and impartially select evaluation factors to reduce the subjective interference of human factors on the model and accurately calculate weightage. The evaluation model is scientific, universally applicable, and can serve various disaster-prone areas with similar disaster-prone characteristics in human living environments.

Author Contributions: All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by [Shucheng Tan, Lifeng Liu, Duanyu Ding, Yongqi Sun, Jun Li]. The first draft of the manuscript was written by [Shaohan Zhang] and all authors commented on previous versions of the manuscript. All authors read and approved the final version.

Funding: This work was supported by the Science and Technology Innovation Team Program (Grant number YNEDUSTIT202202), and the Education Department of Yunnan Province and Famous teacher of Yunnan Province "Xingdian talents support program".

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The DEM and NDVI data are openly available in [the geospatial data cloud platform] at [<https://www.gscloud.cn/>]. The disaster sites and geological data are from the Yunnan Provincial Bureau of Nonferrous Geology. They can be obtained from author Shaohan Zhang upon reasonable request. Water system and road data can be obtained from [the National Catalogue Service for Geographic Information] at [<https://www.webmap.cn/>].

Acknowledgments: Heartful thanks for the data provided by the 317 teams of Yunnan Provincial Bureau of Non-ferrous Geology, and thank you to all the teachers who worked hard on this manuscript.

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

References

1. Wang, Q.; Guo, Y.; Li, W.; He, J.; Wu, Z. Predictive modeling of landslide hazards in Wen County, north-western China based on information value, weights-of-evidence, and certainty factor. *Geomat. Nat. Hazards Risk* **2019**, *10*, 820-835, doi:10.1080/19475705.2018.1549111.
2. Sun, Y.; Ge, Y.; Chen, X.; Zeng, L.; Liang, X. Risk assessment of debris flow along the northern line of the Sichuan-Tibet highway. *Geomat. Nat. Hazards Risk* **2023**, *14*, 2195531, doi:10.1080/19475705.2023.2195531.
3. Kohno, M.; Higuchi, Y.; Ono, Y. Evaluating earthquake-induced widespread slope failure hazards using an AHP-GIS combination. *Nat. Hazards* **2023**, *116*, 1485-1512, doi:10.1007/s11069-022-05725-w.
4. Riegel, R.P.; Alves, D.D.; Schmidt, B.C.; de Oliveira, G.G.; Haetinger, C.; Osório, D.M.M.; Rodrigues, M.A.S.; de Quevedo, D.M. Assessment of susceptibility to landslides through geographic information systems and the logistic regression model. *Nat. Hazards* **2020**, *103*, 497-511, doi:10.1007/s11069-020-03997-8.

5. Devkota, K.C.; Regmi, A.D.; Pourghasemi, H.R.; Yoshida, K.; Pradhan, B.; Ryu, I.C.; Dhital, M.R.; Althuwaynee, O.F. Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya. *Nat. Hazards* **2013**, *65*, 135-165, doi:10.1007/s11069-012-0347-6.
6. Fan, Z.; Gou, X.; Qin, M.; Fan, Q.; Yu, J.; Zhao, J. Information and logistic regression models based coupling analysis for susceptibility of geological hazards. *J. Eng. Geol.* **2018**, *26*, 340-347, doi:10.13544/j.cnki.jeg.2017-052.
7. Luo, L.; Pei, X.; Huang, R.; Pei, Z.; Zhu, L. Landslide susceptibility assessment in Jiuzhaigou scenic area with GIS based on certainty factor and Logistic regression model. *J. Engi. Geol.* **2021**, *29*, 526-535, doi:10.13544/j.cnki.jeg.2019-202.
8. Huang, R.; Xu, X.; Tang, C. *Geoenvironmental assessment and geohazard management*; Science Press: Beijing, 2008.
9. Zhao, X.; Tan, S.; Li, Y. Dongchuan district based on slope unit and combined empowerment method geological hazard risk evaluation. *J. Yunnan Univ. (Nat. Sci. Ed.)* **2021**, *43*, 299-300, doi:10.7540/j.ynu.20200141.
10. He, P.; Tong, L.; Guo, Z.; Liu, C.; Tu, J.; Wang, S.; Xu, J. Evaluation research on the landslide disaster liability in Zhada region of Tibet. *Sci. Technol. Eng.* **2016**, *16*, 193-200, doi:10.3969/j.issn.1671-1815.2016.25.033.
11. Tian, C.S.; Liu, X.L.; Wang, J. Geohazard susceptibility assessment based on CF model and Logistic Regression models in Guangdong. *Hydrogeol. Eng. Geol.* **2016**, *43*, 154-170, doi:10.16030/j.cnki.issn.1000-3665.2016.06.24.
12. Chen, L.; Li, L.; Wu, F.; Xu, Y. Evaluation of the geological hazard vulnerability in the Beiliu City based on GIS and information value model. *Earth Environ.* **2020**, *48*, 471-479, doi:10.14050/j.cnki.1672-9250.2020.48.060.
13. Dong, L. Evaluation of ecosystem service value and its driving force analysis based on landscape pattern: taking Chengdu plain and Longmen mountain transition zone as an example. Doctoral dissertation, Sichuan Normal University, Chengdu, 2017.
14. Wang, Z.W.; Li, D.Y.; Wang, X.G. Zonation of landslide hazards based on weights of evidence model. *Chin. J. Geotech. Eng.* **2007**, *29*, 1268-1273.
15. Wang, C.M.; Huang, J.; Li, Q.; Zhang, S. Evaluation of geological hazard vulnerability in Lyuliang City in Shanxi Province based on coupling of information content model and Logistic regression model. *Water Resour. Hydropower Eng.* **2019**, *50*, 132-138, doi:10.13928/j.cnki.wrahe.2019.03.018.
16. Du, Q.; Fan, W.; Li, K. Geohazard susceptibility assessment by using binary logical regression and information value model. *J. Catastrophol.* **2017**, *32*, 220-226.