Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 12, No. 1, 2024

Landslide Susceptibility Assessment Analysis based on IV and IV-IOE Models: A Case Study of Mao County

Fulu Sun [1, *], Yanan Lv [2]

[1] Institute of Disaster Prevention, Beijing, China
[2] Zhongke Pengyang (Hangzhou) Technology Co., Ltd, Hangzhou, China
*** Corresponding author:** Fulu Sun

Abstract: Taking Mao County as the study area, and based on existing data combined with the characteristics of landslide development in the study area, six evaluation factors including elevation, slope gradient, aspect, land use, NDVI (Normalized Difference Vegetation Index), and rainfall were selected. An independence test was conducted using Pearson correlation analysis to establish an evaluation index system for landslide disaster susceptibility. The study on the susceptibility of landslide geological disasters was carried out using both the Information Quantity Model and the Information Quantity-Entropy Index Model. The evaluation results were divided into five levels—extremely low, low, medium, high, and very high—using the natural break method based on GIS. The susceptibility of landslide geological disasters in the study area was clarified, and the accuracy was tested using the ROC curve. The experimental results show that the AUC values of the two evaluation models are 0.847 and 0.865, respectively, indicating that the Information Quantity-Entropy Index Model is superior to the Information Quantity Model and has stronger robustness in predicting the accuracy of landslide susceptibility.

Keywords: Susceptibility Assessment; Information Quantity Model; Information Quantity-Entropy Model; ROC.

1. Introduction

Landslides are one of the common geological disasters in the world, causing significant losses to construction and the lives and property of people in various countries around the world. Landslides are characterized by high frequency, wide distribution, and strong destructive power. China is one of the countries with the most severe landslide disasters in the world. Natural disasters such as rainfall, earthquakes, and floods often induce a series of landslides. Due to the sudden occurrence of landslide disasters, the precursors are generally not obvious, and most activities are intense, making prediction, forecasting, and prevention difficult, often catching people off guard and causing serious damage and loss. Landslide susceptibility [1] is based on the historical landslide data of the study area, and comprehensively analyzes various factors such as topography, geology, hydrology, climate, and human activities in the study area to determine the probability of landslides occurring in the study area. Carrying out spatial susceptibility evaluation of landslide disasters can understand "where landslides are likely to occur", so as to achieve targeted prevention and control of landslide disasters, and form a scientific, complete, and practical susceptibility evaluation system, which is of great significance for the risk management of landslide disasters, the safety of human life and property, and the future planning of cities.

International research on the evaluation of landslide disasters started earlier. Kumar [2] and others used the support vector machine model and the optimized support vector machine model for susceptibility prediction and research in the Mandakini River Basin in India, and the results showed that the optimized support vector machine model had higher accuracy. Yesilnacar E. and TopalT [3] used the logistic regression model and the neural network model for geological disaster susceptibility evaluation in Hendek, Turkey, and the neural network model was more accurate than the logistic regression model. Kalantar B [4] and others used the logistic regression model, support vector machine model, and artificial neural network model for landslide susceptibility evaluation in the Dodangeh Basin of Mazandaran Province, Iran, indicating that the selection method of the training set samples has a greater impact on the accuracy of the landslide disaster evaluation model. Jibson and Keefer [5] used discriminant analysis and multiple linear regression analysis to analyze and evaluate the landslide susceptibility on the edge of the Mississippi Plain. Youssef Ahmed Mohame [6] and others took the Aba Basin in the Aseer region of Saudi Arabia as the study area, and constructed a support vector machine model, random forest model, multiple adaptive regression spline model, artificial neural network model, quadratic discriminant analysis model, linear discriminant analysis model, and double beta model, a total of seven machine models for landslide susceptibility evaluation, and found that the random forest model had the highest accuracy.

Landslide disaster susceptibility is an important part of geological disaster risk assessment research, as well as an important basis for the evaluation of geological disaster risk and prevention projects. It studies the laws of regional landslide occurrence and distribution under different strata unit combinations, regional geological structural unit characteristics, topography, geomorphology, etc., and combines the existing landslide distribution in the region to define different landslide susceptibility levels. At present, the common methods of landslide susceptibility analysis are mainly divided into two categories: statistical analysis method and machine learning method. The model established by the statistical analysis method has independent variable parameters determined by the original data of the landslide, which can avoid human factors, and the obtained parameter results are more objective. However, in the process of selecting evaluation factors, it is more susceptible to subjective factors. The machine learning method can avoid the problem of interdependence between evaluation factors, but due to the large amount of model calculation and the difficulty in selecting model parameters, the evaluation results are prone to overfitting problems.

2. Data Collection and Processing

2.1. Data Acquisition

Geological disaster impact indicators mainly encompass four major categories: topography, geological environment, meteorological hydrology, and human activitiesd. Based on the available data and in conjunction with the characteristics of landslide development in the study area, this paper selects six factors as the evaluation indicators for the susceptibility of landslide disasters in the study area, which include elevation (a), slope gradient (b), aspect (c), land use (d), Normalized Difference Vegetation Index (NDVI) (e), and rainfall (f). Among these, aspect and land use are discrete factors, while the others are continuous factors.

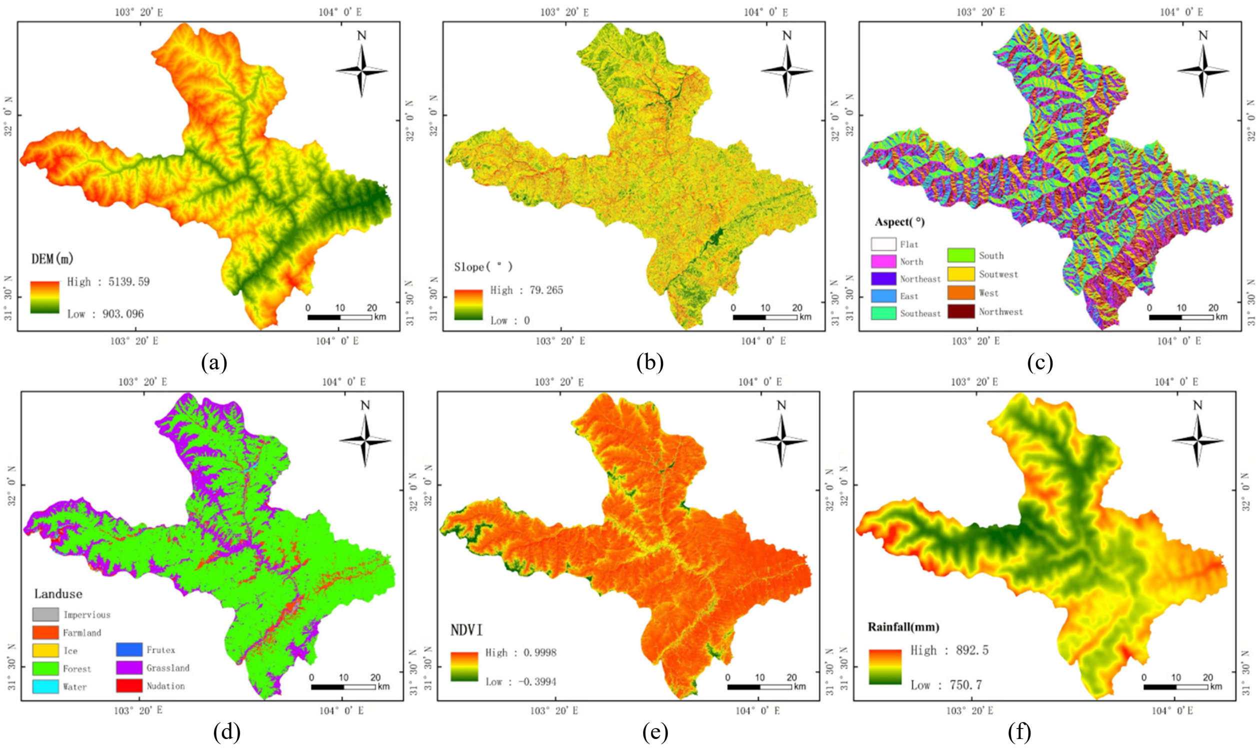

Figure 1. Influencing Factors

2.2. Data Pre-processing

Pearson correlation analysis [7], as an essential method in statistics, is primarily used to quantify the degree and direction of association between two or more variables. This method plays a pivotal role in exploratory data analysis, assisting researchers in revealing potential associations and the strength of the connections between variables. The outcome of the correlation analysis, that is, the correlation coefficient, is a value between -1 and +1 that describes the linear relationship between variables. A correlation coefficient of +1 indicates a perfect positive linear correlation between the two variables; a coefficient of -1 suggests a perfect negative linear correlation; and a coefficient of 0 implies no linear association between the variables under examination. In the realm of correlation analysis, the Pearson product-moment correlation coefficient and the Spearman rank correlation coefficient are the two most commonly used methods. This study selects the Pearson product-moment correlation coefficient as the analytical tool to delve into the degree of association between variables, providing a more accurate and reliable foundation for subsequent statistical analyses.

The Pearson correlation coefficient is a statistical measure used to assess the degree of linear association between two variables. This coefficient was proposed by the statistician Karl Pearson, hence the name. The Pearson correlation coefficient ranges from -1 to 1. If the value of r is 1, it means there is a perfect positive linear relationship between the two variables, where an increase in one variable leads to a proportional increase in the other. Conversely, if the value of r is -1, it indicates a perfect negative linear relationship, where an increase in one variable results in a proportional decrease in the other. If the value of r is 0, it suggests that there is no linear relationship between the two variables. The formula for calculating the Pearson correlation coefficient is as follows:

$$r = \frac{\sum_{i=1}^{n}(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X-\bar{X})^2(Y-\bar{Y})^2}} \tag{1}$$

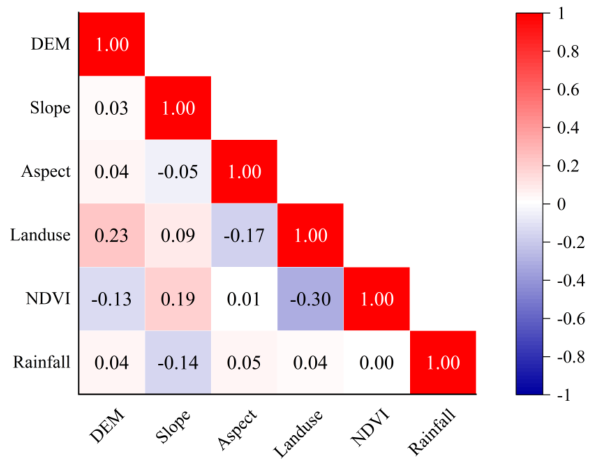

Figure 2. Pearson Correlation Analysis

Here, X and Y denote the sample values of two variables, while $\overline{X}$ and $\overline{Y}$ represent their sample means. The relatedness between the indicators intensifies as the value of the correlation coefficient increases. When the absolute value of the correlation coefficient falls within the range [0, 0.5), it is deemed that there is no correlation or the correlation is weak among the influencing factors.

3. Susceptibility Assessment Research

3.1. Landslide Susceptibility Assessment Based on the IV Model

The concept of the Information Quantity Model can be briefly summarized as: by calculating and analyzing the information quantity of different intervals of each influencing factor, it quantitatively represents the degree of impact of each attribute interval of environmental factors on landslide disasters, and then obtains the comprehensive level of information quantity by superimposing each factor. The core of this model lies in calculating the magnitude of the information quantity contributed by each influencing factor to the occurrence of the disaster [8].

$$I = \ln\left(\frac{n_i / N}{s_i / S}\right) \tag{2}$$

In the formula, n_i represents the number of landslide disaster points in the i-th level area, N represents the total number of landslide disaster points in the study area, s_i represents the area of the i-th level area, and S represents the total area of the study area.

Table 1. IV of Influencing Factors and IV-IOE

Factor	Class	IV	IOE	IV-IOE	Factor	Class	IV	IOE	IV-IOE
	[903.10,2044.96]	1.480		1.031		[0,17.05]	0.622		0.056
	(2044.96,2673.82]	0.611		0.425		(17.05,26.42]	0.246		0.022
DEM	(2673.82,3236.48]	-2.804	0.696	-1.953	Slope	(26.42,34.35]	0.012	0.089	0.001
	(3236.48,3815.68]	-2.116		-1.473		(34.35,43.27]	-0.384		-0.034
	(3815.68,5139.59]	0.000		0.000		(43.27,79.27]	-1.187		-0.106
	Flat	0.000		0.000		Farmland	2.122		2.354
	North	0.419		0.028		Forest	-0.575		-0.638
	Northeast	-0.350		-0.024		Shrubland	0.000		0.000
	East	-0.448		-0.030	Landuse	Grassland	0.062	1.108	0.069
	Southeast	0.043	0.068	0.027		Water	0.000		0.000
	South	-0.122		-0.008		Snow	0.000		0.000
	Southwest	0.022		0.001		Bareland	0.000		0.000
	West	-0.043		-0.003		Impervious	0.000		0.000
	Northwest	-0.343		-0.023		[750.7,782.4]	0.461		0.189
	[-0.4,0.4]	0.067		0.005		(782.4,802.9]	0.239		0.098
	(0.4,0.63]	0.112		0.009	Rainfal	(802.9,821.6]	-0.611	0.065	-0.251
	(0.63,0.78]	0.289	0.018	0.022		(821.6,841.2]	-0.637		-0.262
	(0.78,0.87]	0.276		0.021		(841.2,892.5]	0.392		0.161
	(0.87,1]	-0.399		-0.030					

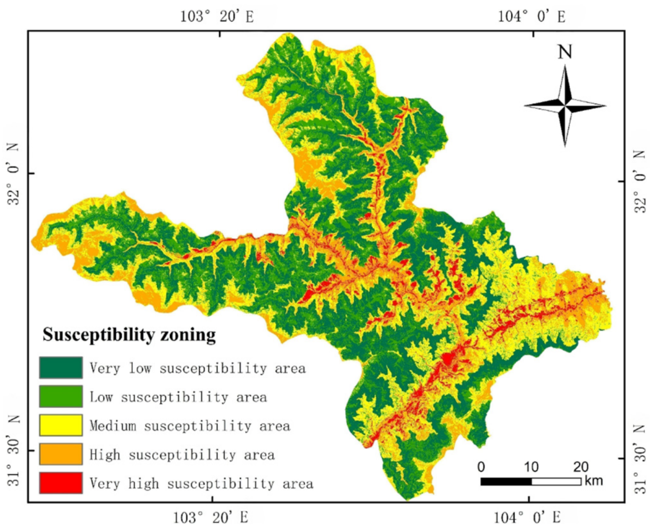

Figure 3. IV Model Evaluation Result Map

The calculated information quantities of each evaluation factor are superimposed. Based on the natural break method, the susceptibility of landslide disasters in the study area is divided into five levels: extremely low susceptibility zone, low susceptibility zone, medium susceptibility zone, high susceptibility zone, and very high susceptibility zone. The following is the landslide susceptibility map based on the information quantity model.

3.2. Landslide Susceptibility Assessment based on the IV-IOE Model

The information quantity model does not fully consider the comprehensive weight distribution of each factor. By using the information entropy method, we can more objectively assign weights to the influencing factors. This method eliminates human factors through quantification, ensuring the fairness and objectivity of the weight distribution. The weight values of the evaluation factors can be calculated using the following formula [9]:

$$I_i = \frac{H_{i\max} - H_i}{H_{i\max}} \tag{3}$$

$$P_i = \frac{1}{S_j}\sum_{j=1}^{S} FR_{ij} \tag{4}$$

$$W_i = I_i \times P_i \quad (5)$$

These formula normalizes the entropy values to obtain the weights, which can then be used to adjust the information quantity values of each factor accordingly. The calculated weighted information quantity of each evaluation factor is superimposed. Based on the natural break method, the susceptibility of landslide disasters in the study area is divided into five levels: extremely low susceptibility zone, low susceptibility zone, medium susceptibility zone, high susceptibility zone, and very high susceptibility zone. The following is the landslide susceptibility map based on the IV-IOE model.

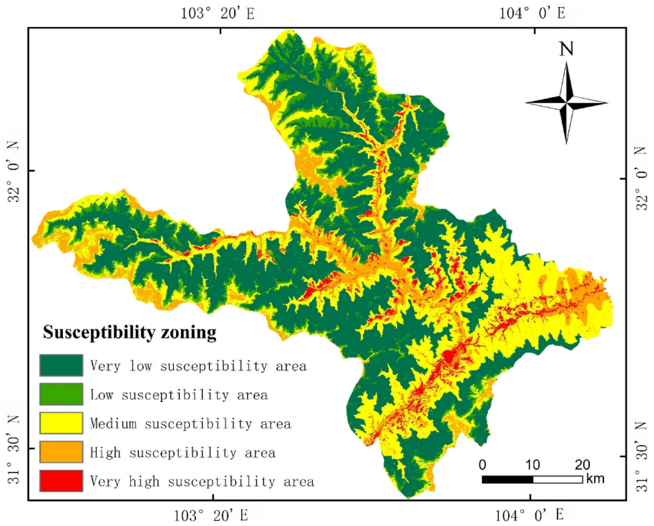

Figure 4. IV-IOE Model Evaluation Result Map

4. Model Precision Verification

The predictive accuracy of the two models was analyzed using the Receiver Operating Characteristic (ROC) curve [10]. The area under the ROC curve (AUC) represents the accuracy of the prediction, with an AUC value greater than 0.7 indicating that the model has good predictive performance. The precision verification results for the Information Quantity Model and the Information Quantity-Entropy Index Model are shown in the figure.

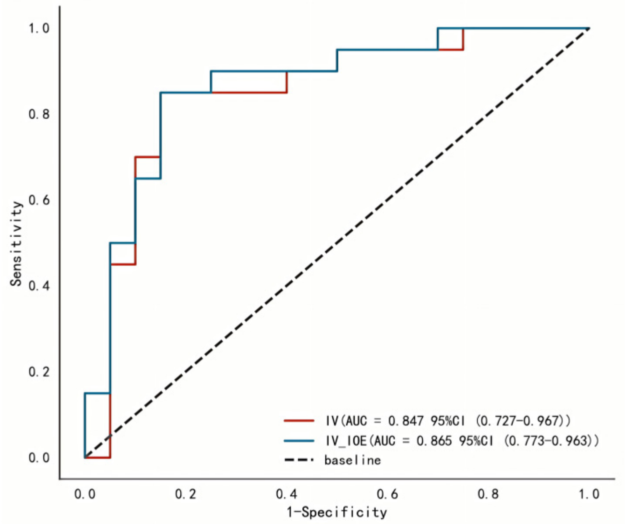

Figure 5. ROC Accuracy Comparison Test

5. Summary

(1) This paper takes Mao County as the study area and selects six influencing factors, including DEM (Digital Elevation Model), slope gradient, aspect, land use, NDVI (Normalized Difference Vegetation Index), and rainfall, to construct an evaluation index system for landslide disaster susceptibility. After Pearson correlation testing, the results show that the correlation between the influencing factors is low, making it suitable for the assessment of landslide geological disaster susceptibility.

(2) The IV (Information Value) model and the IV-IOE (Information Value-Entropy Index) model were used to assess and analyze the susceptibility of landslide disasters in the study area. The AUC values for the IV model and the IV-IOE model are 0.847 and 0.865, respectively. The experimental results indicate that the IV-IOE model has stronger robustness in the predictive accuracy of landslide susceptibility.

References

[1] Wang X ,Bai S .Landslide Susceptibility Mapping and Interpretation in the Upper Minjiang River Basin[J].Remote Sensing, 2023,15(20).

[2] Kumar D, Thakur M, Dubey C S et al. Landslide susceptibility mapping & prediction using Support Vector Machine for Mandakini River Basin, Garhwal Himalaya, India[J]. Geomorphology, 2017,295(Oct.15): 115-125.

[3] Yesilnacar E, Topal T Landslide susceptibility mapping: a comparison of logistic regression and neural networks methods in a medium scale study, Hendek region(Turkey)[J]. Engineering Geology, 2005,79(3-4): 251-266.

[4] Kalantar B, Ueda N, Al-Najjar H A H et al. Landslide susceptibility mapping at Dodangeh watershed, Iran, using LR and ANN models in GIS[C]. SPIE REMOTE SENSING.2018.

[5] Jibson, R .W. and Keefer, D.K.: 1989, Statistical Analysis of Factors Affecting Landslide Distribution inthe New Madrid Seismic Zone, Tennessee and Kentucky, Engineering Geology 27, 509-542.

[6] Youssef Ahmed Mohamed, Pourghasemi Hamid Reza. Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia[J].Geoscience Frontiers, 2021,12(2).

[7] Gogtay, Nithya J., and Urmila M. Thatte. "Principles of correlation analysis." Journal of the Association of Physicians of India 65.3 (2017): 78-81.

[8] SHADMAN ROODPOSHTI M, ARYAL J, SHAHABI H, et al. Fuzzy shannon entropy: A hybrid gis-based landslide susceptibility mapping method[J]. Entropy, 2016, 18(10): 343.

[9] LIU J, WANG J, XIONG J, et al. Hybrid models incorporating bivariate statistics and machine learning methods for flash flood susceptibility assessment based on remote sensing datasets [J]. Remote Sensing, 2021, 13(23): 4945.

[10] SUJATHA E R, SRIDHAR V. Landslide susceptibility analysis: A logistic regression model case study in Coonoor, India [J]. Hydrology, 2021, 8(1): 41.