Study of landslide susceptibility prediction based on information value model: a case study of Ningdu area

Ziyue Yan, Biao Zhang, Jianfei Guo, Jiawu Chen* and Zhongshan Cao
School of Civil Engineering and Architecture, Nanchang University, Nanchang 330031, China.

*Corresponding author e-mail: jiawuchen_ncu@sina.com

Abstract. The landslide susceptibility prediction (LSP) is of great importance to the prevention and control of regional landslide geology disasters. Taking Ningdu County of Jiangxi province as a case, this study obtains a total of 297 landslide locations in the study area and selects ten conditioning factors (elevation, slope aspect, slope, profile curvature, plan curvature, topographic relief, distance to rivers, lithology, NDVI, NDBI). The information value model (IVM) is used to predict the landslide susceptibility and the receiver operating characteristic curve (ROC) is adopted to evaluate the prediction accuracy of IVM. The result shows that the area under ROC (AUC) value of IVM is 0.838. It can be seen that IVM has a good prediction accuracy and also obtains a reasonable distribution characteristics of landslide susceptibility.

Key words: landslide susceptibility prediction; information value model.

1. Introduction
Landslide is one of the most common natural disasters, often causing huge economic losses and casualties [1]. In recent years, the frequency of rainfall landslides in Jiangxi province has gradually increased [2]. Therefore, it is of great practical significance and social value to carry out the study of landslide risk assessment for disaster reduction and prevention.

LSP means to explore the probability of landslide occurrence in a specific area by studying the joint action of several conditioning factors. It generally includes five steps including landslide cataloging, assessment unit division, selection of basic conditioning factors, construction of model and model performance testing [3]. The construction of models is crucial to the LSP, and the relevant models are mainly divided into qualitative and quantitative models. The common qualitative analysis is to evaluate and determine the main factors through expert ratings [4], such as analytic hierarchy process, et al [5-9].

The quantitative models mainly include information value model [10], support vector machine [11-12], logistic regression [13], neural network [14], evidence weight method [15] and other models [16-19]. In the above models, the IVM is widely used in the LSM, which is efficient in modeling and can obtain the weight of subcategories of different factors to the landslide occurrence. Hence, this study applies an IVM to predict the regional landslide susceptibility.

This study takes Ningdu County of Jiangxi province as a case, which considers the topography, hydrological environment, the basic geology and the surface cover factor. The IVM is used to predict
the landslide susceptibility and adopt forecast rate curve to evaluate the model accuracy. Finally obtains accurate result of LSP, so as to provide theoretical guidance and help for the related department in disaster prevention and mitigation.

2. Information value model

Information value, that is, the quality and quantity of obtained landslide information, refers to the specific situation of landslide in the research area [20]. Based on probability theory, information theory and engineering geological analogy method, this paper uses information value to represent the probability relative magnitude of landslide occurrence under the joint action of various factors. In addition, the greater the information value, the greater the probability of landslide occurrence [21]. The original formula is as follows:

\[ I(h, x_1, x_2, \ldots, x_n) = \log_2 \frac{P(h | x_1, x_2, \ldots, x_n)}{P(h)} \quad (1) \]

Where \( h \) represents landslide event, \( x_n \) represents conditioning factors, \( I(h, x_1, x_2, \ldots, x_n) \) represents the information value under the joint action of \( n \) conditioning factors, \( P(h) \) represents the probability of landslide, \( P(h | x_1, x_2, \ldots, x_n) \) represents the probability of landslide under the combination of all factors. According to the definition of conditional probability, equation (1) can be written as:

\[ I(h, x_1, x_2, \ldots, x_n) = I(h, x_1) + I_{x_1}(h, x_2) + \ldots + I_{x_1, x_2, \ldots, x_{n-1}}(h, x_n) \quad (2) \]

Where \( I_{x_1, x_2, \ldots, x_{n-1}}(h, x_n) \) refers to the information value \( x_n \) contributed to the landslide occurrence on the premise that \( x_1, x_2, \ldots, x_{n-1} \) is determined. The area ratio is generally considered to calculate regional landslide prediction, equation (2) can be expressed as:

\[ I(h, x_1, x_2, \ldots, x_n) = \log_2 \frac{S^* / A^*}{S / A} \quad (3) \]

Where \( P(h | x_1, x_2, \ldots, x_n) = \frac{S^*}{A^*}, P(h) = \frac{S}{A} \). \( S^* \) is the unit area where the landslide occurs with combination of the same factors \( x_1, x_2, \ldots, x_n \); \( A^* \) is the total area of landslide in the study area; \( S \) is the unit area with the same combination of factors; \( A \) is the total area of the study area. In most studies, single factor superposition formula is often used to calculate information value [1]. The following equation can be obtained:

\[ I = \sum_{i=1}^{n} I_i = \sum_{i=1}^{n} \log_2 \frac{S^*_i / A^*}{S_i / A} = \sum_{i=1}^{n} \log_2 \frac{N^*_i / M^*}{N_i / M} \quad (4) \]

Where \( I \) represents the prediction information, \( S^*_i \) represents the landslide area occupied by the number \( i \) conditioning factor; \( S_i \) represents the total area occupied by the number \( i \) conditioning factor. \( N^*_i \) represents the number of landslide grids occupied by the number \( i \) conditioning factor;
$M^*$ represents the total number of landslide grids in the study area; $N_i$ represents the total number of grids occupied by the number $i$ conditioning factor; $M$ is the total number of grids in the study area.

3. Study area and data source

3.1. Introduction of Ningdu County
In this paper, Ningdu County of Ganzhou city of Jiangxi province with relatively developed landslides is selected as the study case. The study area is located in between latitudes 26°05' ~ 26°31' N and longitudes 115°40' ~ 116°17'E, with a total area of about 1709 km² (Fig. 1). The study area is rugged, with numerous hills and mountains, which belongs to the subtropical monsoon humid climate, with an annual precipitation of 1500 ~ 1700 mm. At the same time, the rich rock types mainly distributed metamorphic rocks and carbonate rocks and the complicated geology structure provide favorable material conditions for the occurrence of landslide events.

3.2. Landslide evaluation unit division
There are 297 landslides in the study area, which are divided into 2087 landslide grid units in ARCGIS10.2 software. Continuous heavy rainfall and human engineering activities are two important factors for inducing landslides. Evaluation unit division is an important link in the process of LSP. Grid and slope unit are widely used by many scholars. In this study, grid unit is adopted as the evaluation unit due to its strong operability in division and applicability for small and medium-sized scale topographic maps [3]. In addition, this paper is selected 30 m×30 m grids for analysis and research based on the actual situation of the study area.

3.3. Selection of basic conditioning factors
Data sources of this study include:(1) digital elevation data (DEM) with spatial resolution of 30 m, which is used for landslide compilation; (2) field investigation data and landslide catalogue information; (3) geological map of Ningdu County with 1:50,000 measuring scale for extracting lithologic factor; (4)
Remote sensing data in Ningdu County used to extract Normalized Difference Vegetable Index (NDVI) and Normalized Difference Building Index (NDBI).

LSP is to obtain the possibility of regional landslide in the same engineering geology environment by analyzing the spatial distribution of basic conditioning factors and the interaction of them on the landslide without the effect of external inducing factors [22]. Basic conditioning factors are closely related to landslide occurrence, as shown in table 1.

Both slope aspect and slope are extracted from DEM data. The slope is 4.56~14.85 with the frequency ratio greater than 1, which is conducive to the landslide occurrence. Plan curvature is defined as the slope in the slope aspect, which reflects all the ridgelines and valley lines on the surface in the horizontal direction (Fig. 2(c)). Profile curvature is defined as the slope of slope, which represents the degree of change of slope in the vertical direction (Fig. 2(d)). Topographic relief is an index to describe the macroscopic terrain.

In this study, ARCGIS 10.2 software is used to obtain the rivers distribution map of DEM data through depression filling, flow extraction and river connection, and then polycyclic buffer analysis is conducted (Fig. 2(f)). Lithology is an important internal factor of landslide. NDVI represents the degree of surface vegetation coverage, while NDBI refers to the density of surface buildings.

After obtaining the basic conditioning factors, the frequency ratio is used to explain the nonlinear response relationship between the landslide and the basic conditioning factors. The larger the frequency ratio is, the more favorable the landslide occurrence [23]. In this study, the selected basic conditioning factors are divided into 5 classes using the natural discontinuity method in ARCGIS 10.2, and then obtain the number of grids and landslide grids in each interval of factors. The results are shown in table 1.

![Fig. 2 Conditioning factor maps: Slope aspect(a), Slope(b), Plan curvature(c), Profile curvature(d), Topographic relief(e), Distance to rivers(f), Lithology(g), NDVI (h), NDBI (i)](image)
4. Information value model predicts landslide susceptibility

In this paper, ten conditioning factors are selected to establish a landslide susceptibility evaluation index system after testing the correlation between factors, then calculate the information value corresponding to each factor subclass according to Eq(3), and each grid is assigned a total of information value with the superposition of information value of each factor using Eq(4). The larger the...
information value is, the more likely the landslide is to happen, and the size of the information value is use to quantify the possibility of the landslide occurrence.

The partition map of landslide susceptibility is obtained through the natural discontinuity method in ARCGIS 10.2 software (Fig. 3). The landslide susceptibility indexes (LSI) are divided into five classes (table 2): very high (3.527~3.871), high (3.365~3.527), moderate (3.219~3.365), low (3.063~3.219) and very low (2.542~3.063), with the proportions of each class area being 17.8%, 22.9%, 27.7%, 22.4% and 9.2%, respectively. It can be seen from table 2 that the frequency ratio of the IVM increases gradually from very low class to very high class, indicating that the IVM has better landslide prediction accuracy (Fig. 3). On the whole, landslides in the high and very high class area mainly distribute along the rivers, which is consistent with table 1 and the actual situation, further confirming the importance of water in the process of landslide occurrence.

Table 2. Disaster distribution table of each prone area

| Model            | Classes        | Information value zone | Landslide grid number | Percentage of landslide | Total grid number | Percentage of domain | Frequency ratio |
|------------------|----------------|------------------------|-----------------------|-------------------------|-------------------|----------------------|----------------|
| Information value model | very high      | 3.527~3.871            | 949                   | 0.455                   | 337,605           | 0.178                | 2.558          |
|                   | high           | 3.365~3.527            | 389                   | 0.186                   | 434,061           | 0.229                | 0.815          |
|                   | moderate       | 3.219~3.365            | 371                   | 0.178                   | 526,631           | 0.277                | 0.641          |
|                   | low            | 3.063~3.219            | 247                   | 0.118                   | 425,066           | 0.224                | 0.529          |
|                   | very low       | 2.542~3.063            | 131                   | 0.063                   | 175,572           | 0.092                | 0.679          |

Fig. 3 Prediction results of landslide susceptibility of information value model

5. Model accuracy verification

In this study, the accuracy of the IVM is obtained by using the prediction rate curve, and the degree of conformity between the results of model prediction and the actual situation was verified by the proportion of landslide grids in each class area. The prediction rate curve takes the predicted landslide area ratio as the abscissa and the actual landslide area ratio as the ordinate. The accuracy of the model is evaluated by the area (AUC value) enclosed by the curve and abscissa and ordinate. The higher the AUC value, the higher the model accuracy and the better the prediction effect. As shown in Fig. 4, the
AUC value of the IVM is 0.838, indicating that the prediction accuracy of landslide susceptibility of IVM is higher.

Fig. 4 Prediction rate curve of prediction accuracy of landslide susceptibility

6. Summary
Taking Ningdu County as an case, ten conditioning factors are selected to predict the landslide susceptibility by establishing IVM. The following conclusions are obtained:

1) The AUC of LSP by IVM in Ningdu area is 0.833. It shows that the IVM reflects more accurately the distribution rule of landslide in Ningdu area.

2) According to the IVM, the proportion of landslide grid in the very high class zone is 45.5%, while the very low class zone is 6.3%, which indicates that the LSI predicted by the IVM is consistent with the actual situation.

3) The landslide susceptibility map predicted in this paper shows that the high and very high class areas of landslides are mainly distributed along the rivers, indicating that water plays an important role in the landslide occurrence. In addition, elevation and lithology also have a great impact on the landslide occurrence.

References
[1] Faming Huang, Kunlong Yin, Guirong Zhang, et al. Landslide displacement prediction using discrete wavelet transform and extreme learning machine based on chaos theory[J]. Environmental Earth Sciences, 2016, 20(75): 1-18.

[2] Faming Huang, Xiaoyan Luo, Weiping Liu. Stability Analysis of Hydrodynamic Pressure Landslides with Different Permeability Coefficients Affected by Reservoir Water Level Fluctuations and Rainstorms[J]. Water, 2017, 9(7): 450.

[3] Zhu A X, Wang R, Qiao J, et al. An expert knowledge-based approach to landslide susceptibility mapping using GIS and fuzzy logic[J]. Geomorphology, 2014, 214: 128-138.

[4] Wu Y, Li W, Liu P, et al. Application of analytic hierarchy process model for landslide susceptibility mapping in the Gangu County, Gansu Province, China[J]. Environmental Earth Sciences, 2016, 75(5): 422.

[5] Faming Huang, Kunlong Yin, Jinsong Huang, et al. Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine[J]. Engineering Geology, 2017, 223: 11-22.
[6] Chen W, Xie X, Wang J, et al. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility[J]. Catena, 2017, 151: 147-160.

[7] Faming Huang, Chi Yao, Weiping Liu, et al. Landslide susceptibility assessment in the Nantian area of China: a comparison of frequency ratio model and support vector machine[J]. Geomatics, Natural Hazards and Risk, 2018, 9(1): 919-938.

[8] Huang F, Yin K, Yang B, et al. Step-Like Displacement Prediction of Landslide Based on Time Series Decomposition and Multivariate Chaotic Model[J]. Earth Science, 2018, 43(3):887-898.

[9] Faming Huang, Kunlong Yin*, Tao He, et al. Analysis of Influence Factors and Displacement Prediction of Reservoir Landslide-A Case Study of Three Gorges Reservoir, China[J]. Technical Gazette, 2016, 23(02):617-626.

[10] Che V B, Kervyn M, Suh C E, et al. Landslide susceptibility assessment in Limbe (SW Cameroon): A field calibrated seed cell and information value method[J]. Catena, 2012, 92(1):83-98.

[11] Faming Huang, Jinsong Huang, Shuihua Jiang, et al. Landslide displacement prediction based on multivariate chaotic model and extreme learning machine[J]. Engineering Geology, 2017, 218: 173-186.

[12] Faming Huang, Jinsong Huang, Shui-Hua Jiang, et al. Prediction of groundwater levels using evidence of chaos and support vector machine[J]. Journal of Hydroinformatics, 2017, 19(4): 586-606.

[13] Zhang J, Yin K, Wang J, et al. Evaluation of landslide susceptibility for Wanzhou district of Three Gorges Reservoir[J]. Chinese Journal of Rock Mechanics & Engineering, 2016.

[14] Dou J, Yamagishi H, Pourghasemi H R, et al. An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan[J]. Natural Hazards, 2015, 78(3): 1749-1776.

[15] Ilia I, Tsangaratos P. Applying weight of evidence method and sensitivity analysis to produce a landslide susceptibility map[J]. Landslides, 2016, 13(2): 379-397.

[16] Yuanyao Li, Jinsong Huang, Shui-Hua Jiang, et al. A web-based GPS system for displacement monitoring and failure mechanism analysis of reservoir landslide[J]. Scientific reports, 2017, 7(1): 17171.

[17] Faming H, Kunlong Y, Shuihua J, et al. Landslide susceptibility assessment based on clustering analysis and support vector machine[J]. Chinese Journal of Rock Mechanics and Engineering, 2018.

[18] Huang F, Chen L, Yin K, et al. Object-oriented change detection and damage assessment using high-resolution remote sensing images, Tangjiao Landslide, Three Gorges Reservoir, China[J]. Environmental Earth Sciences, 2018, 77(5):183.

[19] Jiang S H, Huang J, Huang Faming*, et al. Modelling of spatial variability of soil undrained shear strength by conditional random fields for slope reliability analysis[J]. Applied Mathematical Modelling, 2018, 63:374–389.

[20] Li, D.; Huang, F.; Yan, L.; Cao, Z.; Chen, J.; Ye, Z. Landslide Susceptibility Prediction Using Particle-Swarm-Optimized Multilayer Perceptron: Comparisons with Multilayer-Perceptron-Only, BP Neural Network, and Information Value Models. Appl. Sci. 2019, 9, 3664.

[21] Ba Q, Chen Y, Deng S, et al. An Improved Information Value Model Based on Gray Clustering for Landslide Susceptibility Mapping[J]. ISPRS International Journal of Geo-Information, 2017, 6(1):18.

[22] Huang Faming, Jing Zhang, Chuangbing Zhou, Yuhao Wang, Jinsong Huang & Li Zhu*. A deep learning algorithm using a fully connected sparse autoencoder neural network for landslide susceptibility prediction[J], Landslides, 2019. DOI: 10.1007/s10346-019-01274-9

[23] Zizheng Guo, Kunlong Yin, Lei Gui, Qingli Liu, Faming Huang, Tengfei Wang. Regional Rainfall Warning System for Landslides with Creep Deformation in Three Gorges using a Statistical Black Box Model[J], Scientific Reports, 2019,9,8962