



# The Open Civil Engineering Journal

Content list available at: [www.benthamopen.com/TOCIEJ/](http://www.benthamopen.com/TOCIEJ/)

DOI: 10.2174/1874149501812010021, 2018, 12, 21-34



## RESEARCH ARTICLE

# Application of GIS-Based Back Propagation Artificial Neural Networks and Logistic Regression for shallow Landslide Susceptibility Mapping in South China-Take Meijiang River Basin as an Example

Qing-hua Gong<sup>1,2,\*</sup>, Jun-xiang Zhang<sup>3</sup> and Jun Wang<sup>1,2</sup>

<sup>1</sup>Guangzhou Institute of Geography, Guangzhou 510070, China

<sup>2</sup>Guangdong Open Laboratory of Geo-spatial Information Technology and Application, Guangzhou 510070, China

<sup>3</sup>School of Tourism, Huangshan University, Huangshan 245021, China

Received: September 15, 2017

Revised: December 11, 2017

Accepted: December 12, 2017

### Abstract:

#### Introduction:

In this study, artificial neural network (ANN) model and logistic regression were applied to analyze susceptibility and identify the main controlling factors of landslide in Meijiang River Basin of Southern China.

#### Methods:

Methods: Eleven variables such as altitude, slope angle, slope aspect, topographic relief, distance to fault, rock-type, soil-type, land-use type, NDVI, maximum rainfall intensity, distance to river were employed as landslide conditioning factors in landslide susceptibility mapping.

Both landsliding and non-landsliding samples were needed as training data for ANN model. 384 landslides and 380 non-landsliding points with no recorded landslides according to field investigation and survey data were chosen as sample data of ANN model. Moreover, the ROC curve was applied to calculate the prediction accuracy.

#### Results:

The validation results showed that prediction accuracy rate of 82.6% exists between the susceptibility map and the location of the initial 384 landsliding samples. However, logistic regression analysis showed that the average correct classification percentage was 75.4%. The prediction results of ANN model in high sensitive zone is more accurate than the logistic regression model.

#### Conclusion:

Therefore, the ANN model is valid when assessing the susceptibility. The main controlling factors were identified from the eleven factors by ANN model. The slope, rock and land use type appeared to be the main controlling factors in landslide formation process in Southern China.

**Keywords:** Shallow landslide, Susceptibility, Artificial neural network, Logistic regression, GIS, South China.

## 1. INTRODUCTION

Landslides result in enormous casualties and huge economic losses in mountainous regions. The recent intensification of land-use changes has raised the level of landslide susceptibility. China is a mountainous country, correspondingly, landslide disasters occur frequently. There is no doubt that this will be a serious threat to the economic

\* Address correspondence to this author at the Guangzhou Institute of Geography 100#, XianLie Road, Yuexiu District, Guangdong Province, Guangzhou City, China, Fax: Tel: +86-13560347467, E-mail: [gqh100608@163.com](mailto:gqh100608@163.com)

and society development, especially in mountainous areas. According to the China Ministry of Land and Natural Resources, the landslide hazards resulted in almost 1000 deaths, hundreds of millions of dollars in direct economic loss, and inestimable indirect loss every year [1]. There were 19,522 geological hazards from January to June in 2010 which leads to 464 people dead or missing and direct economic losses of around 18.6 billion Yuan (about 2.3 billion US dollars). Geological disasters have increased 10 times in the last few years. Massive landslide was induced by the 5·12 Wenchuan earthquake in China in 2008 [2]. The Zhouqu debris flows shocked the whole world on 8 August, 2010, which resulted in 1,144 dead and 627 missing [3]. Landslide susceptibility maps provides urban planners overview information of landslide prone areas [4]. Based on the thorough research on the formation mechanism of landslides, the spatial and temporal variation characteristics of shallow landslides can provide scientific basis for evaluating and managing hazards. In recent years, regional and medium-scale landslide susceptibility mapping have become an important topic for specialists of different disciplines, such as engineering geologists, planners, local administrations or decision makers, etc [5, 6].

Now researchers have made great progress in the landslide susceptibility assessment. Several different approaches can be found in current literature, including direct and indirect heuristic approaches, deterministic, probabilistic and statistical approaches, nonlinear system theory method, as well as coupling analysis method of endogenic and exogenic geological processes. Those methods may be summarized into statistical models and dynamical model [7]. Statistical method includes discriminant analysis [8], principal component analysis, multivariate analysis [9], and logistic regression [10] and so on. With the advanced GIS technology and the factor-overlap method, the multivariate analysis method was drawn into the landslide sensitivity assessment [11, 12]. However, there are some weaknesses in statistical models. Firstly, the result is too heavily dependent on the quality of the data collected; secondly, the statistical model cannot explain the mechanism and, thirdly, regional disparity also exists in the relationship between influence factor and landslide susceptibility, and the method cannot suit for different area. The physical mechanism model based on the limit equilibrium is widely used in geo-technical engineering [13]. The model requires physical mechanics parameters obtained by geotechnical experiment. However, the model parameters are difficult to obtain. The output of the mechanical models provides a basis for the stability analysis of a landslide. The landslide is the result of interaction of meteorological, hydrographic, geology, physiognomy and human environment. Therefore, a single-landslide pattern cannot represent a whole zone or achieve the goal of disaster prevention and mitigation.

To overcome these shortcomings, the ANN method was developed and became popular in the recent year. Many papers concerning landslide susceptibility mapping using this method have been published [14 - 18]. Artificial neural network method can reflect non-linear feature between the factors of landslide formation. The model's precision of the model depends on the data samples. In fact, landslide is determined with a complicated variable nonlinear system by factors like rock deformation, nonlinearities of evolution law and internal factors. The landslide susceptibility prediction should be based on complex, uncertain, and non-linear relationships between the stability of slide rock and conditioning factors. ANNs are information-driven models and universal non-linear function approximations. The ability to learn non-linear functions from data is an important feature in classifying landslide-prone areas [19]. Moreover, a neural network does not require assumptions about the input variable distribution.

In this paper, the application of the ANN methods to landslide susceptibility mapping is studied. This paper introduced the details of the methods as well as the findings of the study. This paper is organized in four major parts. The first part describes the feature about landslide of the study area. The second part determines the statistical correlations between landslide frequency and the physical parameters contributing to the initiation of landslides. The third part describes the methods, development and training of ANNs algorithms. In the fourth part, landslide susceptibility mapping was modeled by incorporating factors in a GIS-based ANN model. The controlling factors were also identified by the model. The fifth part describes logistic regression model applied in the susceptibility mapping.

## 2. STUDY AREA

The ANN model was tested on the Meizhou city which is located at the south-east of china, with a total area 15876.05 km<sup>2</sup> and the elevation ranging from 0 to 1559 m (Fig. 1). There are widespread mountains in the study area with complex geological conditions and strong tectonic activity, which caused serious landslides disasters. Therefore, it is a serious hazard to the economic development in mountain area. Considering the hazard from landslides, it is urgent to identify area which may be prone to landslide.

The study area is typical of the geological features of surf-layer and bedrock in the south of China. The bedrock geology mainly comprises of Paleozoic and Mesozoic strata in the middle of study area (Fig. 2). Volcanic rocks

compose mainly of granite in the northwest part and southeast part of study area, which occupied 40% of gross area. The climate is typical of subtropical and monsoonal, hot and humid in summer mild and dry in winter. Heavy rains and typhoons often bring huge rainfall. The average annual precipitation varies from 1067 to 1727 mm during the period of 1990–2005. The groundwater is composed of pore water, bedrock fissure water and carbonate rock karst-cranny water in the zone. The development of transient perched water table at the interface of the colluviums and residual soils, and the underlying weathered rock during or following periods of intense rainfall, result in the natural steep slopes prone to sliding.

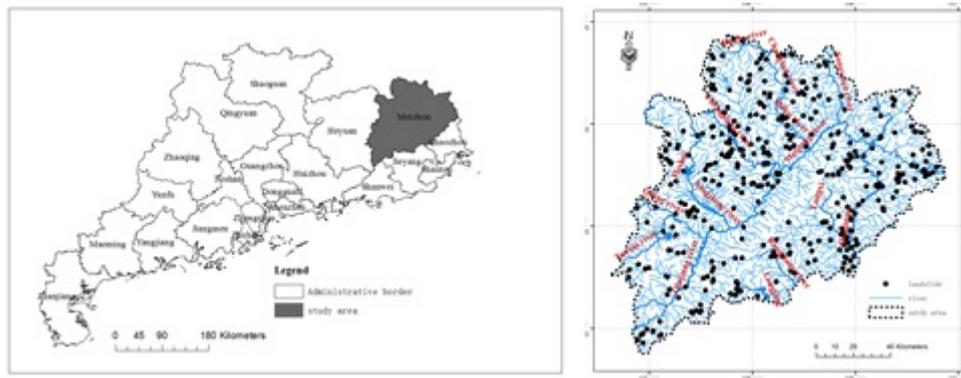


Fig. (1). The Study Area.

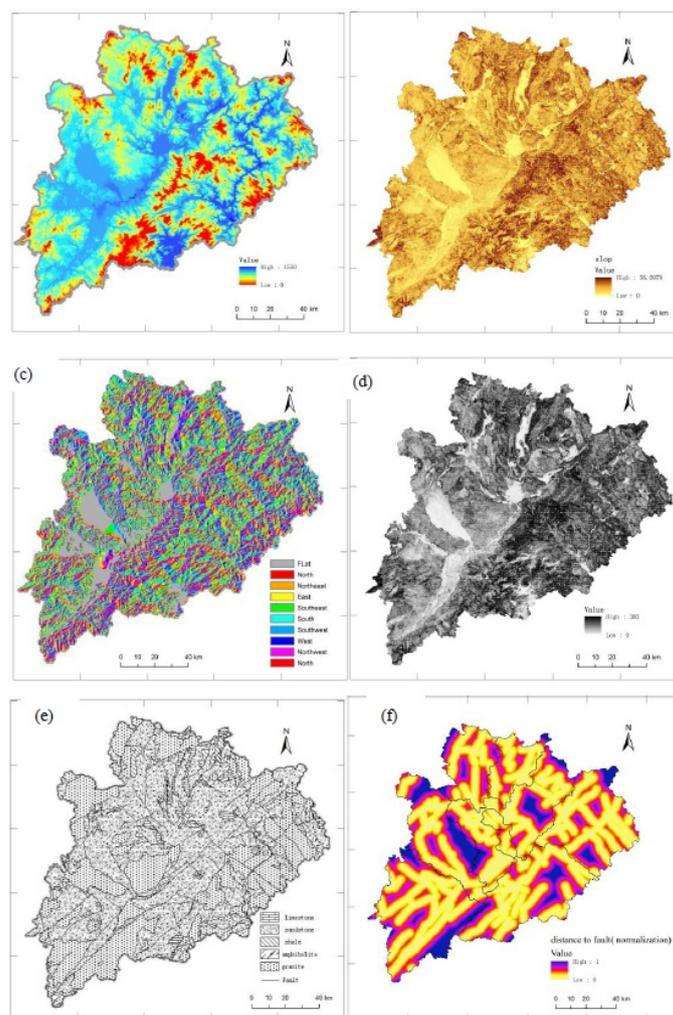


Fig. (2). Input Data Layers (a) altitude; (b) slope; (c) aspect; (d) surface rolling; (e) rock types; (f) distance to faults.

Landslide locations were identified in the study areas by the interpretation of aerial photographs, field surveys and inventory reports. Altogether, 384 landslides were recorded in the landslide inventory and related spatial parameters, such as location, time of failure, and dimensions, available for each record. The scene of landslide disaster is shown in Fig. (3). Landslides type in the study area is mainly the shallow soil landslides. Landslide scale is a small shallow landslide, whereas the landslide thickness is generally 1-2 meters. The feature of the landslide district is highly regional and seasonal. Most of the landslides in the study area are located at the hilly ground area with an elevation of 100-500 meters, which is easy to deposit colluviums. As can be seen from layer structure, the hazards are mainly distributed in the territory of granite rock, which covered with weathered layer and soft metamorphosed rock. From the time of distribution, 97.35% of the landslides occurred during the flood season.

### 3. METHOD AND DATA USED

The use of the ANNs can be a valid alternative in hazard mapping, when the conditioning factors are not approximable by a normal distribution and are strongly correlated. Artificial neural network model is a non-linear model with input layer, hidden layer and output layer. The landslide influencing factors are taken as the input layer, the landslide susceptibility as the output layer. The training and learning of the model are carried out. After the model has been validated, the output from the forecast area can be used to assess the sensitivity of the disaster. The data-learned model has an evaluation function that can calculate the magnitude of the landslide susceptibility.

Landslides result from interdependent spatio-temporal processes, including hydrology rainfall, groundwater, vegetation, soil condition, bedrock, topography, and human activities and so on. These can be subdivided into four groups: geomorphological conditions, geological conditions, land use and landcover, and meteorological and hydrological conditions. The factors utilized in this study are variables of both categorical and continuous data. In this study, both quantitative and qualitative variables were subdivided into proper categories, defined on the basis of the influence that they exert on landslide mechanics, and normalized in the interval 0-1. In this paper, based on index factor status and factor correlation analysis results [20], eleven factors were selected. Detailed description of the factors and data sources in the study area are summarized in Table 1. These parameters were transformed into a spatial vector database, and the thematic layers were converted into a raster grid for application in the neural network modeling.



Fig. (3). The shallow landslides.

Table 1. Basic Data Sets for Landslide Susceptibility.

Classification	Data Layers	Data Type	Format	Scale or Resolution
Geomorphic map	Altitude	Continuous	ARC/INFO GRID	1:25,000
	Slope angle	Continuous	ARC/INFO GRID	1:25,000
	Slope aspect	Categorical	ARC/INFO GRID	1:25,000
	Topographic relief	Continuous	ARC/INFO GRID	1:25,000
Geological map	Distance to faults	Continuous	ARC/INFO GRID	1:25,000
	Rock types	Categorical	ARC/INFO polygon	1:25,000
	Soil types	Categorical	ARC/INFO polygon	1: 100,000
Land use and land cover	Land use type	Categorical	ARC/INFO polygon	1:25,000
	NDVI	Continuous	ARC/INFO GRID	10m*10m
Meteorological and hydrological	maximum rainfall intensity	Continuous	ARC/INFO GRID	10m*10m
	Distance to rivers	Continuous	ARC/INFO GRID	10m*10m

Classification	Data Layers	Data Type	Format	Scale or Resolution
Landslide	Landslide	Categorical	ARC/INFO point	1:25,0000

### 3.1. Geomorphological Parameters

The genesis of geological hazards is highly relevant to topographic conditions factors. This model was used to prepare altitude, slope, aspect, and topographic relief. The topographical and geographical conditions are based on the digital elevation model (DEM). The digital elevation model of the study area was produced by using topographic maps in 1/25,0000 scale. DEM was generated by contour interpolation. The triangulation growth algorithm is used to transform the discrete points on contours into TIN. It was generated using 3D Analyst in the ArcScene module and then convert the correct TIN to grid data. The grid data was resampled to form a grid DEM. The gross error was checked by overlay the contour line interpolated by DEM and the original contour. The DEM was defined in the Gauss Krugar Xian 1980 coordinate system with 5m cell size.

#### 3.1.1. Altitude

Altitude is useful to classify the local relief and locate points of maximum and minimum heights within terrains. Therefore, altitude is one of the topographic factors affecting landslides. The altitude map of study area was divided into five categories with 150m interval. The number of landslide point in each altitude group was presented in Table 2. In the study area, 77% landslides located in plateau hilly areas of less than 300m. This is because plateau hilly areas are beneficial to the soil deposit. The accumulation of loose layer provides the source condition for landslide formation.

Table 2. Data Sets and Their Classes for This Study.

Data Layers	Classes	Landslide Bodies	% in Landslides Bodies	Data Layers	Classes	Landslide Bodies	% in Landslides Bodies
Altitude	<150 m	94	24%	rock types	Granite	130	34%
	150-300 m	205	53%		Sandstone	236	61%
	300-450 m	68	18%		Shale	7	2%
	450-600 m	10	3%		Limestone	11	3%
	>600 m	7	2%	soil types	Waterloggogenic paddy soil	63	16.40%
Slope angle	<15	7	1.20%	Yellow soil	2	0.50%	
	15-20	49	8.10%	Red soil	309	80.50%	
	20-30	226	37.50%	Purple soil	10	2.60%	
	30-40	237	39.30%	Land use type	Cultivated land	110	29%
	>40	84	13.90%		Garden plot	41	11%
Slope aspect	Flat	60	15.60%	Forest land	214	56%	
	N(0°-22.5°,337.5°-360°)	28	7.30%	Construction land	19	5%	
	NE(22.5°-67.5°)	37	9.60%	NDVI	0.3-0.5	8	2%
	E(67.5°-112.5°)	41	10.70%		0.5-0.6	79	21%
	SE(112.5°-157.5°)	48	12.50%		0.6-0.7	154	40%
	S(157.5°-202.5°)	26	6.80%		0.7-0.8	127	33%
	SW(202.5°-247.5°)	73	19%	>0.8	16	4%	
	W(247.5°-292.5°)	30	7.80%	maximum rainfall intensity	<500mm	52	13.50%
NW(292.5°-337.5°)	41	10.70%	500-800mm		260	67.70%	
			800-1000mm		43	11.20%	
Topogr-aphic relief	0-50m	249	65%	>1000mm	29	7.60%	
	50-100m	123	32%	Distance to rivers	0-100m	69	18%
	>100m	13	3%		100-300m	126	32.80%
Distance to faults	<100	7	1.80%		300-500m	67	17.40%
	100-1000m	82	21.40%		500-800m	46	12%
	1000-3000m	159	41.40%		800-100m	23	6%
	3000-5000m	88	22.90%		>1000m	53	13.80%
	5000-10000m	44	11.50%				
>10000m	4	1%					

### 3.1.2. Slope

The main parameter in slope stability analysis is the slope angle. The slope angle is directly related to the landslides. The slope angle map of the study area was divided into five slope categories. The landslide percentage in each slope group is presented in Table 2. This table indicates that most of the landslides occur at slope angle of less than 40°.

### 3.1.3. Aspect

The vegetation, precipitation and temperature are different in sunny slopes and shady slopes. Therefore, this is another important factor in landslide susceptibility maps. In this study, the aspect map (Fig. 2). of study area was produced to show the relationship between aspect and landslide. Aspect regions are classified according to the aspect class as flat (−1°), north(337.5°-360°, 0°-22.5°), Northeast (22.5°-67.5°), east(67.5°-112.5°), southeast (112.5°-157.5°), south(157.5°-202.5°), southwest (202.5°-247.5°), west (247.5°-292.5°) and northwest (292.5°-337.5°). The landslide percentage in each aspect group is presented in Table 2.

### 3.1.4. Topographic Relief

Terrain relief is the relative height difference between the top and bottom of the slope. It provides the necessary effective surface for landslide occurrence and determines the kinetic energy of the slope itself. It can determine the stability of the slope state. Topographic relief has an effect on hydrological and deposition processes [21]. So, it can reflect the change of relief and indicate the degree of surface erosion. Therefore, Topographic relief was selected for landslide susceptibility maps by many researches. The Topographic relief map (Fig. 2) of study area was produced from DEM. The Topographic relief of study area was divided into three groups. The landslide percentage in each group is presented in Table 2. This table indicates that most of the landslides occur at the scope of 0-100m.

## 3.2. Geologic Parameter

### 3.2.1. Distance to Fault

Geological structure factors have a very significant effect on slope stability [21]. Activity faults are the main cause of large-scale landslides. The fault map (Fig. 2). of study area was produced by using regional geological map in 1/250,000 scale. The fractures are generally NS-SW striking, though some are NW-SE striking. The distance to faults was divided into six groups. The landslide percentage in each group is presented in Table 2.

### 3.2.2. Rock Type

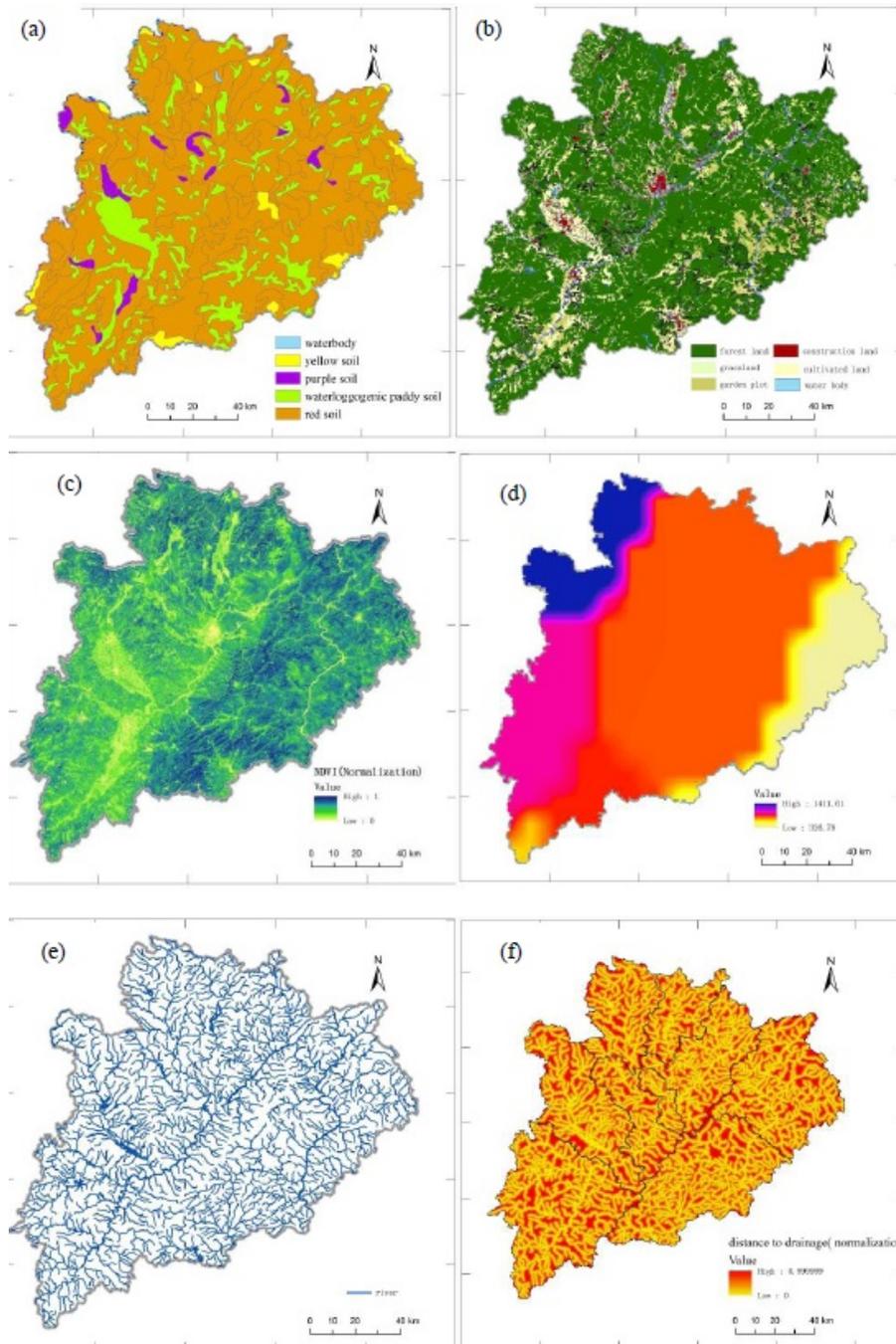
The rock type features are the foundation of the landslide, which can control the development of the landslide and provide material source for the landslide. Because of their different physical and chemical properties, different lithologies have different effects on the landslides [21]. As the chart shows, there are four rock types in the region including limestone, sandstone, shale and granite (Fig. 2). The landslide percentage in each rock type is presented in Table 2. As in the table, most of the landslides occur in the granite and sandstone region.

## 3.3. Soil Types

The effects of soil on slope stability have been widely considered in landslide research [22]. The soil types have an effect on landslide distribution through cohesiveness, thickness and so on. The different soil types present in this region were grouped into a number of types that are homogenous in terms of chemical composition. The digital soil layer is shown in Fig. (4). There are four soil types in the region including waterloggogenic paddy soil, yellow soil, red soil and purple soil. The landslide percentage in each soil type is presented in Table 2. As seen from the table, most of the landslides occur in the red soil.

## 3.4. Land Use

Many studies have revealed the relationship between human activity and slope stability [23]. Rainfall and human activities are very important triggering factors of the landslide relating to geological condition. In this study, land use layer was extracted from Landsat ETM image by using object-based classification method. There are four different land use types in the region including cultivated, land garden plot, forest land and construction land. The landslide percentage in each land use type is presented in Table 2. As seen from the table, most of the landslides occur in the forest land and cultivated land.



**Fig. (4).** Input data layers  
 (a) soil types; (b) land use type; (c) NDVI; (d) 24h heaviest rainfall; (e) river; (f) distance to drainage.

### 3.5. NDVI (Normalized Difference Vegetation Index)

There are two ways for plant roots to improve soil capability of anti-erosion. One is that fine root can keep soil by its net-work, another is increasing slope load [24]. The NDVI map was obtained from LandsatTM satellite image acquired on 15 September 2009. The NDVI map of the study area was divided into five categories. The landslide percentage in each category is presented in Table 2. As seen from the table, most of landslides occur at the scope of 0.6-0.7.

**Table 3. Weights and Thresholds between the Input Layer and the First Hidden Layer.**

–	i=1	i=2	i=3	i=4	i=5	i=6	i=7	i=8	i=9	i=10	i=11	$b_{ij}$
j=1	3.32	-3.96	1.38	-0.28	2.68	-1.60	-3.67	0.46	-5.88	-4.00	-1.09	1.79
j=2	4.87	0.85	-4.40	1.01	-1.50	-4.03	-3.61	-3.20	4.59	1.30	-3.12	-1.4
j=3	-4.84	4.04	-4.40	0.38	0.13	-4.38	-1.32	0.54	-7.56	-2.76	4.11	7.09
j=4	6.53	8.34	-0.32	8.47	5.84	1.71	1.73	-6.23	2.59	-0.97	9.26	-12.3
j=5	4.62	8.00	-1.04	6.91	5.94	-3.10	5.53	1.88	5.58	-3.92	1.85	-5.4
j=6	-3.49	0.68	5.83	-4.06	6.49	1.04	-2.69	-6.07	3.54	9.89	-7.50	-7.6
j=7	-3.68	-6.86	1.54	-0.72	-0.80	3.24	-1.49	-8.14	-7.03	2.40	-4.42	1.58
j=8	6.50	16.09	-6.68	14.48	0.41	-10.	0.82	-18.1	-0.92	-11.7	10.58	-3.1
j=9	5.23	5.21	8.70	-1.66	-4.47	3.92	-2.71	2.71	-3.54	1.93	0.98	-0.7
j=10	3.65	0.68	-7.03	-3.64	-2.43	2.04	-4.03	-4.29	1.43	-2.89	-4.60	0.47
j=11	-5.01	0.90	-2.89	-1.79	4.23	-2.26	-2.18	9.78	-1.86	-5.48	4.00	2.18
j=12	12.16	10.14	-1.37	12.74	2.07	3.38	3.32	-1.22	4.31	4.16	2.70	-14.23
j=13	1.80	-4.16	1.25	-4.29	3.84	7.78	0.28	11.44	-2.66	-0.91	-4.62	2.04
j=14	-0.91	5.00	1.48	4.96	-5.37	-13.5	6.59	-14.9	-3.11	-14.5	1.69	11.46
j=15	-9.39	-11.9	-6.54	-21.3	-1.96	-3.17	-5.54	2.52	-5.99	-6.77	-7.65	19.54

### 3.6. Distance to Rivers

In general, at a closer distance to the river, the erosion is stronger and the probability of the occurrence of landslides is higher [24]. The distance to rivers is represented by the proximity of the rivers and drainages in the area. The river map at the scale of 1:25,000 was obtained from DEM. River system can be divided into four grades. The smallest watershed area is about 20 km<sup>2</sup>. The distance to river of the study area was divided into six groups. The landslide percentage in each group is presented in Table 2.

### 3.7. Rainfall

Slope in certain geological setting and in a certain mechanical environment requires a certain rainfall, rainfall intensity or duration to promote slope damage [24]. Rainfall is one of the important inducing factors to disasters. The maximum rainfall intensity isocline map at scale 10m\*10m was obtained from the meteorological department. There are 12 precipitation stations distributed in the study area. The maximum rainfall intensity isocline map was generated by interpolating data from precipitation stations. The maximum rainfall intensity of study area was divided into four groups. The landslide percentage in each group is presented in Table 2.

## 4. ARTIFICIAL NEURAL NETWORKS

### 4.1. Architecture of Neural Network

In the presented study, the traditional approach of landslide susceptibility mapping by using an artificial neural network model was implemented in a GIS framework. This study sets up a three-layer BP artificial neural network to analyze landslide sensitivity. The three-layer interconnected neural network (Fig. 5). consists of one input layer, two hidden layers and one output layer. In this specific structure network, there are 11 input nodes (respectively for altitude, slope, aspect, Topographic relief, distance to faults, rock types, soil types, land use type, NDVI, Maximum rainfall intensity, and distance to drainage,) and the output layer will have one node, reflecting the disaster situation (Value of 0 or 1).

### 4.2. Training of ANN

According to the material of field investigation and survey, there have been 384 landslides in the study area. The spatial distribution of landslides in the region is shown in Fig. (1). Before running the artificial neural network program, the training site should be selected. The study chose 380 points as the test data where no landslide disasters occur. Furthermore, using the landslides and security point locations, we extracted 11 quantitative data and constructed spatial database by GIS. The spatial correlations between the variables and neural network equations were established in ARCGIS. Then, the landslides and security point data are partitioned into two subsets, such as the “training data” and “test data”. The 300 landslides and 300 security points were selected for training the ANN. And 84 landslides and 80 non-landslide points were used for the prediction testing. The most popular ANN model used in prediction and

regression tasks is the multi-layer perceptron (MLP) with a feed-forward back-error propagation (BP) type of learning algorithm [25]. This learning algorithm was trained with the BP type, which consists of one input layer, two hidden layers, and one output layer.

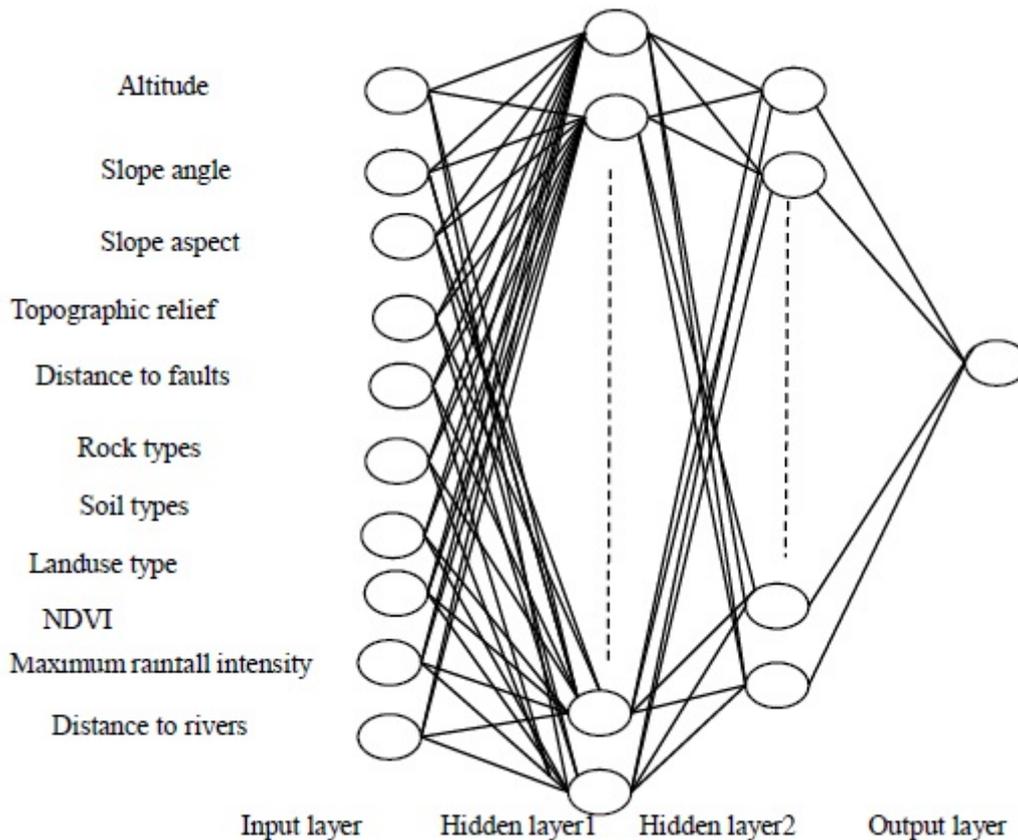


Fig. (5). Neural Network Structure Designed For Landslide Susceptibility Assessment.

Table 4. Weights and Thresholds between the First Hidden Layer and the Second Hidden Layer.

-	j=1	j=2	j=3	j=4	j=5	j=6	j=7	j=8	j=9	j=10	j=11	j=12	j=13	j=14	j=15	b <sub>2k</sub>
k=1	-1.27	-1.06	-1.04	-10.11	-3.00	5.18	-3.38	-7.76	-3.66	-2.45	4.85	4.72	0.79	-9.74	-10.19	0.86
k=2	1.01	-2.69	0.01	-1.71	2.73	3.38	0.47	-11.92	1.16	-0.87	3.74	3.32	6.38	-6.69	1.08	0.32
k=3	0.99	0.60	-3.61	6.41	5.90	0.84	-3.86	-3.20	3.66	1.78	-2.79	6.73	0.01	-4.00	-8.35	-0.89
k=4	0.11	-0.48	1.27	-0.45	0.08	2.85	-0.26	-2.41	2.15	-3.77	0.60	1.33	2.61	-1.01	-3.61	2.61
k=5	-0.73	0.11	-0.45	-2.91	1.71	-2.24	-0.57	9.30	-1.06	-1.94	-1.92	-6.27	-5.96	7.98	3.28	-5.25
k=6	-0.18	-1.81	2.01	4.17	4.71	-1.14	-3.17	4.08	-0.47	-1.00	5.04	-1.37	4.74	0.87	3.71	1.74
k=7	-0.41	-1.38	1.79	-0.85	2.84	-1.43	0.09	-1.05	-0.99	-2.33	1.86	1.31	3.60	1.62	0.62	1.70
k=8	1.27	-1.66	1.42	-3.43	-3.15	-1.87	0.95	-3.44	-4.03	1.54	3.68	-7.35	1.13	1.95	9.25	-1.46

Three-layer feed-forward network was implemented using the MATLAB software package. The normalized transfer function, the training function, the number of hidden layer and the active function for ANN were modeled, simulated, and determined by using neural network toolbox of MATLAB7.11.0. The adaptive learning algorithm was selected in this study which can enable studying-speed faster and it is self-adaptive to data. And the maximum training is 5000 times and the goal of training is 0.001. The weights and threshold of each factor estimated by neural network in this study is shown in Tables 3-5.

Table 5. Weights and Thresholds between the Second Hidden Layer and the Output Layer.

-	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	b <sub>3n</sub>
n=1	14.86	9.49	2.69	-3.61	-4.97	-2.67	-4.55	-7.78	0.27

### 4.3. Accuracy of the ANN Model

Assessing the performance of the landslide susceptibility models is considered to be a crucial step in model selection [26]. In order to evaluate the performance of the produced map, two methods were used. The first method is to calculate the accuracy rate by comparing the pixel values between the landslide inventory map and the final susceptibility maps. According to the results of this process, the accuracy rate is calculated as 0.799. It shows that the produced map represents the reliable results. The second performance index is the AUC values. In practice, the AUC values usually were used for the assessment of relative quality of susceptibility maps [26 - 28]. Thus, the results of the information model were validated by using landslide inventories and the area under the curve (AUC). When the AUC value close to 1, it means high accuracy, while the AUC close to 0.5, indicates accuracy. We use the model result building the Receiver Operating Characteristic curve (ROC) and the AUC could also be calculated. The results of the model show that the AUC is 82.6% respectively shown in Fig. (6). Therefore, the ANN model is valid when assessing the susceptibility.

### 4.4. Analysis on the Main Control Factor

The landslide-susceptibility analysis is a function of a variety of variables that include the altitude, slope, aspect, slope angle, distance to faults, rock types, soil types, land cover, maximum rainfall intensity, NDVI and distance to river. In order to further verify the model and determine the sensitivity of various factors for slope stability, we assume that each variable values is 1 and input it into the model (Single factor effect doesn't exist in reality). We take slope as an example, assuming that when all other factors is 0, only slope, and set it to 1, the input matrix of the nine factors is ,the model calculation results is 0.866. The rest of the factors can be done in the same manner. The single factor conditions of neural network identification results were calculated and normalized. The land use type showed the highest value as 0.294, then the slope is 0.288 and rock type is 0.225. The result displays that the slope, rock types and the land use type are the main controlling factors in the disaster formation process. Landslide formation of internal cause in South China mainly depends on the topography and the geology, while rainfall is the motivating factor. The model results conform to the disaster mechanism in the south China.

## 5. LOGISTIC REGRESSION ANALYSES

In the recent years, logistic regression analysis is one of the most popular multivariate statistical methods. In the recent literature, many studies have been published on the assessment of the landslides by using logistic regression analyses [29, 30]. Logistic regression analysis mainly predicts the probability of occurrence of an event through the multiple regression relationship between a dependent variable and multiple independent variables. In the logistic regression analysis, the dependent variable  $Y$  is a dichotomous variable, the values  $Y = 1$  and  $Y = 0$ , represent landslide and no landslide, respectively. The independent variables are  $X_1, X_2, \dots, X_n$  respectively. The conditional probability of landslide occurring under the independent variables is  $P = P(Y = 1 | X_1, X_2, \dots, X_n)$ . Then logistic regression model can be expressed as:

$$z_i = a_0 + a_1X_{i1} + a_2X_{i2} + \dots + a_nX_{in}$$

$$P_i = \frac{1}{1 + \exp(-z_i)}$$

$Z_i$  is intermediate variable parameter,  $a_0$  is regression constants,  $a_i$  is regression coefficients ( $i=1,2,\dots,n$ ),  $X_{ij}$  is the value of the  $j_{th}$  variable in element  $i$ , and  $P_i$  is regression prediction value of landslide occurrence in unit  $i$ .

The first stage in the application of logistic regression analyses is production of data matrix. For the continuous variable data (altitude, Slope angle, , topographic relief, NDVI, maximum rainfall intensity, distance to rivers), draw the histogram of the frequency distribution of continuous variables, the continuous variable data were normalized in the range of [0, 1] according to the frequency distribution of the histogram. Since the parameter slope aspect, rock types, soil types, land use type are categorical data, they were expressed in binary format with respect to each definitions. Dependent variables of the analyses are also expressed in binary format with respect to presence (1) and absence (0) of landslide or no landslide cell.

The logistic regression analysis was calculated in SPSS software. As a result regression analysis showed that the average correct classification percentage was 75.4%. Hosmer and Lemeshow test showed that the significance (Sig) is less than 0.05. The logistic regression model of landslide risk factors is shown in Table 6.

Table 6. Variables in the Equation.

	Altitude	Slope	Aspect	Topographic Relief	Distance to Faults	Rock Types	Soil Types	Land Use	NDVI	Maximum Rainfall Intensity	Distance to Rivers	Constant
B	0.3	0.77	0.67	-1.29	2.15	0.08	1.06	1.28	0.06	1.12	-0.82	-0.657
S.E.	0.0795	0.0951	0.04	0.07	0.13	0.06	0.12	0.05	0.14	0.09	0.07	1.44

6. RESULT

The data used are shown in Table 2. Various GIS data layers have been illustrated in Fig. (3). For the convenience of computing, all input layers were converted into raster layers. Then, using the raster calculator, the result of each cell is obtained on the basis of the LR model and the established BP model above. The weights and threshold of each factor estimated by neural network in this study are shown in Tables 3-5. At last, the calculated results were reclassified into three categories according to value: lower sensitive zone, medium sensitive zone and high sensitive zone (Fig. 7).

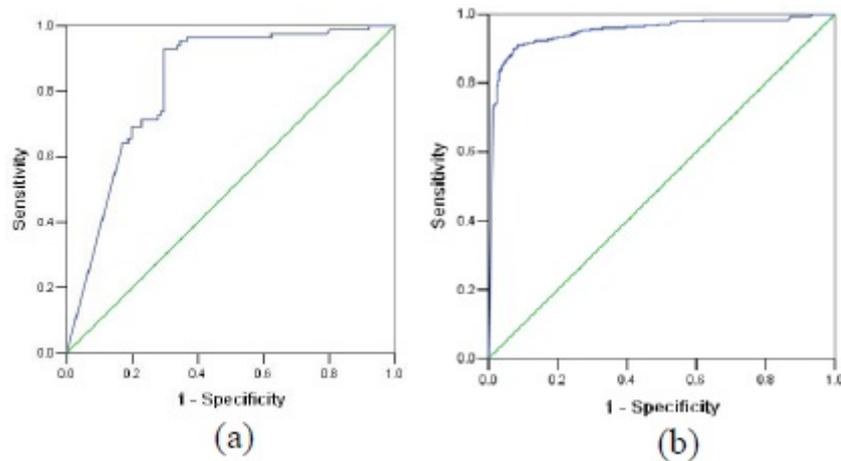


Fig. (6). Receiver Operating Characteristic (ROC) curve(a. training data;b. test data).

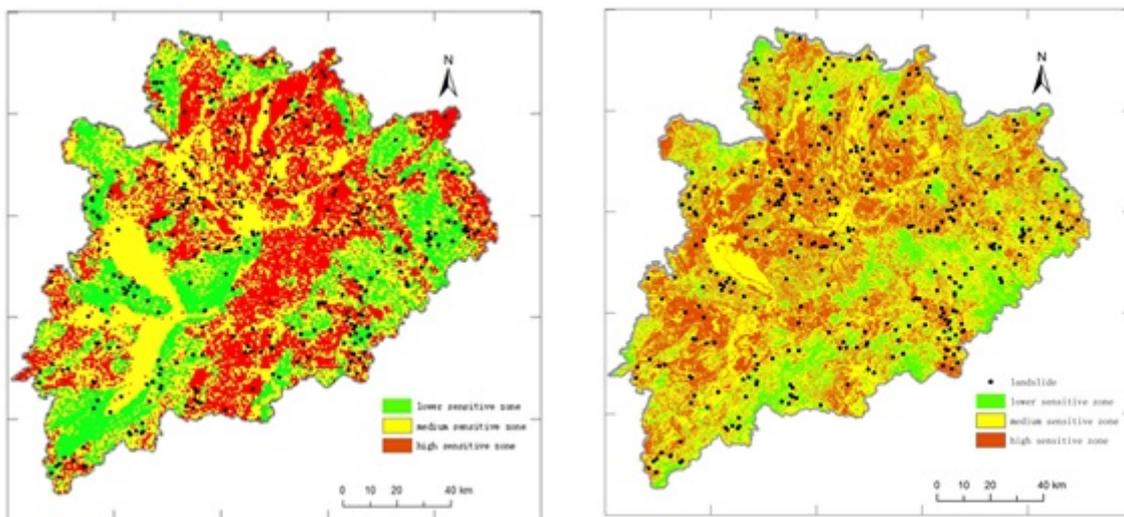


Fig. (7). Landslide Susceptibility Map by BP and LR.

The artificial neural network model and the logistic regression model can be verified mutually. According to the results, the landslide bodies are distributed at each sensitive level. There are 70.8% and 55.21% of the landslides distributed in the high sensitive zone by BP and LR model. The sensitivity level was higher and the bigger proportion of the landslides. The prediction results of ANN model in high sensitive zone is more accurate than the logistic regression

model. However, this does not mean that the artificial neural network model is better than the ANN model in other geological environments.

**Table 7. The statistical table of susceptibility partition and landslide distribution with BP and LR.**

–	Landslide Bodies		% in Landslides Bodies		Area(km <sup>2</sup> )		% in Area	
	ANN	LR	ANN	LR	ANN	LR	ANN	LR
Lower	30	78	7.8	20.31	2314.40	4163.11	14.58	27.88
Medium	82	94	21.4	24.48	8104.74	7286.36	51.05	45.89
High	272	212	70.8	55.21	5456.92	4426.58	34.37	26.23

## 7. DISCUSSION

It is difficult to get a complete and detailed shallow landslide map in short-term because the landslides have the characteristics of small size and wide distribution. Hence, it is needed for landslide susceptibility assessment work under the conditions of incomplete records. We employed ANN model and logistic regression to analyze landslide susceptibility and to select the slope, land use and so on eleven factors to establish susceptibility evaluation index system. The results show that the ANN model is feasible to susceptibility map. The susceptibility zoning map was in line with the actual conditions of the area. It can play an important role in the work of landslide hazard and risk assessment of disasters.

In South China, shallow landslides are commonly triggered by high pore-water pressure which results from high-intensity or short-duration rainfall. The deformation modes of slope in red soil hilly region are mainly shallow landslide. Shallow landslides are preferentially distributed on slopes with high-permeable soils overlaying low-permeable soils (Table 7). The landslides are roughly parallel to the ground surface. The shallow landslides are highly correlated to the landform. The results conform to the landslide characteristic in red hilly region in the South China. The research results show that high sensitive zones in the study area locate at the hilly ground area with elevation from 150 to 300m, northern aspect, slope gradient among 5-10° and topographic relief with 0-40m.

Regarding the application of artificial neural network and logistic regression model, as well as the relative importance and weighting between factors calculated, landslide hazard maps are of great help to planners and engineers when they choose suitable locations to implement development activities. These results can be used as basic data to assist slope management and land-use planning. The models used in the study are valid for generalized planning and assessment purposes, although they may be less useful at the site-specific scale where local geological and geographic heterogeneities may prevail. To make the model more general, more landslide data are needed.

## CONCLUSION

In this study, the neural network model and its cross-application approach was used successfully. Using ANN model established by MATLAB, the landslide susceptibility map was created and verified. The result of verification showed a prediction accuracy of 82.6%. The verification result is of a high value. The conclusion basically matches with the actual situation, so it shows that it is feasible to use the BP network model based on MATLAB neural network toolbox for landslide susceptibility analysis. The results display that slope, rock types and land use type were the main controlling factors in the disaster formation process. The results conform to the disaster mechanism in the South China.

## CONSENT FOR PUBLICATION

Not applicable.

## CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

## ACKNOWLEDGEMENTS

This paper is supported by the National Natural Science Foundation of China (No. 41671506), the Science and Technology Planning Project of Guangdong Province (No. 2012A030200010, No. 2013B031500007, and 2014A020218013, 2014A020219006).

## REFERENCES

- [1] R. Liu, and J.R. Ni, "Landslide and rock fall hazard zonation in china", *J. Basic Sci. Eng.*, vol. 1, pp. 13-15, 2005.
- [2] R.Q. Huang, and W.L. Li, "Research on development and distribution rules of geohazards induced by Wenchuan Earthquake", *Chin. J. Rock Mech. Eng.*, vol. 27, no. 12, pp. 2585-2592, 2008.
- [3] C.Z. Liu, T.B. Miao, and H.Q. Chen, "Basic feature and origin of the "8·8" mountain torrent-debrisflow disaster happened in Zhouqu County, Gansu, China", *Geol. Bull. China*, vol. 30, no. 1, pp. 141-150, 2011.
- [4] S. Wan, "A spatial decision support system for extracting the core factors and thresholds for landslide susceptibility map", *Eng. Geol.*, vol. 108, pp. 237-251, 2009.  
[<http://dx.doi.org/10.1016/j.enggeo.2009.06.014>]
- [5] F. Robin, and C. Jordi, "Guidelines for landslide susceptibility, hazard and risk zoning for land use planning", *Eng. Geol.*, vol. 102, pp. 85-98, 2008.  
[<http://dx.doi.org/10.1016/j.enggeo.2008.03.022>]
- [6] E. Murat, and G. Candan, "Use of fuzzy relations to produce landslide susceptibility map of a landslide prone area (West Black Sea Region, Turkey)", *Eng. Geol.*, vol. 75, pp. 228-250, 2004.
- [7] W.Q. Cong, and T.F. Li, "Research on dynamic predictive model of regional rainfall triggered geologic hazard based on unsaturated flow theory", *Beijing Da Xue Xue Bao Zi Ran Ke Xue Bao*, vol. 2, pp. 212-216, 2008.
- [8] H.X. Lan, F.Q. Wu, and Ch.H. Zhou, "Analysis on susceptibility of GIS based landslide triggering factors in YunNan Xiaojiang watershed", *Chinese Journal of Rock Mechanics and Engineering*, vol. 21, no. 10, pp. 1500-1506, 2012.
- [9] J. Mathew, and V.K. Rawat, "Landslide susceptibility zonation mapping and its validation in part of Garhwal Lesser Himalaya, India, using binary logistic regression analysis and receiver operating characteristic curve method", *Landslides*, vol. 6, pp. 17-26, 2009.  
[<http://dx.doi.org/10.1007/s10346-008-0138-z>]
- [10] S. Lee, "Comparison of landslide susceptibility maps generated through multiple logistic regression for three test areas in Korea", *Earth Surf. Process. Landf.*, vol. 32, pp. 2133-2148, 2007.  
[<http://dx.doi.org/10.1002/esp.1517>]
- [11] G. Gullà, L. Antronico, and P. Iaquineta, "Susceptibility and triggering scenarios at a regional scale for shallow landslides", *Geomorphology*, vol. 99, pp. 39-58, 2008.  
[<http://dx.doi.org/10.1016/j.geomorph.2007.10.005>]
- [12] P. Daniela, and T. Francesco, "Statistical analysis for assessing shallow-landslide susceptibility in South Tyrol (south-eastern Alps, Italy)", *Geomorphology*, vol. 151, pp. 196-206, 2012.
- [13] LI Zh, SH. Z. Zhu, J.G. Wang, and Y. Liu, "Analysis of the stability and sensitivity of DATIAN bay landslide", *J. Disaster Prev. Reduct.*, vol. 27, pp. 5-10, 2011.
- [14] S. Lee, J. Hwang, and I. Park, "Application of data-driven evidential belief functions to landslide susceptibility mapping in Jinbu", *Korea CATENA*, vol. 100, pp. 15-30, 2013.
- [15] K. Daisaku, and B. Joel, "Landslide susceptibility mapping using geological data, a DEM from ASTER images and an Artificial Neural Network (ANN)", *Geomorphology*, vol. 113, pp. 97-109, 2009.  
[<http://dx.doi.org/10.1016/j.geomorph.2009.06.006>]
- [16] P. Biswajeet, and L. Saro, "Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modeling", *Environ. Model. Softw.*, vol. 6, pp. 747-759, 2010.
- [17] C. Jaewon, and O. Hyun-Joo, "Combining landslide susceptibility maps obtained from frequency ratio, logistic regression, and artificial neural network models using ASTER images and GIS", *Eng. Geol.*, vol. 124, pp. 12-23, 2011.
- [18] S.W. He, P. Peng, and D. Lan, "Application of kernel-based Fisher discriminant analysis to map landslide susceptibility in the Qinggan River delta, Three Gorges, China", *Geomorphology*, vol. 171-172, pp. 30-41, 2012.  
[<http://dx.doi.org/10.1016/j.geomorph.2012.04.024>]
- [19] H.Y. Hong, B. Pradhan, and C. Xu, "Spatial prediction of landslide hazard at the Yihuang area (China) using two -class kernel logistic regression, alternating decision tree and support vector machines", *Catena*, vol. 133, pp. 266-281, 2015.  
[<http://dx.doi.org/10.1016/j.catena.2015.05.019>]
- [20] Q.H. Gong, "DEM and GIS based analysis of topographic and geomorphologic factors of shallow landslide in red soil hilly region of south china", *Int. J. Earth Sci. Eng.*, pp. 393-399, 2014.
- [21] M. Wang, and J.P. Qiao, "Application of contributing weights model in regional landslides hazard assessment", *Chin. J. Geol. Hazards Control*, vol. 21, pp. 1-6, 2010.
- [22] S.H. Wang, H.P. Chen, J.Q. Lu, and G. Jing, "Study on the relationship between landslide and soil genesis characteristics in basalts platform region", *Geol. Zheijiang*, vol. 18, no. 1, pp. 76-81, 2002.
- [23] J. Wang, K. Yin, and L. Xiao, "Landslide susceptibility assessment based on GIS and weighted information value: A case study of wanzhou district, Three Gorges Reservoir", *Chin. J. Rock Mech. Eng.*, vol. 33, pp. 797-808, 2014.

- [24] C. Chen, Q. Wang, J.P. Chen, Y.K. Ruan, L.J. Zheng, S.Y. Song, and C.C. Niu, "Landslide susceptibility mapping in vertical Distribution law of precipitation area: Case of the xulong hydropower station reservoir, southwestern china", *Water*, vol. 8, p. 270, 2016. [<http://dx.doi.org/10.3390/w8070270>]
- [25] P. Biswajeet, L. Saro, and F.B. Manfred, "A GIS-based back-propagation neural network model and its cross-application and validation for landslide susceptibility analyses. Computers", *Environ Urban Syst.*, vol. 34, pp. 216-235, 2010. [<http://dx.doi.org/10.1016/j.compenvurbsys.2009.12.004>]
- [26] C. Christos, F. Maria, and P. Christos, "GIS-Based landslide susceptibility mapping on the peloponnese peninsula, greece", *Geoscience*, vol. 4, pp. 176-190, 2014. [<http://dx.doi.org/10.3390/geosciences4030176>]
- [27] C. Christos, F. Maria, P. Christos, and K. Efthimios, "Integrating expert knowledge with statistical analysis for landslide susceptibility assessment at regional scale", *Geosciences (Basel)*, vol. 6, p. 14, 2016. [<http://dx.doi.org/10.3390/geosciences6010014>]
- [28] Q.Q. Wang, D.C. Wang, H. Yong, Zh.H. Wang, L.H. Zhang, and Q.Z. Guo, "Landslide susceptibility mapping based on selected optimal combination of landslide predisposing factors in a large catchment", *Sustainability*, vol. 7, pp. 16653-16669, 2015. [<http://dx.doi.org/10.3390/su71215839>]
- [29] W. Chen, X. Xie, and J. Wang, "A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility", *Catena*, vol. 151, pp. 147-160, 2017. [<http://dx.doi.org/10.1016/j.catena.2016.11.032>]
- [30] C.V. Patriche, R. Pirnau, and A. Grozavu, "A comparative analysis of binary logistic regression and analytical hierarchy process for landslide susceptibility assessment in the dobrov river basin, romania", *Pedosphere*, vol. 26, pp. 335-350, 2016. [[http://dx.doi.org/10.1016/S1002-0160\(15\)60047-9](http://dx.doi.org/10.1016/S1002-0160(15)60047-9)]

---

© 2018 Gong *et al.*

This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Public License (CC-BY 4.0), a copy of which is available at: <https://creativecommons.org/licenses/by/4.0/legalcode>. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.