

Article

Hybrid Integration Approach of Entropy with Logistic Regression and Support Vector Machine for Landslide Susceptibility Modeling

Tingyu Zhang ¹, Ling Han ^{1,*} , Wei Chen ² and Himan Shahabi ³ 

¹ School of Earth Science and Resources, Chang'an University, Key Laboratory of Degraded and Unutilized Land Remediation Engineering, Ministry of Land and Resources, Shaanxi Provincial Key Laboratory of Land Rehabilitation, Xi'an 710064, Shaanxi, China; ztyymy@live.com

² College of Geology & Environment, Xi'an University of Science and Technology, Xi'an 710054, Shaanxi, China; chenwei0930@xust.edu.cn

³ Department of Geomorphology, Faculty of Natural Resources, University of Kurdistan, Sanandaj 66177-15175, Iran; himanshabhi@gmail.com

* Correspondence: 2016027003@chd.edu.cn; Tel.: +86-130-0841-8178

Received: 7 October 2018; Accepted: 7 November 2018; Published: 17 November 2018



Abstract: The main purpose of the present study is to apply three classification models, namely, the index of entropy (IOE) model, the logistic regression (LR) model, and the support vector machine (SVM) model by radial basis function (RBF), to produce landslide susceptibility maps for the Fugu County of Shaanxi Province, China. Firstly, landslide locations were extracted from field investigation and aerial photographs, and a total of 194 landslide polygons were transformed into points to produce a landslide inventory map. Secondly, the landslide points were randomly split into two groups (70/30) for training and validation purposes, respectively. Then, 10 landslide explanatory variables, such as slope aspect, slope angle, altitude, lithology, mean annual precipitation, distance to roads, distance to rivers, distance to faults, land use, and normalized difference vegetation index (NDVI), were selected and the potential multicollinearity problems between these factors were detected by the Pearson Correlation Coefficient (PCC), the variance inflation factor (VIF), and tolerance (TOL). Subsequently, the landslide susceptibility maps for the study region were obtained using the IOE model, the LR–IOE, and the SVM–IOE model. Finally, the performance of these three models was verified and compared using the receiver operating characteristics (ROC) curve. The success rate results showed that the LR–IOE model has the highest accuracy (90.11%), followed by the IOE model (87.43%) and the SVM–IOE model (86.53%). Similarly, the AUC values also showed that the prediction accuracy expresses a similar result, with the LR–IOE model having the highest accuracy (81.84%), followed by the IOE model (76.86%) and the SVM–IOE model (76.61%). Thus, the landslide susceptibility map (LSM) for the study region can provide an effective reference for the Fugu County government to properly address land planning and mitigate landslide risk.

Keywords: landslides; hybrid model; statistical method; machine learning; loess area

1. Introduction

Landslides often occur in mountainous and hilly areas and are one of the most dangerous geological disasters [1]. Landslides can cause huge economic losses and a large number of casualties. According to statistics, almost 1000 people and 4 billion dollars are lost annually in the world [2], and this figure still keeps growing. China is also a region where landslides frequently occur; it has been reported that 7122 geological disasters occurred in 2017, causing 327 deaths, 173 injured, 25 missing, and a loss of 3.54 billion CNY [3]. In addition, in northwestern China, landslides pose a greater threat to

resident security and transportation, because of the harsh environment and population concentration. However, enormous manpower and material resources may be required to control and renovate every landslide. Therefore, predicting landslide occurrence is both valuable and important.

As the first step to predicting landslide occurrences, a landslide susceptibility analysis aims to recognize hazardous and high-risk regions, and a preference for the negative effects of landslides [4]. The landslide susceptibility map (LSM) is the final result of the landslide susceptibility analysis. However, the traditional methods for landslide susceptibility mapping based on field investigation and manual analysis are time-consuming and expensive, and the result is imprecise [5,6]. In recent years, geographical information systems (GIS) have been vigorously developed, which make the preparation of the landslide susceptibility map more convenient, which has great advantages [7]. Meanwhile, there has been a lot of research on the combination of geographical information systems, and statistical and nonstatistical methods to evaluate landslide susceptibility—in terms of the binary statistical method, for example, the frequency ratio (FR) model [8–13], the certainty factor (CF) model [14–17], the statistical index (SI) [18,19], the weights of evidence (WOE) [20–22], and the index of entropy (IOE) model [23,24]. The factor internal coefficient of certainty or weight of evidence is decided by landslide data, but the selection of factors would be influenced by humans. As a multivariate statistical method, the logistic regression (LR) model is extensively applied by many researchers [25–30].

Due to the limitation of statistical models, some machine learning algorithms that can avoid the influence from humans were also introduced and applied for landslide susceptibility analysis, such as artificial neural networks (ANN) [31–33], neuro-fuzzy [34–37], fuzzy logic [38,39], decision trees [40–42], kernel logistic regression (KLR) [43,44], and support vector machines (SVM) [45–47].

Statistical models and machine learning algorithms have their own advantages and disadvantages [48,49]. The internal parameters of the explanatory variables in binary statistical models are determined by landslide data, which can avoid the interference of human factors and be more objective. However, the selection of explanatory variables will receive interference from humans. By contrast, multivariable statistical models and machine learning methods can avoid the problem of factor dependence, but they are less widespread and limited to few cases of study for their intensive computation [50,51]. In recent years, many hybrid models have been used in the literature, such as the fuzzy weight of evidence method [17], adaptive network-based fuzzy inference system (ANFIS) based on frequency ratio (FR–ANFIS) model [52], wavelet packet–statistical (WP–SM) models [53], and integration of support vector machines and the multiboost [54]. According to plenty of research, the hybrid model generally performed better than the original models, so trying to mix different models and apply them to different regions is significant. Therefore, this research assembled the IOE model with the LR and SVM models to form two hybrid models (LR–IOE and SVM–IOE) for landslide susceptibility mapping in the Fugu County of Shaanxi Province, China.

2. Study Area

The Fugu County, whose geographic coordinates are 110°25' to 111°15' east longitude and 38°42' to 39°33' north latitude, covers an area of 3229 Km² (Figure 1). The elevation in the study area is between 761 and 1423 m above sea level, and increases from east to west. The temperate zone with an arid continental monsoon climate is the main climate type in the study region, and the maximum and minimum temperatures in history are 38.9 °C and –24 °C, while the average annual temperature is 9.1 °C. The average annual rainfall is 428.6 mm, and the geographical distribution of rainfall shows a gradual increase from northwest to southwest. Meanwhile, most of the precipitation is concentrated from July to September, accounting for 69% of the annual rainfall. There are 62 rivers with drainage areas above 1 × 10⁷ m² in the study region, and the average annual runoff is 5.911 × 10⁹ m³.

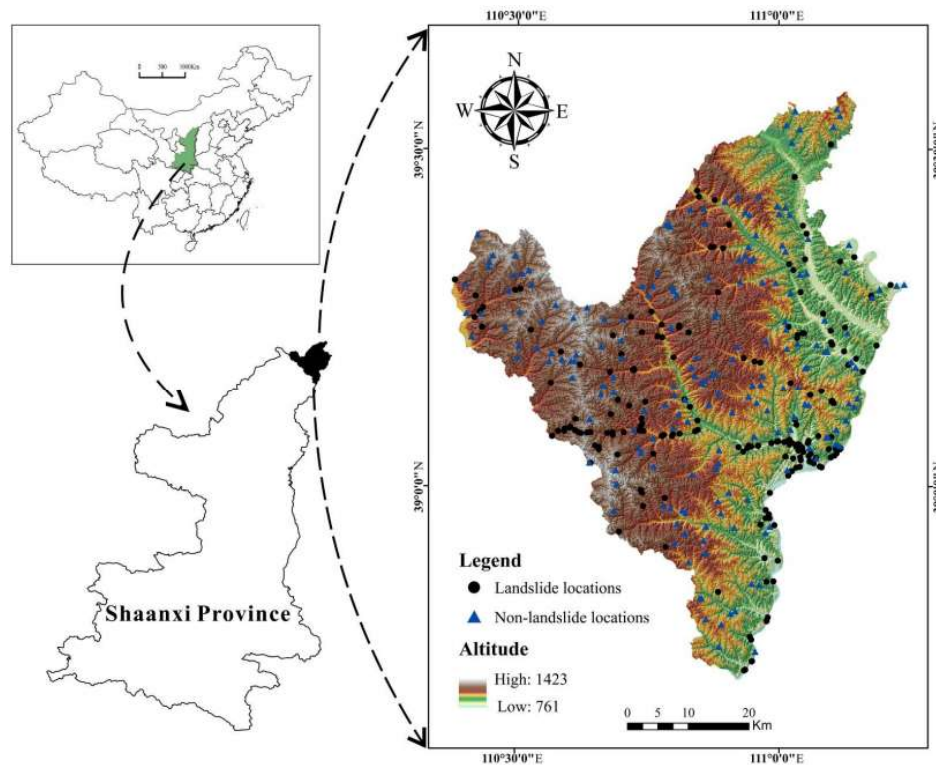


Figure 1. Landslide inventory map and the location of study area.

The overall topography of the study area is high in the northwest and low in the southwest. The main landform types can be divided into four types as follow: Loess girder landform, loess gully landform, canyon hilly landform, and valley terraces. The dip direction of rock formation is roughly southwest–northwest, with a dip angle of approximately 5–8 degrees except for a few areas, which are about 20 degrees. The Carboniferous–Permian strata in the east and the Jurassic strata in the northwest are coal-bearing strata, and the lithology in the study area is shown in Table 1.

Table 1. Lithological units of study area.

Category	Geological Age	Code	Main Lithology
A	Holocene	Q4	Sand, gravel, loess
	Pleistocene	Q3	Loess, gravel
B	Pliocene	N2j	Sandy clay
	Pliocene	N2b	Quartz sand, clay
C	Middle Jurassic	J2y	Siltstone, sandstone, mudstone, shale, coal seam
	Late Jurassic	J1f	Mudstone, glutenite
D	Early Triassic	T3w	Mudstone, shale, coal seam
	Early Triassic	T2-3y	Glutenite, mudstone, shale, siltstone
	Middle Triassic	T2z	Sandstone, mudstone
	Late Triassic	T1h	Medium-fine sandstone, siltstone, mudstone
	Late Triassic	T1l	Sandstone, mudstone
E	Early Permian	P _{2s}	Glutenite, sandstone, mudstone
	Early Permian	P _{2sh}	Mudstone, silty mudstone, sandstone, clay minerals, siliceous
	Late Permian	P _{1sh}	Feldspar quartz sandstone, conglomerate, sandstone, mudstone, shale
	Late Permian	P _{1s}	Mudstone, shale, sandstone, coal seam
F	Carboniferous	C _{2t}	Calcareous sandstone, coal seam, mudstone

Due to the rich coal resources in the study area, the mining industry is developed and the population is concentrated, which caused serious damage to the environment. At the same time, it has also formed massive landslides.

3. Data Used

3.1. Landslide Inventory Map

A landslide inventory map is the first step in a landslide susceptibility analysis and includes historical and newly discovered landslides and their relational information [43], such as the location, the date of occurrence, the extent of landslide phenomena in a region, and the types of mass movements that have left discernable traces [55]. In order to obtain a practical and accurate landslide inventory map, data collection and an adequate field survey were significantly in the current study. A digital elevation model (DEM) of the study region with 30 m resolution was obtained from ASTER GDEM, downloaded from Geospatial Data Cloud [56]. The geological map and mean annual precipitation data were provided by the government of Fugu County. Based on field investigations, a total of 194 landslides polygons, including 162 slides, 29 falls, and 3 debris flows, were drawn according to the depletion zone, and these landslides were triggered by rainfall and excavation. In the study area, the smallest and largest sizes of these landslides were about 39 m² and 13.5 × 10⁴ m², respectively. Because only 12% of landslides are over 10,000 m² in size, landslide polygons were transformed into points using the centroid method and then the landslide inventory map (Figure 1) was obtained in the present study [57,58].

To avoid the overfitting problems in modeling, a total of 194 nonlandslide points were randomly generated and mapped on the landslide inventory map. All of these landslide and nonlandslide points were randomly divided into two groups; namely, the training dataset, including 272 (70%) points, was used to train the models, and the validating dataset, including 116 (30%) points, was used for validation propose.

3.2. Landslide Explanatory Variables

In order to produce the landslide susceptibility map, 10 landslide explanatory variables, namely slope aspect, altitude, slope angle, lithology, mean annual precipitation, distance to roads, distance to rivers, distance to faults, land use, and normalized difference vegetation index (NDVI), were selected to produce data layers representing themselves with a resolution of 30 × 30 m. Slope aspect, altitude, and slope angle maps were extracted from DEM data using ArcGIS software. Land use and NDVI were extracted from GF-2 satellite images gathered from the China Center for Resources Satellite Data and Application. Lithology, distance to roads, mean annual precipitation, distance to rivers, and distance to faults maps were extracted based on existing data.

The slope aspect, which is considered to be a prerequisite condition, was frequently adopted by many works in the literature to produce a landslide susceptibility map [30]. The slope aspect was reclassified into nine groups, based on the equal interval method, as follows: Northwest, west, southwest, south, southeast, east, northeast, north, flat, respectively (Figure 2a).

As it is considered to be another critical factor, the slope angle was widely used by a lot of relevant research [59]. In the current research, the slope angle was divided into the following six categories, based on the Jenks natural break method, as follows: 0°–6.65°, 6.65°–11.40°, 11.40°–16.39°, 16.39°–22.09°, 22.09°–29.45°, 29.45°–60.57° (Figure 2b).

Altitude is also considered a significant factor for landslide susceptibility mapping [1]. Thus, based on the Jenks natural break method, elevation values were classified into the following seven ranges: 761–903 m, 903–984 m, 984–1054 m, 1054–1124 m, 1124–1194 m, 1194–1262 m, and 1262–1423 m (Figure 2c).

The difference of lithology is the basis of landslide formation conditions [60]. According to field investigations and the existing geological data and maps, lithological units were divided into six categories (Table 1) and the lithology map was produced (Figure 2d).

Previous research has indicated that there is a strong correlation between mean annual precipitation and landslide occurrences [61–63]. According to the existing and local observation data, mean annual precipitation is divided into seven classes based on equal interval method as follows: <360 mm/y, 360–380 mm/y, 380–400 mm/y, 400–420 mm/y, 420–440 mm/y, 440–460 mm/y, and >460 mm/y (Figure 2e).

Distance to roads is used as an important landslide explanatory variable to prepare the distance to roads map [64]. In this study, the values of distance to roads were reclassified into five ranges based on equal interval method as follows: <200 m, 200–400 m, 400–600 m, 600–800 m, and >800 m (Figure 2f).

River erosion of slope is considered to be a significant explanatory variable inducing landslides; thus, distance to rivers is employed to be a quantitative index of river erosion [25]. In this study, with 200 m as the interval, the values of distance to rivers were reclassified into five ranges based on equal interval method as follows: <200 m, 200–400 m, 400–600 m, 600–800 m, and >800 m (Figure 2g).

Fault movement is not only the requirement for individual landslide occurrences, but also a controlling factor for regional landslide occurrences [12]. A mass of field surveys indicated that the more fault movement occurred acutely, the more landslides were triggered. In the current research, with 2000 m as the interval, the values of distance to faults were reclassified into five ranges based on equal interval method as follows: <2000 m, 2000–4000 m, 4000–6000 m, 6000–8000 m, and >8000 m (Figure 2h).

Land use in different regions will be different. The use of these land may lead to an asymmetrical distribution of landslides [65]. Thus, land use was also employed to be an explanatory variable in the study region, which was generally divided into five categories as follows: Water, residential areas, bare land, forest/grassland, and farmland (Figure 2i).

NDVI reflects the surface condition and provides a quantitative estimate of vegetation growth and biomass. This is depending on the biomass, the position within the hillslope profile, the root-zone depth and possibility to crack rocks and to prevent or ease water infiltration [66,67]. Therefore, NDVI is also considered to be a pivotal explanatory variable. The computational formula of NDVI is defined as follows:

$$\text{NDVI} = \frac{\text{NIR} - \text{R}}{\text{NIR} + \text{R}}, \quad (1)$$

where R stands for the red part of electromagnetic spectrum, while NIR represents the infrared part of electromagnetic spectrum. Using the Jenks natural break method, the NDVI values were reclassified into five categories as follows: −0.39 to −0.019, −0.019 to 0.063, 0.063–0.134, 0.134–0.216, and 0.216–0.607 (Figure 2j).

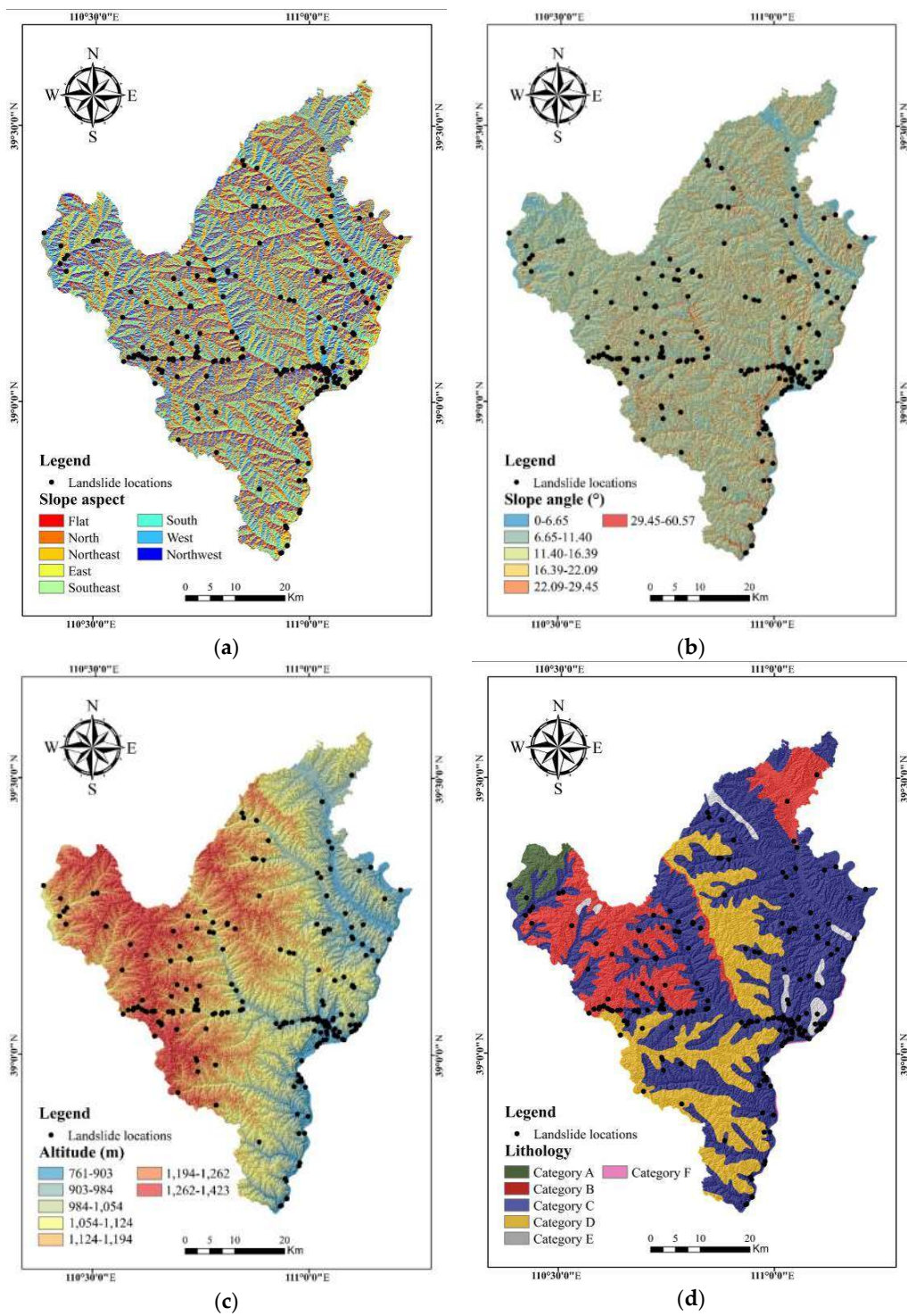


Figure 2. Cont.

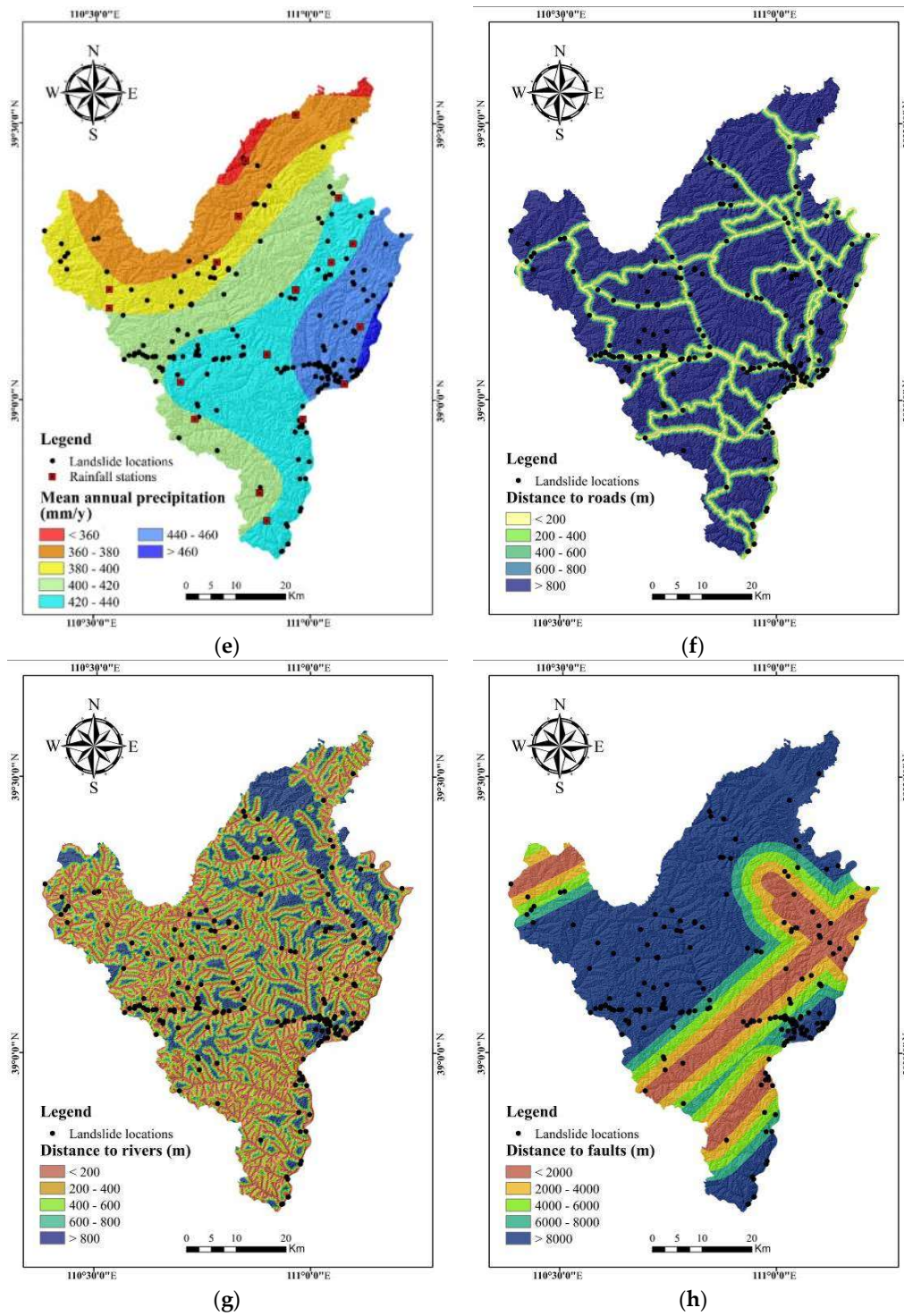


Figure 2. Cont.

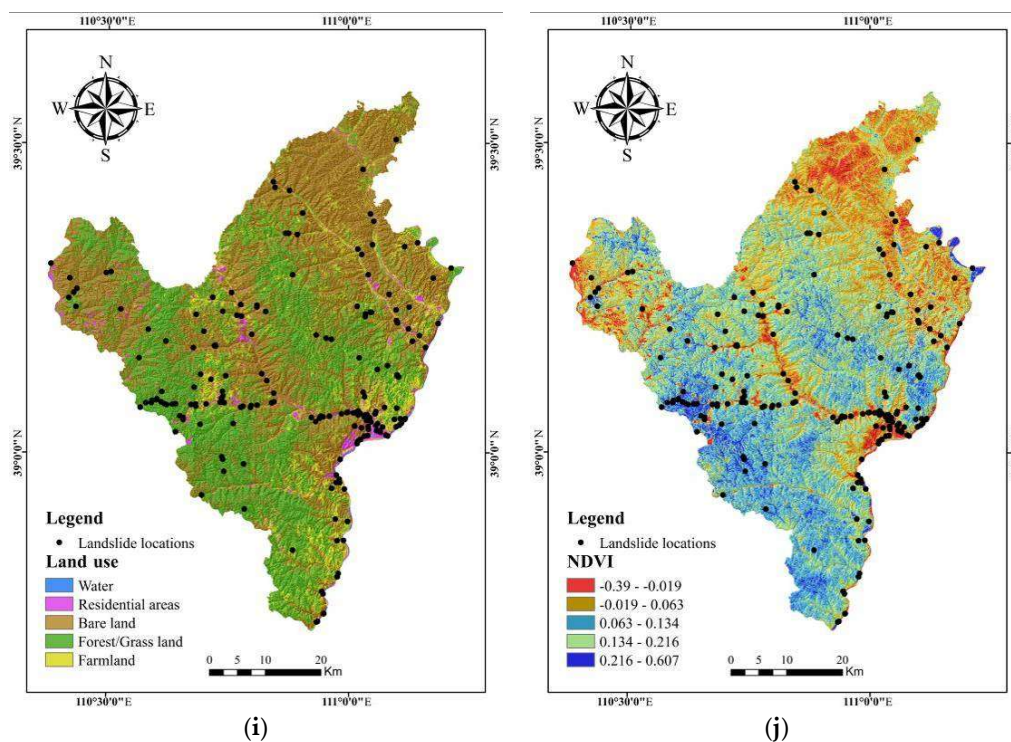


Figure 2. Landslide explanatory variable maps involving: (a) Slope aspect; (b) slope angle; (c) altitude; (d) lithology; (e) mean annual precipitation; (f) distance to roads; (g) distance to rivers; (h) distance to faults; (i) land use; (j) normalized difference vegetation index (NDVI).

4. Methodologies

4.1. Multicollinearity Diagnosis

In the study region, not all explanatory variables have a positive impact on the classification results. Multicollinearity problems may exist between explanatory variables, which may lead to an overfit in modeling. Thus, the Pearson correlation coefficient (PCC), the variance inflation factor (VIF), and tolerance (TOL) were introduced to detect the potential multicollinearity problems [68].

The essence of PCC is a statistical linear correlation coefficient, and its analysis is usually used to measure the linear relationship between distance variables. For two sets of samples X_i ($i = 1, 2, 3, \dots, n$) and Y_j ($j = 1, 2, 3, \dots, n$), the PCC between them can be expressed as:

$$\text{PCC} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \sum_{j=1}^n (y_j - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{j=1}^n (y_j - \bar{y})^2}}, \quad (2)$$

where x_i and y_j are variable values for X_i and Y_j . \bar{x} and \bar{y} are the average of X_i and Y_j , respectively. In general, the greater the absolute value of PCC is, the higher the risk of multicollinearity between the landslide explanatory variables [69], and a PCC of >0.7 indicates a multicollinearity problem [70].

The VIF and TOL are two important indexes for a multicollinearity diagnosis. VIF refers to the ratio of the variance when there is multicollinearity between the conditioning factors and the variance when there is no multicollinearity, and the tolerance is the reciprocal of VIF [71]. In general, the larger the VIF values and the smaller the tolerances values are, the stronger the multicollinearity between the conditioning factors. In this study, the explanatory variables with $\text{VIF} > 2$ or $\text{TOL} < 0.4$ should be abandoned [72].

4.2. Index of Entropy (IOE) Method

The first classification model applied in the present study is the index of entropy (IOE) model, which is a bivariate statistic model; the IOE is also used to be the input data to build the hybrid models in the subsequent modeling. The entropy means the degree of unsteadiness and indeterminacy of a system, and also indicates that elements in a natural environment are the most related development for mass movement [23]. In addition, the entropy represents the degree of different explanatory variables that affect the development of landslides in a landslide susceptibility analysis. The weight values (W_j) of each landslide explanatory variable are determined by the following equations [73]:

$$FR_{ij} = \frac{y_{ij}}{x_{ij}}, \quad (3)$$

$$S_{ij} = \frac{FR_{ij}}{\sum_{j=1}^{N_j} FR_{ij}}, \quad (4)$$

$$M_j = -\sum_{i=1}^{N_j} S_{ij} \log_2 S_{ij}, j = 1, 2, 3, \dots, n, \quad (5)$$

$$M_{jmax} = \log_2 N_j, \quad (6)$$

$$I_j = \frac{M_{jmax} - M_j}{M_{jmax}}, \quad (7)$$

$$W_j = I_j \times FR_{ij}, \quad (8)$$

where FR_{ij} is the frequency ratio value; x and y represent the percentage of domain and percentage of landslides, respectively; S_{ij} stands for the probability density; entropy values are represented by M_j and M_{jmax} ; N_j means the number of categories or ranges of each explanatory variables; and I_j is the information parameters.

Then, the final weight values are calculated by SPSS software. Because these three explanatory variables (aspect, lithology, and land use) are generated from vector graphics with no attribute values, the FR values of aspect, lithology, and land use were used as input data for the computation of W_j . Finally, the landslide susceptibility map for the IOE model is produced using the following equation:

$$LSI_{IOE} = \sum_{j=1}^n \frac{e}{f_j} \times C \times W_j, \quad (9)$$

where LSI_{IOE} stands for the sum of all the categories; j represents the number of explanatory variable maps; e means the number of classes within explanatory variable maps with the greatest number of groups; f_j is the number of classes within particular explanatory variable maps; and C indicates the value of the categories after secondary classification [74].

4.3. Integration of Logistic Regression and Index of Entropy Model

The logistic regression (LR) model is employed to integrate with the IOE to build a new hybrid model, namely, the LR–IOE model in this study. Logistic regression is a commonly used statistical analysis method for regression analysis of binary classification dependent variables. The superiority of the LR model is that independent variables can be discrete or continuous and there is no need to satisfy the normal distribution [75]. In a logistic regression analysis, the dependent variable has values

of 0 and 1, representing nonlandslide occurrences and landslide occurrences, respectively. The LR model can be expressed as the following equation:

$$P = \frac{\exp(Z)}{1 + \exp(Z)}, \quad (10)$$

where P stands for the probability of landslide occurrences, whose value ranges from 0 to 1; Z is calculated by the following equation with the output values range from $-\infty$ to $+\infty$:

$$Z = B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_n, \quad (11)$$

where n is the number of independent variables; B_i ($i = 1, 2, 3, \dots, n$) is the logistic regression coefficient and X_i are the values of the n explanatory variables; and B_0 is a constant.

Because the values of S_{ij} were obtained from the IOE model and the dimension of S_{ij} is uniform, it can avoid the linear correlation between landslides and explanatory variables and also reduce the noise in modeling. In this study, the 10 explanatory variables were reclassified with the corresponding S_{ij} values. Then, the values of S_{ij} were regarded as the input data to build the hybrid model (LR–IOE) through the forward stepwise method to calculate B_0 and B_i .

4.4. Integration of Support Vector Machine and Index of Entropy Model

The basic theory of the support vector machine is to transform the input space into high-dimensional space through an inner product function using the training data [76]. The support vectors are defined as the training samples that have the smallest distance from the optimal hyperplane [40]. In this study, SVM is designed to solve binary classification problems, which means that the positive and negative samples exist at the same time.

Consider a set of training vectors x_i ($i = 1, 2, 3, \dots, n$), and x_i consists of two types denoted as $y_i = \pm 1$ [77]. SVM aims to search an n -dimensional hyperplane distinguishing the two categories; meanwhile, ensure that these two classes are farthest from the hyperplane. Using mathematical formulas, this can be expressed as follows:

$$P = \frac{1}{2} \|w\|^2, \quad (12)$$

followed by constraints:

$$y_i((w \times x_i) + k) \geq 1 \quad (13)$$

where $\|w\|$ stands for the norm of hyperplane normal; k is a constant. By applying the Lagrangian multiplier (λ_i), the cost function can be written as:

$$L = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \lambda_i (y_i((w \times x_i) + k) - 1). \quad (14)$$

In addition, slack variable ξ_i is applied to solve the nonseparable problems [76]; thus, Equations (12) and (13) can be modified as:

$$y_i((w \times x_i) + k) \geq 1 - \xi_i, \quad (15)$$

$$L = \frac{1}{2} \|w\|^2 - \frac{1}{vn} \sum_{i=1}^n \xi_i, \quad (16)$$

where v stands for misclassification, with values ranging from 0 to 1. In addition, by introducing a kernel function, the nonlinear decision boundary can be calculated. In the current research, the following kernel function, namely, the radial basis function (RBF), which is considered to be one of

the most powerful kernels [78], is selected to calculate LSI_{SVM} and produce landslide susceptibility map. The radial basis function is shown as follows:

$$K(x_i, x_j) = \exp(-\delta \|x_i - x_j\|^2), \delta > 0, \quad (17)$$

where δ accounts for the width of the Gaussian kernel function [19].

Similarly, the S_{ij} was used to be the input data for the SVM model and then build the new hybrid model (SVM-IOE).

4.5. The ROC Curve

To test the performance of LSMs obtained by the three models, the receiver operating characteristics (ROC) curve was applied. Based on a series of different dichotomies (cutoffs or decision thresholds), the ROC curve plots 1—specificity as X-axis and sensitivity as Y-axis, which can be expressed as:

$$X - \text{axis} = 1 - \text{specificity} = 1 - \left[\frac{TN}{TN + FP} \right], \quad (18)$$

$$Y - \text{axis} = 1 - \text{sensitivity} = \frac{TP}{TP + FN}, \quad (19)$$

where TP represents true positive, TN is true negative, FP is false positive, and FN is false negative [79]. The quality of these three models predicting the occurrences or non-occurrences of landslide can be measured by the area under the ROC curve (AUC) [9]. The AUC values range from 0 to 1; in addition, if the AUC value is closer to 1, it indicates that the accuracy of model prediction is higher. Conversely, if AUC value is less than 0.5, and closer to 0, it indicates that the model prediction has no practical value [80].

5. Results

5.1. Assessment of Explanatory Variables

In this study, the training dataset was used to evaluate explanatory variables and the Pearson correlation coefficient between pairs of explanatory variables was calculated (Table 2). It can be seen from the results that the lowest PCC value is -0.009 , which happened between altitude and NDVI, and the highest PCC value happened between slope aspect and distance to rivers (0.368). All PCC values are less than 0.7.

The calculation results of VIF and TOL are shown in Table 3. It can be observed that the maximum VIF value is 1.926 and the minimum TOL value is 0.519, which means all the explanatory variables can be applied for landslide susceptibility modeling.

Table 2. Pearson correlation coefficient between pairs of explanatory variables.

Explanatory Variables	Slope Aspect	Slope Angle	Altitude	Lithology	Mean Annual Precipitation	Distance to Roads	Distance to Rivers	Distance to Faults	Land Use
Slope aspect	1								
Slope angle	0.037	1							
Altitude	0.116	0.003	1						
Lithology	0.165	0.170	0.010	1					
Mean annual precipitation	0.140	0.100	−0.021	0.025	1				
Distance to roads	0.280	0.067	0.079	0.048	0.205	1			
Distance to rivers	0.368	0.104	0.112	−0.010	0.004	0.160	1		
Distance to faults	0.320	0.054	−0.070	0.075	0.024	0.034	0.119	1	
Land use	0.123	−0.116	0.087	0.053	0.287	0.050	0.084	0.019	1
NDVI	0.038	0.011	−0.009	0.179	0.146	−0.065	−0.055	0.047	0.082

Table 3. VIF and tolerances for explanatory variables.

Explanatory Variables	VIF	Tolerances
Slope angle	0.657	1.523
Slope aspect	0.962	1.040
Altitude	0.790	1.265
Distance to rivers	0.687	1.455
Distance to roads	0.573	1.746
Distance to faults	0.909	1.100
NDVI	0.770	1.298
Land use	0.910	1.099
Lithology	0.519	1.926
Mean annual precipitation	0.611	1.637

5.2. Result of IOE Model

The calculation method of W_j has already been described in Section 4.2, Equations (3)–(8), and the results are shown in Table 4. The FR_{ij} values shown in Table 4 were used as the input data for slope aspect, lithology, and land use. For the remaining explanatory variables, the original (continuous) data were used as input data to compute the IOE values. Based on the obtained results, the landslide susceptibility index for the IOE model (LSI_{IOE}) was calculated using Equation (9) and was written as follows:

$$LSI_{IOE} = (\text{slope aspect} \times 0.084) + (\text{slope angle} \times 0.064) + (\text{altitude} \times 0.874) + (\text{lithology} \times 0.119) + (\text{mean annual precipitation} \times 0.232) + (\text{distance to roads} \times 0.517) + (\text{distance to rivers} \times 0.127) + (\text{distance to faults} \times 0.030) + (\text{land use} \times 0.974) + (\text{NDVI} \times 0.303) \quad (20)$$

In the end, all of the 10 explanatory variables were used to build the IOE model, and LSI_{IOE} values range from -10.37 to 11.67 . LSI_{IOE} values reflect the probability of landslide occurrence. In other words, the closer the values of LSI_{IOE} are to 11.67 , the higher the probability of landslide occurrence, and the values of LSI_{IOE} are close to -10.37 , indicating that the probability of occurrence of a landslide is lower. Then, the natural break method was applied to classify the final LSM produced by the IOE model into four categories, which were low (-10.37 to -4.33), moderate (-4.33 to -1.65), high (-1.65 to 1.64), and very high (1.64 to 11.67) (Figure 3a). Additionally, the area percentage of low, moderate, high, and very high regions is 31.24%, 16.39%, 33.23%, and 19.14%, respectively.

Table 4. Spatial relationship between each landslide explanatory variable and landslide by the index of entropy (IOE) model.

Explanatory Variables	Classes	No. of Pixels in Domain	% Percentage of Domain	No. of Landslide	% Percentage of Landslides	FR_{ij}	S_{ij}	M_j	M_{jmax}	I_j	W_j	B_i
Slope aspect	Flat	736	0.021	0	0.000	0.000	0.000	2.870	3.170	0.095	0.084	0.061
	North	436,175	12.234	9	6.569	0.537	0.067					
	Northeast	478,233	13.413	21	15.328	1.143	0.143					
	East	453,979	12.733	9	6.569	0.516	0.065					
	Southeast	435,974	12.228	32	23.358	1.910	0.239					
	South	492,245	13.806	15	10.949	0.793	0.099					
	Southwest	471,646	13.229	25	18.248	1.379	0.173					
	West	413,514	11.598	13	9.489	0.818	0.103					
	Northwest	382,820	10.737	13	9.489	0.884	0.111					
Slope angle (°)	0–6.65	434,598	12.190	16	11.679	0.958	0.135	2.445	2.585	0.054	0.064	0.043
	6.65–11.40	954,012	26.758	31	22.628	0.846	0.119					
	11.40–16.39	937,524	26.296	25	18.248	0.694	0.098					
	16.39–22.09	640,546	17.966	28	20.438	1.138	0.161					
	22.09–29.45	349,550	9.804	14	10.219	1.042	0.147					
	29.45–60.57	249,092	6.987	23	16.788	2.403	0.339					
Altitude (m)	761–903	71,702	2.011	26	18.978	9.437	0.675	1.577	2.807	0.438	0.874	−0.252
	903–984	354,938	9.955	26	18.978	1.906	0.136					
	984–1054	796,328	22.335	27	19.708	0.882	0.063					
	1054–1124	851,004	23.869	26	18.978	0.795	0.057					
	1124–1194	989,546	27.755	28	20.438	0.736	0.053					
	1194–1262	487,438	13.672	4	2.920	0.214	0.015					
	1262–1423	14,366	0.403	0	0.000	0.000	0.000					
Lithology	Category A	80,805	2.266	1	0.730	0.322	0.109	1.963	2.585	0.240	0.119	−0.013
	Category B	650,270	18.239	14	10.219	0.560	0.189					
	Category C	2,029,316	56.918	115	83.942	1.475	0.497					
	Category D	736,194	20.649	6	4.380	0.212	0.072					
	Category E	65,704	1.843	1	0.730	0.396	0.134					
	Category F	3033	0.085	0	0.000	0.000	0.000					

Table 4. Cont.

Explanatory Variables	Classes	No. of Pixels in Domain	% Percentage of Domain	No. of Landslide	% Percentage of Landslides	FR_{ij}	S_{ij}	M_j	M_{jmax}	I_j	W_j	B_i
Mean annual precipitation (mm/y)	<360	63,468	1.780	2	1.460	0.820	0.081	2.357	2.807	0.160	0.232	0.239
	360–380	630,456	17.683	5	3.650	0.206	0.020					
	380–400	537,282	15.070	20	14.599	0.969	0.096					
	400–420	850,900	23.866	22	16.058	0.673	0.066					
	420–440	999,895	28.045	44	32.117	1.145	0.113					
	440–460	451,402	12.661	39	28.467	2.248	0.222					
	>460	31,919	0.895	5	3.650	4.077	0.042					
Distance to roads (m)	<200	385,498	10.812	77	56.204	5.198	0.617	1.609	2.322	0.307	0.517	−0.533
	200–400	311,580	8.739	20	14.599	1.670	0.198					
	400–600	282,125	7.913	9	6.569	0.830	0.099					
	600–800	248,289	6.964	4	2.920	0.419	0.050					
	>800	2,337,830	65.571	27	19.708	0.301	0.036					
Distance to rivers (m)	<200	1,108,722	31.097	86	62.774	2.019	0.501	1.956	2.322	0.158	0.127	−0.269
	200–400	881,383	24.721	26	18.978	0.768	0.191					
	400–600	642,145	18.011	12	8.759	0.486	0.121					
	600–800	389,497	10.925	7	5.109	0.468	0.116					
	>800	543,575	15.246	6	4.380	0.287	0.071					
Distance to faults (m)	<2000	526,624	14.771	19	13.869	0.939	0.190	2.251	2.322	0.030	0.030	0.110
	2000–4000	459,271	12.882	10	7.299	0.567	0.115					
	4000–6000	431,651	12.107	14	10.219	0.844	0.171					
	6000–8000	344,339	9.658	20	14.599	1.512	0.307					
	>8000	1,803,437	50.583	74	54.015	1.068	0.217					
Land use	Water	13,266	0.372	0	0.000	0.000	0.000	1.258	2.322	0.458	0.974	0.061
	Residential areas	86,117	2.415	25	18.248	7.555	0.711					
	Bare land	178,0712	49.945	71	51.825	1.038	0.098					
	Forest/Grassland	1,317,845	36.963	17	12.409	0.336	0.032					
	Farmland	367,382	10.304	24	17.518	1.700	0.160					
NDVI	−0.39 to −0.019	278,430	7.809	40	19.197	3.739	0.577	1.779	2.322	0.234	0.303	−0.354
	−0.019 to 0.063	988,700	27.731	38	27.737	1.000	0.154					
	0.063–0.134	1,233,777	34.605	43	31.387	0.907	0.140					
	0.134–0.216	837,512	23.491	12	8.759	0.373	0.058					
	0.216–0.607	226,903	6.364	4	2.920	0.459	0.071					

B_0 is 2.345.

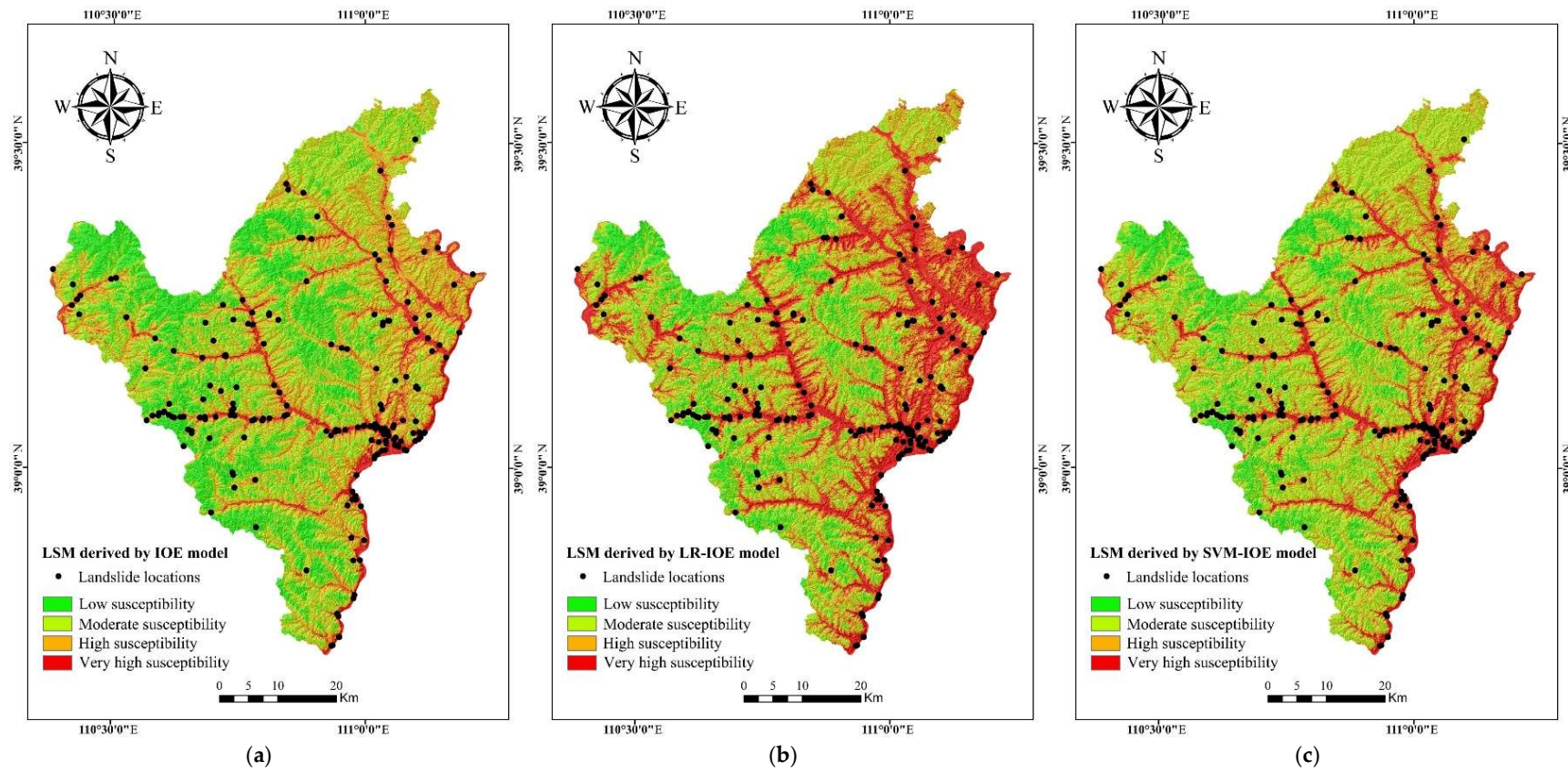


Figure 3. Landslide susceptibility map derived from: (a) The IOE model; (b) logistic regression (LR)–IOE model; (c) support vector machine (SVM)–IOE model.

5.3. Result of LR–IOE Model

The calculation method of Z has already been described in Section 4.2, Equations (3)–(8). The S_{ij} values shown in Table 4 were used as the input data for all 10 explanatory variables through the reclassification method to build the LR–IOE model and to compute B_0 and B_i using SPSS software. Based on the results, Equation (11) can be written as follows:

$$Z = 2.345 + (\text{slope aspect} \times 0.061) + (\text{slope angle} \times 0.043) + (\text{altitude} \times -0.252) + (\text{lithology} \times -0.013) + (\text{mean annual precipitation} \times 0.239) + (\text{distance to roads} \times -0.533) + (\text{distance to rivers} \times -0.269) + (\text{distance to faults} \times 0.110) + (\text{land use} \times 0.061) + (\text{NDVI} \times -0.354) \quad (21)$$

Subsequently, the LSI_{LR-IOE} values were obtained, which range from 0.016 to 0.983. LSI_{LR-IOE} values reflect the probability of landslide occurrence. In other words, the closer the values of LSI_{LR-IOE} are to 1, the higher the probability of landslide occurrence, and the values of LSI_{LR-IOE} are close to 0, indicating that the probability of landslide occurrence is lower. Similarly, the natural break method was applied to classify the final LSM produced by the LR–IOE model into four categories: Low (0.016–0.248), moderate (0.248–0.445), high (0.445–0.688), and very high (0.688–0.983) (Figure 3b). In addition, the area percentage of low, moderate, high, and very high is 16.77%, 33.06%, 21.05%, and 29.12%, respectively.

5.4. Result of SVM–IOE Model

In the current research, the parameters of the radial basis function were selected by the grid search method with 10-fold cross validation, and then the entropy was regarded as the input data to calculate the $LSI_{SVM-IOE}$ values based on SVM–IOE model. The $LSI_{SVM-IOE}$ values range from 0.061 to 0.984. The closer the values are to 1, the higher the probability of landslide occurrence, and the values of $LSI_{SVM-IOE}$ are close to 0, indicating that the probability of landslide occurrence is lower. Then, the natural break method was applied to classify the final LSM produced by the SVM–IOE model into four categories: Low (0.061–0.271), moderate (0.271–0.437), high (0.437–0.658), and very high (0.658–0.984) (Figure 3c). The area percentage of low, moderate, high, and very high is 15.08%, 29.56%, 33.39%, and 21.97%, respectively.

5.5. Validation of Landslide Susceptibility Maps

In the current study, the ROC curve was used to validate and compare the performance of the IOE, LR–IOE, and SVM–IOE models. The final AUC values represent the success and prediction rate derived from the training and validating dataset, respectively.

In the end, for success rate results, the AUC values for the IOE, LR–IOE, and SVM–IOE models were observed to be 0.8743, 0.9011, and 0.8653, respectively (Figure 4a). That is to say, the training accuracy of the susceptibility maps is 87.43%, 90.11%, and 86.53%, respectively. In terms of prediction rate results, the AUC values for the IOE, LR–IOE, and SVM–IOE models were found to be 0.7686, 0.8184, and 0.7661, respectively (Figure 4b). In other words, the prediction accuracy of the susceptibility maps is 76.86%, 81.84%, and 76.61%, respectively.

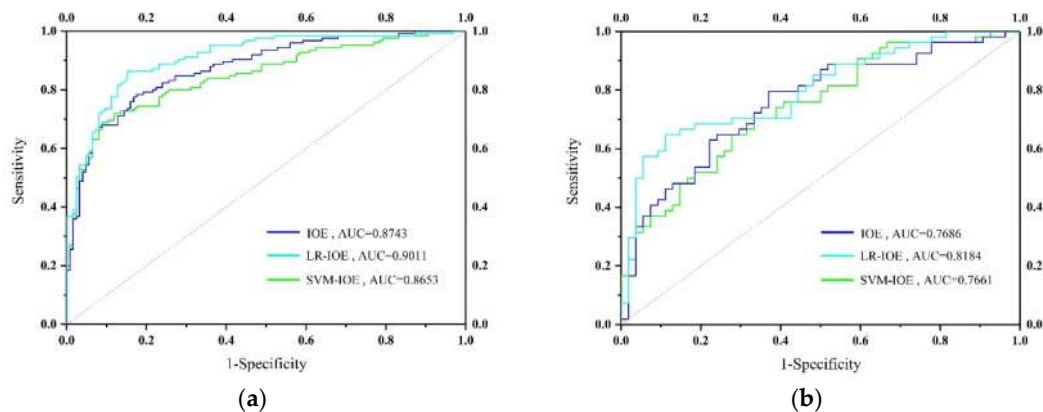


Figure 4. Receiver operating characteristics (ROC) curves of models: (a) Training dataset; (b) validating dataset.

Generally, the results of both the success rate and prediction rate express reasonable and practical accuracies in the current research. However, the LR-IOE model shows the best result for the current study.

6. Discussion

Spatial prediction of landslides is a critical process in the study of landslides and the accuracy of prediction will be affected by the models that we used, and the input data extracted from explanatory variables. However, there is no definitive conclusion about the methods used to select and evaluate explanatory variables. Therefore, it is necessary to investigate the methods which will help us to obtain reasonable conclusions. In this study, we calculated the IOE and PCC to assess 10 explanatory variables, and evaluated three classification models, namely, IOE, LR-IOE, and SVM-IOE, for landslide susceptibility mapping.

According to PCC values (Table 2), all 10 factors are less than 0.7, which means these 10 factors cannot generate noise in landslide susceptibility modeling. From the index of entropy (Table 4), we can see the residential areas have the highest value (7.555), which means that most landslides occurred in this region. We believe that the reason for this condition is the concentration of population and the fact that human engineering activities are intense in this area. Similarly, the closer to the road, the higher the frequency of landslides that occurred was. For the slope aspect, most landslides occurred on south-facing slopes; the reason for this condition may be the climate, and the same results were also reported by the authors of [37] (p. 82). The category C (Siltstone, sandstone, mudstone, shale, coal seam, glutenite) in lithology is the region where the largest number of landslides has occurred. This may be due to the softness of sandstone and siltstone structures and strong weathering erosion. In the case of slope angle and mean annual precipitation, the rate of landslide occurrence is roughly proportional to them. The reason may be that a large amount of water infiltrate increases the water content and weight of the rock and soil mass and increases the sliding force of the rock and soil mass, and the steeper the slope, the stronger the slip force of the rock and soil mass. Interestingly, with the values of distance to faults, distance to rivers, distance to roads, altitude, and NDVI increasing, the IOE is gradually decreasing. The reason for this phenomenon is that road construction usually causes instability, while roads in the study region are generally built at low altitudes and away from faults. The root of the vegetation is conducive to the stability of the soil, while the erosion of the rivers will affect the stability of the slope. These conditions are roughly the same as those observed in the field.

In this study, the selection of explanatory variables was based on previous studies and field observations, which will cause interference from human factors. In addition, although we calculated all the W_j values for the 10 explanatory variables, it is not clear how much the method developed in the work is sensitive to the number of the classes and to the choice of the breaking points. Therefore, this is the focus of future research.

As shown in Figure 4, we can see the AUC value of the LR–IOE model is the highest among the three models, whether it is for the success or prediction rate, which means that the LR–IOE model performs best in landslide susceptibility mapping in this study. However, the AUC value of the SVM–IOE model is the lowest, which may be due to the fact that the SVM–IOE model is more dependent on the selection of the kernel function, and there is no objective way to solve it.

In terms of the proportion of the final susceptibility mapping results (Figure 5), it can be observed that the proportion of high and very high regions obtained by the three models is about 52%. Among them, the LR–IOE model has the lowest result (50.17%), which implies an efficient result corresponding to the LR–IOE model, and it can also improve the efficiency of decision-making and reduce costs.

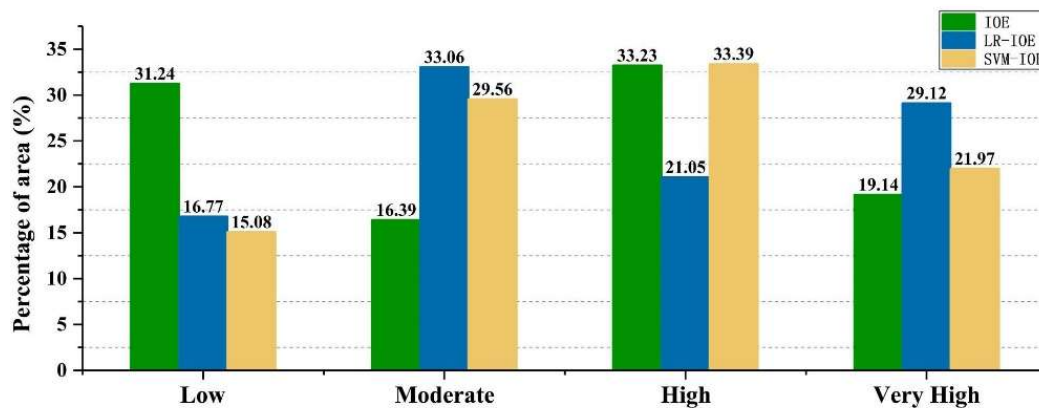


Figure 5. Percentages of different landslide susceptibility classes for the three models.

7. Conclusions

In this present study, the IOE model, LR–IOE model, and SVM–IOE model were used to obtain landslide susceptibility maps for the Fugu County of Shaanxi Province, China. Ten explanatory variables, namely, altitude, slope aspect, mean annual precipitation, slope angle, lithology, distance to roads, land use, distance to rivers, distance to faults, and NDVI, were selected and the potential multicollinearity problem among them was detected by PCC, VIF, and TOL. The results of the analysis showed that there are no potential multicollinearity problems between these 10 factors and they are available for landslide susceptibility modeling. A total of 194 landslides, including landslides recognized from extensive field investigations and historical landslide records, and 194 nonlandslide points were also randomly generated. To build the models, 272 (70%) landslide and nonlandslide points were randomly selected and the remaining 116 (30%) landslide and nonlandslide points were applied for validating purposes. A natural break method was used to split the study region into four categories: Low, moderate, high, and very high. In the end, the performance of the achieved landslide susceptibility maps was evaluated using AUC values.

In terms of the success rate presented by the AUC values, the LR–IOE model has the highest training accuracy (90.11%), followed by the IOE model (87.43%) and the SVM–IOE model (86.53%). As for the prediction rate, the LR–IOE model has the highest training accuracy (81.84%), followed by the IOE model (76.86%) and the SVM–IOE model (76.61%). Thus, the results prove that these three models present good performance in landslide susceptibility mapping. The LR–IOE model performed best for this research and is more suitable for landslide susceptibility mapping in the study area.

The results of this study provide available information for the engineers, decision makers, and urban planners in this study region.

Author Contributions: T.Z. established the model and wrote the main manuscript text. L.H. guided the work and analysis. W.C. and H.S. contributed to the adjustment of the article structure. This paper was prepared using the contributions of all authors. All authors have read and approved the final manuscript.

Funding: This research was funded by National Key Research and Development Program of China, Ecological Safety Guarantee Technology and Demonstration Channel and Slope Treatment Project in Loess Hilly and Gully Area, grant number 2017YFC0504700.

Acknowledgments: We thank the Shaanxi Provincial Key Laboratory of Land Rehabilitation for data used in this study. Special thanks are given to Zhou Zhao, the associate professor of Xi'an University of Science and Technology. We also thank China Center for Resources Satellite Data and Application.

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

References

1. Akgun, A.; Erkan, O. Landslide susceptibility mapping by geographical information system-based multivariate statistical and deterministic models: In an artificial reservoir area at northern Turkey. *Arab. J. Geosci.* **2016**, *9*, 1–15. [[CrossRef](#)]
2. Petley, D. Global patterns of loss of life from landslides. *Geology* **2012**, *40*, 927–930. [[CrossRef](#)]
3. National Statistics on Geological Disasters in 2017. Available online: www.jianzai.gov.cn (accessed on 25 October 2018).
4. Brabb, E.E. Innovative approaches to landslide hazard mapping. In Proceedings of the IV International Symposium on Landslides, Toronto, Canada, 23–31 August 1985; Volume 1, pp. 307–324.
5. Yin, K. The computer-assisted mapping of landslide hazard zonation. *Hydrogeol. Eng. Geol.* **1993**, *5*, 21–23.
6. Brabb, E.E. The San Mateo County California Gis Project for Predicting the Consequences of Hazardous Geologic Processes. In *Geographical Information Systems in Assessing Natural Hazards*; Springer: Dordrecht, The Netherlands, 1995.
7. Pourghasemi, H.R.; Pradhan, B.; Gokceoglu, C. Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. *Nat. Hazards* **2012**, *63*, 965–996. [[CrossRef](#)]
8. Yilmaz, I. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat-Turkey). *Comput. Geosci.* **2009**, *35*, 1125–1138. [[CrossRef](#)]
9. Pradhan, B. Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches. *J. Indian Soc. Remote Sens.* **2010**, *38*, 301–320. [[CrossRef](#)]
10. Akinci, H.; Doğan, S.; Kılıçoğlu, C.; Temiz, M.S. Production of landslide susceptibility map of Samsun (Turkey) city center by using frequency ratio method. *Int. J. Phys. Sci.* **2011**, *6*, 1015–1025.
11. Mondal, S.; Maiti, R. Integrating the analytical hierarchy process (AHP) and the frequency ratio (FR) model in landslide susceptibility mapping of Shiv-Khola watershed, Darjeeling Himalaya. *Int. J. Disaster Risk Sci.* **2013**, *4*, 200–212. [[CrossRef](#)]
12. Vakhshoori, V.; Zare, M. Landslide susceptibility mapping by comparing weight of evidence, fuzzy logic, and frequency ratio methods. *Geomatics* **2016**, *7*, 1–21. [[CrossRef](#)]
13. Dev, T.; Tae, I.; Ha, D. GIS-based landslide susceptibility mapping of Bhotang, Nepal using frequency ratio and statistical index methods. *J. Korean Soc. Surv. Geod. Photogramm. Cartogr.* **2017**, *35*, 357–364.
14. 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. [[CrossRef](#)]
15. Pourghasemi, H.R.; Pradhan, B.; Gokceoglu, C.; Mohammadi, M.; Moradi, H.R. Application of weights-of-evidence and certainty factor models and their comparison in landslide susceptibility mapping at Haraz watershed, Iran. *Arab. J. Geosci.* **2013**, *6*, 2351–2365. [[CrossRef](#)]
16. Wang, Q.; Li, W.; Chen, W.; Bai, H. GIS-based assessment of landslide susceptibility using certainty factor and index of entropy models for the Qianyang County of Baoji City, China. *J. Earth Syst. Sci.* **2015**, *124*, 1–17. [[CrossRef](#)]
17. Hong, H.; Chen, W.; Xu, C.; Youssef, A.M.; Pradhan, B.; Bui, D.T. Rainfall-induced landslide susceptibility assessment at the Chongren area (China) using frequency ratio, certainty factor, and index of entropy. *Geocarto Int.* **2017**, *32*, 139–154. [[CrossRef](#)]

18. Bui, D.T.; Lofman, O.; Revhaug, I.; Dick, O. Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. *Nat. Hazards* **2011**, *59*, 1413–1444. [[CrossRef](#)]
19. Pourghasemi, H.R.; Moradi, H.R.; Aghda, S.M.F. Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. *Nat. Hazards* **2013**, *69*, 749–779. [[CrossRef](#)]
20. Polykretis, C.; Chalkias, C. Comparison and evaluation of landslide susceptibility maps obtained from weight of evidence, logistic regression, and artificial neural network models. *Nat. Hazards* **2018**, *93*, 1–26. [[CrossRef](#)]
21. Ilija, I.; Tsangaratos, P. Applying weight of evidence method and sensitivity analysis to produce a landslide susceptibility map. *Landslides* **2016**, *13*, 379–397. [[CrossRef](#)]
22. Chen, W.; Li, H.; Hou, E.; Wang, S.; Wang, G.; Panahi, M.; Li, T.; Peng, T.; Guo, C.; Niu, C. GIS-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models. *Sci. Total Environ.* **2018**, *634*, 853. [[CrossRef](#)] [[PubMed](#)]
23. Pourghasemi, H.R.; Mohammady, M.; Pradhan, B. Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. *Catena* **2012**, *97*, 71–84. [[CrossRef](#)]
24. Razavizadeh, S.; Solaimani, K.; Massironi, M.; Kaviani, A. Mapping landslide susceptibility with frequency ratio, statistical index, and weights of evidence models: A case study in northern Iran. *Environ. Earth Sci.* **2017**, *76*, 499. [[CrossRef](#)]
25. Manzo, G.; Tofani, V.; Segoni, S.; Battistini, A.; Catani, F. GIS techniques for regional-scale landslide susceptibility assessment: The Sicily (Italy) case study. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 1433–1452. [[CrossRef](#)]
26. Chen, W.; Shahabi, H.; Shirzadi, A.; Hong, H.Y.; Akgun, A.; Tian, Y.Y.; Liu, J.Z.; Zhu, A.X.; Li, S.J. Novel hybrid artificial intelligence approach of bivariate statistical-methods-based kernel logistic regression classifier for landslide susceptibility modeling. *Bull. Eng. Geol. Environ.* **2018**, 1–23. [[CrossRef](#)]
27. Trigila, A.; Iadanza, C.; Esposito, C.; Scarascia-Mugnozza, G. Comparison of logistic regression and random forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy). *Geomorphology* **2015**, *249*, 119–136. [[CrossRef](#)]
28. Saro, L.; Woo, J.S.; Kwanyoung, O.; Mounjin, L. The spatial prediction of landslide susceptibility applying artificial neural network and logistic regression models: A case study of Inje, Korea. *Open Geosci.* **2016**, *8*, 117–132. [[CrossRef](#)]
29. Tsangaratos, P.; Ilija, I. Comparison of a logistic regression and naïve bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size. *Catena* **2016**, *145*, 164–179. [[CrossRef](#)]
30. Mandal, S.; Mandal, K. Modeling and mapping landslide susceptibility zones using GIS based multivariate binary logistic regression (LR) model in the Rorachu river basin of eastern Sikkim Himalaya, India. *Model. Earth Syst. Environ.* **2018**, *4*, 69–88. [[CrossRef](#)]
31. Lin, H.M.; Chang, S.K.; Wu, J.H.; Juang, C.H. Neural network-based model for assessing failure potential of highway slopes in the Alishan, Taiwan Area: Pre- and post-earthquake investigation. *Eng. Geol.* **2009**, *104*, 280–289. [[CrossRef](#)]
32. Conforti, M.; Pascale, S.; Robustelli, G.; Sdao, F. Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo river catchment (northern Calabria, Italy). *Catena* **2014**, *113*, 236–250. [[CrossRef](#)]
33. Aditian, A.; Kubota, T. Causative factors optimization using artificial neural network for GIS-based landslide susceptibility assessments in Ambon, Indonesia. *Int. J. Eros. Control Eng.* **2017**, *10*, 120–129. [[CrossRef](#)]
34. Oh, H.J.; Pradhan, B. Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area. *Comput. Geosci.* **2011**, *37*, 1264–1276. [[CrossRef](#)]
35. Lee, M.J.; Park, I.; Lee, S. Forecasting and validation of landslide susceptibility using an integration of frequency ratio and neuro-fuzzy models: A case study of Seorak mountain area in Korea. *Environ. Earth Sci.* **2015**, *74*, 413–429. [[CrossRef](#)]
36. Chen, W.; Pourghasemi, H.R.; Panahi, M.; Kornejady, A.; Wang, J.; Xie, X. Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. *Geomorphology* **2017**, *297*, 69–85. [[CrossRef](#)]

37. Chen, W.; Panahi, M.; Tsangaratos, P.; Shahabi, H.; Ilia, I.; Panahi, S.; Li, S.; Jaafari, A.; Ahmad, B.B. Applying population-based evolutionary algorithms and a neuro-fuzzy system for modeling landslide susceptibility. *Catena* **2019**, *172*, 212–231. [[CrossRef](#)]
38. Anbalagan, R.; Kumar, R.; Lakshmanan, K.; Parida, S.; Neethu, S. Landslide hazard zonation mapping using frequency ratio and fuzzy logic approach, a case study of Lachung valley, Sikkim. *Geoenviron. Disasters* **2015**, *2*, 1–17. [[CrossRef](#)]
39. Tsangaratos, P.; Loupasakis, C.; Nikolakopoulos, K.; Angelitsa, V.; Ilia, I. Developing a landslide susceptibility map based on remote sensing, fuzzy logic and expert knowledge of the island of Lefkada, Greece. *Environ. Earth Sci.* **2018**, *77*, 363. [[CrossRef](#)]
40. Pradhan, B. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput. Geosci.* **2013**, *51*, 350–365. [[CrossRef](#)]
41. Bui, D.T.; Pradhan, B.; Lofman, O.; Revhau, I. Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and naïve bayes models. *Math. Probl. Eng.* **2012**, 2012.
42. Lombardo, L.; Cama, M.; Conoscenti, C.; Märker, M.; Rotigliano, E. Binary logistic regression versus stochastic gradient boosted decision trees in assessing landslide susceptibility for multiple-occurring landslide events: Application to the 2009 storm event in Messina (Sicily, southern Italy). *Nat. Hazards* **2015**, *79*, 1621–1648. [[CrossRef](#)]
43. Hong, H.; Pradhan, B.; Xu, C.; Bui, D.T. Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. *Catena* **2015**, *133*, 266–281. [[CrossRef](#)]
44. Chen, W.; Xie, X.; Peng, J.; Wang, J.; Duan, Z.; Hong, H. GIS-based landslide susceptibility modelling: A comparative assessment of kernel logistic regression, Naïve-Bayes tree, and alternating decision tree models. *Geomat. Nat. Hazards Risk* **2017**, *8*, 950–973. [[CrossRef](#)]
45. Colkesen, I.; Sahin, E.K.; Kavzoglu, T. Susceptibility mapping of shallow landslides using kernel-based gaussian process, support vector machines and logistic regression. *J. Afr. Earth Sci.* **2016**, *118*, 53–64. [[CrossRef](#)]
46. Lee, S.; Hong, S.M.; Jung, H.S. A support vector machine for landslide susceptibility mapping in Gangwon province, Korea. *Sustainability* **2017**, *9*, 48. [[CrossRef](#)]
47. Huang, Y.; Zhao, L. Review on landslide susceptibility mapping using support vector machines. *Catena* **2018**, *165*, 520–529. [[CrossRef](#)]
48. Chen, W.; Shahabi, H.; Shirzadi, A.; Li, T.; Guo, C.; Hong, H.; Li, W.; Pan, D.; Hui, J.; Ma, M. A Novel Ensemble Approach of Bivariate Statistical Based Logistic Model Tree Classifier for Landslide Susceptibility Assessment. *Geocarto Int.* **2018**, 1–32. [[CrossRef](#)]
49. Chen, W.; Xie, X.; Peng, J.; Himan, S.; Hong, H.; Bui, D.T. GIS-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method. *Catena* **2018**, *164*, 135–149. [[CrossRef](#)]
50. Eeckhaut, M.V.D. Statistical modelling of Europe-wide landslide susceptibility using limited landslide inventory data. *Landslides* **2012**, *9*, 357–369. [[CrossRef](#)]
51. Trigila, A.; Frattini, P.; Casagli, N.; Catani, F.; Crosta, G.; Esposito, C.; Iadanza, C.; Lagomarsino, D.; Mugnozza, G.S.; Segoni, S. Landslide Susceptibility Mapping at National Scale: The Italian Case Study. In *Landslide Science and Practice*; Springer: Berlin, Germany, 2015; pp. 287–295.
52. Aghdam, I.N.; Pradhan, B.; Panahi, M. Landslide susceptibility assessment using a novel hybrid model of statistical bivariate methods (FR and WOE) and adaptive neuro-fuzzy inference system (ANFIS) at southern Zagros mountains in Iran. *Environ. Earth Sci.* **2017**, *76*, 237. [[CrossRef](#)]
53. Moosavi, V.; Niazi, Y. Development of hybrid wavelet packet-statistical models (WP-SM) for landslide susceptibility mapping. *Landslides* **2016**, *13*, 1–18. [[CrossRef](#)]
54. Pham, B.T.; Jaafari, A.; Prakash, I.; Bui, D.T. A novel hybrid intelligent model of support vector machines and the multiboost ensemble for landslide susceptibility modeling. *Bull. Eng. Geol. Environ.* **2018**, *47*, 1–22. [[CrossRef](#)]
55. Guzzetti, F.; Mondini, A.C.; Cardinali, M.; Fiorucci, F.; Santangelo, M.; Chang, K.T. Landslide inventory maps: New tools for an old problem. *Earth Sci. Rev.* **2012**, *112*, 42–66. [[CrossRef](#)]
56. Geospatial Data Cloud. Available online: <https://http://www.gscloud.cn/> (accessed on 25 September 2018).

57. Dou, J.; Yamagishi, H.; Pourghasemi, H.R.; Yunus, A.P.; Song, X.; Xu, Y. An integrated artificial neural network model for the landslide susceptibility assessment of Osado island, Japan. *Nat. Hazards* **2015**, *78*, 1749–1776. [[CrossRef](#)]
58. Chen, W.; Peng, J.; Hong, H.; Shahabi, H.; Pradhan, B.; Liu, J. Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren county, Jiangxi province, China. *Sci. Total Environ.* **2018**, *626*, 230. [[CrossRef](#)] [[PubMed](#)]
59. Nakamura, H.; Kubota, T. Landslide susceptibility from the viewpoint of its slope angle and geology. *J. Jpn. Landslide Soc.* **2010**, *23*, 6–12_1. [[CrossRef](#)]
60. Westen, C.J.V.; Rengers, N.; Soeters, R. Use of geomorphological information in indirect landslide susceptibility assessment. *Nat. Hazards* **2003**, *30*, 399–419. [[CrossRef](#)]
61. Dahal, R.K.; Hasegawa, S.; Nonomura, A.; Yamanaka, M.; Masuda, T.; Nishino, K. GIS-based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping. *Environ. Geol.* **2008**, *54*, 311–324. [[CrossRef](#)]
62. Youssef, A.M. Landslide susceptibility delineation in the Ar-Rayth area, Jizan, kingdom of Saudi Arabia, using analytical hierarchy process, frequency ratio, and logistic regression models. *Environ. Earth Sci.* **2015**, *73*, 1–20. [[CrossRef](#)]
63. Chang, S.K.; Lee, D.H.; Wu, J.H.; Juang, C.H. Rainfall-based criteria for assessing slump rate of mountainous highway slopes: A case study of slopes along Highway 18 in Alishan, Taiwan. *Eng. Geol.* **2011**, *118*, 63–74. [[CrossRef](#)]
64. Erener, A.; Mutlu, A.; Düzgün, H.S. A comparative study for landslide susceptibility mapping using GIS-based multi-criteria decision analysis (MCDA), logistic regression (LR) and association rule mining (ARM). *Eng. Geol.* **2016**, *203*, 45–55. [[CrossRef](#)]
65. Bourenane, H.; Guettouche, M.S.; Bouhadad, Y.; Braham, M. Landslide hazard mapping in the Constantine city, northeast Algeria using frequency ratio, weighting factor, logistic regression, weights of evidence, and analytical hierarchy process methods. *Arab. J. Geosci.* **2016**, *9*, 1–24. [[CrossRef](#)]
66. Jaafari, A.; Najafi, A.; Pourghasemi, H.R.; Rezaeian, J.; Sattarian, A. GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. *Int. J. Environ. Sci. Technol.* **2014**, *11*, 909–926. [[CrossRef](#)]
67. Youssef, A.M.; Pourghasemi, H.R.; Pourtaghi, Z.S.; Alkatheeri, M.M. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah basin, Asir region, Saudi Arabia. *Landslides* **2016**, *13*, 839–856. [[CrossRef](#)]
68. Su, Q.; Zhang, J.; Zhao, S.; Wang, L.; Liu, J.; Guo, J. Comparative assessment of three nonlinear approaches for landslide susceptibility mapping in a coal mine area. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 228. [[CrossRef](#)]
69. Jiang, P.; Chen, J. Displacement prediction of landslide based on generalized regression neural networks with k-fold cross-validation. *Neurocomputing* **2016**, *198*, 40–47. [[CrossRef](#)]
70. Bui, D.T.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **2016**, *13*, 361–378.
71. Menard, S. Applied Logistic Regression Analysis. *Technometrics* **2002**, *38*, 192.
72. Bai, S.B.; Wang, J.; Lü, G.N.; Zhou, P.G.; Hou, S.S.; Xu, S.N. GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China. *Geomorphology* **2010**, *115*, 23–31. [[CrossRef](#)]
73. Al-Abadi, A.M.; Al-Temmeme, A.A.; Al-Ghanimy, M.A. A gis-based combining of frequency ratio and index of entropy approaches for mapping groundwater availability zones at Badra–Al al-Gharbi–Teeb areas, Iraq. *Sustain. Water Resour. Manag.* **2016**, *2*, 265–283. [[CrossRef](#)]
74. Bednarik, M.; Magulová, B.; Matys, M.; Marschalko, M. Landslide susceptibility assessment of the Kra’ovany–Liptovský Mikuláš railway case study. *Phys. Chem. Earth* **2010**, *35*, 162–171. [[CrossRef](#)]
75. Atkinson, P.M.; Massari, R. Generalised linear modelling of susceptibility to landsliding in the central Apennines, Italy. *Comput. Geosci.* **1998**, *24*, 373–385. [[CrossRef](#)]
76. Vapnik, V.N. *Statistics for Engineering and Information Science*; Springer: New York, NY, USA, 2000.

77. Xu, C.; Dai, F.; Xu, X.; Yuan, H.L. GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang river watershed, China. *Geomorphology* **2012**, *145–146*, 70–80. [[CrossRef](#)]
78. Chen, W.; Yan, X.; Zhao, Z.; Hong, H.; Bui, D.T.; Pradhan, B. Spatial prediction of landslide susceptibility using data mining-based kernel logistic regression, naive Bayes and RBFNetwork models for the Long County area (China). *Bull. Eng. Geol. Environ.* **2018**, 1–20. [[CrossRef](#)]
79. Chen, W.; Shirzadic, A.; Shahabi, H.; Ahmade, B.B.; Shuai, Z.; Hong, H.; Ning, Z. A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naïve Bayes tree classifiers for a landslide susceptibility assessment in Langao County, China. *Geomatics Nat. Hazards Risk* **2017**, *8*, 1–23. [[CrossRef](#)]
80. Chen, W.; Zhang, S.; Li, R.; Himan, S. Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve bayes tree for landslide susceptibility modeling method. *Sci. Total Environ.* **2018**, *644*, 1006–1018. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).