Application of an Integrated Model Based on Bivariate and Multivariate Method in Landslide Susceptibility Mapping

Han Hu
Jilin University
Changming Wang (wangcm@jlu.edu.cn)
Jilin University
Zhu Liang
Jilin University

Research Article

Keywords: Landslide Susceptibility, Certainty factor, Frequency ratio, Information value, GIS, Logistic regression

Posted Date: November 24th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-107923/v1

Application of an integrated model based on bivariate and multivariate method in landslide susceptibility mapping

Han Hu[ab], **Changming Wang**[a*], **Zhu Liang**[a]

College of Construction Engineering, Jilin University, Changchun 130026, People's Republic of China[a]

School of Geomatics and Prospecing Engineering, Jilin Jianzhu University, Changchun 130118, People's Republic of China[b]

Corresponding author[*]: Changming Wang

E-mail address: wangcm@jlu.edu.cn

Abstract: Landslides usually result in human losses and economic damages in mountainous areas especially for Himalayan areas. Landslide susceptibility mapping (LSM) is a key approach for avoiding hazard and risk. This study aims to explore an improved model combining multivariate and bivariate statistical methods for LSM. Four models were established as logistic regression (LR), LR integrated with certain factor (CF), LR integrated with frequency ratio (FR) and LR integrated with information value method (IV) and their performance was compared in LSM. Firstly, a landslide inventory map with 313 determined landslide events was prepared and 12 predisposing factors were selected. Secondly, the dataset was randomly divided into two parts, 75% of which was used for modeling and 25% for validation. Finally, area under the curve (AUC) and statistical metrics were applied to validate and compare the performance of the models. Results show that the performance of IVLR model is the best (AUC 0.792 and accuracy=78.8%). Besides, the LSM constructed by IVLR model did a reasonable job at predicting the distribution of

susceptible areas. It identified the major factors and intervals of high susceptibility that profile curvature greater than 0.1, less than 2 km from the stream, maximum elevation difference greater than 1200 m and rainfall between 440 and 450 mm were prone to landslide. The conclusion reveals that the quality of LSM can be improved by comparing and combining the bivariate and multivariate methods, which serve as a more effective guide for land use planning in the study area or other highlands where landslides are frequent.

Key words: Landslide Susceptibility; Certainty factor; Frequency ratio; Information value; GIS; Logistic regression

1. Introduction

Landslide is a sudden geological phenomenon widely distributed across the world, causing direct or indirect damages to property and injuries or fatalities of people residing in the area[1,2,3]. The frequency and scale of landslide outbreaks in China are far beyond than that of other countries in the world[4,5]. The prevention measures need to identify the existing landslides for spatial zonation[6]. Generally, damages can possibly be controlled by prediction where disasters may occur in the future[7]. Therefore, landslide susceptibility mapping (LSM) is considered as an effective approach to avoid hazards and risk.

The approaches for landslide susceptibility modeling can be broadly classified as qualitative (knowledge-driven methods or physically based methods) and quantitative (data-driven methods) [8,9]. Data-driven methods can be categorised as bivariate methods (like frequency ratio (FR), certainty factor (CF) and information value (IV)) and multivariate methods (like logistic regression (LR), factor analysis and cluster analysis)[10,11,12,13]. Qualitative methods are mostly

43 subjective and limited to apply in the small-scale areas. In recent years, geographic information
44 system (GIS) and computing techniques increasingly developed and statistical-based methods are
45 becoming popular. New machine learning methods like AdaBoost, extremely randomized tree and
46 support vector machine have also been noticed due to their ability of solving the problem of
47 non-linear geo-environmental issues without necessary assumptions compared to traditional
48 multivariate methods[14,15]. However, the relationship between landslide occurrences and landslide
49 related factors are determined by training rather than inference, which is also called "the black
50 box" operation and the methods lack credibility and interpretability in practice[16].

51 Bivariate methods are famous for its simplicity and applicability to explore the correlations
52 between landslides and predisposing factors, although multivariate methods generally perform
53 better in terms of accuracy[17,18,19]. The FR, CF and IV methods not only produce landslide
54 susceptibility maps but also serve to explore the vulnerabilities to landslide failure of individual
55 landslide predisposing factors by the corresponding values calculated for each interval. However,
56 the potential relationship between the various conditioning factors is complex having different
57 impacts on landslide susceptibility. Bivariate methods cannot determine the relative weights
58 between different factors. Therefore, it is necessary to optimize the bivariate methods.

59 LR model as one of the representatives of multivariate methods has been well verified in the
60 LSM[20,21]. It models between various dependent variables and multiple independent variables
61 which is applicable in LSM. However, the final mathematical expression cannot comprehensively
62 express the correlation between different intervals of various factors and landslides occurring.
63 Therefore, it is meaningful to combine the bivariate and multivariate statistical methods to better
64 analyze the conditioning factors and we aim to compare the performance of three ensembles

models (CFLR, FRLR and IVLR) for the most suitable one for LSM. The Louza County in Southeastern Tibet is selected as the study area where shallow landslides are frequent due to topographic and geological conditions. The value of AUC, statistical metrics and kappa coefficient were combined to assess the comparison of performance of these models.

2. Materials

2.1 Study area

Luoza county is situated in Shannan city, Southeastern Tibet. It ranges from 90°59'E to 91°15'E of longitude and 28°26'N to 29°56'N of latitudes **(Fig. 1)**. It has a population of more than 4500, covering an area of 870.9 km^2. The study area belongs to temperate semi-humid climate zone and the annual rainfall is 454 mm which concentrated mainly in July to September.

Topographically, it is a typical highland area and the highest elevation is 6696 while the lowest is3488 m and slope angle ranges from 0°to73°. During our field investigation, four lithologies were common: gray sandstone from Cretaceous Gucuo (K1g), pellet micrite from Jurassic (J_{2n}), limestone from Triassic (T_3) and Quaternary deposits. The landslides in the area belong to rain-induced landslide, which are a threat to local villagers and economies commonly (**Fig. 2** and **Fig. 3**).

2.2 Data preparation

2.2.1 Landslide inventory

An important assumption in the statistically-based methods for LSM is that landslides have

more chance to occur again under the conditions which led to the landslides before or at present [22]. Therefore, a complete and accurate landslide inventory which shows the locations of landslides is essential[7]. Inventory data comes from remote sensing image **(Fig. 4)**, historical records and extensive field survey. All landslides are bounded by polygons containing the whole landslide perimeter and 313 landslide polygons were identified. Model training and testing was based on commonly applied 1:1 sampling strategy[23]. Therefore, the total number of samples is 626, including 313 landslide locations and 313 non-landslides samples. Non-landslide samples were optionally selected far from the landslide areas.

2.2.2 Conditioning factors

Depending on the characteristics of the study region, the availability, reliability and practicality of the data, 12 conditioning factors were selected: annual rainfall (F1), maximum elevation difference (F2), altitude (F3), plan curvature (F4), profile curvature (F5), slope angle (F6), topographic wetness index (F7), distance to roads (F8), distance to faults (F9), distance to streams (F10), slope aspect (F11) and lithology (F12)[24].

Rainfall is the unique triggering factor considered and it has been applied for many times. The thematic map was generated by ordinary kriging interpolation in ArcGIS and 12 precipitation stations near the study area provided related information. The thematic was reclassified into 4 classes **(Fig. 5a)**.

Maximum elevation difference reflects potential kinetic energy of a slope unit, which was calculated in ArcGIS 10.2. The thematic map was reclassified into 6 classes by 300 m intervals **(Fig. 5b)**. Altitude has influence on both rainfall and vegetation[25,26] and elevation in the area was

105 divided into four subclasses by 500 m intervals **(Fig. 5c)**. Curvatures are essential to the geometry
106 of slopes, which provides valuable information about erosion and deposition[27] . Both the plan and
107 profile curvatures were reclassified into six classes **(Fig. 5d and 5e)**. TWI is another
108 morphometric parameter that represents basic terrain[28], which was divided into six categories **(Fig.**
109 **5g)**. Slope angle is another considerable factor, which controls shear strength on potential slide
110 surface and the subsurface flow[29] and was reclassified into five classes by 10° intervals **(Fig. 5f)**.
111 Six topographical related factors were extracted from the digital elevation model (DEM) with the
112 spatial resolution of 30 m.

113 Faults act as potential weak planes in slopes which could reduce bulk-rock strength and
114 distance to fault were constructed with six classes for <2000, 2000~4000, 4000~6000, 6000~8000,
115 8000~10000 and >100000m **(Fig. 5i)**. Similarly, six subclasses of distance to road and distance to
116 river were created in an interval of 2000 m **(Fig. 5h and 5j)**.

117 Lithology map with four classes was constructed for Shale with siltstone, Quaternary
118 deposits, Shale with limestone and Pellet micrite **(Fig. 5k)**. It is an important factor in slope
119 stability and was also adopted by several researchers[30]. Slope aspect can initiate great impact on
120 microclimate and has previously been used by researchers[17,31]. It was attributed to eight subclasses
121 and are shown in **Fig. 5l**.

122 The lithology as well as the faults information were extracted using an existing 1:50,000
123 geological map. Roads and rivers network information were obtained from Landsat 8 LOI images.

124 Distance to fault, distance to road and distance to stream were calculated with the Euclidean
125 Distance ArcGIS Tool that measures the distance in meters from each raster unit of the area to the
126 closest vector segment.

2.2.3 Choice of mapping units

Three types of mapping units are commonly applied in LSM: grid cells, slope units and unique-condition units[32,33]. The choice of mapping unit is controversial while grid cells are the most popular[34]. Slope units are better in reflecting the geomorphological and geological condition of a landslide, which comprises the source, transport and accumulation areas of a landslide. Finally, the study area is divided into 2060 slope units with the hydrologic analysis tool in ArcGIS and necessary artificial corrections are accompanied according to remote sensing image.

3. Methods

3.1 LR model

LR model establishes the nonlinear probabilistic function of binary dependent variables [35,36]. The data type is not limited and nominal, continuous or a combination of both are feasible[37], the equation for which can be shown as follow:

$$\mathrm{p} = \frac{1}{1+e^{-y}} \tag{1}$$

where p represents the probability of an event ranging from 0 to 1; y represents a linear fitting function as showed below:

$$\mathrm{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_nx_n \tag{2}$$

Where b_0 is the intercept of the model, $b_1, b_2, ...,b_n$ are the partial regression coefficients and $x_1, x_2, ..., x_n$ are the variables.

LR was modeled in SPSS software and forward stepwise method was applied for exclude the non-significant variables. The values of 12 evaluation factors of all units were extracted as

147 independent variables while dependent variables represent the occurrence of landslide event i.e. 1
148 represents occurrence and 0 represents non-occurrences. The significant values of all variables
149 retained in the last step of the analysis were less than 0.05 and no variables were added.

150 3.2 CF method

151 CF is a bivariate statistical method which is commonly used in analyzing the probabilistic
152 relationship between the dependent and independent variables. Accordingly, the classification of
153 landslide-related factors should be determined first with continuous factor values.

154 CF method is a probability function firstly introduced by Shortliffe and Buchanan[10] in 1975
155 and then Heckerman[38] improved it. The CF method assumes the same as other statistical methods
156 that the conditions for future geological disasters are the same as those for the past. The method
157 for the calculation of the CF is shown as follow:

158
$$C_F = \begin{cases} \dfrac{PP_a - PP_s}{PP_a(1-PP_s)}, PP_a \geq PP_s \\ \dfrac{PP_a - PP_s}{PP_s(1-PP_a)}, PP_a < PP_s \end{cases} \quad (3)$$

159 Where PP_a represents the ration of the area of landslide of the a-th conditioning factor in a specific
160 interval to the area of a-th factor; PP_s represents the ratio of the total number (or area) of landslide
161 to the total study area.

162 The value of CF ranges from -1~1 on the basis of equation 3. A positive CF value indicates
163 that the occurrence of landslide is highly certain and the geological environmental conditions are
164 prone to geological disasters. The higher the value, the higher will be the certainty. On the contrary,
165 negative values represent low certainty of landslide.

166 CFLR model was established in SPSS, which considers the CF value of the units as

independent variable and the occurrence of landslide as the dependent variable.

3.3 IV method

The IV method uses the frequency or density of landslides to reflect the magnitude of the hazards of different influencing factors and their sub-intervals. It was first proposed by Yin and Yan[39] and the equation 4 shows the method for calculating the information values:

$$\mathrm{I}_{(A_{i-j})} = In\frac{n_{i-j}/s_{i-j}}{n/s} \tag{4}$$

Where i=1，2，3，…，n; j = 1，2，3，…，m; n_{i-j} represents the area of landslide of the i-th conditioning factor in j-th interval; s_{i-j} represents the area of the i-th conditioning factor in j-th interval; n represents the total area of landslide and s represents the total area.

The values from IV method can either be positive or negative. If it is positive, it indicates that the factor is conducive to the occurrence of landslide in a specific interval: the greater the IV, the higher the possibility of landslide, and vice versa.

Similarly, the IVLR model was established in SPSS, taking the IV value of the units as independent variable and the occurrence of landslide as the dependent variable.

3.4 FR method

The equation allows calculating the FR values of a certain level of a conditioning factor[40]:

$$\mathrm{FR_i} = \frac{\frac{landslide\_cells_i}{landslide\_cells_{tot}}}{\frac{no\_landslide\_cells_i}{no\_landslide\_cells_{tot}}} \tag{5}$$

where i indicates the i-th class for each considered variable.

If the FR >1, a determined correlation exists between the landslide occurrence and factor

class and if the FR <1 than there will be a reflection of weak correlation. The IV, FR and CF method has produced the corresponding indexes of each class of the 12 control factors as shown in **Table 1**.

Finally, the FRLR model was established by taking the FR values of the units as independent variable and the occurrence of landslide as the dependent variable.

3.5 Models evaluation

Models will be unpersuasive without scientific validation. The elaboration of the predictive capability requires a splitting of the available data into training and test data. 75% of the datasets were randomly chosen for modeling training and 25% for testing.

Three statistical metrics as accuracy, sensitivity and specificity are applied to evaluate the performance of the models.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

where True Positives (TP) represents the number of landslide units predicted correctly as unstable, True Negatives (TN) represents the number of non-landslide units predicted correctly as stable, False Positives (FP) represents the number of non-landslide units predicted incorrectly as unstable and False Negatives (FN) represents the number of landslide units predicted incorrectly stable. The area under the receiver operating characteristic curve (AUROC) is another indicator to assess the evaluation effect of LSM.

In this study, both AUC and three statistical evaluation measures were combined to evaluate and validate the models established by LR, CFLR, IVLR and FRLR.

4. Results and verification

4.1 Performance and comparison of different models

Z-score normalization method was applied to standardize the data and to eliminate the impact of different dimensions (units) before modeling. Besides, a correlation analysis was conducted to test collinearity among the independent variables. The variance inflation factor (VIF) is a common applied index[41]. If VIF is greater than 5 or 10, it indicates that there is severe collinearity between the selected variables.

Table 2 showes the VIF values of the chosen independent variables and indicates that no multicollinearity exists among the chosen variables. SPSS also provides the test indexes that reflect the overall goodness of fitting of the model: -2LL, CSR^2 and NR^2. The Cox and Snell R square values and the Nagelkerke R square value indicated that the independent variables can explain the dependent variables, having values of 57.7% and 50.3%, 63.3% and 74.4%, 66.2% and 68.2%, 67.2% and 69.5% for LR, CFLR, FRLR and IVLR models respectively (**Table 4**).

The IVLR model achieves the highest value of sensitivity (81.6%), followed by the FRLR model (sensitivity=80.9%), the CFLR model (sensitivity=76.8%) and LR model (sensitivity =74.2%) as shown in **Table 5**. The FRLR model showed the best performance for specificity with the value of 84.2%, followed by IVLR model (83.5%), CFLR model (81.5%) and LR model (80.7%). The IVLR model performed the best in terms of accuracy and ROC with the values of 82.6% and 0.829 (**Table 6**) . However, LR model performed the worst with the values of 77.3% and 0.781. The FRLR model also performed well with the value of 82.4% and 0.820, followed by CFLR model (78.7% and 0.800).

Verification dataset is more important and valuable to evaluate the generalization ability of a model. It was found that IVLR model performed the best with the highest values of sensitivity, specificity, accuracy and AUC as 76.8%, 81.6%, 78.8% and 0.792 respectively (**Fig.6**). The FRLR model also performed well with the values of 77.5%, 78.5%, 78% and 0.766 respectively, which are close to the values obtained from IVLR model. The LR model followed by CFLR model remained the worst with the values of 73.3%, 75.9%, 74.6%, 0.72 and 74.1%, 80.7%, 76.9%, 0.757 respectively (**Table 7 and Table 8**).

The performance of the models was declining in verification especially for the CFLR model which indicated that the model was over-fitting and generalization ability was doubtful. It was noticed that hybrid models were better than the single model in terms of prediction capacity. However, there was a certain gap between the three hybrid models. The improvement of CFLR model was not obvious compared to LR model. The performance of FRLR model and IVLR were close and better than CFLR model.

4.2 Evaluation of influencing factors

4.2.1 Application of bivariate methods

The IV method was applied to ensure the relationship between the influencing factors and the occurrence of the landslides and the results are shown in **Table 1**. As for rainfall, the percentages of landslide area for 440~450 mm and >450 mm were 52.05% and 38.81% respectively, which means that 90% of landslide areas were distributed among the two class of rainfall and the IV value of these two classes were 0.44 and 0.36 while the FR and CF values were 1.55 and 1.43, 0.8 and 0.67 respectively. For maximum elevation difference, the IV values ranged

from -1.13~0.59 in which the IV and CF values of 600~900 m, 900~1200 m and >1200 m were all positive as 0.39, 0.54 and 0.59, 0.73, 0.95 and 1 respectively. For elevation, the maximum IV, FR and CF values were found for the class <4500 m as 0.52, 1.75 and 0.92 respectively, accounting for 52.61% of the landslide area. Besides, the minimum value of IV, FR and CF were found for the class >5500m as -3.1, 0.29 and -0.98respectively. For plan curvature, the percentages of landslide areas of the classes of -0.1~0 and 0~0.1 were 34.09 % and 30.66 %, while the IV, FR and CF values were -0.38 and 0.44, 1.56 and 0.81 respectively. For profile curvature, the percentages of landslide areas of the classes of 0.1~0.2 was 74.79 % with the IV, FR and CF values of 0.17, 1.18 and 0.35. Besides, the IV values of the classes of 0.1~0.2, 0.2~0.3 and >0.3 were the same as 0.59 while 1.8 for FR and 1.18 for CF. For slope angle, the class <10 and 10~20 had a negative IV and CF values of -3.89 and -1.27, -0.99 and -0.85 respectively. While the FR values of these two classes were less than 1 as 0.02 and 0.28. For the class 20~30, 30~40 and >40, the IV, FR and CF values were 0.14, 1.08 and 0.95, 1.15, 2.94 and 2.59, 0.29, 0.9 and 0.98 respectively and are prone to landslide. It can be seen that the TWI class of 0.1~0.2 and 0.2~0.3 were also prone to landslide.

For distance to road, the class <2000 m showed the highest IV, FR and CF values of 0.46, 1.59 and 0.83 with a total 47.4 % of landslide areas. Similarly, landslides were prone to the class <2000 m of the fault, and <2000 m of the stream. As for the slope aspect, the North face achieved the highest IV, FR and CF values of 0.57, 1.77 and 0.98 respectively. As for lithology, landslide was mainly disturbed in Quaternary deposits and Shale with siltstone, accounting for 65.73% and 26.21% of landslide area respectively. The IV, FR and CF values of Shale with siltstone were 0.37, 1.45 and 0.7 showing obvious contribution towards landslide susceptibility.

Three bivariate methods were applied to explore the relationship between the occurrence of

landslide and the conditioning factors in this study, the results of which were basically consistent in predicting the significant and important intervals of each factor.

4.2.2 Application of integrated models

The coefficients of generalized linear models reflect the relative importance of variables. The logistic regression equation is as follow:

$$logit(p) = \ln\left(\frac{p}{1-p}\right) = 0.219 + 0.519 \times F1 + 0.95 \times F2 + 0.458 \times F3 + 1.307 \times F5 + 0.466 \times F9$$

$$logit(p) = \ln\left(\frac{p}{1-p}\right) = -3.592 + 0.649 \times F1 + 1.39 \times F2 - 0.574 \times F3 + 1.701 \times F5 + 0.637 \times F9$$

$$logit(p) = \ln\left(\frac{p}{1-p}\right) = 0.897 + 0.591 \times F1 + 0.942 \times F2 + 1.477 \times F5 + 0.856 \times F10 \quad (7)$$

Where F1 represents the bivariate value of each classification interval in rainfall; F2 represents the bivariate value of each classification interval maximum elevation difference; F5 represents the bivariate value of each classification interval in profile curvature; F3 represents the bivariate value of each classification interval in elevation; F9 represents the bivariate value of each classification interval in distance to fault; F10 represents the IV value of each classification interval in distance to stream.

The integrated models finally selected the factors that have significant influence on the model fitting by rainfall, elevation, maximum elevation difference, profile curvature, distance to fault and distance to stream (**Table 3**). As for CFLR and FRLR models, both of them involved the factors F1, F2, F3, F5 and F9 and the coefficients of F5 reached the highest as 1.307 and 1.701 respectively. Besides, the coefficient of elevation was positive (0.458) for CFLR model and negative (-0.574) for FRLR model. A negative coefficient means that it is not conducive in the

occurrence of landslides and was not appropriate for analysis in this study.

As for IVLR model, the coefficients of the selected factors were all positive and the profile curvature reached 1.477 as the highest, which illustrate that it had the highest influence on the occurrence of landslide. The coefficient of rainfall was the smallest as 0.591, followed by distance to stream as 0.856 and maximum elevation difference as 0.942. According to the regression coefficient of the selected factors and the IV values of each classification interval, landslides are most likely to occur in areas with profile curvature greater than 0.1 within 2 km from the stream, maximum elevation difference greater than 1200 m and rainfall between 440 and 450 mm.

Three integrated models were established in this study to explore the relative importance of conditioning factors, the results of which were obviously different. On the other hand, aspect was not involved in the integrated models while appeared in LR model (**Table 3**).

4.3 Landslide susceptibility mapping

The aforementioned analyses indicated that the IVLR model shows a prominent fitting and generalization capability in predicting the landslide susceptibility compared to the other 3 models presented in this study. Therefore, it is determined as the most suitable model and applied to the calculate the landslide susceptibility index for the whole study area.

The probability P of the occurrence of landslide in the whole study area was determined based on the four models (LR, CFLR, FRLR and IVLR). The equal spacing principle is used to reclassify the landslide susceptibility index into five levels: very low (0~0.2), low (0.2~0.4), moderate (0.4~0.6), high (0.6~0.8) and very high (0.8~1) .

Fig. 7 show the distribution of landslide susceptible classes and the area percentage of each

311 class of each map is summarized in **Fig. 8**. As for LR model, very low, low, moderate, high or
312 very high susceptible class occupied 27.32 %, 20.85 %, 12.55 %, 21.74 % and 17.53 % of the
313 study area respectively (**Fig. 8**). In case of CFLR model, the corresponding area percentages were
314 18.93 %, 23.27 %, 17.43 %, 24.27 % and 16.10 % respectively. Similarly, five reclassified classes
315 of FRLR model accounted for 16.79 %, 21.08 %, 21.69 %, 24.64 % and 15.8 % respectively of the
316 entire area. LSM constructed by IVLR model was also divided into five classes as very low
317 (19.3 %), low (23.8 %), moderate (15.2 %), high (26.8 %) and very high (14.9 %). Therefore, the
318 overall differences were obvious among the four models.

319 A regular landslide susceptibility map should meet two rules: **(1)** the determined landslide
320 locations should appear in the high or very high-susceptibility class area as much as possible and
321 **(2)** the very high-susceptibility class area should occupy only a small proportion (Bui et al., 2012).
322 It was noticed that the landslide samples were mainly located in the dark (purple or red) areas and
323 the non-landslide points in the light (green or yellow) areas for IVLR model. Besides, LSM had
324 the smallest percentage of very high susceptible class compared to the others. The very-high
325 susceptibility areas of landslide are mainly distributed around the Yarlung Zangbo river and its
326 tributaries in the study area. River network curves and shapes geomorphology scour eroded slopes
327 in a great extent[42]. The areas near stream were densely populated with human activities and the
328 occurrence of landslide threats lives and property.

329 The performance of FRLR was also excellent in terms of prediction capability. However,
330 the percentage of moderate susceptible area was the largest among the models as 21.69%. The
331 predicted units as the moderate class were impalpable. Besides, the percentage of low or very low
332 susceptible areas was combinedly smallest as 37.87 % which was contrary to previous research.

Therefore, the LSM constructed by IVLR model was more analytical and receivable.

5. Discussion

Ensemble algorithms as bagging, stacking and boosting have been applied in LSM and the accuracy was exceeded up to 85% or 90% in previous studies[43,44]. New machine learning methods and deep learning emphasize the optimization, and accordingly the multiple parameters involved need to be tuned before application, which is difficult to implement especially for the non-professionals[45].Traditional statistical methods establish mathematical equations to explore the relationship between landslide-related factors and landslide occurrence, which are more acceptable. In this study, the IVLR model also performed well with satisfactory prediction capability (AUC = 0.792 and accuracy = 78.8%). Three integrated models performed better than the normal LR model in terms of accuracy, which indicated that the combination was effective. Previous researchers have applied bivariate and multivariate statistical methods and compared their performance in LSM[46,47]. Although CF, FR and IV have similarity in both principles and results but their performance varies when combined with LR model. Each method has its own strengths and weaknesses and generally its performance varies with different study areas[9]. Some researches indicated that the bivariate methods perform better than multivariate methods, while others support the multivariate methods[17,48]. However, it is believed that the integrated models have more accurate results than the result of an individual classifier, which has slightly better generalization ability than that of random guessing[13]. Therefore, it is recommended to compare various models for the selection of most suitable one on the basis of robustness and reasonability.

Accuracy is of major consideration for LSM but it should not be the only focus. Identifying

the major conditioning factors responsible for landslide occurrence is also important which helps in further engineering guidance. The determination of subjective weight and objective weight helps to distinguish the contribution of these factors and analytic hierarchy process (AHP) and factor analysis (FA) are the two commonly used methods without prior conditions[49,50]. CF, FR and IV are three commonly used bivariate statistical methods which directly reveals the correlation between landslide locations and the influencing factors[51,52]. Besides, factors at different interval range have different susceptibility to landslide, and bivariate methods can distinguish the difference. LR describes the contribution of the conditioning factors through modeling, the coefficients of which reflect the relative importance of different factors and the sign of the coefficients indicate the positive or negative effect of the factors on landslide[53]. Although there is no special requirement for data distribution, the LR model needs to convert nominal variables into dummy variables, which makes the regression model more complex. Therefore, the bivariate and multivariate methods are complementary up to some extent and it is worth combining them for a more reasonable and comprehensive analysis to provide a better way to analyze the major factors in details.

6. Conclusions

In the current study, four models based on bivariate and multivariate methods as LR, CFLR, FRLR and IVLR were explored and their performance is compared in LSM in Luoza county and the following conclusions can be drawn:

The IVLR model performed the best in terms of accuracy and the landslide susceptibility map constructed by IVLR model was reasonable and analytical. It indicated that landslides are

more likely to occur in areas with profile curvature greater than 0.1, within 2 km from the stream, maximum elevation difference greater than 1200 m and rainfall between 440 and 450 mm. The combination of bivariate and multivariate methods not only improves the prediction accuracy but analyze the major conditioning factors in details. It is desirable to improve the advancement of the application by combining multiple methods considering that some methods are complementary up in some ways. The conclusion of the current study is helpful for landslide risk mitigation in highlands and provides idea for non-professionals who fail to optimize the new machine learning methods.

However, there are also some limitations of the present study:

1. There are more possibilities for the combination of different models and further exploration is needed to improve prediction accuracy obviously;

2. Models need to be validated more reliably;

3. The selection of non-landslide locations should be more accurate to improve the purity of samples.

Reference

1.Das, I., Sahoo, S., Van Westen, C., Stein, A., Hack, R., 2010. Landslide susceptibility assessment using logistic regression and its comparison with a rock mass classification system, along a road section in the northern Himalayas (India). Geomorphology 114, 627–637.

2.Achour Y, BoumezbeurA, Hadji R, Chouabbi A, Cavaleiro V, Bendaoud EA (2017) Landslide susceptibility mapping using analytic hierarchy process and information value methods along a highway road section in Constantine, Algeria. Arab J Geosci 10 (194). https://doi.

org/10.1007/s12517-017-2980-6

3.Arabameri, A., Pradhan, B., Rezaei, K., Lee, S., Sohrabi, M. An ensemble model for landslide susceptibility mapping in a forested area. Geocarto International, 1-26, 2019.

4.Ni, H. Y., Zheng, W. M., Li, Z. L., Ba, R. J.: Recent catastrophic debris flows in Luding county, SW China: geological hazards, rainfall analysis and dynamic characteristics, Nat. Hazards, 55,523–542, 2016.

5.Liqiang Tong, Wensheng Qi, Guoying An,Chunling Liu. Remote sensing survey of major geological disasters in the Himalayas[J]. Journal of engineering geology, 2019, 27(03):496.

6.Rai PK, Mohan K, Kumar VK (2014) Landslide hazard and its mapping using remote sensing and GIS. J Sci Res BHU 58 ISSN 0447-9483

7.Liang Zhu, Wang Changming, Han Songling, Kaleem Ullah Jan Khan, and Liu Yiao. Classification and susceptibility assessment of debris flow based on a semi-quantitative method combination of the fuzzy C-means algorithm, factor analysis and efficacy coefficient.Nat. Hazards Earth Syst. Sci., 20, 1287–1304, 2020a https://doi.org/10.5194/nhess-20-1287-2020

8.Corominas J, van Westen C, Frattini P, Cascini L, Malet J-P, Fotopoulou S, Catani F, Van Den Eeckhaut M, Mavrouli O, Agliardi F, Pitilakis K, Winter MG, Pastor M, Ferlisi S, Tofani V, Hervás J, Smith JT (2014) Recommendations for the quantitative analysis of landslide risk. Bull Eng Geol Environ 73:209–263. doi:10.1007/s10064-013-0538-8

9.Reichenbach,P.,Rossi,M.,Malamud,B.D.,et al.,2018.A Review of Statistically-Based Landslide Susceptibility Models. Earth-Science Reviews, 180(5): 60-91. doi:https://doi.org/x 2018. 03. 001

10.Shortliffe E H，Buchanan B G. A model of inexact reasoning in medicine [J]. Mathematical Biosciences，1975，23:351 －379.

11.Korup O, Stolle A (2014) Landslide prediction from machine learning. Geol Today 30:26–33.doi:10.1111/gto.12034

12.Chang K-T, Merghadi A, Yunus AP, Pham BT, Dou J. Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques. Scientific reports. 2019;9(1):12296. doi:10.1038/s41598-019-48773-2

13.Liang Zhu , Wang Changming, Zhang Zhi-Min and Kaleem-Ullah-Jan Khan. A comparison of statistical and machine learning methods for debris flow susceptibility mapping. Stoch Environ Res Risk Assess (2020b) https://doi.org/10.1007/s00477-020-01851-8

14.Aditian A, Kubota T, Shinohara Y (2018) Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. Geomorphology 318: 101-111. https://doi.org/10.1016/j.geomorph.2018.06.006.

15.Liang Zhu, Wang Changming and Kaleem-Ullah-Jan Khan. Application and comparison of different ensemble learning machines combining with a novel sampling strategy for shallow landslide susceptibility mapping. Stoch Environ Res Risk Assess (2020c). https://doi.org/10.1007/s00477-020-01893-y

16.Merghadi, Abdelaziz, Ali P. Yunus, Jie Dou, Jim Whiteley, Binh ThaiPham, Dieu Tien Bui, Ram Avtar, and Boumezbeur Abderrahmane. 2020. Machine Learning Methods for Landslide Susceptibility Studies: A Comparative Overview of Algorithm Performance. Earth-Science Reviews 207 (August): N.PAG. doi:10.1016/j.earscirev.2020.103225.

17.Youssef AM (2015) Landslide susceptibility delineation in the Ar-Rayth Area, Jizan, Kingdom of Saudi Arabia, by using analytical hierarchy process, frequency ratio, and logistic regression models. Environ Earth Sci. doi:10.1007/s12665-014-4008-9, Article on line first

18.Li, Langping, Hengxing Lan, Changbao Guo, Yongshuang Zhang, Quanwen Li, and Yuming Wu. 2017. A Modified Frequency Ratio Method for Landslide Susceptibility Assessment. Landslides 14 (2): 727–41. doi:10.1007/s10346-016-0771-x.

19.Chen, Z., Liang, S., Ke, Y., Yang, Z., Zhao, H., 2019. Landslide susceptibility assessment using evidential belief function, certainty factor and frequency ratio model at Baxie River basin, NW China. Geocarto International 34, 348–367.

20.Gorsevski P V,Gessler P E,Foltz R B.Spatial prediction of landslide hazard using logistic regression and GIS [J].Transactions in GIS, 2000, 10 (3) : 395-415

21.Tian Chunshan, Liu Xilin, Wang Jia. Geohazard susceptibility assessment based on CF model and Logistic regression models in Guangdong[J].HYDROGE＆ ENGINEERING, 2016, 43（06）：154-161+170

22.Varnes, D.J., 1984. Landslide hazard zonation: a review of principles and practice. Commission on Landslides of the IAEG, UNESCONatural Hazards No. 3 (61 pp.).

23.Kornejady A, Ownegh M, Bahremand A. Landslide susceptibility assessment using maximum entropy model with two different data sampling methods[J]. Catena 2017, 152: 144-162.

24.Van Westen CJ, Castellanos E, Kuriakose SL (2008) Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview. Eng Geol 102(3–4):112–131

25.Hong, H., Pradhan, B., Xu, C., Bui, D.T., 2015. Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and

462 support vector machines. Catena 133, 266–281.
463 26.Maher Ibrahim Sameen,Biswajeet Pradhan,Dieu Tien Bui,Abdullah M. Alamri. Systematic
464 sample subdividing strategy for training landslide susceptibility models[J]. Catena, 2019.
465 27.Zevenbergen, L.W., Thorne, C.R., 1987. Quantitative analysis of land surface topography.
466 Earth Surf. Proc. land. 12, 47–56.
467 28.Evans, I.S., 1979. An Integrated System of Terrain Analysis and Slope Mapping. Final Report
468 on Grant DA-ERO-591-73-G0040. University of Durham, England. Freund, Y., Schapire,
469 R.E., 1997. A decision-theoretic generalization of on-line learning and an application to
470 boosting. J. Comput. Syst. Sci. 55, 119–139.
471 29.Magliulo, P., DiLisio, A., Russo, F., Zelano, A., 2008. Geomorphology and landslide
472 susceptibility assessment using GIS and bivariate statistics: a case study in southern Italy. Nat.
473 Hazards 47, 411–435.
474 30.Pourghasemi, H.R., Moradi, H.R., Aghda, S.F., 2013. Landslide susceptibility mapping by
475 binary logistic regression, analytical hierarchy process, and statistical index models and
476 assessment of their performances. Nat. Hazards 69, 749–779.
477 31.Pradhan B (2010) Landslide susceptibility mapping of a catchment area using frequency ratio,
478 fuzzy logic and multivariate logistic regression approaches. J Indian Soc Remote Sens
479 38(2):301–320
480 32.Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V., Reichenbach, P., 1991. GIS
481 techniques and statistical models in evaluating landslide hazard. Earth Surf. Process. Landf.
482 16 (5), 427–445. http://dx.doi.org/10.1002/esp.3290160505.
483 33.Carrara, A., Cardinali, M., Guzzetti, F., Reichenbach, P., 1995. GIS technology in mapping

484 landslide hazard. In: Carrara, A., Guzzetti, F. (Eds.), Geographical Information Systems in
485 Assessing Natural Hazards. Kluwer Academic Publisher, Dordrecht, The Netherlands, pp.
486 135–175.
487 34.Guzzetti, F., Galli, M., Reichenbach, P., Ardizzone, F., Cardinali, M., 2006a. Landslide hazard
488 assessment in the Collazzone area, Umbria, central Italy. Nat. Hazard. Earth Syst. Sci. 6,
489 115–131. http://dx.doi.org/10.5194/nhess-6-115-2006. Guzzetti, F., Reichenbach, P.,
490 Ardizzone, F., Cardinali, M., Galli, M., 2006
491 35.Hosmer DW, Lemeshow S (1989) Applied regression analysis. Wiley,New York, p 307
492 36.Atkinson P, Massari R (1998) Generalised linear modelling of susceptibility to landsliding in
493 the central Apennines, Italy. Comput Geosci 24(4):373–385
494 37.Lee, S., 2005. Application of logistic regression model and its validation for landslide
495 susceptibility mapping using GIS and remote sensing data. Int. J. Remote Sens.26 (7),
496 1477–1491.
497 38.Heckerman, D (1986) Probabilistic interpretations for mycin's certainty factors. Machine
498 Intelligence and Pattern Recognition Vol.4, 167-196.
499 39.Yin KL, Yan TZ (1988) Statistical prediction model for slope instability of metamorphosed
500 rocks.In:Bonnard, C(ed)Proceedings of the 5th international symposium on landslides,
501 Lausanne Balkema, Rotterdam, pp 1269–1272
502 40.Korup O, Stolle A (2014) Landslide prediction from machine learning. Geol Today 30:26–
503 33. doi:10.1111/gto.12034
504 41.Bui DT, Pradhan B, Lofman O, Revhaug I, Dick OB (2012) Landslide susceptibility
505 assessment in the Hoa Binh Province of Vietnam: a comparison of the Levenberg-Marquardt

and Bayesian regularized neural networks. Geomorphology.
doi:10.1016/j.geomorph.2012.04.023
42.Pourghasemi, H.R., Mohammady, M., Pradhan, B., 2012. Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. Catena 97, 71–84.
43.Chen Wei, Li Yang. GIS-based evaluation of landslide susceptibility using hybrid computational intelligence models. 2020, 195.
44.Dou, J., Yunus, A.P., Bui, D.T. et al. Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. Landslides 17, 641–658 (2020).
45.Pourghasemi, H. R., Kariminejad, N., Amiri, M., Edalat, M., Zarafshar, M., Blaschke, T., & Cerda, A. (2020). Assessing and mapping multi-hazard risk susceptibility using a machine learning technique. Scientific Reports, 10(1), 1–9. https://doi.org/10.1038/s41598-020-60191-3
46.Franny G. MURILLO-GARCíA, Stefan STEGER, Irasema ALCáNTARA-AYALA.Landslide susceptibility: a statistically-based assessment on a depositional pyroclastic ramp[J].Journal of Mountain Science, 2019 ,16(03):561-580. DOI: 10.1007/s11629-018-5225-6
47.Zhang Yi-xing,Lan Heng-xing, Li Lang-ping,Wu Yu-ming,Chen Jun-hui,Tian Nai-man. Optimizing the frequency ratio method for landslide susceptibility assessment: A case study of the Caiyuan Basin in the southeast mountainous area of China[J].Journal of Mountain Science,2020,17(02):340-357.
48.Chen T, Niu R, Jia X (2016) A comparison of information value and logistic regression models

528 in landslide susceptibility mapping by using GIS. Environmental Earth Sciences 75(10): 1-16.
529 DOI: 10.1007/s12665-016-5317-y
530 49.Mandal B, Mandal S. Analytical hierarchy process (AHP) based landslide susceptibility
531 mapping of Lish river basin of eastern Darjeeling Himalaya, India. Advances in Space
532 Research. 2018;62(11):3114-3132. doi:10.1016/j.asr.2018.08.008
533 50.Tang Yaming, Feng Fan , Guo Zizheng, Feng Wei, Li Zhengguo, Wang Jiayun, Sun Qiaoyin,
534 Ma Hongna, Li Yane. Integrating principal component analysis with statistically-based
535 models for analysis of causal factors and landslide susceptibility mapping: A comparative
536 study from the loess plateau area in Shanxi (China)[J]. Journal of Cleaner Production, 2020,
537 277.
538 51.Korup O, Stolle A (2014) Landslide prediction from machine learning. Geology Today 30(1):
539 26-33. http://doi.org/10.1111/gto.12034
540 52.Tay LT, Lateh H, Hossain MK, Kamil AA (2014) Landslide hazard mapping using a poisson
541 distribution: a case study in Penang Island, Malaysia. In Landslide Science for a Safer
542 Geoenvironment (pp.521–525). Springer International Publishing
543 53.Pourghasemi HR, Moradi HR, Aghda SMF. (2013) Landslide susceptibility mapping by binary
544 logistic regression, analytical hierarchy process, and statistical index models and assessment
545 of their performances. Natural hazards 69(1): 749-779. DOI: 10.1007/s11069-013-0728-5

546 Acknowledgements

547 This work was supported by the National Natural Science Foundation of China (Grant No.
548 **41972267**, **41977221**, and **41572257**).

Author contributions

Han Hu and Zhu Liang were responsible for modeling and writing manuscript. Changming Wang was responsible for the review.

Competing interests

The authors declare no competing interests.

Table 1 Spatial relationship between landslide conditioning factors and landslides

Conditioning factor	Zone	Landslide (%)	Non-landslide (%)	IV	FR	CF
Rainfall(mm)	<430	1.41	43.89	-2.66	0.77	-0.97
	430~440	7.72	32.90	-0.9	0.41	-0.77
	440~450	52.05	10.55	0.44	1.55	0.8
	>450	38.81	12.67	0.36	1.43	0.67
Maximum elevation difference (m)	<300	4.61	26.42	-1.13	0.32	-0.83
	300~600	33.41	60.41	-0.31	0.74	-0.45
	600~900	45.29	12.31	0.39	1.48	0.73
	900~1200	15.80	0.86	0.54	1.72	0.95
	>1200	0.90	0.00	0.59	1.8	1
Elevation(m)	<4500	52.61	4.46	0.52	1.75	0.92
	4500~5000	38.51	23.70	0.19	1.53	0.38
	5000~5500	8.60	57.98	-1.27	0.84	-0.85

	>5500	0.28	13.85	-3.1	0.29	-0.98
	<-0.1	7.35	1.64	0.42	1.53	0.78
	-0.1~0	34.09	69.61	-0.38	0.68	-0.51
	0~0.1	30.66	5.91	0.44	1.56	0.81
Plan curvature	0.1~0.2	5.08	0.06	0.58	1.78	0.99
	0.2~0.3	0.01	1.12	-4.54	0.01	-1
	>0.3	0.24	0.17	0.14	1.15	0.29
	<0	9.03	51.19	-1.12	0.33	-0.82
	0~0.1	74.79	48.81	0.17	1.18	0.35
Profile curvature	0.1~0.2	13.76	0.00	0.59	1.8	1
	0.2~0.3	1.33	0.00	0.59	1.8	1
	>0.3	1.08	0.00	0.59	1.8	1
	<10	0.09	9.44	-3.89	0.02	-0.99
	10~20	6.32	42.83	-1.27	0.28	-0.85
Slope angle (°)	20~30	62.50	44.52	0.14	1.15	0.29
	30~40	22.52	-10.95	1.08	2.94	0.9
	>40	1.91	-0.73	0.95	2.59	0.98
	<0	0.01	0.06	-0.82	0.50	-0.66
	0~0.1	0.14	0.25	-0.28	0.86	-0.25
TWI	0.1~0.2	0.39	0.26	0.16	1.33	0.49
	0.2~0.3	0.25	0.20	0.09	1.24	0.38

	0.3~0.4	0.10	0.14	-0.18	0.95	-0.09
	>0.4	0.11	0.09	0.08	1.23	0.37
	<2000	47.40	7.97	0.46	1.59	0.83
	2000~4000	25.17	11.87	0.27	1.31	0.53
Distance to	4000~6000	15.09	14.04	0.03	1.03	0.07
road (m)	6000~8000	5.83	14.40	-0.5	0.6	-0.6
	8000~10000	3.22	12.63	-0.83	0.44	-0.74
	>10000	3.19	39.21	-1.79	0.17	-0.92
	<2000	48.10	10.97	0.42	1.52	0.77
	2000~4000	28.91	13.73	0.27	1.3	0.52
Distance to	4000~6000	14.56	17.27	-0.08	0.92	-0.16
fault (m)	6000~8000	6.04	26.93	-0.93	0.39	-0.78
	8000~10000	1.40	15.98	-1.73	0.18	-0.91
	>10000	1.00	15.12	-1.99	0.14	-0.93
	<2000	65.88	36.35	0.22	1.25	0.45
	2000~4000	22.58	28.34	-0.11	0.9	-0.2
Distance to stream	4000~6000	7.52	17.96	-0.48	0.62	-0.58
(m)	6000~8000	2.11	9.31	-0.92	0.4	-0.77
	8000~10000	1.90	5.24	-0.58	0.56	-0.64
	>10000	0.30	2.43	-1.43	0.24	-0.88
Aspect	North (A1)	0.81	0.02	0.57	1.77	0.98

	East (A2)	19.64	12.34	0.18	1.2	0.37
	Northeast (A3)	2.57	16.29	0.33	1.39	0.64
	Southeast (A4)	38.84	42.57	-0.04	0.96	-0.09
	South (A5)	30.04	35.23	-0.07	0.93	-0.15
	Southwest (A6)	0.88	0.76	-0.06	0.95	-0.12
	West (A7)	7.21	8.15	0.06	1.06	0.13
Lithology	Quaternary deposits (L1)	65.73	67.72	-0.01	0.99	-0.02
	Shale with siltstone (L2)	26.21	7.96	0.37	1.45	0.7
	Pellet micrite (L3)	7.69	17.18	-0.44	0.64	-0.55
	Shale with limestone (L4)	0.37	7.13	-2.21	0.11	-0.95

555 **Table 2** Multicollinearity diagnosis indexes for variables

Variables	VIF
Rainfall	2.645
Maximum elevation difference	1.668
Elevation	4.346
Plan curvature	1.058
Profile curvature	0.03
Slope angle	1.952

TWI	1.444
Distance to road	2.554
Distance to fault	1.611
Distance to stream	1.756

556 **Table 3** Coefficients of the logistic regression models

Parameters/coefficients	LR	CFLR	FRLR	IVLR
Rainfall	0	0.519	0.649	0.591
Maximum elevation difference	0.861	0.95	1.39	0.942
Elevation	0.542	0.458	-0.574	0
Plan curvature	-3.19	0	0	1.477
Profile curvature	0.982	1.307	1.701	0
Slope angle	0	0	0	0
TWI	0	0	0	0
Distance to road	0	0	0	0
Distance to fault	-0.396	0.466	0.637	0
Distance to stream	0	0	0	0.856
Lithology	0	0	0	0
Aspect		0	0	0
A1	-23.616			
A2	0.152			

A3	1.158			
A4	0.493			
A5	0.551			
A6	-0.057			
Constant	-0.453	0.219	-3.592	0.897

Table 4 Statistics of the logistic regression models

	LR	CFLR	FRLR	IVLR
−2 Log likelihood	51.099	61.851	63.044	74.598
Cox & Snell R^2	0.577	0.633	0.662	0.672
Nagelkerke R^2	0.503	0.744	0.682	0.695

Table 5 Models' performance using training dataset

Indexes	LR	CFLR	FRLR	IVLR
TP（%）	83.4	82.6	84.6	81.0
TN（%）	71.3	74.9	80.2	84.2
FP（%）	16.6	17.4	15.4	19.0
FN（%）	28.7	25.1	19.8	15.8
Sensitivity（%）	74.2	76.8	80.9	81.6
Specificity（%）	80.7	81.5	84.2	83.5
Accuracy（%）	77.3	78.7	82.4	82.6

Table 6 ROC analysis of the three models using training data.

Models	AUC	Standard Error	95% Confidence Interval

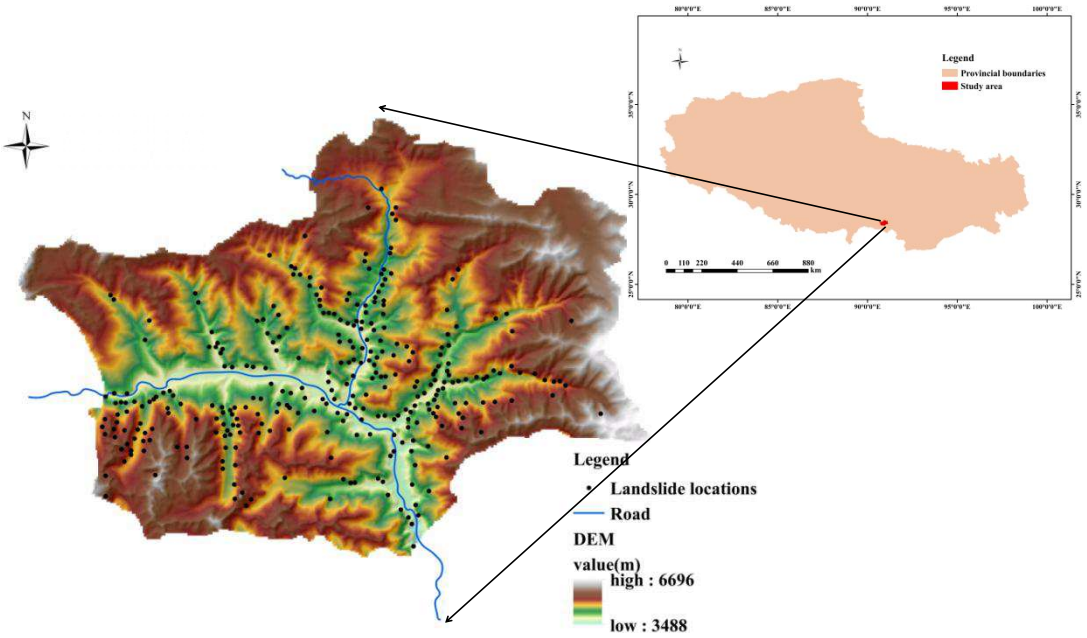
LR	0.781	0.021	0.736-0.882
CFLR	0.800	0.021	0.760-0.841
FRLR	0.820	0.020	0.781-0.860
IVLR	0.829	0.019	0.791-0.867

560 **Table 7** Models' performance using verification dataset

Indexes	LR	CFLR	FRLR	IVLR
TP（%）	76.9	83.1	79.4	82.5
TN（%）	72.5	71.0	76.8	75.4
FP（%）	23.1	16.9	20.6	17.5
FN（%）	27.5	29.0	23.2	24.6
Sensitivity（%）	73.3	74.1	77.5	76.8
Specificity（%）	75.9	80.7	78.5	81.6
Accuracy（%）	74.6	76.9	78.0	78.8

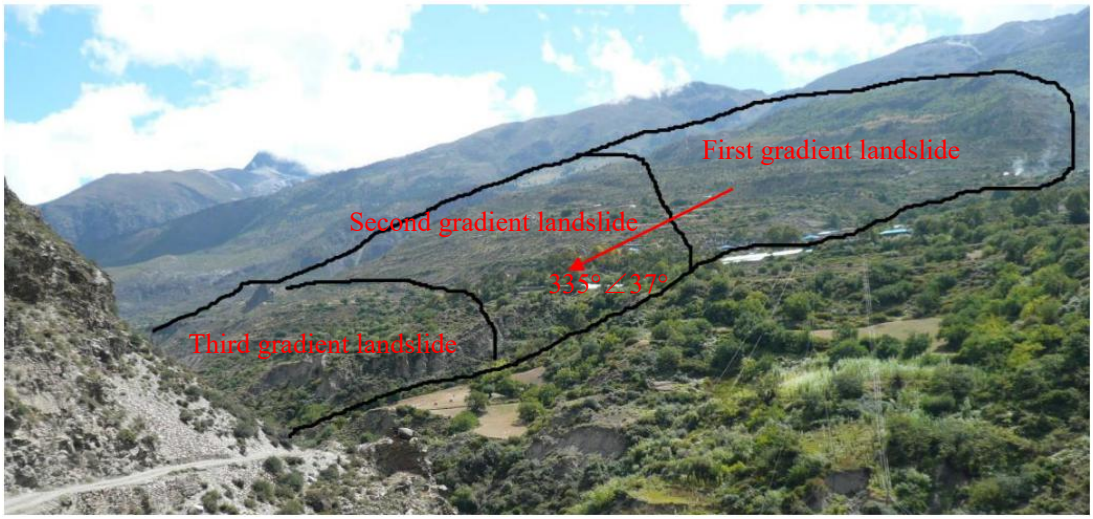
561 **Table 8** ROC analysis of the three models using verification data.

Models	AUC	Standard Error	95% Confidence Interval
LR	0.720	0.046	0.629-0.810
CFLR	0.757	0.043	0.672-0.842
FRLR	0.766	0.043	0.682-0.852
IVLR	0.792	0.041	0.712-0.872



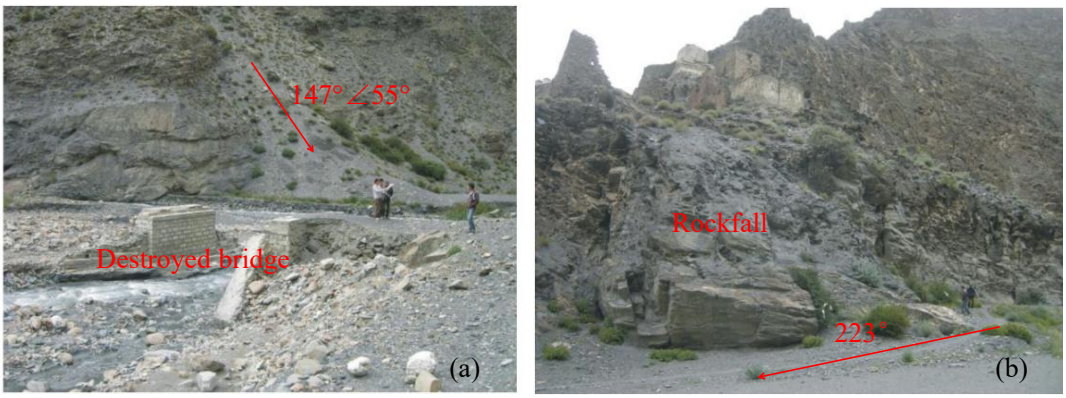
562

563 **Fig. 1.** Location map of the study area showing landslide inventory.



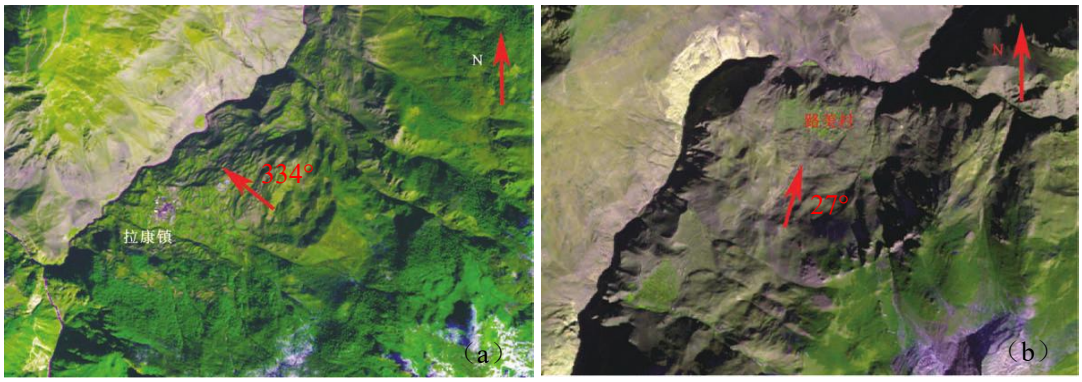
564

565 **Fig. 2.** Multistage landslide in Lakang county.



566

567 **Fig. 3.** Multistage landslide in Degacuo village.



568

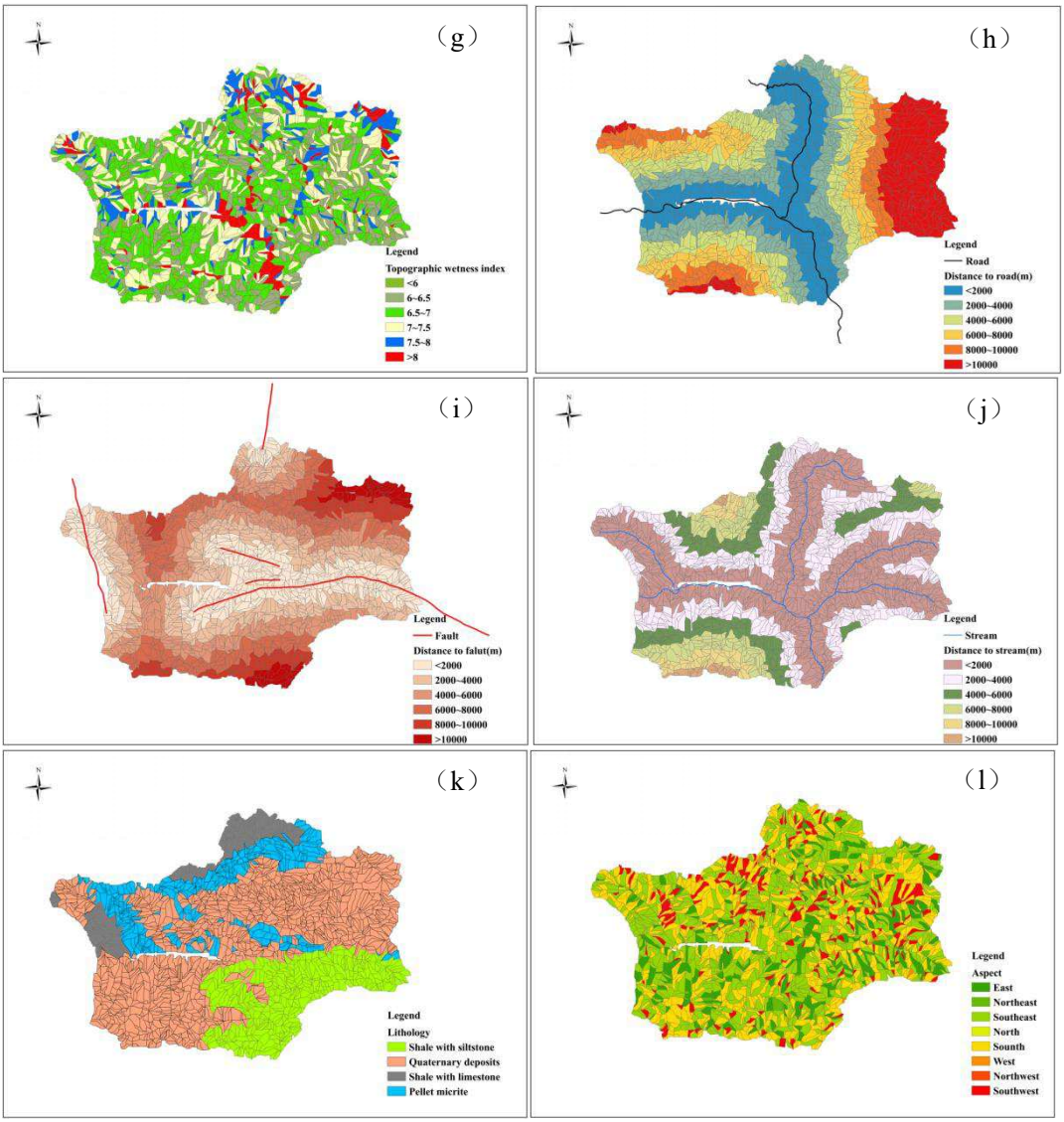
569 **Fig.4.** Stereo remote sensing map of landslides in Luoza county (Tong et al., 2019): **(a)** Landslide in

570 Lakang county; (b) Landslide in Lumei village.

571

572

573



574

575

576

577 **Fig.5.** Study area thematic maps: （**a**）Rainfall; （**b**）MED Slope angle DTR; （**c**）Altitude aspect;

578 （**d**）Plan curvature;（**e**）Profile curvature;（**f**）Slope;（**g**）TWI;（**h**）DTR;（**i**）DTF;（**j**）DTS;

579 （**k**）Lithology;（**l**）Slope aspect.

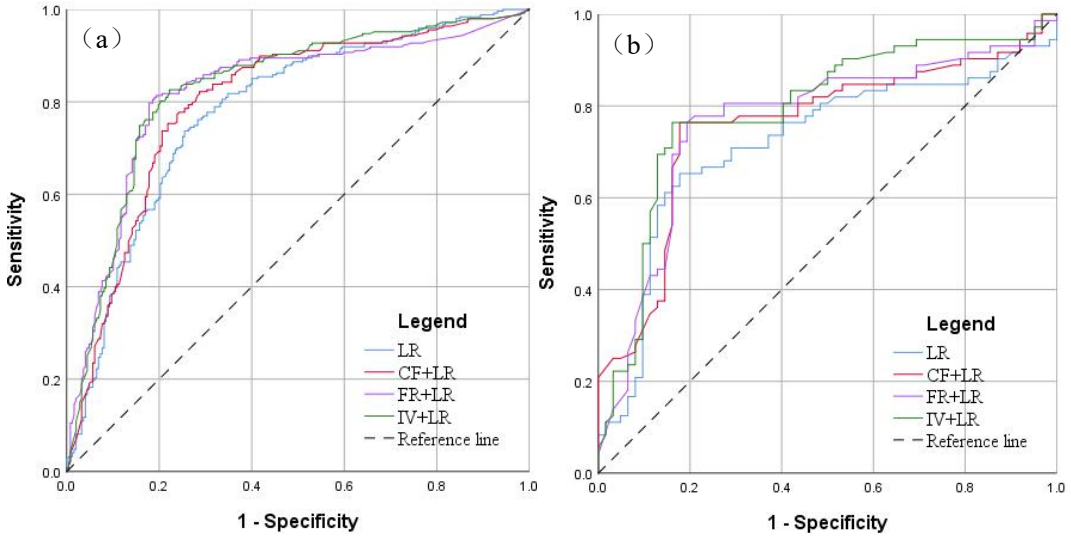


Fig.6. Analysis of ROC curve for the landslide susceptibility map: **(a)** Success rate curve of landslide using the training dataset; **(b)** Prediction rate curve of landslide using the validation dataset.

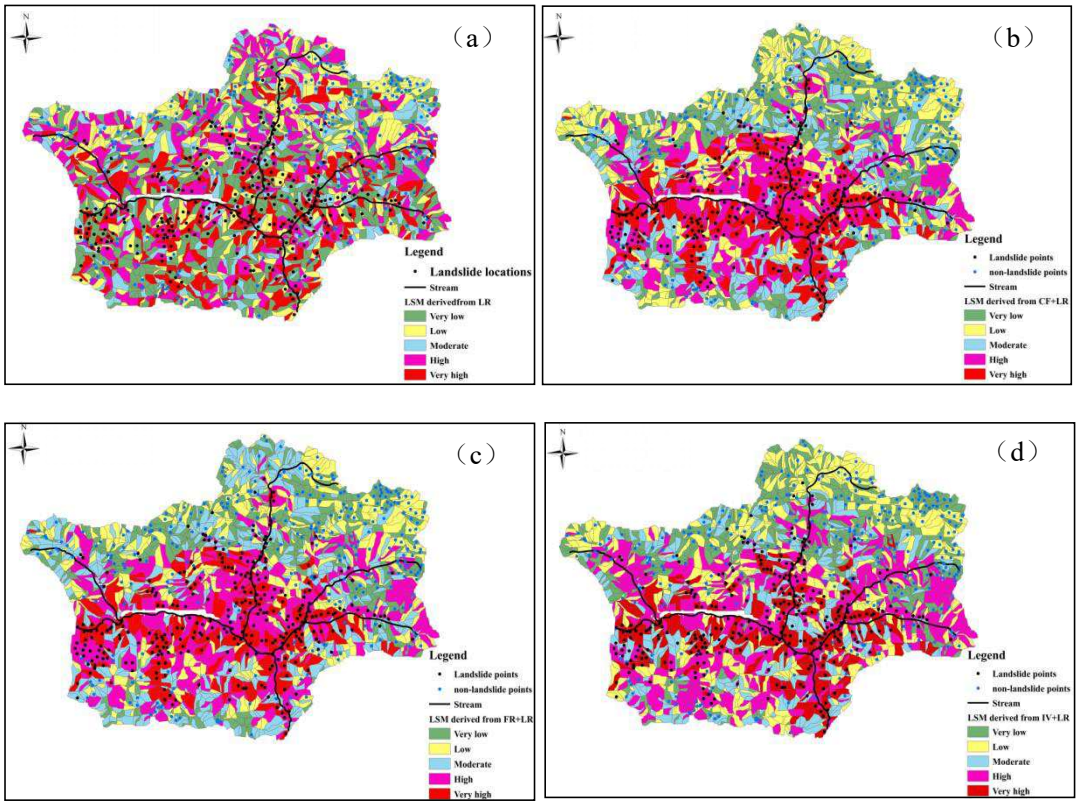


Fig. 7 Landslide susceptibility maps: **(a)** LR model; **(b)** CFLR model; **(c)** FRLR model; **(d)** IVLR model.

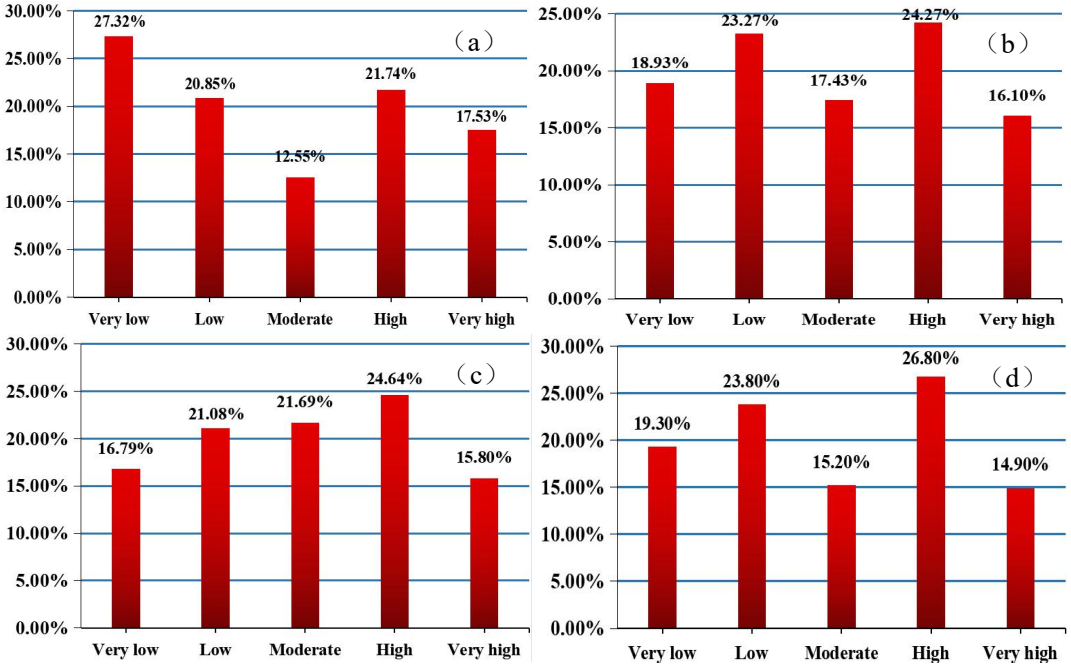


Fig.8. Percentages of areas in different susceptibility classes for landslide: **(a)** LR model; **(b)** CFLR model; **(c)** FRLR model; **(d)** IVLR model.

Figures

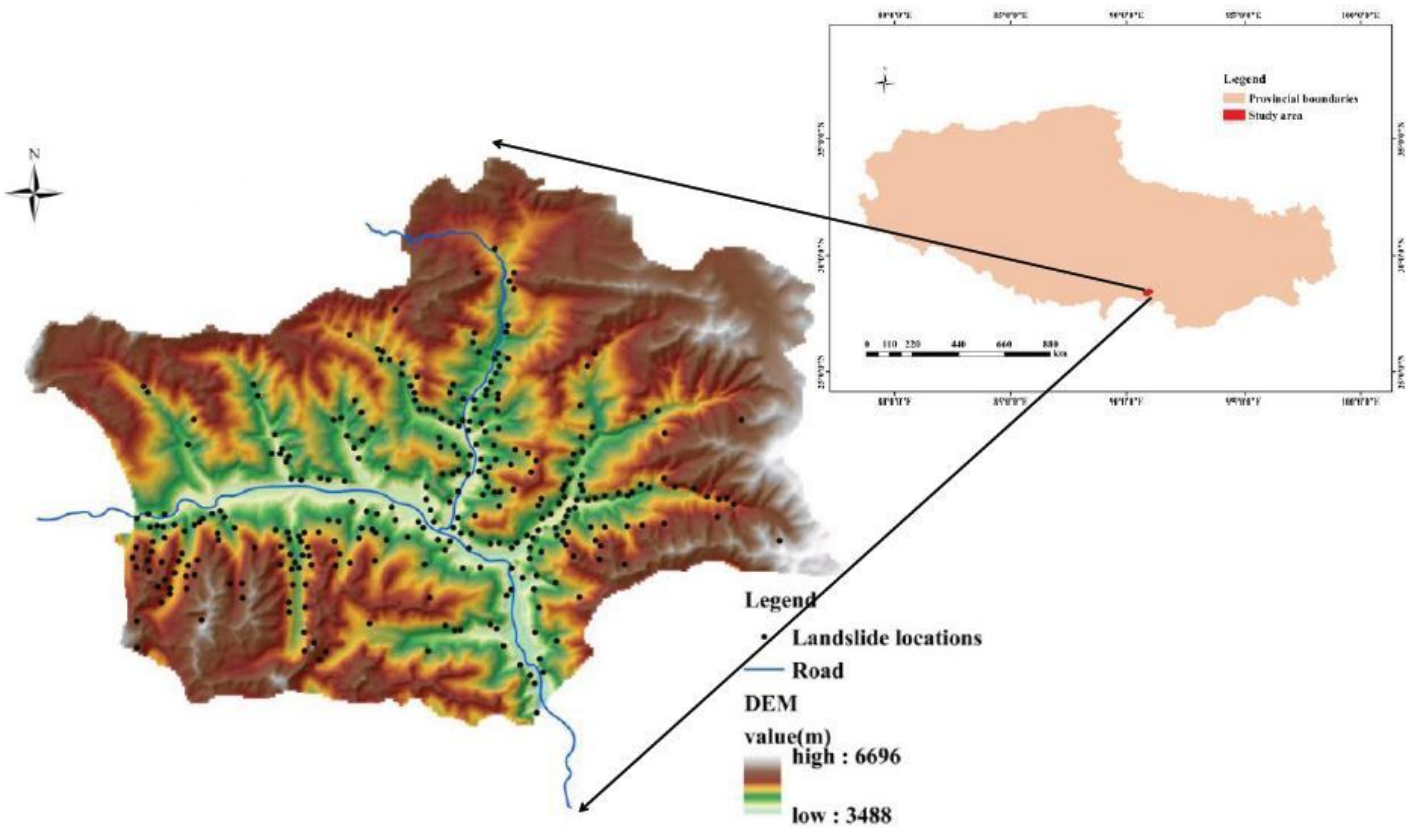


Figure 1

Location map of the study area showing landslide inventory. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

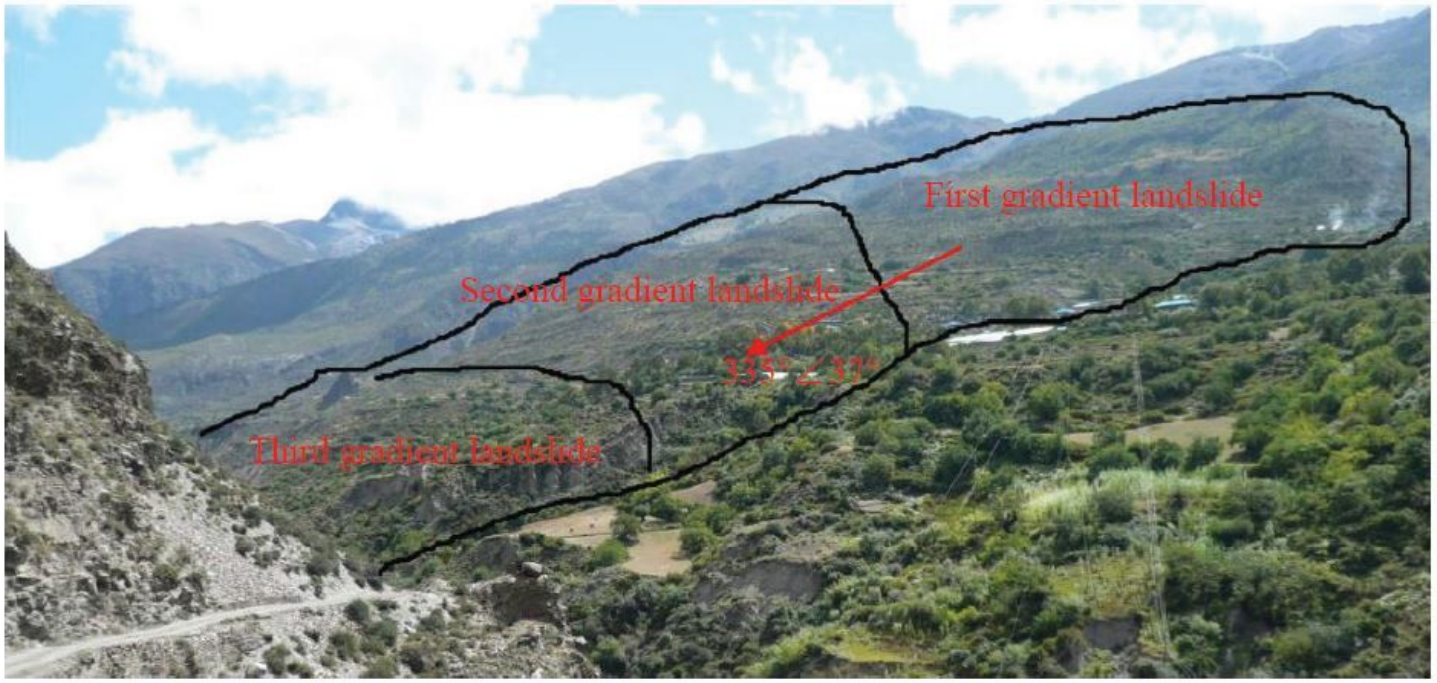


Figure 2

Multistage landslide in Lakang county. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

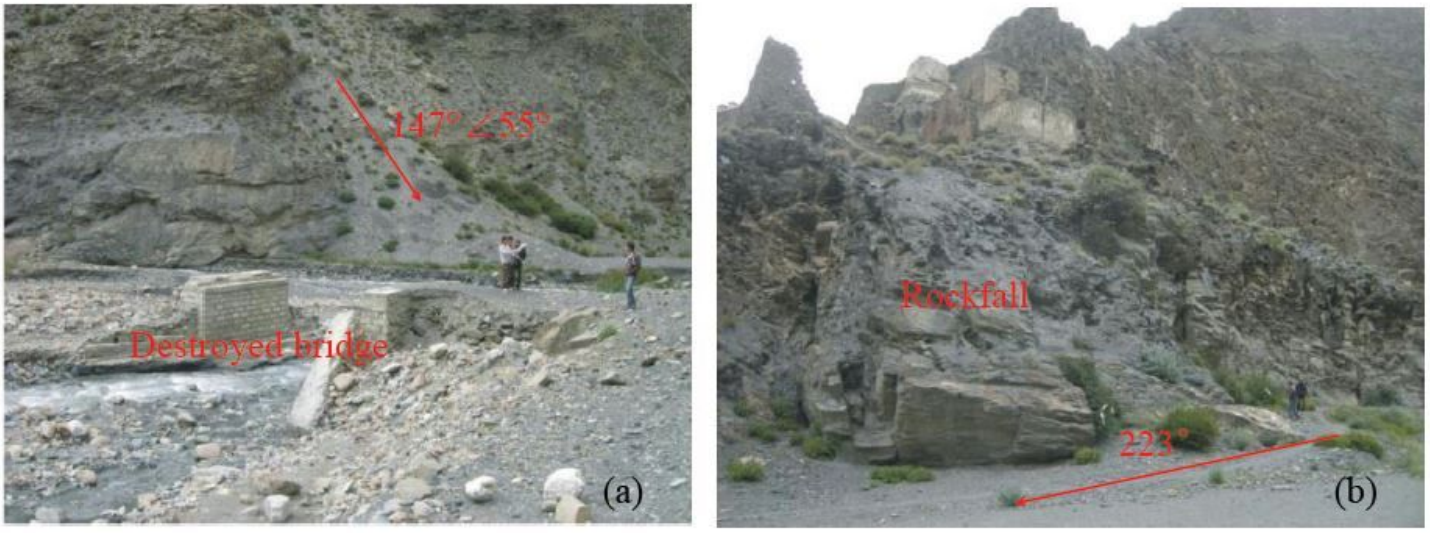


Figure 3

Multistage landslide in Degacuo village. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research

Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

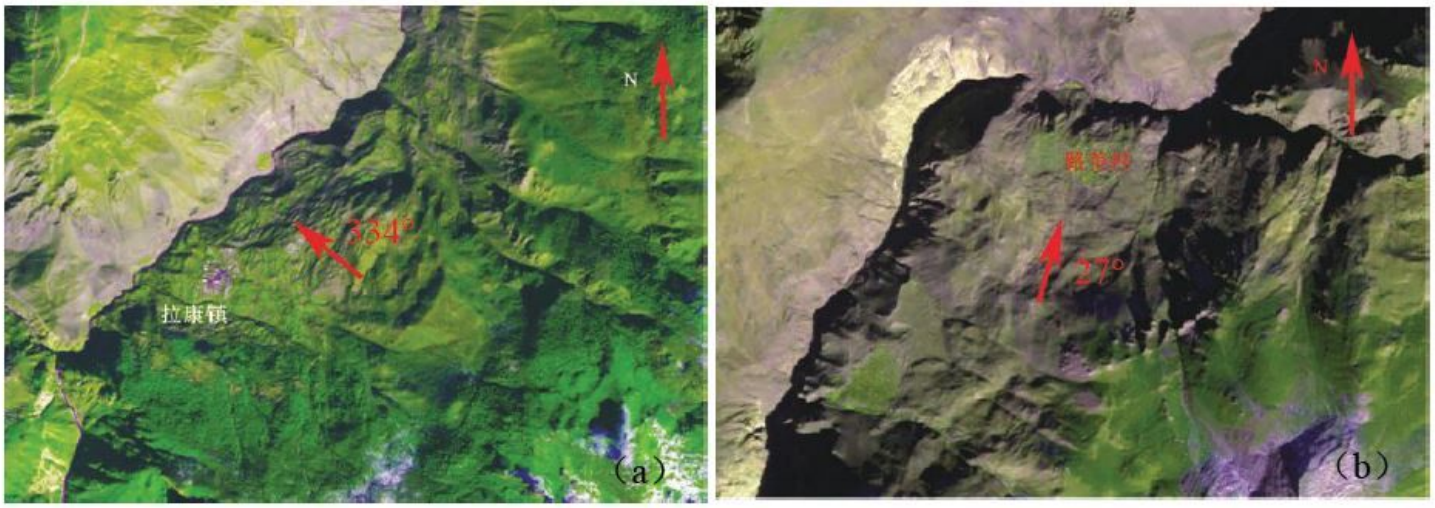


Figure 4

Stereo remote sensing map of landslides in Luoza county (Tong et al., 2019): (a) Landslide in Lakang county; (b) Landslide in Lumei village. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

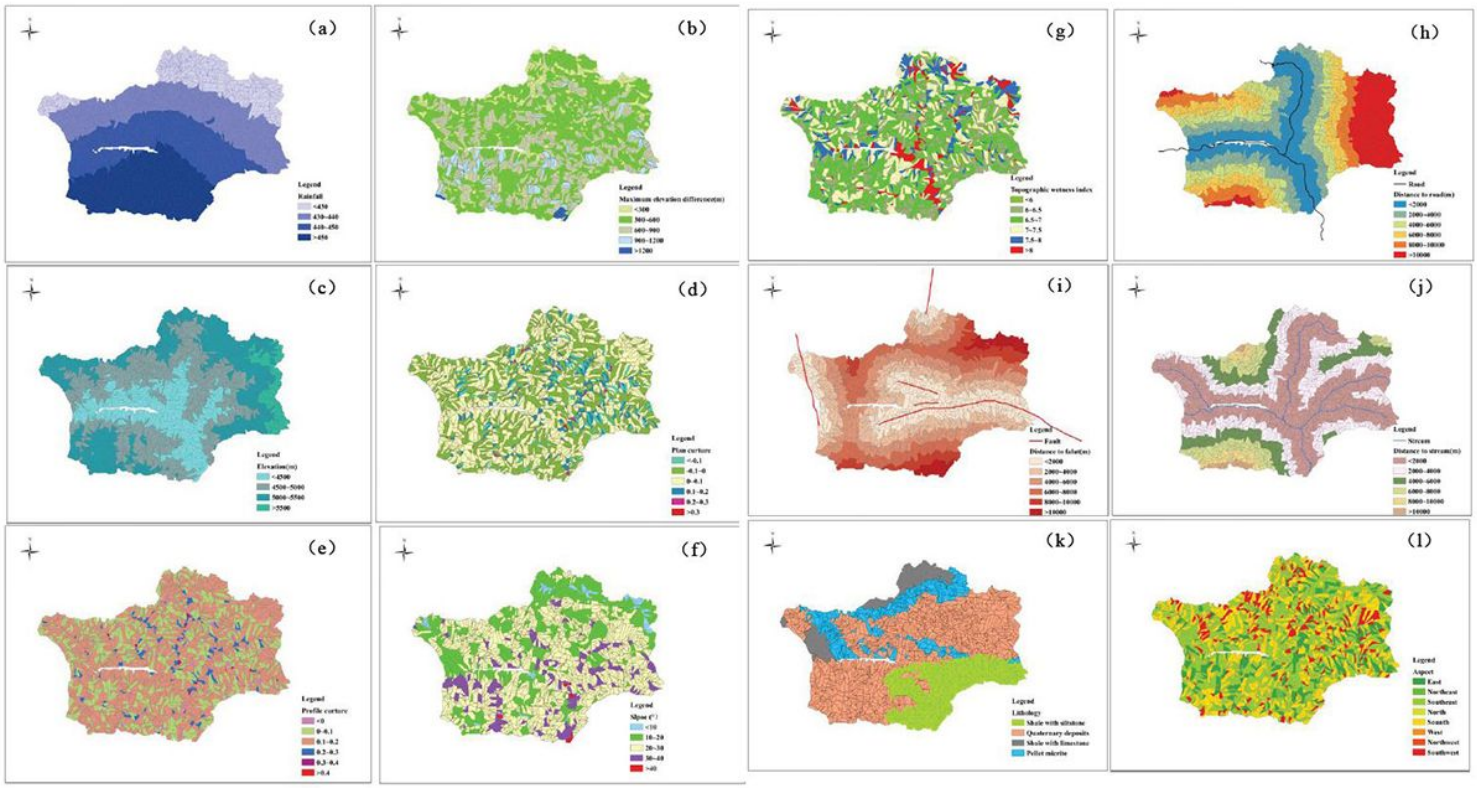


Figure 5

Study area thematic maps: (a) Rainfall; (b) MED Slope angle DTR; (c) Altitude aspect; (d) Plan curvature; (e) Profile curvature; (f) Slope; (g) TWI; (h) DTR; (i) DTF; (j) DTS; (k) Lithology; (l) Slope aspect. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

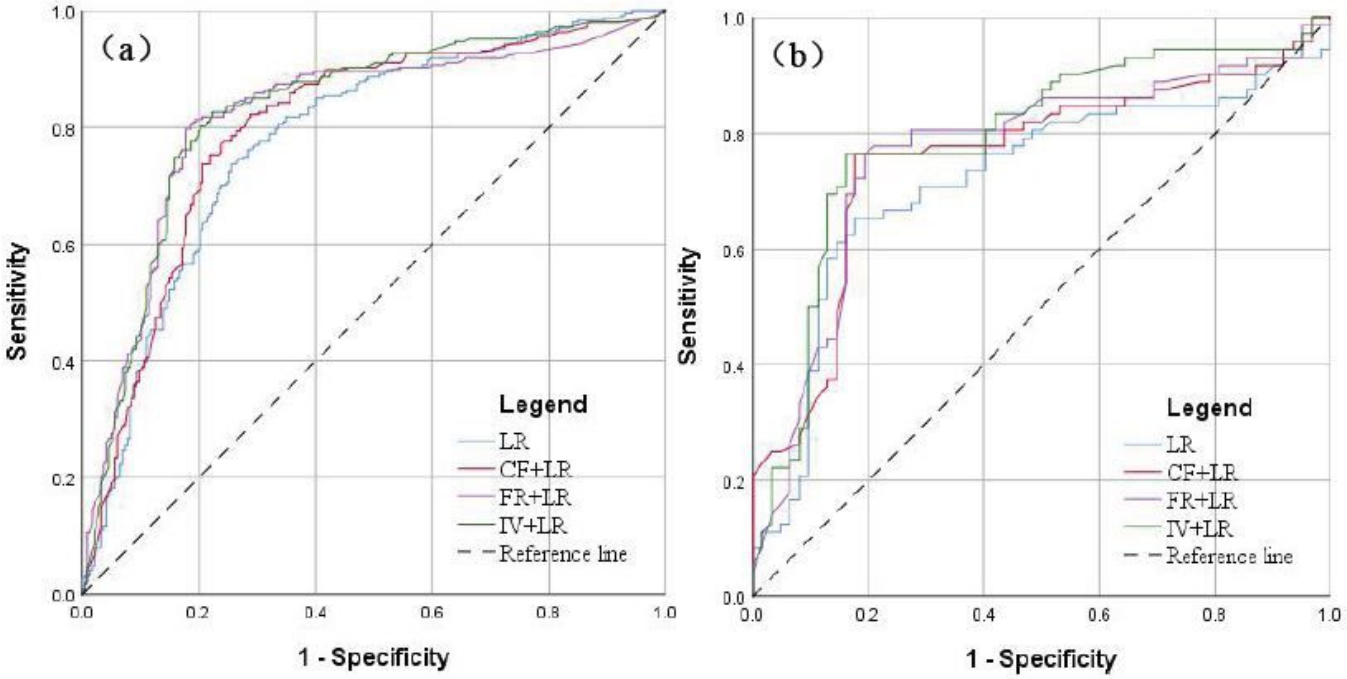


Figure 6

Analysis of ROC curve for the landslide susceptibility map: (581 a) Success rate curve of landslide using the training dataset; (b) Prediction rate curve of landslide using the validation dataset.

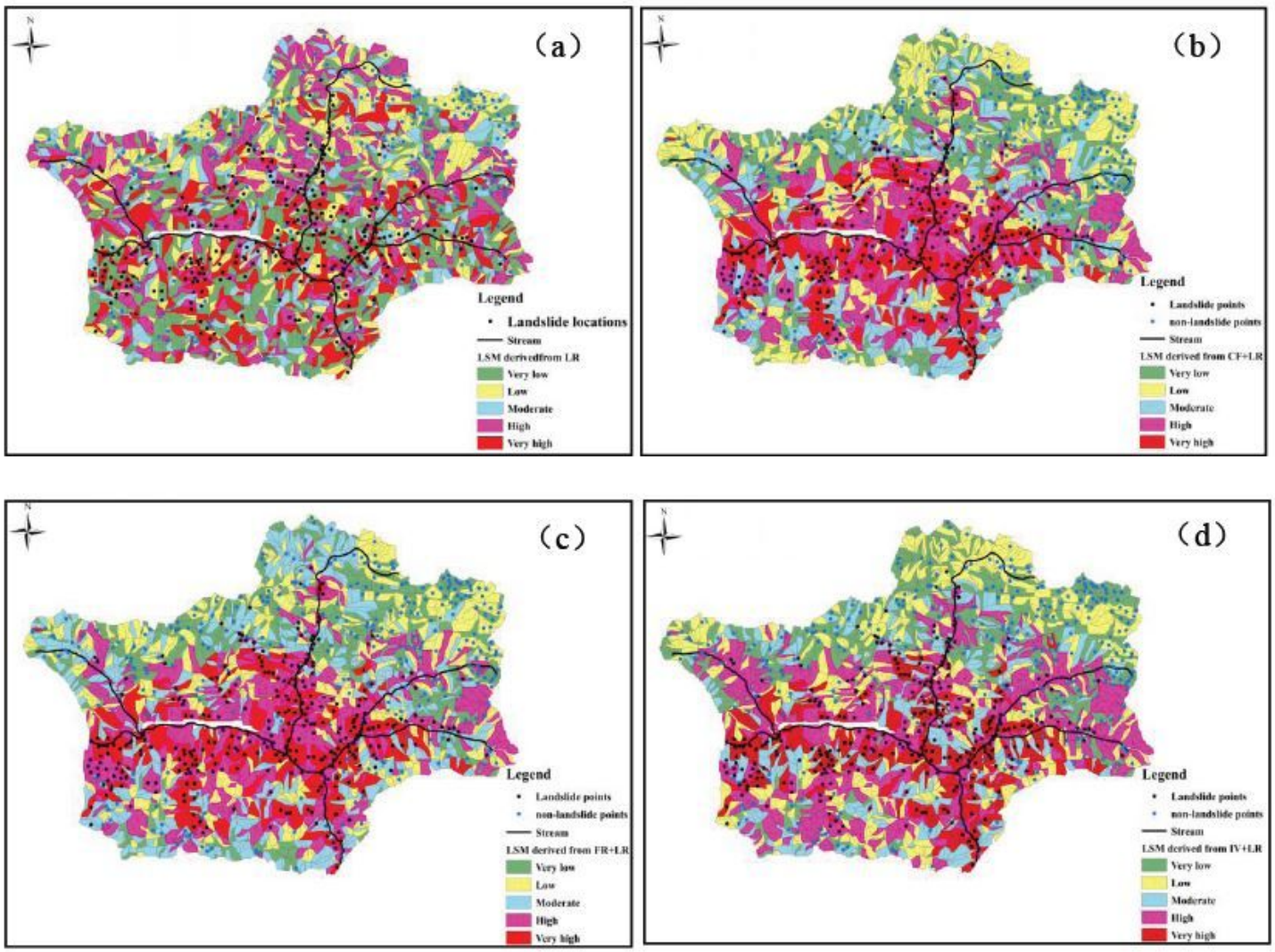


Figure 7

Landslide susceptibility maps: (a) LR model; (b) CFLR model; (c) FRLR model; (d) IVLR model. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

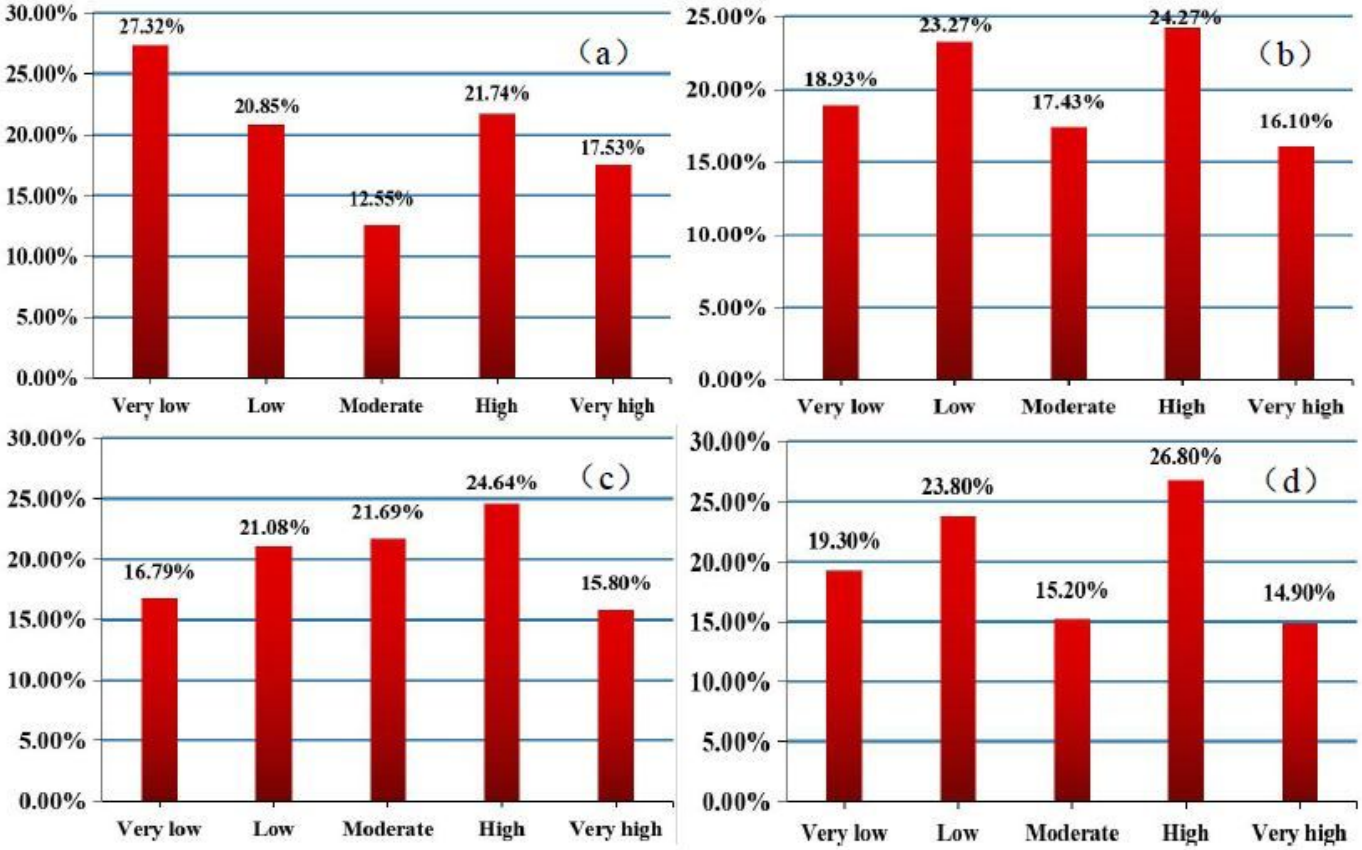


Figure 8

Percentages of areas in different susceptibility classes for landslide: (a) LR model; (b) CFLR model; (c) FRLR model; (d) IVLR model.