remote sensing

MDPI

Article

# Integration of Information Theory, K-Means Cluster Analysis and the Logistic Regression Model for Landslide Susceptibility Mapping in the Three Gorges Area, China

**Qian Wang [1], Yi Wang [1,*], Ruiqing Niu [1] and Ling Peng [2]**

[1] Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan 430074, China; cug_qian_wang@163.com (Q.W.); rqniu@163.com (R.N.)
[2] China Institute of Geo-Environment Monitoring, Beijing 100081, China; pengl@mail.cigem.gov.cn
* Correspondence: cug.yi.wang@gmail.com; Tel.: +86-27-6788-3251

Received: 23 July 2017; Accepted: 6 September 2017; Published: 11 September 2017

**Abstract:** In this work, an effective framework for landslide susceptibility mapping (LSM) is presented by integrating information theory, K-means cluster analysis and statistical models. In general, landslides are triggered by many causative factors at a local scale, and the impact of these factors is closely related to geographic locations and spatial neighborhoods. Based on these facts, the main idea of this research is to group a study area into several clusters to ensure that landslides in each cluster are affected by the same set of selected causative factors. Based on this idea, the proposed predictive method is constructed for accurate LSM at a regional scale by applying a statistical model to each cluster of the study area. Specifically, each causative factor is first classified by the natural breaks method with the optimal number of classes, which is determined by adopting Shannon's entropy index. Then, a certainty factor (CF) for each class of factors is estimated. The selection of the causative factors for each cluster is determined based on the CF values of each factor. Furthermore, the logistic regression model is used as an example of statistical models in each cluster using the selected causative factors for landslide prediction. Finally, a global landslide susceptibility map is obtained by combining the regional maps. Experimental results based on both qualitative and quantitative analysis indicated that the proposed framework can achieve more accurate landslide susceptibility maps when compared to some existing methods, e.g., the proposed framework can achieve an overall prediction accuracy of 91.76%, which is 7.63–11.5% higher than those existing methods. Therefore, the local scale LSM technique is very promising for further improvement of landslide prediction.

**Keywords:** landslide susceptibility; logistic regression; causative factors; K-means cluster; Three Gorges area

## 1. Introduction

As the water level in the reservoir fluctuates periodically, the famous Three Gorges Reservoir is characterized by plentiful active and reactivated landslides with different scales, which seriously threaten the local people's lives and property. Up to 2009, more than 3800 landslides have been recorded along this reservoir [1]. Therefore, it is very significant to perform landslide susceptibility mapping (LSM) to dynamically monitor the unstable areas.

The spatial forecasting of landslides is more efficient and economical by integrating geographical information systems (GIS) and statistical analysis, compared to the traditional field geological surveying, and can provide an effective solution for landslide mitigation and management [2].

This method has been widely documented in recent literature stating that remote sensing (RS) can be used for landslide investigation. According to the three most comprehensive reviews [3–5], landslide susceptibility and hazard assessment is one of the three hot topics in landslide investigation using RS. Over the last three decades, many effective methods have been developed to investigate the role of RS and GIS for producing landslide hazard zoning maps. These techniques are mainly divided into two categories with different theoretical bases, i.e., qualitative and quantitative [6]. The qualitative techniques are characterized by subjective assessments that describe the probability of landslide occurrences based on expert experience and knowledge of landslide formation mechanism(s) [7], including the analytical hierarchy process (AHP) [8–10], fuzzy mathematics [11], multi-criteria evaluation [12,13], weighted linear combination (WLC) [14,15] and ordered weighted average [16,17]. The quantitative techniques represent landslide occurrences by exploiting mathematical models to perform LSM on a continuous scale [14,18]. Landslides are typically complex processes triggered by various causative factors, which have geomorphological, geological, hydrological, terrestrial, meteorological or geotechnical properties. The quantitative methods are commonly divided into two types, bivariate and multivariate. To estimate the weights for each variable in the bivariate methods, each causative factor map is combined with a landslide distribution map. Various techniques can be used with the bivariate methods, such as favorability functions [19–21], information value [22,23], weights of evidence [24–29], the frequency ratio [30–32] and the Dempster–Shafer method [2,33]. However, failure to consider the correlation of the causative factors is the main shortcoming of such methods [34]. The multivariate methods assess the relationships between the landslide distribution and a series the causative factors [35]. Specifically, all of the causative factors are resampled for each terrain mapping unit (TMU), and the events of landslides are estimated through the resulting matrix, which can be analyzed with logistic regression (LR) [18,36–38] using multiple regression [39–43], discriminant analysis [44] or principle component analysis (PCA) techniques [45,46]. Apart from these statistical methods, data mining and machine learning techniques have drawn much attention for LSM including decision tree (DT) [47,48], random forests [49–51], neural networks [52–56], support vector machine (SVM) [57,58] and Bayesian network (BN) approaches [59]. Nevertheless, it is improper to adopt all of the causative factors for LSM because the problem of overfitting always occurs, and the model generalization is not well respected, without considering the issue of data dimensionality [60]. Therefore, screening the factors through feature selection using filtering methods [11,61] or wrapper methods [62–64] is a common step for producing more accurate landslide susceptibility maps. However, the mentioned dimensionality reduction techniques are burdened with a high computational cost. Furthermore, the same set of selected factors is exploited throughout the entire study area, without taking the spatial dependence between TMUs into account.

Landslide are triggered by many causative factors at a local scale, and the impact of these factors is closely related to geographic locations and the nearest neighborhood. Recently, Das et al. [65] proposed to obtain landslide susceptibility maps using homogeneous susceptibility units (HSUs), which is an effective local-scale analysis method. In this work, we develop an alternative framework to solve the previously-mentioned issues by integrating the techniques of information theory, K-means cluster analysis and statistical models. The proposed framework consists of the following steps. First, each causative factor used in the study area is classified by the natural breaks method with the corresponding optimal number of classes, which is determined by using Shannon's entropy index. Then, a certainty factor (CF) for each class of factors is estimated. It is observed that the impact of each causative factor may occur at a local scale for a certain study area in practice [66]. To address this fact, each TMU in the study area is assigned with an appropriate combination of the causative factors represented by a unique binary encoding, according to expert experience and knowledge of the CFs. By performing the K-means cluster analysis on the new encoded TMUs, the spatial dependence between these units is considered. Therefore, the TMUs where the landslides are affected by a similar set of the causative factors are aggregated together. The final binary-encoded centroid of each cluster is employed for choosing the optimal combination of the causative factors shared by all of the TMUs

in the same cluster. Next, the weights of the selected factors for each cluster are computed, and the proposed predictive method is constructed for accurate LSM at a regional scale by applying an LR model to each cluster of the study area. On this basis, a global landslide susceptibility map is created by integrating the regional maps. The proposed framework was validated in the Zigui-Badong section of the Three Gorges area by using the LR method implemented by SPSS Clementine 12.0, which can effectively integrate remote sensing datasets with field surveying data. IBM SPSS Statistics 19.0 was used for the computation of the proposed framework, and ESRI ArcGIS 10.0 was used for producing the resultant maps.

## 2. Methodology

### *2.1. Study Area*

#### 2.1.1. General Characteristics and Geological Setting

The study area is located in the Zigui-Badong section of the middle and lower reaches of the Yangtze River and covers 446.32 km$^2$ in the southwest of Hubei Province. Its latitudes and longitudes lie between 30°54′59″N to 31°03′32″N and 110°18′44″E to 110°52′30″E, respectively, while its highest point reaches 2000 m above sea level, as shown in Figure 1. This study area belongs to the subtropical monsoon climate zone and is characterized by abundant rainfall and humidity. The average annual precipitation for the period from 2001–2010 in Badong County and Zigui County in Hubei Province is 1069.2 mm and 944.5 mm, respectively, while the highest annual precipitation reached 1148.7 mm in 2008. Furthermore, most of the rainfall in this area is concentrated from May–September of each year, accounting for 70% of the annual precipitation.

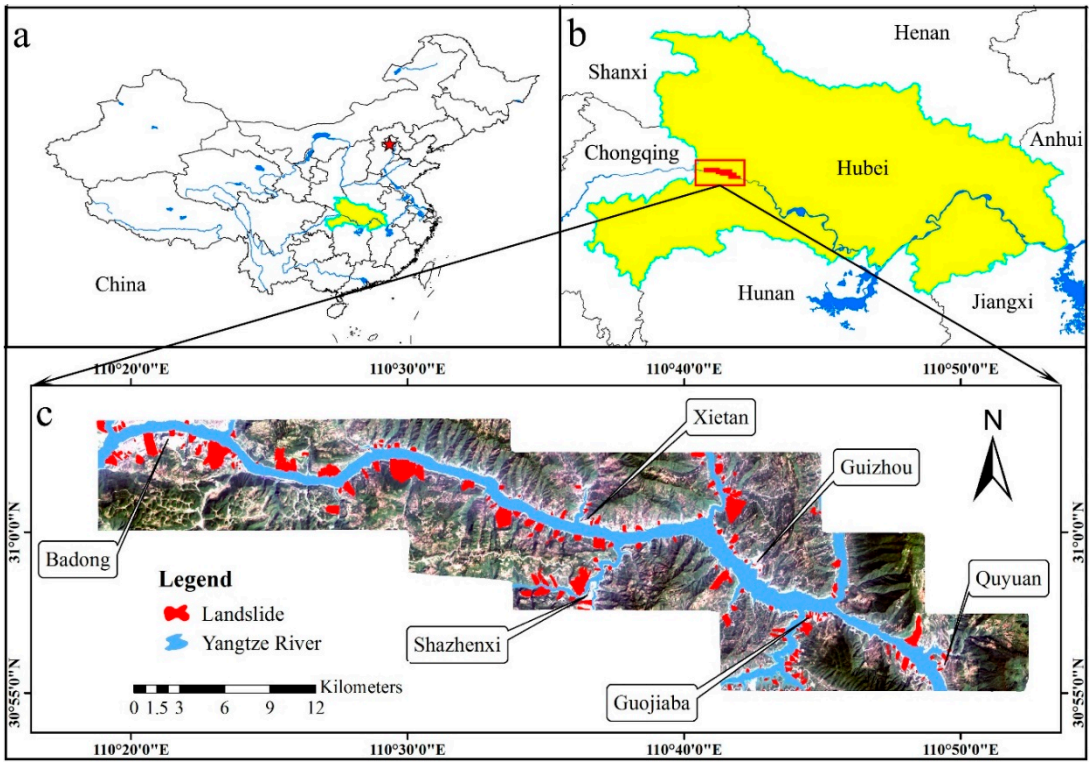



**Figure 1.** Location of the study area. Site maps of (**a**) China, (**b**) Hubei province and (**c**) the study area (a true color composite image using Bands 4, 3 and 2 of Landsat-8 OLI data).

All forms of rock, including igneous, sedimentary and metamorphic, can be observed in this area. The strata from the Middle Triassic to Jurassic are mostly composed of sandstone, shale, mudstone

and marlstone, which are the main components of natural hazards. The geological map of the study area is shown in Figure 2. Several main faults and lineaments can be identified in this figure, including the Xiannvshan Fault, the Jiuwanxi Fault, the Niukou Fault and the Xiangluping Fault, which may promote or mitigate landslides.

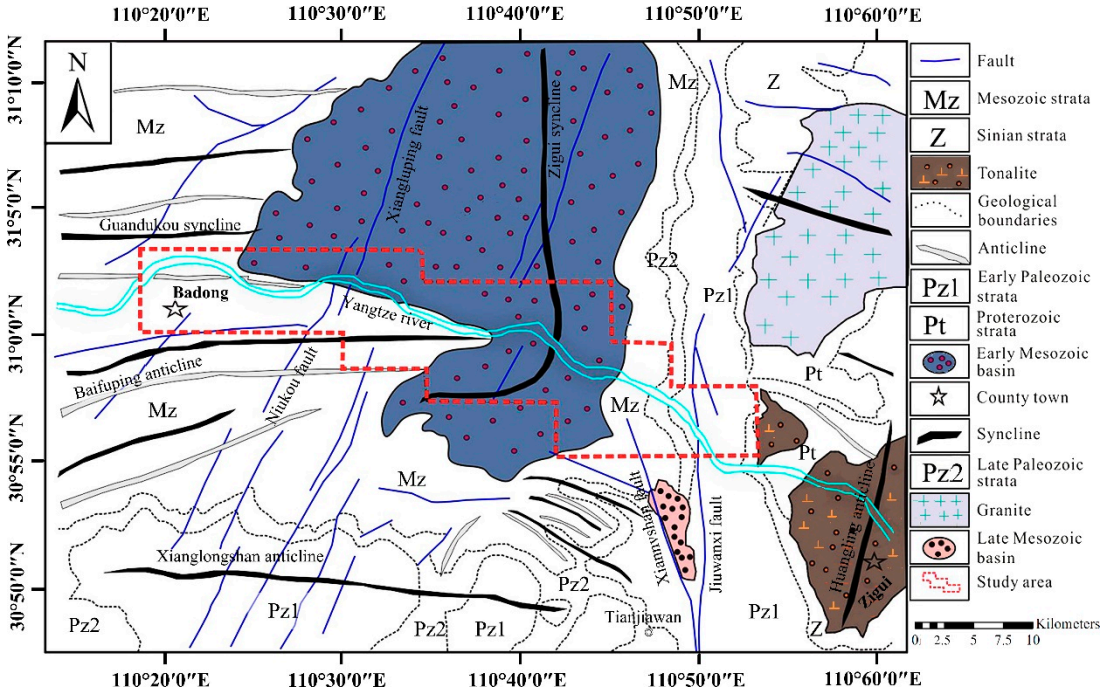



**Figure 2.** A geological map of the study area.

### 2.1.2. Slope Failures and Causative Factors

The geological conditions and human activities in the study area, such as urbanization, deforestation and construction of the reservoir, have caused widespread distribution of landslides, which have brought a serious threat to the lives and property of the local residents. The landslide inventory map of the study area was constructed by using Google Earth 7.1 along with extensive field surveys, historical and bibliographical landslide data. Next, 202 landslide polygons were identified and mapped with total areas of 23.40 km$^2$, covering 5.24% of the study area, as shown in Figure 1c. It can also be observed from Figure 1c that the area of these landslides varies widely, e.g., the largest Fanjiaping landslide has an area of 1.51 km$^2$, while the smallest Kuihua street landslide is only 2068.8 m$^2$.

In this work, the Yangtze River was excluded from the study area because the values of the Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model (ASTER GDEM) data always change dramatically at the junction between this river and its sides [67]. It is known that landslide occurrences are greatly relevant to causative factors. Considering the landslide distribution and the characteristics of the study area, a total of 17 causative factors was selected for the LSM, including the four main categories of geological, geomorphological, hydrological and land cover factors. To extract these causative factors for landslide prediction, some ancillary data were used, including:

√ A Landsat-8 OLI image obtained on 14 April 2015, with the path/row number of 125/38. To perform feature extraction, we have performed a series of operations on this multispectral image. This process includes radiometric correction to avoid radiometric errors or distortions over the whole image, geometric correction to avoid geometric distortion due to Earth's rotation and other imaging conditions from the image and atmospheric correction to remove the effects of

the atmosphere on the reflectance values of the image. Meanwhile, Bands 4 and 5 of the image are used for computing the normalized difference vegetable index (NDVI), whereas Bands 3 and 6 of the image are used for computing the normalized difference water index (NDWI).

√ The 1:50,000-scale geological maps provided by Hubei Geological Bureau for the exaction of geological factors, including lithology and distance to fault.

√ ASTER GDEM Version 2 (V2) data, representing the surface in raster format, for the extraction of geomorphological and hydrological factors, including elevation, distance to rivers, the terrain roughness index (TRI), the terrain position index (TPI), slope gradient, catchment area, catchment slope, terrain curvature, the topographic wetness index (TWI), terrain surface convexity, terrain surface texture, slope aspect and slope form.

The selection of the TMU is very significant for LSM. In this work, grid cell terrain units were exploited to model the landslide susceptibility of the study area, and a value was assigned to each grid cell unit per causative factor. The landslide map and other factor layers were extracted with grid cells having a spatial resolution of 28.5 × 28.5 m, to match the remote sensing data considered here.

### 2.2. The Proposed Framework

The flowchart of the proposed framework is shown in Figure 3. In the following subsections, we present the foundations of our framework.

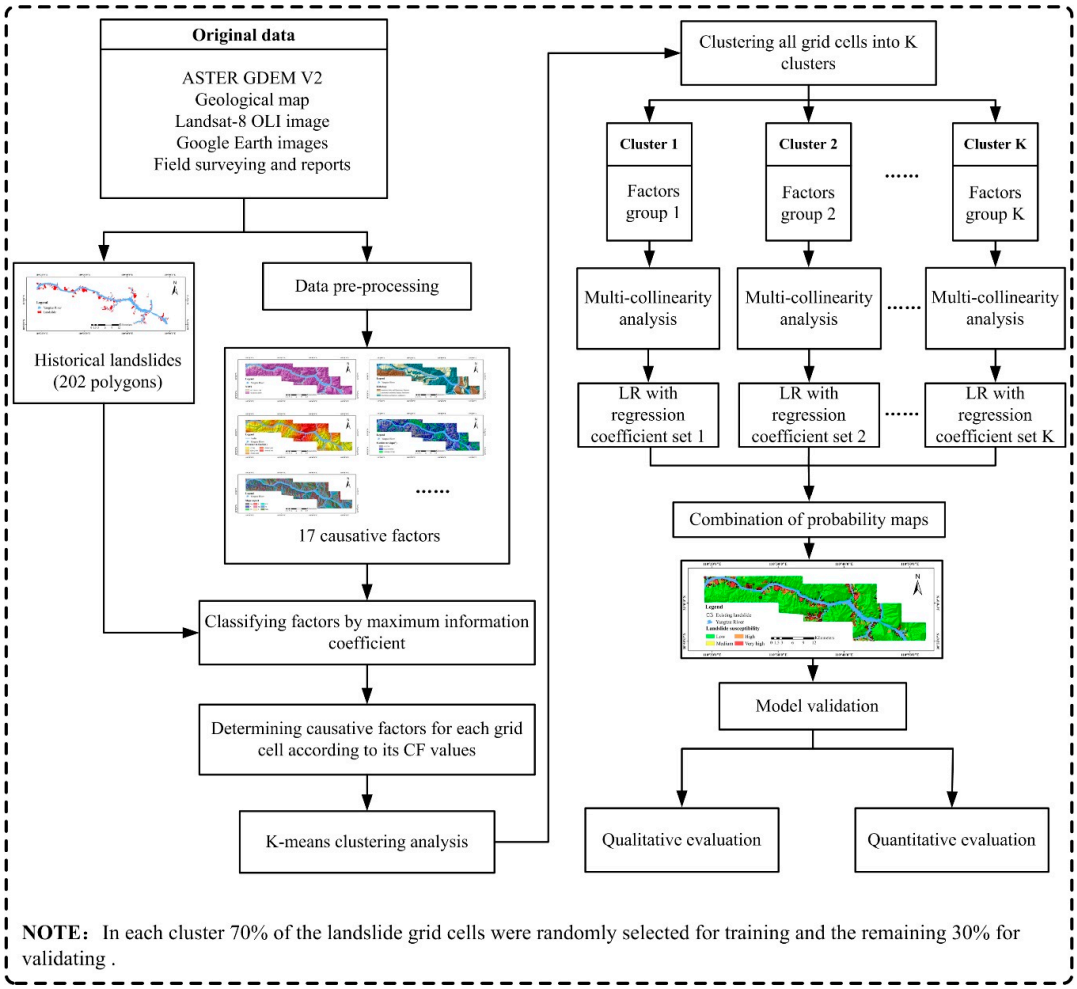



**Figure 3.** Illustration of the proposed framework for landslide susceptibility mapping (LSM). LR, logistic regression.

#### 2.2.1. Information Coefficient Based on Shannon's Entropy Index

Shannon's entropy model has been commonly used to measure the amount of information in a signal or event [68], and landslide occurrence can be estimated using this approach. Shannon's entropy model was used to estimate the density of landslides within each class per factor. With respect to the $i$-th class of the $j$-th factor, let $p_{ij}$ denote the probability density, $A_{ij}$ and $B_{ij}$ the area percentage and the landslide percentage, respectively, $C_j$ the total number of classes of the $j$-th factor and $M_j$ and $M_{jmax}$ the entropy value of the $j$-th factor and its maximum, respectively. The calculation of the information coefficient $I_j$ can be implemented using a series of formulas given below [69,70]:

$$p_{ij} = \frac{B_{ij}}{A_{ij}} \tag{1}$$

$$p_{ij} = \frac{p_{ij}}{\sum_{j=1}^{c_j} p_{ij}} \tag{2}$$

$$M_j = -\sum_{i=1}^{c_j} p_{ij} \log_2 p_{ij} \tag{3}$$

$$M_{jmax} = \log_2 c_j \tag{4}$$

$$I_j = \frac{M_{jmax} - M_j}{M_{jmax}} \tag{5}$$

Since the entropy value is constantly above or equal to 0, the range of the information coefficient is restricted to the domain of [0, 1]. Specifically, the amount of extracted information increases as the information coefficient ranges from 0–1.

#### 2.2.2. Certainty Factor

The CF method has been commonly used for landslide prediction because it is capable of dealing with the challenge of the combination of different vector layers, the heterogeneity and the uncertainty of the input data. The certainty factor can be expressed as follows [19,71]:

$$CF = \begin{cases} \dfrac{PP_a - PP_s}{PP_a(1 - PP_s)} & PP_a \geq PP_s \\ \dfrac{PP_a - PP_s}{PP_s(1 - PP_a)} & PP_a < PP_s \end{cases} \tag{6}$$

where $PP_a$ is the condition probability of a landslide event occurring in a certain class $a$, while $PP_s$ is the prior probability of a landslide event occurring throughout the study area. From Equation (6), a certainty is defined in the range of $[-1, 1]$. If the CF value is larger than zero, the certainty of a landslide occurrence is high, while if this value is smaller than 0, the certainty of a landslide occurrence is low. In particular, if a CF value equals 0, this means that no indication is available about the contribution of a certain class for a causative factor.

#### 2.2.3. K-means Clustering Analysis

K-means is widely used for solving clustering problems due to its efficiency and simple implementation [72]. The main idea of this algorithm is to group a given dataset into K clusters, in which each data point is assigned to the cluster with the nearest mean, thus serving as the centroid of the cluster [73]. The iterative process is performed on the input dataset for re-clustering of all of the data points and updating the location of the centroids until these centroids do not change any more. This algorithm is applied to minimize the following objective function [74]:

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n} \left\| x_j^i - c_i \right\|^2 \tag{7}$$

where $x_j^i$ and $c_i$ represent the $j$-th data point and the $i$-th cluster center, respectively. $\left\| x_j^i - c_i \right\|$ means the $L^2$ norm of $\left( x_j^i - c_i \right)$.

#### 2.2.4. Multicollinearity Analysis

To estimate the correlation between the causative factors, multicollinearity analysis has received extensive attention. Multicollinearity refers to a statistical phenomenon in which there exists a high relationship between two or more predictor variables in a multiple regression model [75]. When those variables are highly correlated, it is difficult to obtain their respective coefficients accurately. To detect multicollinearity, two diagnostic indices are commonly used, tolerance (TOL) and the variance inflation factor (VIF) [76]. Let $X = \{X_1, X_2, \ldots, X_N\}$ define a given independent variable set and $R_j^2$ denote the coefficient of determination when the $j$-th independent variable $X_j$ is regressed on all other predictor variables in the model. The VIF value is computed as follows:

$$\text{VIF} = 1 / \left( 1 - R_j^2 \right) \tag{8}$$

The TOL measure is the reciprocal of the VIF value and represents the degree of linear correlation between independent variables. From Equation (8), if $R_j^2 = 0$, then VIF = TOL = 1, meaning that $X_j$ is not linearly related to the others; if $R_j^2$ is close to 1, then VIF $\to \infty$ and TOL = 0, indicating that $X_j$ is highly related to the others. If the VIF value is above the threshold value of 5 or 10 [75], the corresponding regression coefficients are collinear and should be removed from the predictive model.

#### 2.2.5. Logistic Regression

Logistic regression is a multivariate statistical method to establish the relationship between a dependent variable and several independent variables [6,35,38,77–79]. In recent years, the logical regression model has been commonly used for LSM due to its simplicity and effectiveness [18,58,80–82]. The main idea of such a method is to perform maximum likelihood estimation to obtain the probability of landslide occurrence after each independent variable is converted to a logical variable. The simplified logical regression model can be quantitatively expressed as follows:

$$p = 1 / (1 + e^{-z}) \tag{9}$$

where $p$ denotes the probability of landslide occurrence and ranges between 0 and 1 and $z$ is the linear combination:

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_N X_N \tag{10}$$

where $\beta_0$ is the intercept of the model and $(\beta_1, \beta_2, \ldots, \beta_N)$ are the regression coefficients representing the impact of $X = \{X_1, X_2, \ldots, X_N\}$ mentioned previously on the logit $z$.

### *2.3. Objective Evaluation Measures*

To objectively assess the predictive methods, two measures were utilized. The first one is overall prediction accuracy, evaluating prediction correctness, and defined as follows:

$$p = \frac{a + b}{S} \times 100\% \tag{11}$$

where $a$ and $b$ mean the numbers of correctly-predicted landslide and non-landslide TMUs in the final susceptibility maps, respectively, and $S$ indicates the total number of grid cells in the study area. According to Equation (11), this measure can be directly applied to the LMS of the entire study area to evaluate the global LR model. If this measure in Equation (11) is used to assess the proposed framework, it should be measured in each cluster as follows:

$$p = \frac{\sum_{i=1}^{K} (a_i + b_i)}{S} \times 100\%, i = 1, 2, \ldots, \mathrm{K} \quad (12)$$

where $a_i$ and $b_i$ are the numbers of correctly-predicted landslide and non-landslide grid cells in the $i$-th cluster, respectively.

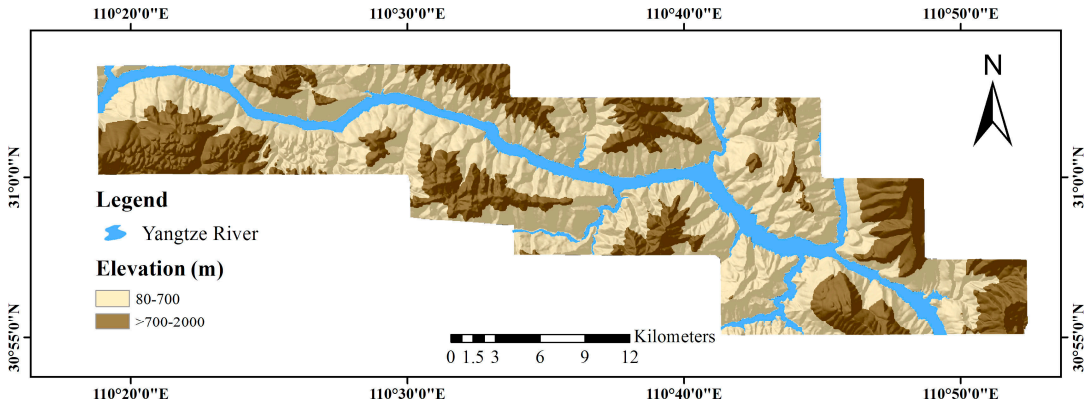
The second one is the commonly-used receiver operating characteristic (ROC) and the area under the ROC curve (AUC). Since a test with perfect discrimination always produces a curve passing through the upper left corner of the plot, the closer the ROC curve is to the upper left corner, the more accurate are the landslide predictive results [83,84]. The AUC value ranges from 0.5–1, and it is close to 1, representing that the model is perfectly reasonable for prediction [85].

## 3. Results

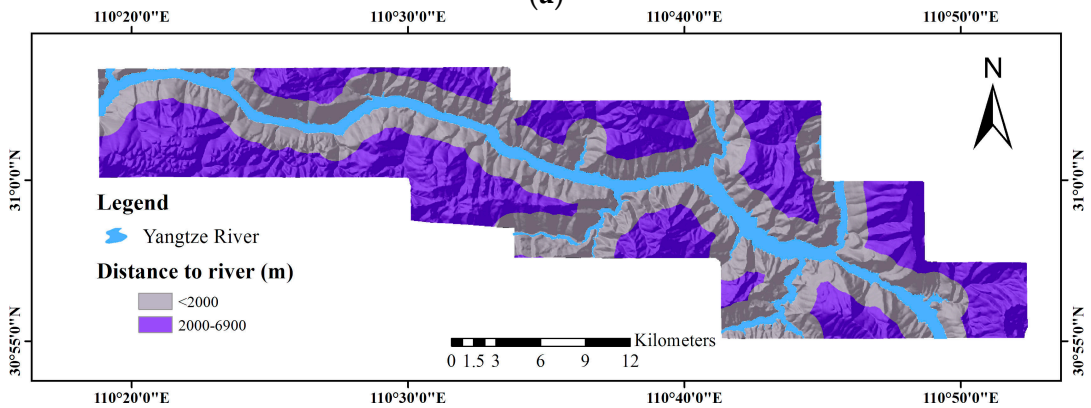
### *3.1. The Construction of the Proposed Framework*

#### 3.1.1. Choosing the Number of Classes for Each Causative Factor

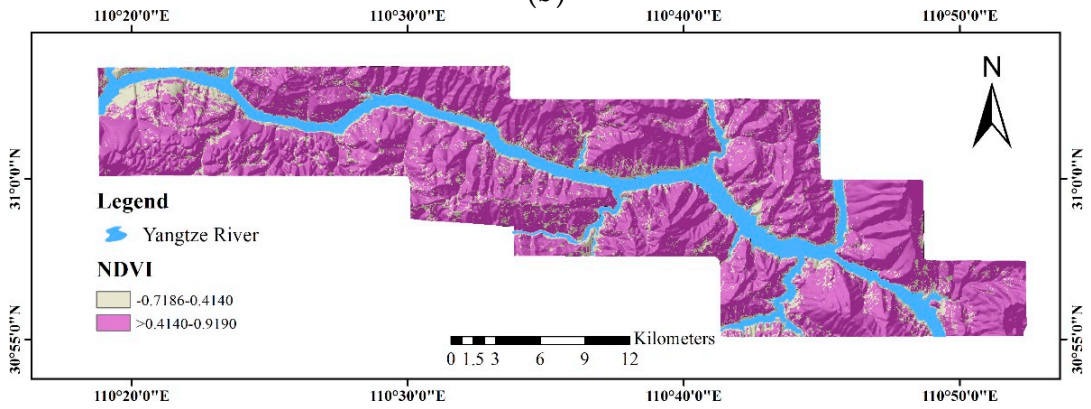
This step is to determine the number of classes for each factor by maximizing its information coefficient. In this work, 14 continuous factors were classified into 2–6 classes using the natural breaks method, except for three categorical factors of slope aspect, lithology and slope form. Based on the approach in [86], the information coefficients of each causative factor considered here are computed and listed in Table 1. It can be concluded from this table that there are six causative factors with the greatest information coefficients when divided into two classes, i.e., elevation, distance to river, NDVI, NDWI, catchment area and terrain surface texture. The three causative factors of slope gradient, catchment slope and TWI have the highest information coefficients of 0.1229, 0.1350 and 0.1798, respectively, when divided into three classes. The causative factors of TRI and Terrain surface convexity maximized their information coefficients to 0.2472 and 0.0933, respectively, when divided into four classes. Only the causative factor of distance to fault was divided into five classes with the highest information coefficient. Finally, the two causative factors of TPI and terrain curvature can obtain the maximum information coefficients of 0.1089 and 0.0933, respectively, when divided into six classes. In this table, the term NC represents non-calculable, which means that there is no landslide grid cell in a class of the factors. Specifically, $p_{ij}$ in Equations (1) and (2) would be zero when there is no landslide grid cell in the $i$-th class of the $j$-th factor. In this case, $\log_2 p_{ij}$ cannot be calculated. There may be some specific operations to avoid this problem, but we did not compute the corresponding information coefficient in this work and assigned it as "NC". Furthermore, the two most influential factors are elevation and distance to river with information coefficients above 0.8, which means that these two causative factors are crucial for the LSM of the study area. In contrast, the two least-influential causative factors are distance to fault and Catchment area, with information coefficients of below 0.06. The classification results of the 14 continuous factors with the optimal number of classes are shown in Figure 4a–n, while the classification maps of the other three categorical factors of lithology, slope aspect and slope form are illustrated in Figure 4o–q, respectively. In this work, the categorical factor of slope form is classified as concave/concave (V/V), elongated/concave (GE/V), convex/concave (X/V), concave/even (V/GR), elongated/even (GE/GR), convex/even (X/GR), concave/convex (V/X), elongated/convex (GE/X).



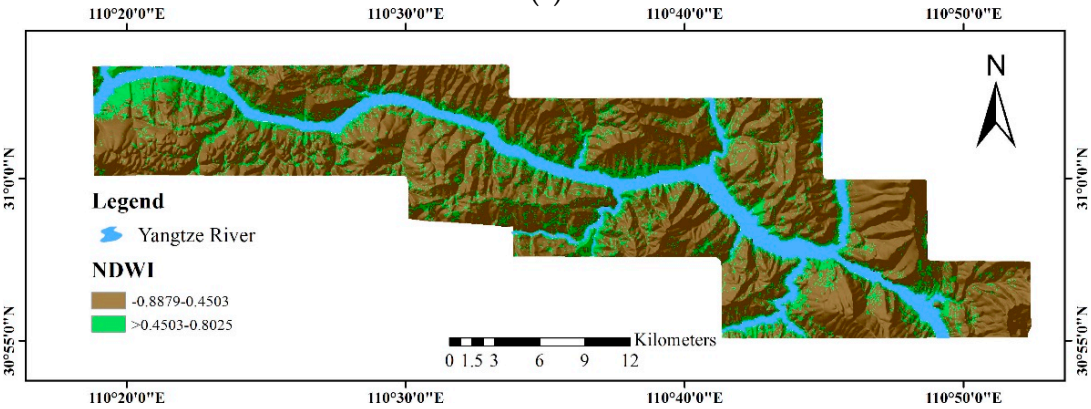

(**a**)




(**b**)




(**c**)




(**d**)

**Figure 4.** *Cont.*

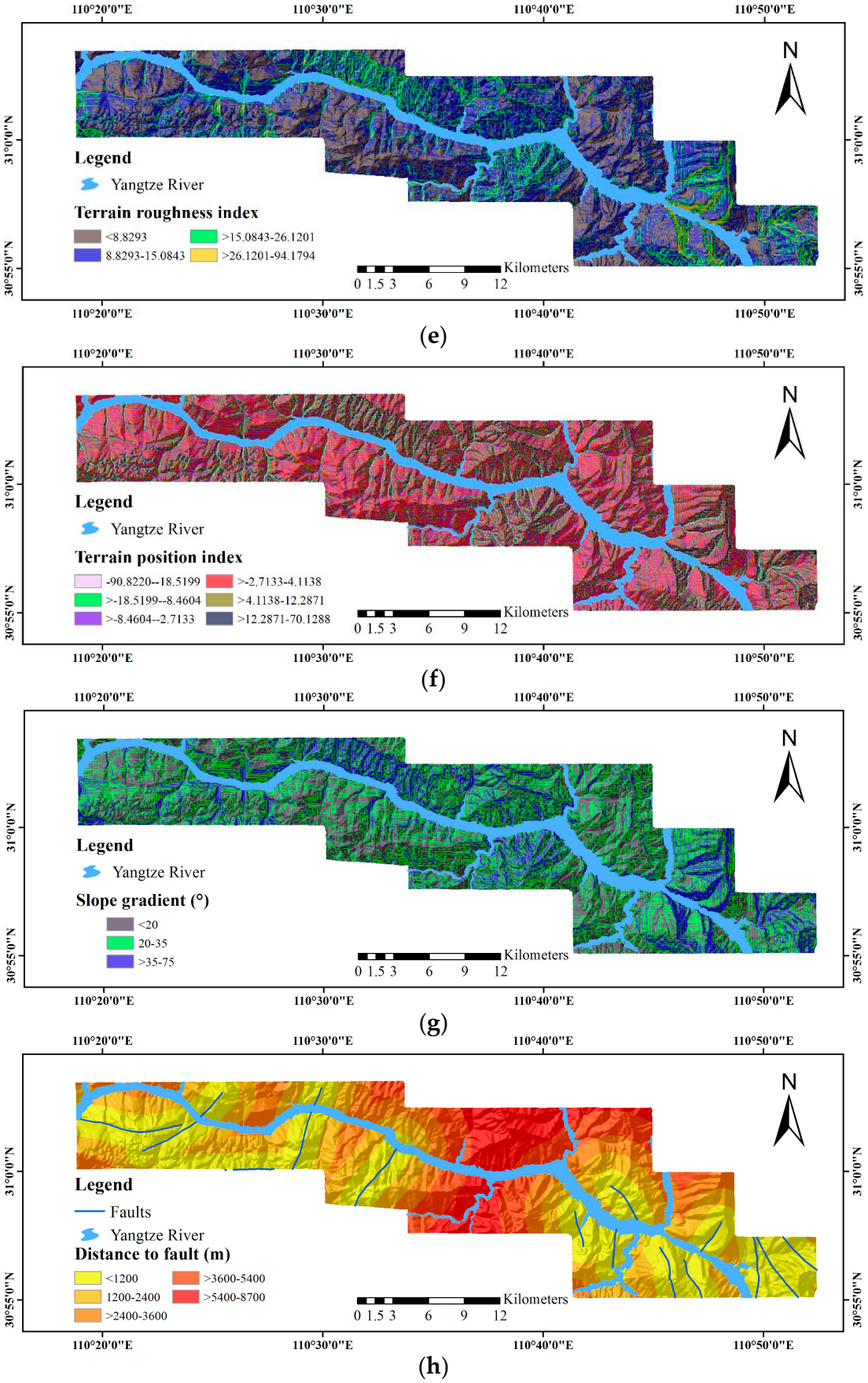



**Figure 4.** *Cont.*

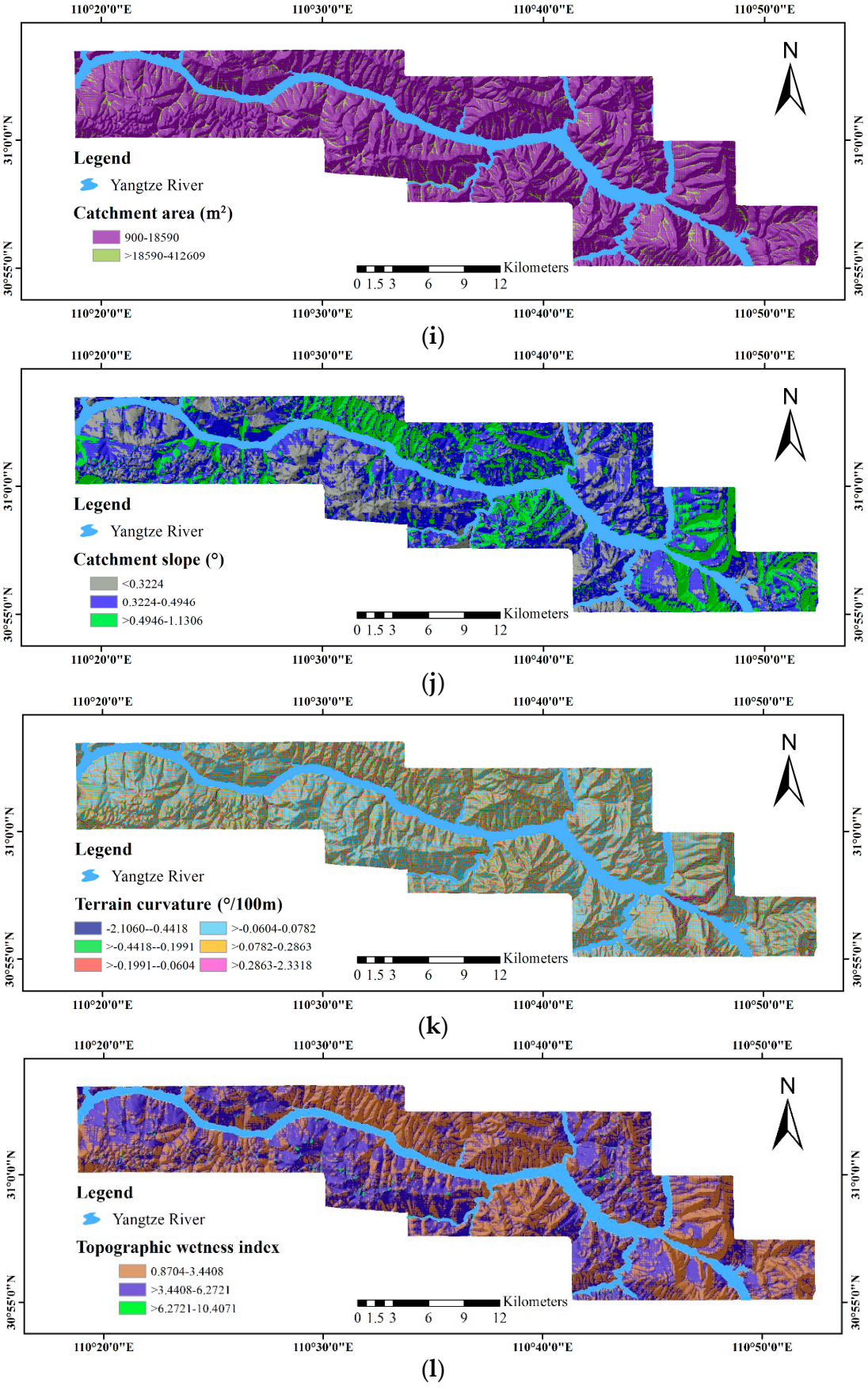
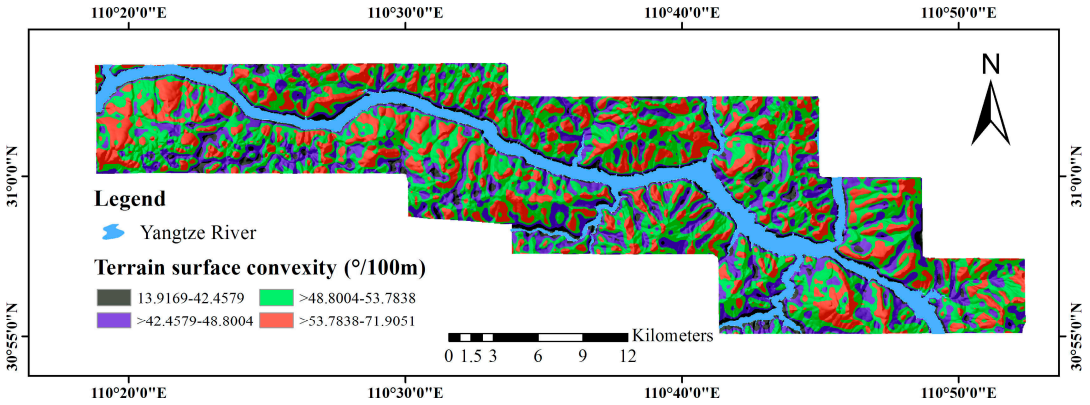
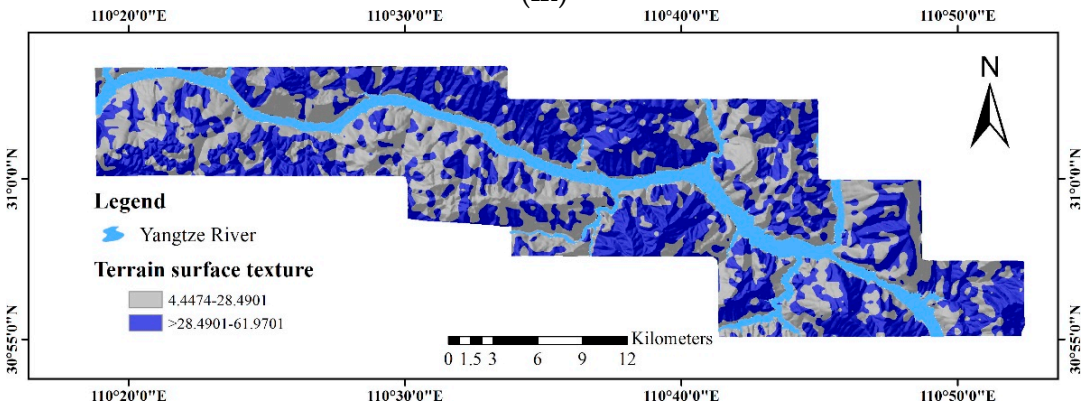


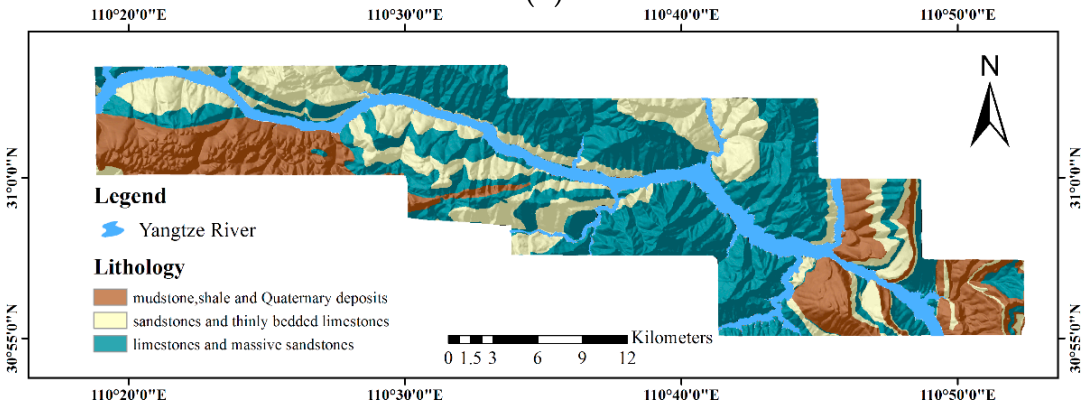

**Figure 4.** *Cont.*



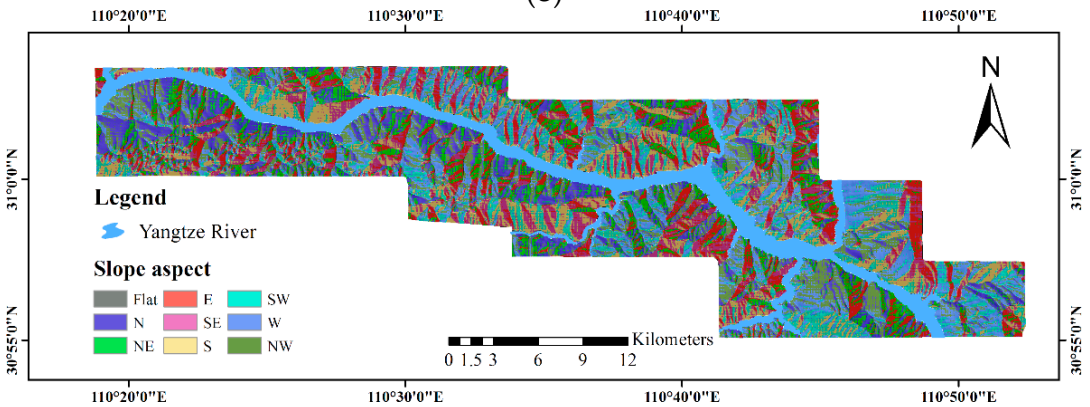

(m)




(n)




(o)




(p)

**Figure 4.** *Cont.*

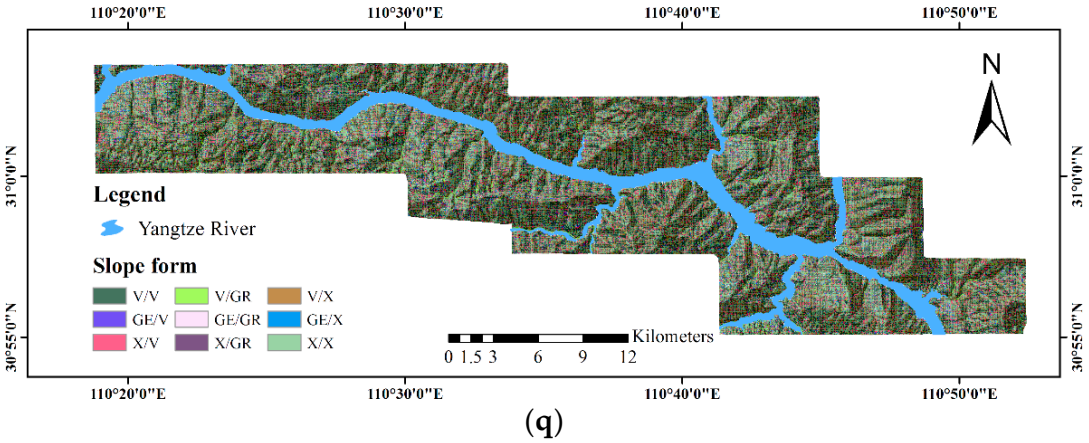



**Figure 4.** Maps showing classification using different causative factors. (**a**) Elevation, (**b**) distance to river, (**c**) NDVI, (**d**) NDWI, (**e**) TRI, (**f**) TPI, (**g**) slope gradient, (**h**) distance to fault, (**i**) catchment area, (**j**) catchment slope, (**k**) terrain curvature, (**l**) TWI, (**m**) terrain surface convexity, (**n**) terrain surface texture, (**o**) lithology, (**p**) slope aspect and (**q**) slope form.

**Table 1.** Information coefficients of each causative factor. The highest ones in each factor are indicated in bold and underlined. TRI, terrain roughness index; TPI, terrain position index; TWI, topographic wetness index.

| Causative Factor | Information Coefficient | | | | |
|---|---|---|---|---|---|
| | 2 Classes | 3 Classes | 4 Classes | 5 Classes | 6 Classes |
| Elevation | **0.8655** | 0.3912 | 0.4209 | NC | NC |
| Distance to river | **0.8040** | 0.4542 | 0.3068 | 0.2743 | NC |
| NDVI | **0.2898** | 0.1578 | 0.1223 | 0.1172 | 0.1053 |
| NDWI | **0.2776** | 0.1439 | 0.1486 | 0.1428 | 0.1175 |
| TRI | 0.2343 | 0.2270 | **0.2472** | 0.2398 | NC |
| TPI | 0.0011 | 0.0715 | 0.0889 | 0.1021 | **0.1089** |
| Slope gradient | 0.1144 | **0.1229** | 0.1140 | 0.1123 | 0.1180 |
| Distance to fault | 0.0193 | 0.0037 | 0.0048 | **0.0322** | 0.0181 |
| Catchment area | **0.0531** | 0.0386 | 0.0294 | 0.0315 | 0.0360 |
| Catchment slope | 0.1175 | **0.1350** | 0.1230 | 0.1279 | 0.1258 |
| Terrain curvature | 0.0080 | 0.0384 | 0.0640 | 0.0808 | **0.0933** |
| TWI | 0.0858 | **0.1798** | 0.1690 | 0.1424 | 0.1539 |
| Terrain surface convexity | 0.0582 | 0.0847 | **0.0933** | 0.0781 | 0.0729 |
| Terrain surface texture | **0.3848** | 0.3495 | 0.3139 | 0.3040 | 0.2940 |

### 3.1.2. Selecting Causative Factors for Each Grid Cell

Based on the above analysis in Section 2.2.2, the CF values were divided into five categories for analyzing the possibility of landslide occurrences, as shown in Table 2. According to Equation (6), for a landslide grid cell, if at least one of its causative factors has a negative CF value, this factor has a negative impact on the landslide prediction of this cell; whereas for a non-landslide grid cell, if at least one of its causative factors has a positive CF value, a similar conclusion can be reached. In both situations, the corresponding factor should not be considered for landslide prediction of this grid cell. In this work, after computing a CF value for each class of all of the causative factors considered here, we were able to obtain an optimal combination of the causative factors for each grid cell, and the selected factors were used as the independent variables of the LR model.

The classes and the corresponding CF values of each causative factor are listed in Table 3 along with the percentages of landslide and class. Elevation is a key factor in landslide occurrences. In Table 3, the CF value of elevation is positive in the range of 80~700 m, which means that landslides always occur in this range of elevation. In this class, the area of landslides accounts for 99% of the total area.

Conversely, the CF value is negative and close to −1 in the range of >700~2000 m, indicating that the probability of landslide occurrences is very low. Distance to river is a commonly-used factor in the evaluation of landslide susceptibility, reflecting the impact of the reservoir water on the landslide. As shown in Table 3, the impact of the reservoir water on the landslides becomes weaker as the distance from the water system increases. From Figure 4a,b, we can also observe that if the distance to river becomes large, the elevation of the same position will be higher. In such terrain, the loose stacking layers are very thin, which is adverse to the development of soil landslides. As slope gradient may affect the slope stability through modulating the surrounding engineering geological conditions, it is another key factor of LSM. Table 3 shows that slope gradient has the highest and lowest CF values of 0.2967 and −0.692, respectively. Moreover, the landslide occurrence in the study area decreases as the slope gradient increases, and very few landslides occur when the slope gradient is higher than 35°, as shown in Table 3. In general, the probability of landslide occurrence should increase with the slope gradient. However, in the study area, the sites with a high slope gradient are mostly distributed in high-elevation areas. As mentioned above, it is difficult to cause slope failures in such places. Slopes with different orientations usually receive different intensities of solar radiation, which affects the distribution of pore water pressure and the physical and mechanical characteristics of rocks and soil. Table 3 shows that the CF value of slope aspect is in the range of [0.2, 1] in the two directions of north and northeast, indicating that the north-facing slopes are more susceptible to landslide occurrences. Apart from these factors, stratigraphic lithology is an important intrinsic factor and the foundation for the development of landslides and can determine the type and scale of landslide occurrences. Table 3 shows that the CF value of lithology is positive in the area with soft and hard sandstone or limestone with thin bedrocks, which is prone to landslide occurrences. Conversely, the value is negative for mudstone, shale, Quaternary deposits, hard limestone or thick sandstone, which means that the area with such lithological characteristics is not conducive to landslide occurrences. Since the term "mudstone, shale and Quaternary deposits" is one class name of lithology, it means that mudstone may not be the only factor to contribute to the stability of a certain area when the CF value of this area is negative. For instance, it was recorded that Quaternary deposits are negative for slope failure [10].

**Table 2.** Categories of the certainty factor (CF) value in terms of slope stability.

| Category | Value Range | Description | Stability |
|---|---|---|---|
| 1 | $CF < -0.6$ | Basically no landslides occurred | stable |
| 2 | $-0.6 \leq CF < -0.2$ | Landslides are less likely to occur | relatively stable |
| 3 | $-0.2 \leq CF < 0.2$ | Uncertain whether landslides will occur | uncertain |
| 4 | $0.2 \leq CF < 0.6$ | Landslides are more likely to occur | unstable |
| 5 | $CF \geq 0.6$ | The possibility of landslides is great | extremely unstable |

**Table 3.** The classes and their CF values of each causative factor.

| Causative Factor | Classes | Percentage of Landslide | Percentage of Class | CF |
|---|---|---|---|---|
| Elevation | 80~700 | 99.00 | 65.50 | 0.3599 |
| | >700~2000 | 1.00 | 34.50 | −0.9728 |
| Distance to river | <2000 | 96.80 | 48.57 | 0.5300 |
| | 2000~6900 | 3.20 | 51.43 | −0.9412 |
| NDVI | −0.7186~0.4140 | 16.24 | 4.46 | 0.7713 |
| | >0.4140~0.9190 | 83.76 | 95.54 | −0.1301 |
| NDWI | −0.8879~0.4503 | 88.29 | 96.79 | 0.0929 |
| | >0.4503~0.8025 | 11.71 | 3.21 | 0.7718 |
| Catchment area | 900~18,590 | 96.13 | 91.52 | −0.0624 |
| | >18,590~412,609 | 3.87 | 8.48 | 0.4134 |

**Table 3.** *Cont.*

| Causative Factor | Classes | Percentage of Landslide | Percentage of Class | CF |
|---|---|---|---|---|
| Terrain surface texture | 4.4474~28.4901 | 82.07 | 45.10 | 0.4792 |
| | >28.4901~61.9701 | 17.93 | 54.90 | −0.6869 |
| Slope gradient | <20 | 48.77 | 35.17 | 0.2967 |
| | 20~35 | 44.61 | 44.25 | 0.0086 |
| | >35~75 | 6.62 | 20.58 | −0.6920 |
| Lithology | mudstone, shale and Quaternary deposits | 3.94 | 22.48 | −0.8660 |
| | sandstones and thinly bedded limestones | 51.82 | 26.97 | 0.5153 |
| | limestones and massive sandstones | 44.24 | 50.55 | −0.1257 |
| TWI | 0.8704~3.4408 | 23.12 | 51.64 | −0.5676 |
| | >3.4408~6.2721 | 76.49 | 47.49 | 0.4033 |
| | >6.2721~10.4071 | 0.39 | 0.87 | −0.5626 |
| Catchment slope | <0.3224 | 36.43 | 28.02 | 0.2457 |
| | 0.3224~0.4946 | 56.68 | 47.52 | 0.1719 |
| | >0.4946~1.1306 | 6.89 | 24.46 | −0.7308 |
| TRI | <8.8293 | 58.80 | 39.26 | 0.3535 |
| | 8.8293~15.0834 | 35.76 | 41.89 | −0.1542 |
| | >15.0834~26.1201 | 5.26 | 16.50 | −0.6947 |
| | >26.1201~94.1794 | 0.18 | 2.35 | −0.9277 |
| Terrain surface convexity | >13.9169~42.4579 | 12.13 | 5.10 | 0.6168 |
| | >42.4579~48.8004 | 31.42 | 25.02 | 0.2167 |
| | >48.8004~53.7838 | 41.06 | 43.07 | −0.0495 |
| | >53.7838~71.9051 | 15.39 | 26.81 | −0.4412 |
| Distance to fault | <1200 | 24.54 | 28.50 | −0.1464 |
| | 1200~2400 | 22.32 | 27.37 | −0.1940 |
| | >2400~3600 | 29.61 | 22.88 | 0.2418 |
| | >3600~5400 | 20.84 | 16.01 | 0.2463 |
| | >5400~8700 | 2.69 | 5.24 | −0.5011 |
| TPI | −90.8220~−18.5199 | 0.52 | 2.13 | −0.7665 |
| | >−18.5199~−8.4604 | 6.04 | 10.30 | −0.4283 |
| | >−8.4604~−2.7133 | 26.97 | 22.60 | 0.1724 |
| | >−2.7133~4.1138 | 48.00 | 35.04 | 0.2873 |
| | >4.1138~12.2871 | 17.17 | 23.56 | −0.2835 |
| | >12.2871~70.1288 | 1.30 | 6.37 | −0.8060 |
| Terrain curvature | −2.1060~−0.4418 | 0.21 | 1.07 | −0.8173 |
| | >−0.4418~−0.1991 | 3.83 | 7.44 | −0.5006 |
| | >−0.1991~−0.0604 | 20.45 | 20.08 | 0.0194 |
| | >−0.0604~0.0782 | 56.76 | 45.41 | 0.2128 |
| | >0.0782~0.2863 | 17.65 | 22.28 | −0.2178 |
| | >0.2863~2.3318 | 1.10 | 3.72 | −0.7184 |
| Slope aspect | Flat | 0.26 | 0.57 | −0.5560 |
| | North | 23.54 | 14.83 | 0.3936 |
| | North–East | 15.95 | 12.57 | 0.2249 |
| | East | 9.67 | 11.86 | −0.1940 |
| | South–East | 7.27 | 10.62 | −0.3292 |
| | South | 13.52 | 12.47 | 0.0826 |
| | South-West | 6.43 | 10.86 | −0.4234 |
| | West | 9.46 | 14.52 | −0.3620 |
| | North–West | 13.90 | 11.70 | 0.1690 |
| Slope form | V/V | 29.74 | 28.71 | 0.0367 |
| | GE/V | 2.70 | 1.63 | 0.4238 |
| | X/V | 10.60 | 11.15 | −0.0522 |
| | V/GR | 4.05 | 3.54 | 0.1340 |
| | GE/GR | 1.30 | 0.58 | 0.5896 |
| | X/GR | 3.27 | 3.05 | 0.0728 |
| | V/X | 13.05 | 13.84 | −0.0605 |
| | GE/X | 3.72 | 2.37 | 0.3867 |
| | X/X | 31.57 | 35.13 | −0.1100 |

### 3.1.3. Clustering Grid Cells into Different Groups

Although each grid cell in the study area has an optimal combination of causative factors based on the analysis mentioned in Section 3.1.2, it is unrealistic and difficult to perform LSM for each grid cell using the predictive model. In this work, the K-means clustering algorithm in IBM SPSS Statistics 19 was adopted to group all of the grid cells into different clusters according to the nearest neighbor principle. To this end, each grid cell in the study area was first assigned a unique 17-digit binary encoding, i.e., "1" denotes that the corresponding factor was selected for this grid cell for the landslide prediction, whereas "0" represents that the corresponding factor was excluded. Then, all binary encoded grid cells were used as input variables in the K-means algorithm for clustering. Consequently, the study area was divided into K clusters, and all of the grid cells have the greatest similarity in the same cluster. Eventually, the optimal combination of causative factors is selected and represented by the final centroid of each cluster shared by all of the cells in the same cluster. In our experiments, we perform the K-means algorithm to divide all of the grid cells into three clusters (K = 3), and the optimal combinations of the causative factors for each cluster are shown in Table 4, where the abbreviations SE and RC denote "selected" and "regression coefficient", respectively, and the symbol "√" indicates that the causative factor is selected for the corresponding cluster. Specifically, all of the causative factors are used as independent variables in the LR model when the study area is not classified using the K-means algorithm.

**Table 4.** The optimal combinations of the causative factors for each cluster and their regression coefficients with the proposed framework when K = 3. SE, selected; RC, regression coefficient.

| **Causative Factor/Intercept** | **No Clustering** | | **K = 3** | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **Cluster 1** | | **Cluster 2** | | **Cluster 3** | |
| Intercept | −8.125 | | −8.783 | | −17.097 | | 1.415 | |
| | SE | RC | SE | RC | SE | RC | SE | RC |
| Elevation | √ | 1.838 | | | | | | |
| Distance to river | √ | 2.801 | √ | 4.473 | | | | |
| NDVI | √ | −0.597 | √ | 0.709 | | | | |
| NDWI | √ | 0.276 | | | | | | |
| Catchment area | √ | −0.156 | | | | | | |
| Terrain surface texture | √ | 1.295 | √ | 0.906 | √ | 0.908 | | |
| Slope gradient | √ | 0.255 | | | √ | 5.476 | √ | 4.256 |
| Lithology | √ | −1.191 | | | | | | |
| TWI | √ | 0.671 | | | √ | 4.470 | √ | −13.190 |
| Catchment slope | √ | 0.400 | | | √ | 3.643 | | |
| TRI | √ | 1.340 | | | | | √ | 3.859 |
| Terrain surface convexity | √ | 0.836 | | | | | √ | 13.433 |
| Distance to fault | √ | 0.286 | | | | | | |
| TPT | √ | 0.573 | | | √ | 1.819 | | |
| Terrain curvature | √ | −0.648 | | | √ | −1.322 | √ | −0.346 |
| Slope aspect | √ | −0.297 | | | √ | −0.898 | | |
| Slope form | √ | 0.012 | √ | −0.184 | | | | |

### 3.1.4. Multicollinearity Analysis of the Selected Causative Factors

After the K-means cluster analysis, a multicollinearity analysis using IBM SPSS Statistics 19.0 was performed for the selected causative factors. The VIF and TOL values of the causative factors for each cluster with K = 3 are listed in Table 5. According to this table, there was no serious multicollinearity between the causative factors in each cluster. For instance, all of the TOL values are higher than 0.4, which is above the commonly-used critical value of 0.1, while the greatest VIF value is less than 2.5, which indicates that the selected causative factors are independent of each other.

**Table 5.** The multicollinearity analysis of the causative factors for each cluster corresponding to Table 4. TOL, tolerance; VIF, variance inflation factor.

| Causative Factor | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|
| | TOL/VIF | | |
| Elevation | | | |
| Distance to river | 0.940/1.064 | | |
| NDVI | 0.944/1.059 | | |
| NDWI | | | |
| Catchment area | | | |
| Terrain surface texture | 0.977/1.023 | 0.944/1.060 | |
| Slope gradient | | 0.648/1.543 | 0.402/2.489 |
| Lithology | | | |
| TWI | | 0.741/1.350 | 0.408/2.451 |
| Catchment slope | | 0.612/1.634 | |
| TRI | | | 0.401/2.494 |
| Terrain surface convexity | | | 0.538/1.860 |
| Distance to fault | | | |
| TPI | | 0.775/1.291 | |
| Terrain curvature | | 0.812/1.231 | 0.907/1.102 |
| Slope aspect | | 0.961/1.041 | |
| Slope form | 0.999/1.001 | | |

### 3.2. Validation and Comparison

In this step, the proposed method was compared with several commonly-used methods, including: (1) the LR method, which is a representative of statistical models; (2) the SVM model, which is representative of machine learning methods; (3) the DT method modelling with the C5.0 algorithm, which is representative of data mining techniques. These methods can be used with both remote sensing images and field surveys and were performed using SPSS Clementine 12.0. To apply these methods to the LSM of the study area, 70% of the landslide grid cells were randomly selected for training the LR, SVM and DT methods, and the remaining landslide grid cells were used for validation. For the proposed framework, the same proportion of the training-validation samples were randomly selected in each cluster. As mentioned in Section 3.1.3, the study area can be clustered by the K-means algorithm for obtaining an optimal combination of the causative factors for each cluster. Meanwhile, the regression coefficients of each cluster per causative factor were computed using the SPSS Clementine 12.0, and the regional LR model with K = 3 (LR_K3) was constructed for comparison. Therefore, the LR, SVM and DT methods were applied to the entire study area for LSM, whereas the LR_K3 method was performed in the different clusters of the study area for accurate LSM at a regional scale. To make the resultant maps more readable, we divided the probability values using the natural breaks method in ESRI ArcGIS 10.0 into four susceptibility zones, i.e., low, medium, high and very high. The landslide susceptibility maps of all of the methods used here are illustrated in Figure 5, which shows that most of the previously investigated landslides are distributed in high or very high susceptibility zones in the maps of all of the predictive methods. However, many grid cells are unreliably categorized by the LR, SVM and DT methods as high or very high susceptibility classes, because landslides occur infrequently in these study areas. In contrast, the map generated by the LR_K3 method is consistent with the actual distribution of landslides, as shown in Figure 1.

The overall accuracies in terms of landslide prediction by all of the methods, which were measured using Equations (11) and (12), are listed in Table 6. The LR_K3 method achieved the best overall accuracy of 85.32% when compared to that of the LR, SVM and DT methods with 80.26%, 83.74% and 84.13%, respectively. The success and prediction rate curves achieved by the different methods are shown in Figure 6. Specifically, the success power of the predictive methods was evaluated by using the training samples, and we can draw similar conclusions as for the overall accuracy mentioned above, i.e., the LR_K3 method achieved better success power with an AUC value of 96.8% compared with that of the LR, SVM and DT methods at 90.4%, 92.3% and 93.4%, respectively. On the other hand, the

predictive performance of the methods considered here was evaluated by using the validation samples, and the observation was consistent with the conclusion on the comparison of the success power of the predictive methods, i.e., the LR_K3 method has superior prediction ability with AUC values of 96.1% in comparison with the LR, SVM and DT methods with AUC values of 90%, 91.5% and 92%, respectively.

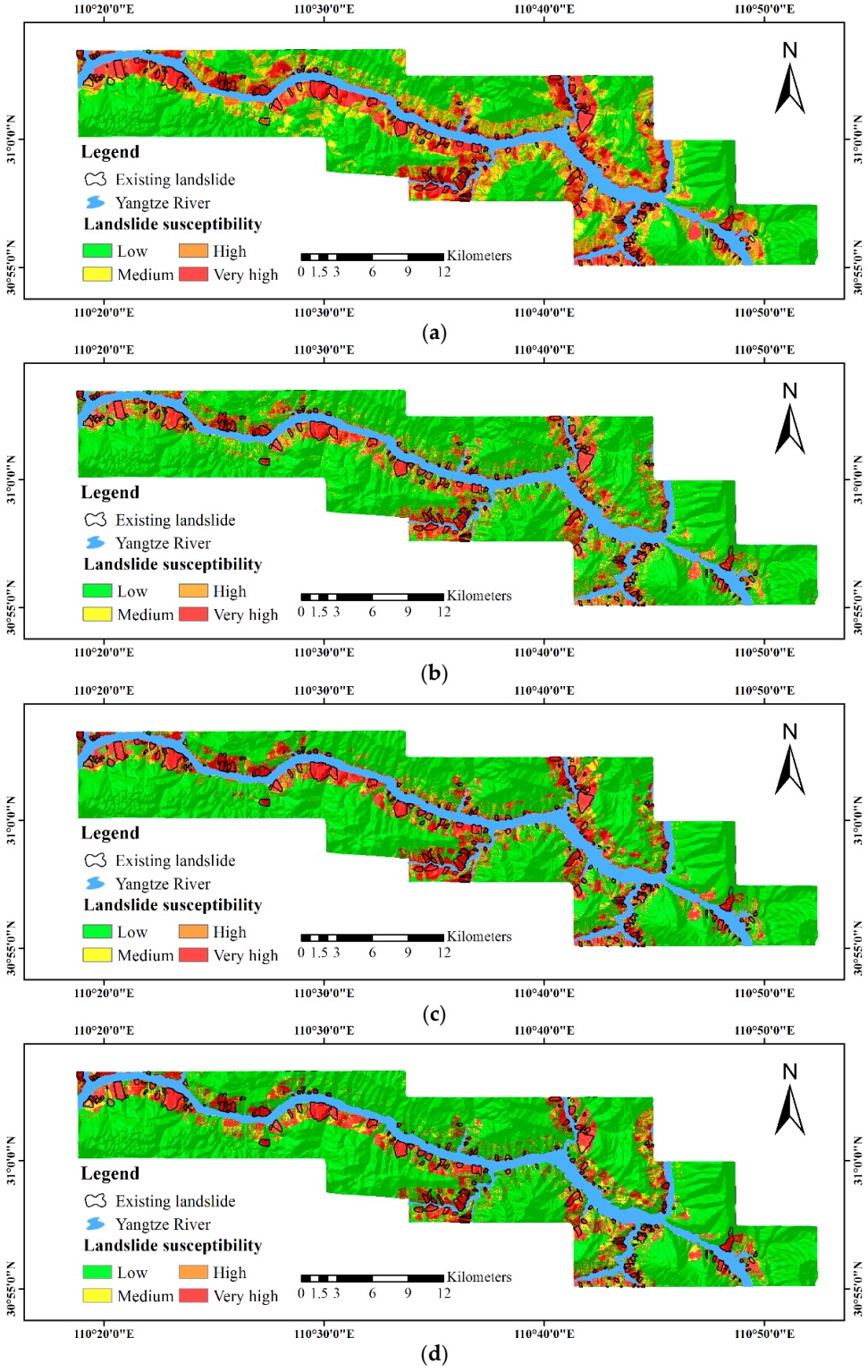



**Figure 5.** The landslide susceptibility maps of the study area produced by the four methods. (**a**) Logistic regression; (**b**) support vector machine; (**c**) decision tree; (**d**) the proposed framework (K = 3).

**Table 6.** Overall accuracies of all of the predictive methods.

| Methods | Overall Accuracy |
|---|---|
| LR | 80.26% |
| SVM | 83.74% |
| DT | 84.13% |
| LR_K3 | 85.32% |

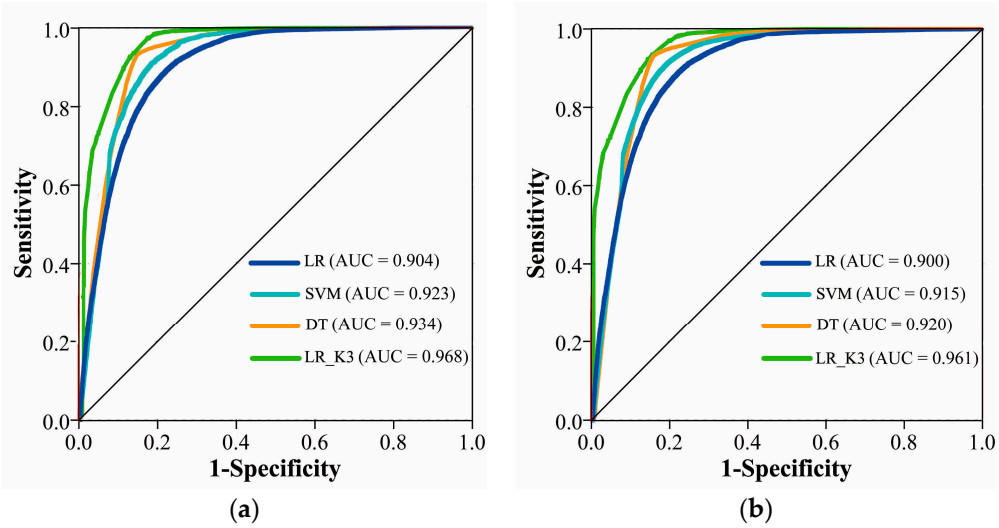



**Figure 6.** ROC curves based on the randomly-selected training-validation samples produced by the four methods. (**a**) The success rate curve; (**b**) the prediction rate curve.

## 4. Discussion

From the above analyses, we can observe that the number of clusters for the division of the study area is greatly significant for landslide prediction. In this section, the impact of K on the predictive performance of the proposed framework is first discussed. Then, to better describe the aim of this work, we provide qualitative/quantitative analysis of the correlation between landslide susceptibility and urban planning.

### *4.1. Impact of K*

Our experimental results reported that there is no grid cell in at least one cluster when the study area was divided into K ($K \geq 5$) clusters. Therefore, we perform the K-means algorithm with $K = \{2, 3, 4\}$. The optimal combinations of the causative factors for each cluster with $K = \{2, 4\}$ are shown in Table 7. Tables 4 and 7 show that the two factors catchment area and distance to fault were not selected, no matter the number of clusters in the study area, indicating that these two factors are not critical for LSM, which is consistent with the conclusion that these two factors have the smallest information coefficients, as mentioned in Section 3.1.1. The VIF and TOL values of the causative factors for each cluster with $K = \{2, 4\}$ are listed in Table 8. According to Tables 5 and 8, there was no serious multicollinearity between the causative factors in each cluster with different values of K, because all of the TOL values are higher than 0.4, and the highest VIF value is less than 2.5. Therefore, the proposed regional LR framework can obtain more accurate landslide susceptibility maps using the selected causative factors.

The statistics of all of the regional LR models are given in Table 9, which shows that both of the two measures $-2$ ln likelihood and goodness of fit of the proposed framework are smaller than those of the traditional LR model using all of the causative factors, validating the improved fitness of the proposed methods. Furthermore, the smallest pseudo $R^2$ measure is 0.252 using the proposed

framework with K = 3, which is appropriate for LSM of the study area. To validate the proposed framework, we constructed a regional LR model with K = {2, 3, 4} (LR_K2, LR_K3 and LR_K4) for comparison, and the landslide susceptibility maps of the three LR-based methods are illustrated in Figure 7. As K was increased from 2–4, more accurate landslide susceptibility maps were obtained, comparing the results in Figure 7a–c. For instance, the high or very high susceptibility zones in Figure 7c produced by the LR_K4 method mainly correlate with the actual landslide areas, whereas most of the non-landslide areas in this map are categorized as medium or low susceptibility zones.

**Table 7.** The optimal combinations of the causative factors for each cluster and their regression coefficients with the proposed framework when K = 2 and 4.

| Causative Factor/Intercept | K = 2 | | | | K = 4 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Cluster 1 | | Cluster 2 | | Cluster 1 | | Cluster 2 | | Cluster 3 | | Cluster 4 | |
| Intercept | −3.600 | | −0.508 | | −17.628 | | −30.784 | | −50.719 | | −12.063 | |
| | SE | RC | SE | RC | SE | RC | SE | RC | SE | RC | SE | RC |
| Elevation | | | | | √ | 3.321 | √ | 4.392 | √ | 20.754 | | |
| Distance to river | √ | 4.473 | | | √ | 7.241 | | | √ | 13.396 | √ | 3.392 |
| NDVI | √ | 0.709 | | | | | | | | | | |
| NDWI | | | | | | | √ | 3.447 | | | | |
| Catchment area | | | | | | | | | | | | |
| Terrain surface texture | √ | 0.906 | √ | 0.908 | | | | | √ | 17.378 | √ | 2.841 |
| Slope gradient | | | √ | 5.476 | | | | | | | | |
| Lithology | | | | | | | √ | 1.350 | √ | 3.513 | | |
| TWI | | | √ | 4.470 | | | √ | 14.754 | √ | −3.863 | √ | 0.400 |
| Catchment slope | | | √ | 3.643 | | | | | | | √ | 3.990 |
| TRI | | | | | | | | | | | | |
| Terrain surface convexity | | | | | √ | 7.693 | | | √ | 31.940 | | |
| Distance to fault | | | | | | | | | | | | |
| TPT | | | √ | 1.819 | | | √ | 3.560 | √ | −2.669 | √ | 0.778 |
| Terrain curvature | | | √ | −1.322 | √ | −1.286 | | | | | | |
| Slope aspect | | | √ | −0.898 | | | | | √ | −1.151 | | |
| Slope form | √ | −0.184 | | | | | | | | | √ | −0.519 |

**Table 8.** The multicollinearity analysis of the causative factors for each cluster corresponding to Table 7.

| Causative Factor | K = 2 (TOL/VIF) | | K = 4 (TOL/VIF) | | | |
|---|---|---|---|---|---|---|
| | Cluster 1 | Cluster 2 | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
| Elevation | | | 0.569/1.756 | 0.861/1.162 | 0.464/2.154 | |
| Distance to river | 0.934/1.070 | | 0.563/1.777 | | 0.406/2.462 | 0.985/1.016 |
| NDVI | | | | | | |
| NDWI | 0.915/1.093 | | | 0.969/1.032 | | |
| Catchment area | | | | | | |
| Terrain surface texture | | 0.975/1.026 | | | 0.670/1.493 | 0.946/1.057 |
| Slope gradient | | | | | | |
| Lithology | 0.977/1.023 | | | 0.852/1.173 | 0.864/1.157 | |
| TWI | | 0.789/1.268 | | 0.949/1.054 | 0.774/1.291 | 0.627/1.594 |
| Catchment slope | | 0.806/1.240 | | | | 0.632/1.581 |
| TRI | | | | | | |
| Terrain surface convexity | | | 0.958/1.043 | | 0.780/1.282 | |
| Distance to fault | | | | | | |
| TPI | 0.975/1.026 | | | 0.931/1.074 | 0.912/1.097 | 0.638/1.567 |
| Terrain curvature | | 0.908/1.101 | 0.998/1.002 | | | |
| Slope aspect | | 0.997/1.003 | | | 0.879/1.138 | |
| Slope form | | | | | | 0.685/1.460 |

**Table 9.** Summary statistics of the logistic regression models.

| Clusters | | Statistics | | | |
|---|---|---|---|---|---|
| | | −2ln Likelihood | −2ln L0 | Goodness of Fit | Pseudo $R^2$ |
| No clustering | | 68,666.849 | 36,819.425 | 69,332.399 | 0.463 |
| K = 2 | Cluster 1 | 15,559.511 | 8542.172 | 215.389 | 0.451 |
| | Cluster 2 | 27,105.308 | 10,245.806 | 2833.666 | 0.622 |
| K = 3 | Cluster 1 | 21,731.018 | 1738.480 | 1887.318 | 0.920 |
| | Cluster 2 | 62,526.249 | 46,769.634 | 3962.812 | 0.252 |
| | Cluster 3 | 12,830.029 | 3412.788 | 2805.024 | 0.734 |
| K = 4 | Cluster 1 | 24,014.293 | 17,722.546 | 1954.458 | 0.262 |
| | Cluster 2 | 16,054.908 | 5025.186 | 19,513.817 | 0.687 |
| | Cluster 3 | 10,818.555 | 2704.639 | 2940.536 | 0.750 |
| | Cluster 4 | 11,601.296 | 7865.679 | 34,575.656 | 0.322 |

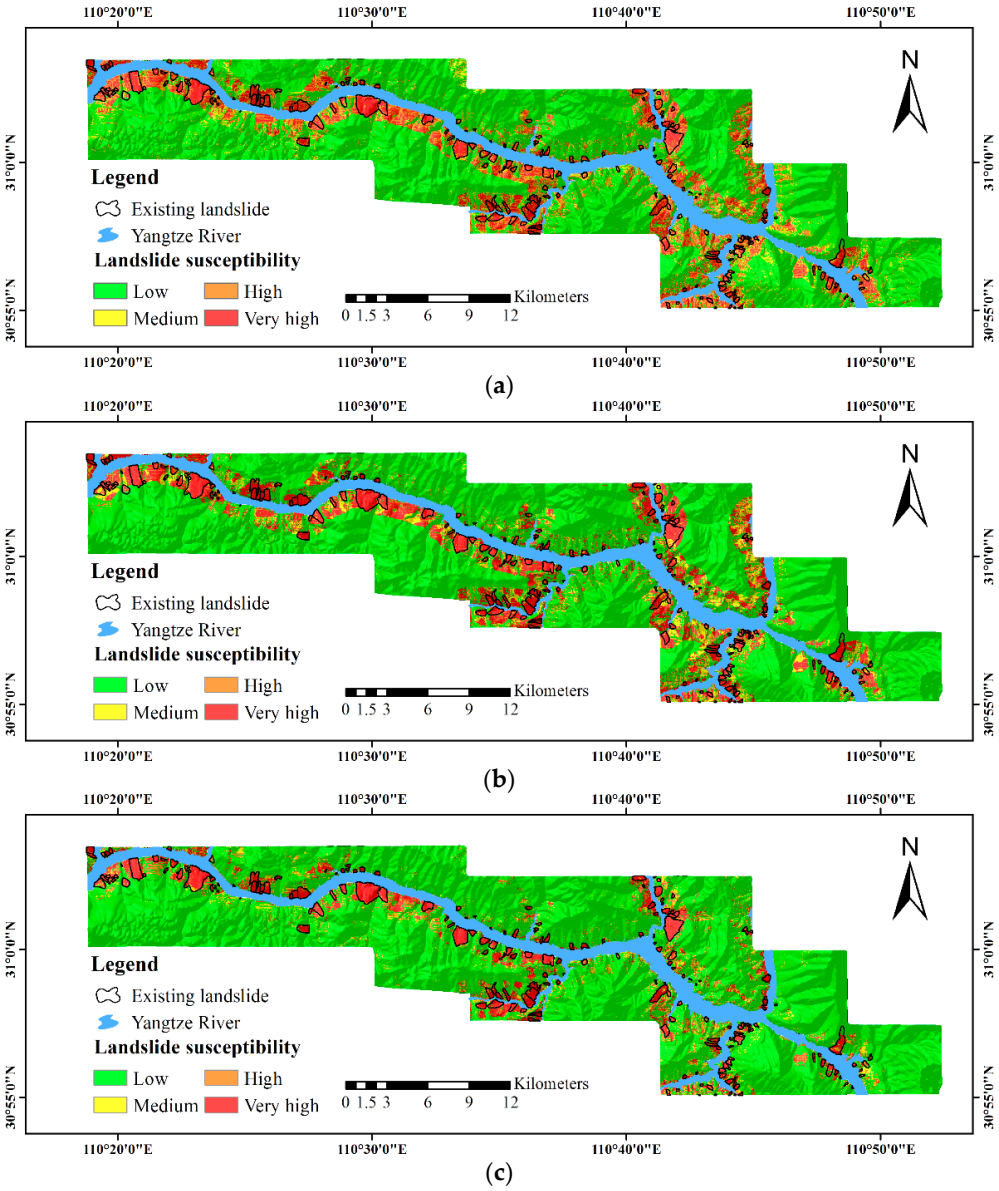



**Figure 7.** The landslide susceptibility maps of our study area produced by the proposed framework with different values of K. (**a**) K = 2; (**b**) K = 3; (**c**) K = 4.

The overall accuracies in terms of landslide prediction by the three constructed methods are listed in Table 10. More accurate landslide susceptibility maps can be obtained as K is increased from 2–4. For instance, the proposed LR_K4 method can obtain the best prediction accuracy of 91.76%, which is 11.5%, 9.91% and 6.44% higher than that of the global LR, LR_K2 and LR_K3 methods, respectively. Furthermore, the success and prediction rate curves achieved by the three methods are shown in Figure 8. The ROC analysis in this figure indicated that all of the curves achieved by the proposed methods were better than that of the LR method. In addition, Figure 8 shows that the AUC value achieved using the proposed framework was better as K was increased from 2–4. It should also be noted that the AUC values of the LR_K4 method can be higher than 0.98.

**Table 10.** Overall accuracies of the LR_K2, LR_K3 and LR_K4 methods.

| Methods | Overall Accuracy |
|---|---|
| LR_K2 | 81.85% |
| LR_K3 | 85.32% |
| LR_K4 | 91.76% |

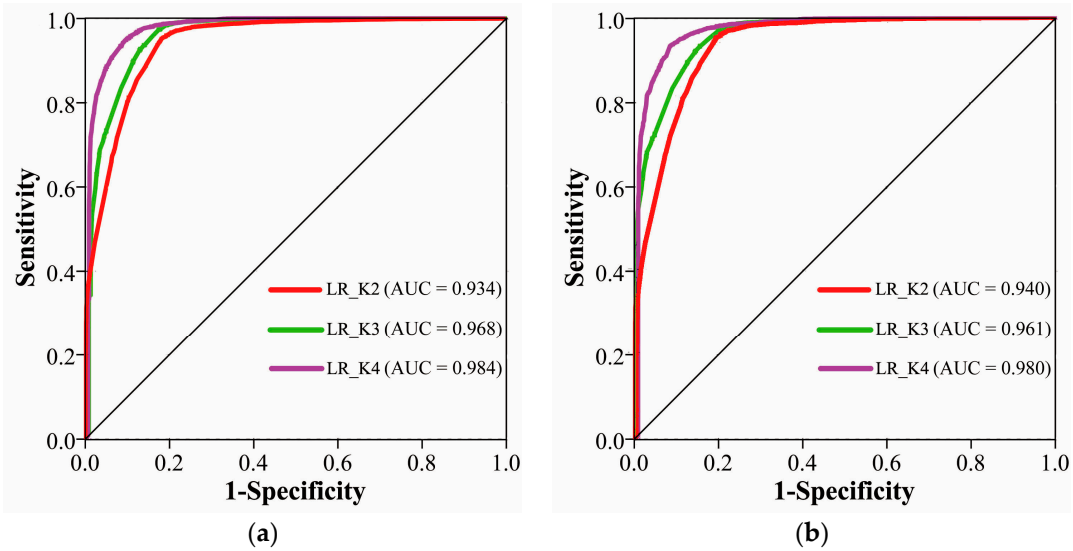



**Figure 8.** The impact of K on the ROC curves of the proposed framework using randomly selected training-validation samples. (**a**) The success rate curve; (**b**) the prediction rate curve.

### *4.2. The Suitability for Urban Development*

During development of an urban environment, landslide susceptibility maps and other physical factors of the study area should be considered by the decision makers and planners since the geology and geomorphology of an area are very significant for urban sustainability [87,88]. In this subsection, the integration technique of AHP and GIS is performed to encourage the evaluation and the selection of suitable areas for urban development of the study area. To assess the suitability for different land uses, the geomorphological, geological and geographical causative factors, along with the landslide hazards were considered. To obtain the potential suitability map for urban development, the causative factors of elevation, distance to river, distance to main towns, landslide susceptibility map, slope gradient and slope aspect were used in this study. The landslide susceptibility map was obtained by the proposed framework with K = 4. The rating of the classes of each causative factor was based on a five-grade scale ranging from 0–4, which has been widely used by other researchers [89–91]. A grade of zero indicates the most favorable conditions for slope failure, while a grade of four describes the most stable conditions for urban development. The selected factors, their classes, ratings and weighting coefficients are listed in Table 11.

**Table 11.** The selected factors, their classes, ratings and weighting coefficients.

| Factors | Potential Rating | | | | | Weights $W_i$ |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | |
| Elevation (m) | >1000 | >750–1000 | >500–750 | 250–500 | <250 | 0.080 |
| Distance to river (m) | >4000 | >3000–4000 | >2000–3000 | 1000–2000 | <1000 | 0.078 |
| Distance to main towns (m) | >4000 | >3000–4000 | >2000–3000 | 1000–2000 | <1000 | 0.212 |
| Landslide susceptibility map | Very high | High | Medium | | Low | 0.320 |
| Slope gradient (°) | >25 | >15–25 | >10–15 | 5–10 | <5 | 0.246 |
| Slope aspect | N | NE, NW | E, W | SE, SW | S, Flat | 0.064 |

The suitability map for urban development of the study area is shown in Figure 9, and this map was classified into the following four categories using the natural breaks method: low, medium, high and very high suitability. Regarding the spatial distribution of the four categories, the areas of very high suitability for urban development are located mostly around the main towns in the study area. Specifically, such areas for each city are as follows:

- near the county of Badong, southwest south and southeast of this county.
- near the county of Xietan, west, northwest, north, northeast and east of this county.
- near the county of Shazhenxi, northwest, north, northeast and south of this county.
- near the county of Guizhou, north, northeast, east and southeast of this county.
- near the county of Guojiaba, south and southeast of this county.
- near the county of Xiangxi, north, northeast and east of this county.
- near the county of Quyuan, northwest, north, northeast, east and southeast of this county.

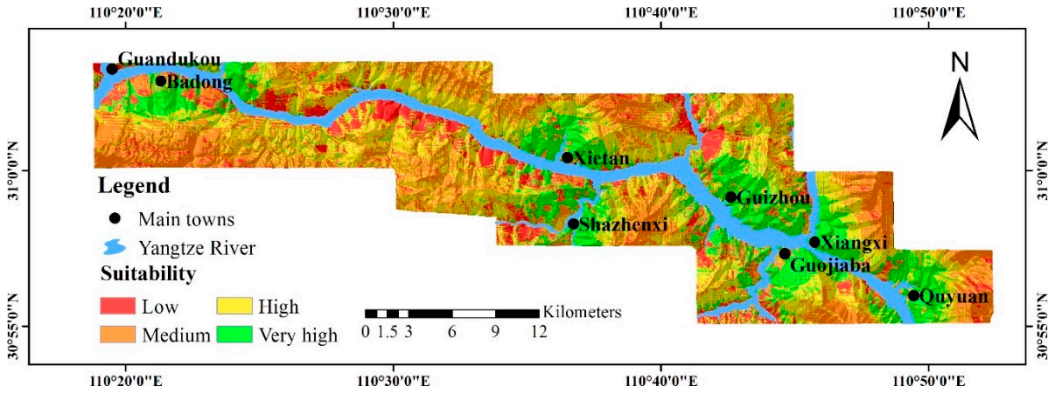



**Figure 9.** The potential suitability for urban development.

## 5. Conclusions

Landslides are the leading natural hazards in the Three Gorges area, as the water level in the reservoir fluctuates periodically, and they pose a serious threat to life and property. To avoid risks and mitigate damage caused by landslides, accurate susceptibility maps are critically significant for land management and land use planning. To better perform LSM, we presented an effective framework through integrating the techniques of information theory, K-means cluster analysis and an LR model. In this work, a total of 17 causative factors were used to construct the LR model, and the impacts of these factors should be closely related to geographic locations and the nearest neighborhood. The major achievement of this work is the grouping of the study area into several clusters to ensure that landslides in each cluster are affected by the same set of selected causative factors. Based on this idea, the proposed predictive method was constructed for accurate LSM at a regional scale by applying a suitable LR model

to each cluster of the study area. In each cluster, 70% of the landslide grid cells were randomly selected for training the LR model, and the remaining cells were used for validation purposes. The experimental results indicated that the proposed framework can demonstrate superior prediction performance when compared with the traditional LR, SVM and DT methods. Furthermore, the predictive methods used in this work were comprehensively assessed in terms of their overall prediction accuracy and using ROC analysis. These objective measures showed that the proposed framework can produce more accurate landslide susceptibility maps with an overall prediction accuracy above 90%. Additionally, this framework is capable of achieving a more reliable success rate and prediction rate curves with AUC values above 98%. Further, to better describe the correlation between landslide susceptibility and urban planning, a potential suitability map for urban development was obtained using the landslide susceptibility map and the geological and geomorphological causative factors. In the future, other statistical models or machine learning methods can be embedded into the proposed framework for better prediction performance and a comprehensive comparison.

**Acknowledgments:** This work was supported by the National Natural Science Foundation of China (61271408). The authors are grateful to the Headquarters of Prevention and Control of Geo-Hazards in Area of the Three Gorges Reservoir for providing data and materials. The authors would like to thank the handling editor, Norman Kerle, and the anonymous reviewers for their valuable comments and suggestions, which significantly improved the quality of this paper.

**Author Contributions:** Qian Wang prepared the data layers, figures and tables and performed the experiments and analyses. Yi Wang has supervised the research, finished the first draft of the manuscript, edited and reviewed the manuscript and contributed to the model construction and verification. Ruiqing Niu and Ling Peng discussed some key issues on the proposed model and provided useful suggestions for improving our work.

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

## References

1. Liu, C.; Liu, Y.; Wen, M.; Li, T.; Lian, J.; Qin, S. Geo-hazard initiation and assessment in the three gorges reservoir. In *Landslide Disaster Mitigation in Three Gorges Reservoir, China*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 3–40.
2. Tangestani, M.H. A comparative study of Dempster-Shafer and fuzzy models for landslide susceptibility mapping using a GIS: An experience from Zagros Mountains, SW Iran. *J. Asian Earth Sci.* **2009**, *35*, 66–73. [CrossRef]
3. Mantovani, F.; Soeters, R.; Van Westen, C.J. Remote sensing techniques for landslide studies and hazard zonation in Europe. *Geomorphology* **1996**, *15*, 213–225. [CrossRef]
4. Metternicht, G.; Hurni, L.; Gogu, R. Remote sensing of landslides: An analysis of the potential contribution to geo-spatial systems for hazard assessment in mountainous environments. *Remote Sens. Environ.* **2005**, *98*, 284–303. [CrossRef]
5. Scaioni, M.; Longoni, L.; Melillo, V.; Papini, M. Remote sensing for landslide investigations: An overview of recent achievements and perspectives. *Remote Sens.* **2014**, *6*, 9600. [CrossRef]
6. 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.
7. Regmi, N.R.; Giardino, J.R.; Vitek, J.D. Assessing susceptibility to landslides: Using models to understand observed changes in slopes. *Geomorphology* **2010**, *122*, 25–38. [CrossRef]
8. Kamp, U.; Growley, B.J.; Khattak, G.A.; Owen, L.A. GIS-based landslide susceptibility mapping for the 2005 kashmir earthquake region. *Geomorphology* **2008**, *101*, 631–642. [CrossRef]
9. Yalcin, A. GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): Comparisons of results and confirmations. *CATENA* **2008**, *72*, 1–12. [CrossRef]
10. Zhang, G.; Cai, Y.; Zheng, Z.; Zhen, J.; Liu, Y.; Huang, K. Integration of the statistical index method and the analytic hierarchy process technique for the assessment of landslide susceptibility in Huizhou, China. *CATENA* **2016**, *142*, 233–244. [CrossRef]
11. Gokceoglu, M.E.C. Assessment of landslide susceptibility for a landslide-prone area (north of Yenice, NW Turkey) by fuzzy approach. *Environ. Geol.* **2002**, *41*, 720–730. [CrossRef]

12. Jiang, H.; Eastman, J.R. Application of fuzzy measures in multi-criteria evaluation in GIS. *Int. J. Geogr. Inf. Sci.* **2000**, *14*, 173–184. [CrossRef]
13. Pradhan, A.M.S.; Kim, Y.T. Evaluation of a combined spatial multi-criteria evaluation model and deterministic model for landslide susceptibility mapping. *CATENA* **2016**, *140*, 125–139. [CrossRef]
14. Akgun, A.; Dag, S.; Bulut, F. Landslide susceptibility mapping for a landslide-prone area (Findikli, NE of Turkey) by likelihood-frequency ratio and weighted linear combination models. *Environ. Geol.* **2008**, *54*, 1127–1143. [CrossRef]
15. Ayalew, L.; Yamagishi, H.; Marui, H.; Kanno, T. Landslides in Sado Island of Japan: Part II. GIS-based susceptibility mapping with comparisons of results from two methods and verifications. *Eng. Geol.* **2005**, *81*, 432–445. [CrossRef]
16. Feizizadeh, B.; Blaschke, T.; Nazmfar, H. GIS-based ordered weighted averaging and Dempster-Shafer methods for landslide susceptibility mapping in the Urmia Lake Basin, Iran. *Int. J. Digit. Earth* **2012**, *7*, 688–708. [CrossRef]
17. Ferretti, V.; Pomarico, S. Ecological land suitability analysis through spatial indicators: An application of the analytic network process technique and ordered weighted average approach. *Ecol. Indic.* **2013**, *34*, 507–519. [CrossRef]
18. Ayalew, L.; Yamagishi, H. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, central Japan. *Geomorphology* **2005**, *65*, 15–31. [CrossRef]
19. Chung, C.-J.F.; Fabbri, A.G. The representation of geoscience information for data integration. *Nat. Resour. Res.* **1993**, *2*, 122–139. [CrossRef]
20. Chung, C.; Fabbri, A. Three bayesian prediction models for landslide hazard. In Proceedings of the International Association for Mathematical Geology 1998 Annual Meeting, Ischia, Italy, 3–7 October 1998.
21. Chung, C.-J.F.; Fabbri, A.G. Probabilistic prediction models for landslide hazard mapping. *Photogramm. Eng. Remote Sens.* **1999**, *65*, 1389–1399.
22. Saha, A.K.; Gupta, R.P.; Sarkar, I.; Arora, M.K.; Csaplovics, E. An approach for GIS-based statistical landslide susceptibility zonation—With a case study in the Himalayas. *Landslides* **2005**, *2*, 61–69. [CrossRef]
23. Guillard, C.; Zezere, J. Landslide susceptibility assessment and validation in the framework of municipal planning in Portugal: The case of Loures municipality. *Environ. Manag.* **2012**, *50*, 721–735. [CrossRef] [PubMed]
24. Pradhan, B.; Lee, S. Landslide susceptibility assessment and factor effect analysis: Backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environ. Model. Softw.* **2010**, *25*, 747–759. [CrossRef]
25. Pradhan, B.; Oh, H.-J.; Buchroithner, M. Weights-of-evidence model applied to landslide susceptibility mapping in a tropical hilly area. *Geomat. Nat. Hazards Risk* **2010**, *1*, 199–223. [CrossRef]
26. Lee, S.; Choi, J. Landslide susceptibility mapping using GIS and the weight-of-evidence model. *Int. J. Geogr. Inf. Sci.* **2004**, *18*, 789–814. [CrossRef]
27. Regmi, N.R.; Giardino, J.R.; Vitek, J.D. Modeling susceptibility to landslides using the weight of evidence approach: Western Colorado, USA. *Geomorphology* **2010**, *115*, 172–187. [CrossRef]
28. Blahut, J.; van Westen, C.J.; Sterlacchini, S. Analysis of landslide inventories for accurate prediction of debris-flow source areas. *Geomorphology* **2010**, *119*, 36–51. [CrossRef]
29. Hong, H.; Ilia, I.; Tsangaratos, P.; Chen, W.; Xu, C. A hybrid fuzzy weight of evidence method in landslide susceptibility analysis on the Wuyuan area, China. *Geomorphology* **2017**, *290*, 1–16. [CrossRef]
30. Pradhan, B.; Lee, S. Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models. *Environ. Earth Sci.* **2009**, *60*, 1037–1054. [CrossRef]
31. Lee, S.; Pradhan, B. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides* **2006**, *4*, 33–41. [CrossRef]
32. Pradhan, B.; Chaudhari, A.; Adinarayana, J.; Buchroithner, M.F. Soil erosion assessment and its correlation with landslide events using remote sensing data and GIS: A case study at Penang Island, Malaysia. *Environ. Monit. Assess.* **2012**, *184*, 715–727. [CrossRef] [PubMed]
33. Pourghasemi, H.; Pradhan, B.; Gokceoglu, C.; Moezzi, K.D. A comparative assessment of prediction capabilities of Dempster-Shafer and weights-of-evidence models in landslide susceptibility mapping using GIS. *Geomat. Nat. Hazards Risk* **2013**, *4*, 93–118. [CrossRef]

34. Corominas, J.; van Westen, C.; Frattini, P.; Cascini, L.; Malet, J.P.; Fotopoulou, S.; Catani, F.; Van Den Eeckhaut, M.; Mavrouli, O.; Agliardi, F.; et al. Recommendations for the quantitative analysis of landslide risk. *Bull. Eng. Geol. Environ.* **2014**, *73*, 209–263. [CrossRef]
35. Zêzere, J.L.; Pereira, S.; Melo, R.; Oliveira, S.C.; Garcia, R.A.C. Mapping landslide susceptibility using data-driven methods. *Sci. Total Environ.* **2017**, *589*, 250–267. [CrossRef] [PubMed]
36. Chen, Z.; Wang, J. Landslide hazard mapping using logistic regression model in Mackenzie Valley, Canada. *Nat. Hazards* **2007**, *42*, 75–89. [CrossRef]
37. Can, T.; Nefeslioglu, H.A.; Gokceoglu, C.; Sonmez, H.; Duman, T.Y. Susceptibility assessments of shallow earthflows triggered by heavy rainfall at three catchments by logistic regression analyses. *Geomorphology* **2005**, *72*, 250–271. [CrossRef]
38. Tsangaratos, P.; Ilia, 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]
39. Gorum, T.; Gonencgil, B.; Gokceoglu, C.; Nefeslioglu, H.A. Implementation of reconstructed geomorphologic units in landslide susceptibility mapping: The Melen Gorge (NW Turkey). *Nat. Hazards* **2008**, *46*, 323–351. [CrossRef]
40. Tunusluoglu, M.C.; Gokceoglu, C.; Nefeslioglu, H.A.; Sonmez, H. Extraction of potential debris source areas by logistic regression technique: A case study from Barla, Besparmak and Kapi mountains (NW Taurids, Turkey). *Environ. Geol.* **2007**, *54*, 9–22. [CrossRef]
41. Nefeslioglu, H.A.; Gokceoglu, C.; Sonmez, H. An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. *Eng. Geol.* **2008**, *97*, 171–191. [CrossRef]
42. Choi, J.; Oh, H.-J.; Lee, H.-J.; Lee, C.; Lee, S. Combining landslide susceptibility maps obtained from frequency ratio, logistic regression, and artificial neural network models using aster images and GIS. *Eng. Geol.* **2012**, *124*, 12–23. [CrossRef]
43. Akgun, A.; Kincal, C.; Pradhan, B. Application of remote sensing data and GIS for landslide risk assessment as an environmental threat to Izmir city (west Turkey). *Environ. Monit. Assess.* **2012**, *184*, 5453–5470. [CrossRef] [PubMed]
44. Dong, J.-J.; Tung, Y.-H.; Chen, C.-C.; Liao, J.-J.; Pan, Y.-W. Discriminant analysis of the geomorphic characteristics and stability of landslide dams. *Geomorphology* **2009**, *110*, 162–171. [CrossRef]
45. Mihai, B.; Sandric, I.; Savulescu, I.; Chitu, Z. Detailed mapping of landslide susceptibility for urban planning purposes in carpathian and subcarpathian towns of Romania. In *Cartography in Central and Eastern Europe: CEE 2009*; Gartner, G., Ortag, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 417–429.
46. Yesilnacar, E.; Süzen, M.L. A land-cover classification for landslide susceptibility mapping by using feature components. *Int. J. Remote Sens.* **2006**, *27*, 253–275. [CrossRef]
47. Saito, H.; Nakayama, D.; Matsuyama, H. Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: The Akaishi mountains, Japan. *Geomorphology* **2009**, *109*, 108–121. [CrossRef]
48. 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]
49. Catani, F.; Lagomarsino, D.; Segoni, S.; Tofani, V. Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues. *Nat. Hazards Earth Syst.* **2013**, *13*, 2815–2831. [CrossRef]
50. Vorpahl, P.; Elsenbeer, H.; Märker, M.; Schröder, B. How can statistical models help to determine driving factors of landslides? *Ecol. Model.* **2012**, *239*, 27–39. [CrossRef]
51. Pourghasemi, H.R.; Kerle, N. Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran. *Environ. Earth Sci.* **2016**, *75*, 185. [CrossRef]
52. Chen, C.-H.; Ke, C.-C.; Wang, C.-L. A back-propagation network for the assessment of susceptibility to rock slope failure in the eastern portion of the southern cross-island highway in Taiwan. *Environ. Geol.* **2009**, *57*, 723–733. [CrossRef]
53. Pradhan, B.; Lee, S. Landslide risk analysis using artificial neural network model focussing on different training sites. *Int. J. Phys. Sci.* **2009**, *4*, 1–15.

54. Pradhan, B.; Youssef, A.; Varathrajoo, R. Approaches for delineating landslide hazard areas using different training sites in an advanced artificial neural network model. *Geospat. Inf. Sci.* **2010**, *13*, 93–102. [CrossRef]
55. Arnone, E.; Francipane, A.; Scarbaci, A.; Puglisi, C.; Noto, L.V. Effect of raster resolution and polygon-conversion algorithm on landslide susceptibility mapping. *Environ. Model. Softw.* **2016**, *84*, 467–481. [CrossRef]
56. Pham, B.T.; Tien Bui, D.; Prakash, I.; Dholakia, M.B. Hybrid integration of multilayer perceptron neural networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *CATENA* **2017**, *149*, 52–63. [CrossRef]
57. Tien Bui, D.; Pradhan, B.; Lofman, O.; Revhaug, I. Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and naïve bayes models. *Math. Probl. Eng.* **2012**, *2012*, 1–26. [CrossRef]
58. Yilmaz, I. Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: Conditional probability, logistic regression, artificial neural networks, and support vector machine. *Environ. Earth Sci.* **2010**, *61*, 821–836. [CrossRef]
59. Tien Bui, D.; Pradhan, B.; Lofman, O.; Revhaug, I.; Dick, O.B. Landslide susceptibility assessment in the Hoa Binh Province of Vietnam: A comparison of the levenberg-marquardt and bayesian regularized neural networks. *Geomorphology* **2012**, *171–172*, 12–29. [CrossRef]
60. Hughes, G. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory* **1968**, *14*, 55–63. [CrossRef]
61. Kavzoglu, T.; Kutlug Sahin, E.; Colkesen, I. Selecting optimal conditioning factors in shallow translational landslide susceptibility mapping using genetic algorithm. *Eng. Geol.* **2015**, *192*, 101–112. [CrossRef]
62. Guzzetti, F.; Carrara, A.; Cardinali, M.; Reichenbach, P. Landslide hazard evaluation: A review of current techniques and their application in a multi-scale study, central Italy. *Geomorphology* **1999**, *31*, 181–216. [CrossRef]
63. Pradhan, B.; Lee, S.; Buchroithner, M.F. A GIS-based back-propagation neural network model and its cross-application and validation for landslide susceptibility analyses. *Comput. Environ. Urban Syst.* **2010**, *34*, 216–235. [CrossRef]
64. 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]
65. Das, I.; Stein, A.; Kerle, N.; Dadhwal, V. Probabilistic landslide hazard assessment using homogeneous susceptible units (HSU) along a national highway corridor in the northern Himalayas, India. *Landslides* **2011**, *8*, 293–308. [CrossRef]
66. Erener, A.; Düzgün, H.S.B. Improvement of statistical landslide susceptibility mapping by using spatial and global regression methods in the case of More and Romsdal (Norway). *Landslides* **2010**, *7*, 55–68. [CrossRef]
67. Yu, X.; Wang, Y.; Niu, R.; Hu, Y. A combination of geographically weighted regression, particle swarm optimization and support vector machine for landslide susceptibility mapping: A case study at Wanzhou in the Three Gorges Area, China. *Int. J. Environ. Res. Public Health* **2016**, *13*, 487. [CrossRef] [PubMed]
68. Shannon, C.E. A mathematical theory of communication. *ACM SIGMOBILE Mobile Comput. Commun. Rev.* **2001**, *5*, 3–55. [CrossRef]
69. Constantin, M.; Bednarik, M.; Jurchescu, M.C.; Vlaicu, M. Landslide susceptibility assessment using the bivariate statistical analysis and the index of entropy in the Sibiciu Basin (Romania). *Environ. Earth Sci.* **2010**, *63*, 397–406. [CrossRef]
70. Bednarik, M.; Magulová, B.; Matys, M.; Marschalko, M. Landslide susceptibility assessment of the Kra'ovany-Liptovský Mikuláš railway case study. *Phys. Chem. Earth Parts A/B/C* **2010**, *35*, 162–171. [CrossRef]
71. Kanungo, D.P.; Sarkar, S.; Sharma, S. Combining neural network with fuzzy, certainty factor and likelihood ratio concepts for spatial prediction of landslides. *Nat. Hazards* **2011**, *59*, 1491–1512. [CrossRef]
72. Likas, A.; Vlassis, N.; Verbeek, J.J. The global K-means clustering algorithm. *Pattern Recogn.* **2003**, *36*, 451–461. [CrossRef]
73. Ding, C.; He, X. K-means clustering via principal component analysis. In Proceedings of the Twenty-First International Conference on Machine Learning, Banff, AB, Canada, 4–8 July 2004.
74. Alsabti, K.; Ranka, S.; Singh, V. An efficient K-means clustering algorithm. *Electr. Eng. Comput. Sci.* **1997**, *43*.
75. O'brien, R.M. A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.* **2007**, *41*, 673–690. [CrossRef]
76. Marquaridt, D.W. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics* **1970**, *12*, 591–612. [CrossRef]

77. Das, I.; Sahoo, S.; van Westen, C.; Stein, A.; Hack, R. 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* **2010**, *114*, 627–637. [CrossRef]
78. 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]
79. Nandi, A.; Shakoor, A. A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. *Eng. Geol.* **2010**, *110*, 11–20. [CrossRef]
80. Dai, F.; Lee, C.; Ngai, Y.Y. Landslide risk assessment and management: An overview. *Eng. Geol.* **2002**, *64*, 65–87. [CrossRef]
81. Gorsevski, P.V.; Gessler, P.E.; Boll, J.; Elliot, W.J.; Foltz, R.B. Spatially and temporally distributed modeling of landslide susceptibility. *Geomorphology* **2006**, *80*, 178–198. [CrossRef]
82. Akgun, A.; Sezer, E.A.; Nefeslioglu, H.A.; Gokceoglu, C.; Pradhan, B. An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. *Comput. Geosci.* **2012**, *38*, 23–34. [CrossRef]
83. Kundu, S.; Saha, A.; Sharma, D.; Pant, C. Remote sensing and GIS based landslide susceptibility assessment using binary logistic regression model: A case study in the Ganeshganga watershed, Himalayas. *J. Indian Soc. Remote Sens.* **2013**, *41*, 697–709. [CrossRef]
84. Sdao, F.; Lioi, D.; Pascale, S.; Caniani, D.; Mancini, I. Landslide susceptibility assessment by using a neuro-fuzzy model: A case study in the Rupestrian heritage rich area of Matera. *Nat. Hazards Earth Syst.* **2013**, *13*, 395. [CrossRef]
85. Fawcett, T. An introduction to ROC analysis. *Pattern Recogn. Lett.* **2006**, *27*, 861–874. [CrossRef]
86. Tsangaratos, P.; Ilia, I.; Hong, H.; Chen, W.; Xu, C. Applying information theory and GIS-based quantitative methods to produce landslide susceptibility maps in Nancheng county, China. *Landslides* **2016**, 1–21. [CrossRef]
87. Bathrellos, G.D.; Gaki-Papanastassiou, K.; Skilodimou, H.D.; Papanastassiou, D.; Chousianitis, K.G. Potential suitability for urban planning and industry development using natural hazard maps and geological-geomorphological parameters. *Environ. Earth Sci.* **2012**, *66*, 537–548. [CrossRef]
88. Bathrellos, G.D.; Skilodimou, H.D.; Chousianitis, K.; Youssef, A.M.; Pradhan, B. Suitability estimation for urban development using multi-hazard assessment map. *Sci. Total Environ.* **2017**, *575*, 119–134. [CrossRef] [PubMed]
89. Dai, F.; Lee, C.; Zhang, X. GIS-based geo-environmental evaluation for urban land-use planning: A case study. *Eng. Geol.* **2001**, *61*, 257–271. [CrossRef]
90. Svoray, T.; Bar, P.; Bannet, T. Urban land-use allocation in a mediterranean ecotone: Habitat heterogeneity model incorporated in a GIS using a multi-criteria mechanism. *Landsc. Urban Plan.* **2005**, *72*, 337–351. [CrossRef]
91. Wang, X.; Zhong, X.; Liu, S.; Liu, J.; Wang, Z.; Li, M. Regional assessment of environmental vulnerability in the Tibetan Plateau: Development and application of a new method. *J. Arid Environ.* **2008**, *72*, 1929–1939. [CrossRef]