

Article

Modeling and Assessment of Landslide Susceptibility of Dianchi Lake Watershed in Yunnan Plateau

Guangshun Bai ^{1,2,3} , Xuemei Yang ⁴, Zhigang Kong ^{1,2,3,*}, Jieyong Zhu ^{1,2,3}, Shitao Zhang ^{1,2,3,*} and Bin Sun ^{1,2,3} 

¹ Faculty of Land Resource Engineering, Kunming University of Science and Technology, Kunming 650093, China; baiguangshun@kust.edu.cn (G.B.); zhujieyong@kust.edu.cn (J.Z.); binsun@stu.kust.edu.cn (B.S.)

² Key Laboratory of Geohazard Forecast and Geocological Restoration in Plateau Mountainous Area, Ministry of Natural Resources of People's Republic of China (MNR), Kunming 650093, China

³ Key Laboratory of Geohazard Forecast and Geocological Restoration in Plateau Mountainous Area in Yunnan Province, Kunming 650093, China

⁴ Yunnan Gaozheng Geo-Exploration Co., Ltd., Kunming 650041, China; yangxuemeilj@foxmail.com

* Correspondence: zhigangkong@kust.edu.cn (Z.K.); taogezhang@hotmail.com (S.Z.)

Abstract: The nine plateau lake watersheds in Yunnan are important ecological security barriers in the southwest of China. The prevention and control of landslides are important considerations in the management of these watersheds. Taking the Dianchi Lake watershed as a typical research area, a comprehensive modeling and assessment process of landslide susceptibility was put forward. The comprehensive process was based on the weight of evidence (WoE) method, and many statistical techniques were integrated, such as cross-validation, multi-quantile cumulative Student's comprehensive weight statistics, independence testing, step-by-step modeling, ROC analysis, and ROC-based susceptibility zoning. In this paper, fourteen models with high accuracy and validity were established, and the AUC reached 0.83–0.87 and 0.85–0.88, respectively. In addition, according to the susceptibility zoning map compiled via the optimal model, 80% of landslides can be predicted in the very-high- and high-susceptibility areas, which only account for 19.58% of the study area. Finally, this paper puts forward strategies for geological disaster prevention and ecological restoration deployment.

Keywords: landslide susceptibility assessment; weight of evidence (WoE) method; modeling; Dianchi Lake watershed; Yunnan Plateau



Citation: Bai, G.; Yang, X.; Kong, Z.; Zhu, J.; Zhang, S.; Sun, B. Modeling and Assessment of Landslide Susceptibility of Dianchi Lake Watershed in Yunnan Plateau. *Sustainability* **2023**, *15*, 15221. <https://doi.org/10.3390/su152115221>

Academic Editors: Wen Zhang, Chen Cao, Jie Dou and Peihua Xu

Received: 24 August 2023

Revised: 11 October 2023

Accepted: 23 October 2023

Published: 24 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Yunnan Province is an important ecological security barrier in the southwest of China, and lakes are important ecological regions. The Yunnan Province government is promoting the ecological protection and restoration of nine plateau lake watersheds (Figure 1). The prevention of landslides is one of the important goals of these activities. In order to support this ongoing project, it is necessary to evaluate landslide susceptibility by taking the lake watershed as the unit, choosing the geological environment and human activity factors that may affect or control landslide susceptibility in the watershed, understanding the distribution of landslide susceptibility in the watershed, and guiding the formulation of targeted prevention and control countermeasures. Dianchi Lake is the largest of the nine plateau lakes in Yunnan Province, and Kunming City is in this watershed, where human engineering activities are relatively strong. The research on landslide susceptibility assessment (LSA) in the whole Dianchi Lake watershed is still relatively lacking. Generally speaking, it is reasonable to choose Dianchi Lake watershed as the study area.

Landslides in plateau mountainous areas will always be a problem because they affect people's lives, destroy the surface, and cause economic losses [1]. Identifying the dangerous areas related to landslides is an important part of disaster management [2],

and it is also an important foundation for promoting human safety, infrastructure development, and ecological environment protection [1]. LSA describes the spatial probability of landslides [3,4]. On the regional scale, the landslide susceptibility modeling method based on statistics is considered to be appropriate [2,5–7].

Statistical LSA is a supervised dichotomy problem, which can be solved via different classification methods [7]. About 163 different data-driven methods are applied to LSA [8], such as weight of evidence (WoE) [9,10], naïve Bayes (NB) [6,11], logistic regression (LR) [12–15], discriminant analysis (DA) [3,16], supported vector machines (SVM) [17,18], random forest (RF) [13,19–21], artificial neural networks (ANN) [22–24], and many others. These methods have their advantages and disadvantages. When dealing with sparse landslide datasets, a simple algorithm can usually provide better results [25]. In addition, the analysis should be kept as simple as possible so that we might obtain a deeper understanding of the effect when testing the new model [7].

In this study, we choose the WoE method, which is a moderately complicated data-driven method based on statistics [9,10]. It is a well-known and widely used statistical method that is used to estimate the relationship between observation data (landslide training inventory) and potential control factors (geological and geomorphological factors) [10,26]. It is widely used in landslide susceptibility mapping (LSM) [1,2,6,7,11,17,27–31] because it is easy to understand and has a strict mathematical foundation and theoretical system. Although WoE has been frequently used in LSA in recent decades [1,2,6,7,9,11,17,27–31], establishing how to optimize the modeling process to improve the accuracy and validation of the model is a problem worth exploring. Because WoE only uses discrete data, continuous raster data need to be classified [2,4,10]. However, there is no standardized factor data classification method. This is the other problem worth discussing. In addition, how to reduce the statistical errors caused by the randomness of landslides and related factors is also worth studying because landslides usually do not happen by accident [4,6,32].

This study focuses on LSA and LSM in Dianchi Lake watershed. It has outstanding application value, which aims to enhance our ability to assess the susceptibility of landslides and improve the corresponding consulting services for stakeholders involved in disaster reduction. Concerning the research content, on the one hand, the characteristics of landslide sensitive factors were clarified; on the other hand, the spatial distribution of landslide susceptibility was clarified, which provides important technical support for guiding ecological restoration and the deployment of landslide prevention and mitigation methods in plateau lake watersheds. Concerning the technical aspect, this paper puts forward a comprehensive process of LSA based on the WoE method, including (1) data preparation; (2) optimizing the compilation of datasets for factor classification based on a cumulative Student's comprehensive weight (sC) curve and WoE statistics; (3) screening modeling factors based on the cross-validation theory and AUC of single-factor analysis; (4) optimizing a high-performance model based on the step-by-step modeling; and (5) dividing landslide susceptibility areas based on ROC. This paper obtains the results of the LSM with excellent fitting and prediction performance (both AUC reached 0.87) based on the above process. The spatial distribution map of landslide susceptibility classification was compiled, and the strategies of geological disaster prevention were put forward.

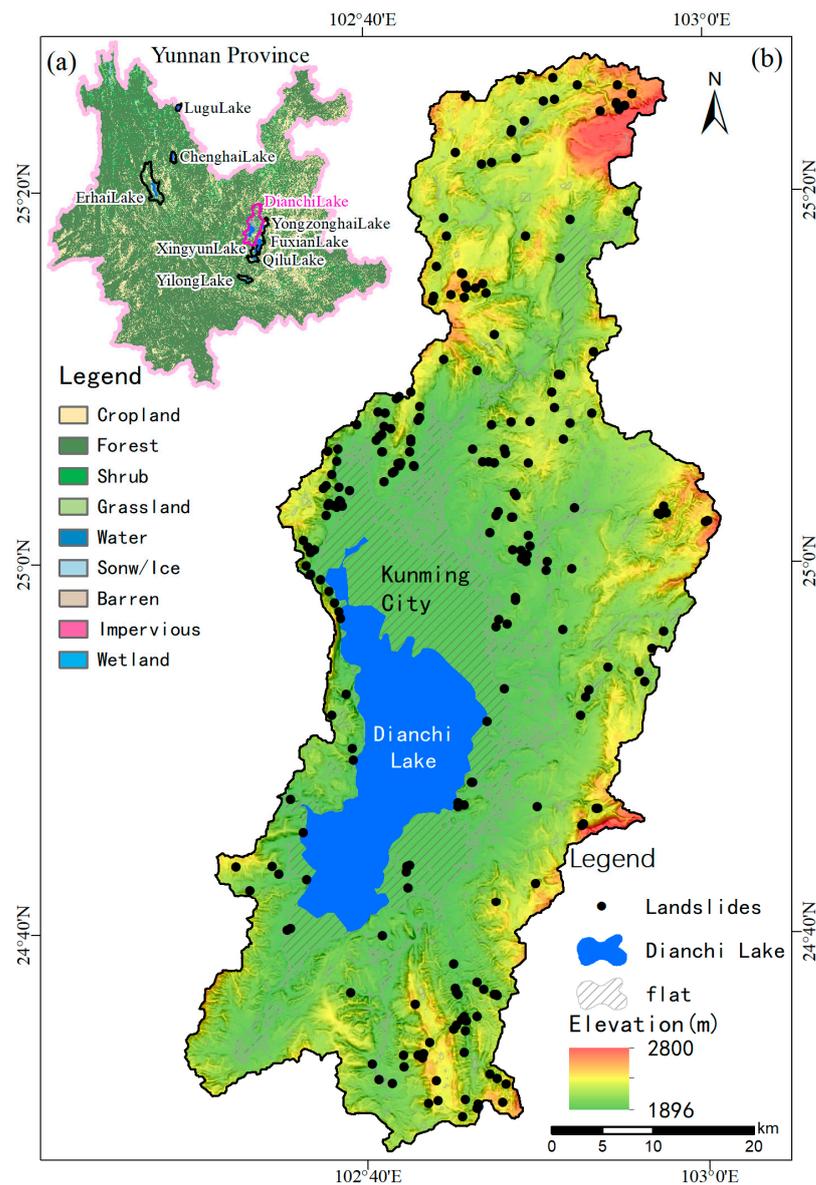


Figure 1. Study area. (a) The distribution of nine plateau lake watersheds in Yunnan, and the location of the Dianchi Lake watershed. The base map is the distribution map of land coverage types in Yunnan Province in 2020 [33]. (b) The distribution map of landslide points in Dianchi Lake watershed. The black points are landslides under investigation, the blue blocks are the water surface, and the gray diagonal lines are the areas with the attribute of “flat” [34,35]. The bottom picture is rendered via elevation and hill shade.

2. Study Area and Data

2.1. Study Area

The study area is the whole watershed of Dianchi Lake (Figure 1), covering an area of 2906.44 km². The climate belongs to the subtropical plateau monsoon climate, which is divided into a dry season and a rainy season. Most of the rainfall is between May and October, with an average annual rainfall of 1000 mm and an average temperature of 14.8 °C [36]. The water surface of Dianchi Lake and reservoir, the center of Kunming basin, the sub-basins, and some flat hilltops are not susceptible to landslides. Therefore, according to the result of the DEM classification [34,35], the actual analysis area in this paper is 2206.29 km² after deducting 700.15 km² of the “flat” category. This area is a lake basin and the mountainous terrain in the central part of Yunnan Plateau. The lake is

located in the southcentral region, rugged mountains are in the north, east, and south, and Xishan mountain, which has steep fault cliffs, is in the west. The elevation of this area ranges from 1896 m at the surface of Dianchi Lake to about 2800 m in the mountainous area, with a height difference of over 900 m. The steep mountainous terrain around the basin, the continuous and rapid river cutting, the heavy rainfall in the rainy season, and the man-made influence of downward slope cutting during road construction make this area prone to slope failure. According to the official regional geological survey report, the strata of different times were merged according to lithology. The loose gravel soil, sandstone, mudstone, shale, siltstone, basalt, limestone, and metamorphic rocks are distributed in the study area (Figure 2).

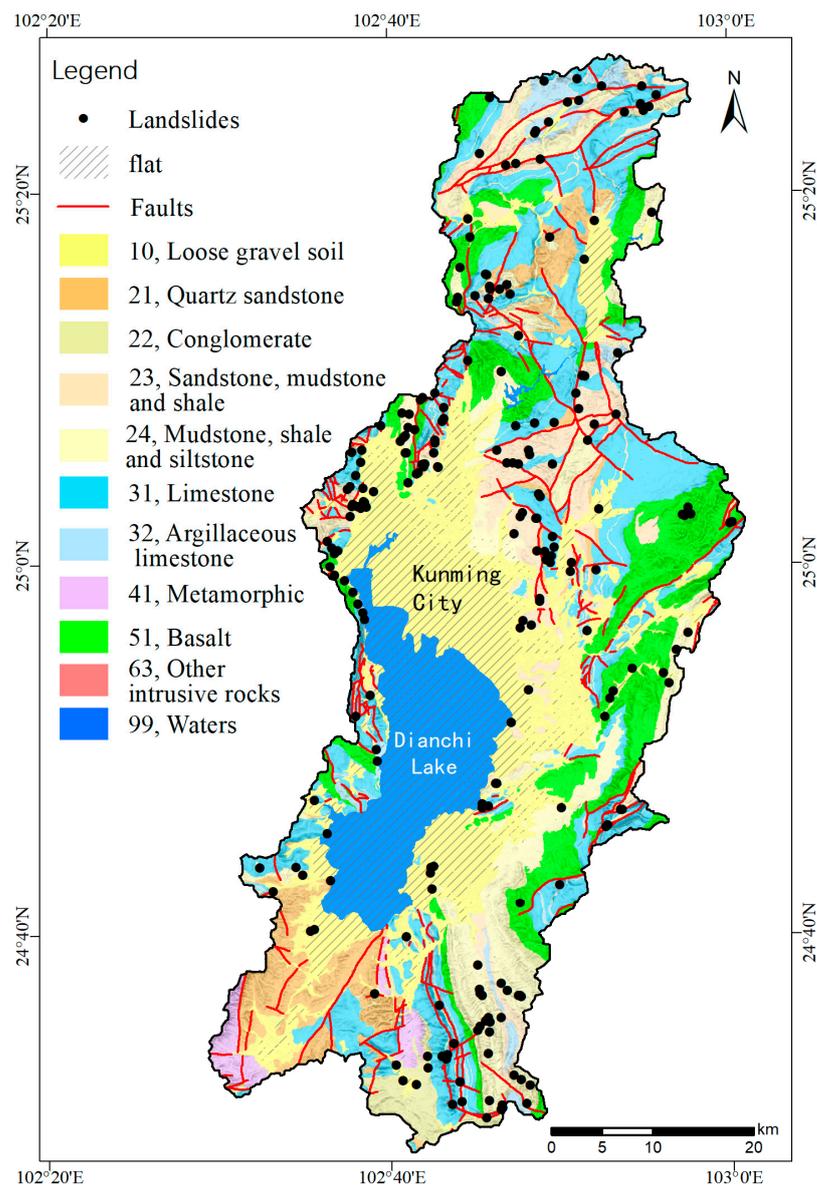


Figure 2. The geological map of lithology and faults.

2.2. Landslides and Data Preparation Based on Random Sampling

The study area has a good working history in landslide surveys, and it is the key monitoring and prevention area of landslides in Yunnan Province. Through field investigation, the historical landslides list was checked and revised, and a total of 228 landslides were included in the landslides list analyzed in this paper (Figure 1).

We adopted the cross-validation technique to prepare the data. The cross-validation technique is a basic technique employed to evaluate the uncertainty of statistical data and models using test datasets that do not involve model training [4,37,38]. Figure 3 briefly summarizes the compilation process of the landslide dataset. (1) We divided all landslide data (ALL) into a training dataset (TRN) containing 158 landslides and a test dataset (TST) containing 70 landslides using random sampling tools. TRN and TST are not duplicated, which account for about 70% and 30% of ALL, respectively. TRN is used to calibrate the model, and TST is used to evaluate the performance of the model. (2) To estimate the model variables that depend on the sample size, we used the random sampling tool to generate 100 random sub-samples with TST size from TRN, and some landslides in different random sub-samples were allowed to be repeated, forming a training data subset trn.

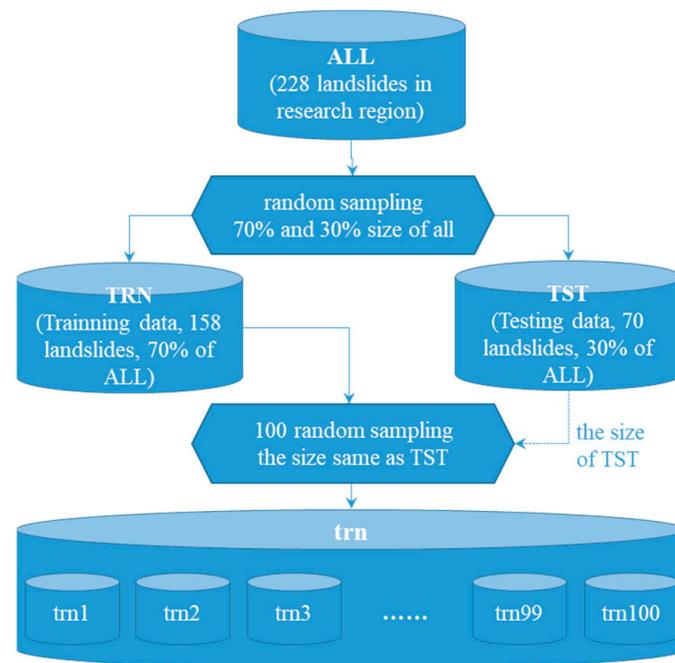


Figure 3. Process flowchart of cross-validation landslide dataset compilation based on random sampling.

2.3. Factor Data

According to the characteristics of the study area, the availability data, and previous research, landslide control factors can be roughly divided into different groups [6,11]. Table 1 lists the landslide control factors which have been compiled. At first, we did not rule out available or main factors that are easy to deduce, as these factors may help explain landslide susceptibility. These factors have been used in many other studies, and descriptions of these factors can be found in a large number of studies [2,11,19,39–42]; so, this article did not elaborate upon this further. In Table 1, we briefly explained the significance of these factors in LSA.

Table 1. Sources and significances of the factors.

No.	General Category	Factors	Significance	Source and Compilation Method
1	Geologic	Distance to faults (dF)	Destruction of the stability of the rock mass structure	The fault structural lines came from the 1:200,000 geological map of Kunming; using QGIS to compile Euclidean distance grid
2		Lithology (Lth)	Lithological types of slope rock and soil	1:200,000 geological map of Kunming

Table 1. Cont.

No.	General Category	Factors	Significance	Source and Compilation Method
3	Land cover	CLCD	The 30 m annual land cover dataset in China	The 30 m annual land cover dataset and its dynamics in China 2019 (CLCD) [33]
4		Land cover (LC)	The 10 m land cover	ESA WorldCover 10 m 2020 v100 [43]
5		Normalized difference vegetation index (NDVIlog)		China 30 m Annual NDVI Maximum Dataset (2021) [44] as the log value
6	Anthropogenic	Distance to roads (dRD)	Road cutting or vehicle vibration	Data come from OSM (OpenStreetMap, 2021); using QGIS to compile the Euclidean distance grid
7	Morphometric terrain parameters	Elevation (Elv)	Climate, vegetation, and potential energy	NASADEM [45], the resolution of which is ~30 m
8		Aspect (Asp)	Solar insolation, flora and fauna distribution and abundance [1]	Compilation using SAGA GIS via DEM [45]
9	Morphometric terrain parameters	Plan curvature (CPlan)	Converging, diverging flow, soil water content, and soil characteristics [1]	Compilation using SAGA GIS via DEM [45], with value $\times 10^6$
10		Profile curvature (CProf)	Flow acceleration, erosion/deposition, and geomorphology [1]	Compilation using SAGA GIS via DEM [45], with value $\times 10^6$
11		Tangential curvature (CTang)	Erosion/deposition [1]	Compilation using SAGA GIS via DEM [45], with value $\times 10^6$
12		Topographic Position Index (TPI)	Quantifies topographic heterogeneity and erosion [46]	Compilation using SAGA GIS via DEM [45]
13		Terrain Ruggedness Index (TRI)	Quantifies topographic heterogeneity and erosion [47]	Compilation using LSAT PM [4] via DEM [45]
14		Roughness (Rou)	Quantifies topographic heterogeneity and erosion	Compilation using LSAT PM [4] via DEM [45]
15		Relative slope position (RSP)		Compilation using LSAT PM [4] via DEM [45]
16		Slope (SL)	Stress field is related to slope	Compilation using SAGA GIS via DEM [45]
17	Water-related	Flow path length (FPL)	River erosion	Compilation using SAGA GIS via DEM [45]
18		Flow Accumulation (FALog)	Runoff velocity, runoff volume, and potential energy	Compilation using SAGA GIS via DEM [45] as the log value
19		Height above nearest drainage (HAND)	River erosion, runoff velocity, runoff volume, and potential energy [48,49]	Compilation using SAGA GIS via DEM [45]
20		Horizontal HAND (HANDH)	River erosion, runoff velocity, runoff volume, and potential energy [48,49]	Compilation using SAGA GIS via DEM [45]
21		Vertical HAND (HANDV)	River erosion, runoff velocity, runoff volume, and potential energy [48,49]	Compilation using SAGA GIS via DEM [45]
22		Distance to channel network (dCN)	River erosion.	Compilation using SAGA GIS via DEM [45]
23		Stream power index (SPIlog)	River erosion [50]	Compilation using SAGA GIS via DEM [45] as the log value
24		Topographic wetness index (TWI)	Moisture content of soil [50–52]	Compilation using SAGA GIS via DEM [45]
25		SAGA Wetness Index (TWSAGA)	Moisture content of soil [52,53]	Compilation using SAGA GIS via DEM [45]

3. Methods

3.1. Weights-of-Evidence Method (WoE)

The weights of a single factor are superimposed on the linear model to obtain the complete landslide sensitivity model [1,10,26,28]. WoE was first introduced in the late 1980s, and it was used for the application of geological science based on GIS, mainly for the mapping of mineral potential [10,26,54–56]. D is defined as the unit with landslides, \bar{D} as the unit without landslides, B as the unit in the evidence factor area, \bar{B} as the unit outside the evidence factor area, $P(\cdot|\cdot)$ as the conditional probability symbol, and $N(\cdot\cdot\cdot)$ as the grid pixels number. WoE considers two kinds of weights and posterior probability [2,6,10,11,26,55,56]:

$$W^+ = \ln \frac{P(B|D)}{P(B|\bar{D})} = \ln \left(\frac{N(B \cap D)}{N(B \cap D) + N(\bar{B} \cap D)} / \frac{N(B \cap \bar{D})}{N(B \cap \bar{D}) + N(\bar{B} \cap \bar{D})} \right), \quad (1)$$

$$W^- = \ln \frac{P(\bar{B}|D)}{P(\bar{B}|\bar{D})} = \ln \left(\frac{N(\bar{B} \cap D)}{N(B \cap D) + N(\bar{B} \cap D)} / \frac{N(\bar{B} \cap \bar{D})}{N(B \cap \bar{D}) + N(\bar{B} \cap \bar{D})} \right). \quad (2)$$

The weight symbols W^+ and W^- do not represent the mathematical meaning of numerical values; rather, they represent the presence (positive) and absence (negative) of feature classes in a given raster cell. According to the above formula, a positive logical value indicates the positive impact of a given variable, a negative logical value indicates the negative impact, and a logical value of zero represents no influence.

The posterior probability is an index of susceptibility, the higher numerical value means higher susceptibility, and a lower numerical value means lower susceptibility. The formula for calculating the posterior probability is as follows: $P = O / 1 + O = \exp(F) / (1 + \exp(F))$, $F = \sum_{i=0}^n W_i^{K(i)} + \ln O(D)$, $O(D) = N(D) / N(\bar{D})$, where $K(i)$ is “+” when the i -th evidence factor layer exists, and $K(i)$ is “-” when it does not exist; W_i is the weight of the existence or non-existence of the i -th evidential factor.

In order to evaluate the spatial correlation strength between single factors, landslide, and the performance of the model, the receiver operating characteristic curve (ROC) algorithm is used in this paper, which is a technique employed to visualize and evaluate the performance of the classifier by describing the ratio of the true positive rate (sensitivity) to the false positive rate (1-specificity) [57]. The area under the ROC curve (AUC) provides a quantitative index by which to compare the advantages and disadvantages. Generally speaking, the AUC is excellent when it is greater than 0.8, the AUC is good when it is 0.7–0.8, the AUC is moderate when it is 0.6–0.7, and the AUC is common when it is smaller than 0.6.

3.2. Main Analysis Process

In this paper, a comprehensive evaluation process of landslide susceptibility based on WoE is proposed, which mainly includes (Figure 4): (1) data preparation (see Sections 2.2 and 2.3 for details); (2) optimizing the compilation of datasets for factor classification (see Section 3.4 for details); (3) screening modeling factors; (4) step-by-step modeling to optimize high-performance model; and (5) dividing landslide susceptibility level zones based on the ROC of model.

3.3. WoE Statistical Process

Our WoE statistical process integrates cross-validation technology and traditional WoE statistical technology (Figure 5). It can solve the statistical error caused via the randomness of landslides and factors [4,6,32], and it has a deeper understanding than the traditional WoE statistics which only use all the landslide data. In this paper, trn (containing 100 subsets) was used for statistics. For each factor, the statistical process was repeated 100 times. We calculate the mean weight of each factor category (WoE_trn)

and its corresponding statistical values, such as variance and standard deviation. ROC is used to evaluate the classification ability of each factor for each statistical data point graphically. There are two advantages to this statistical process [6]: first, based on its estimated variance, it can better represent the general uncertainty of the sensitivity model; second, for classified data, it can determine whether the significance weight has accidental characteristics, or whether it can be reproduced from different random samples, which is more likely to be causality.

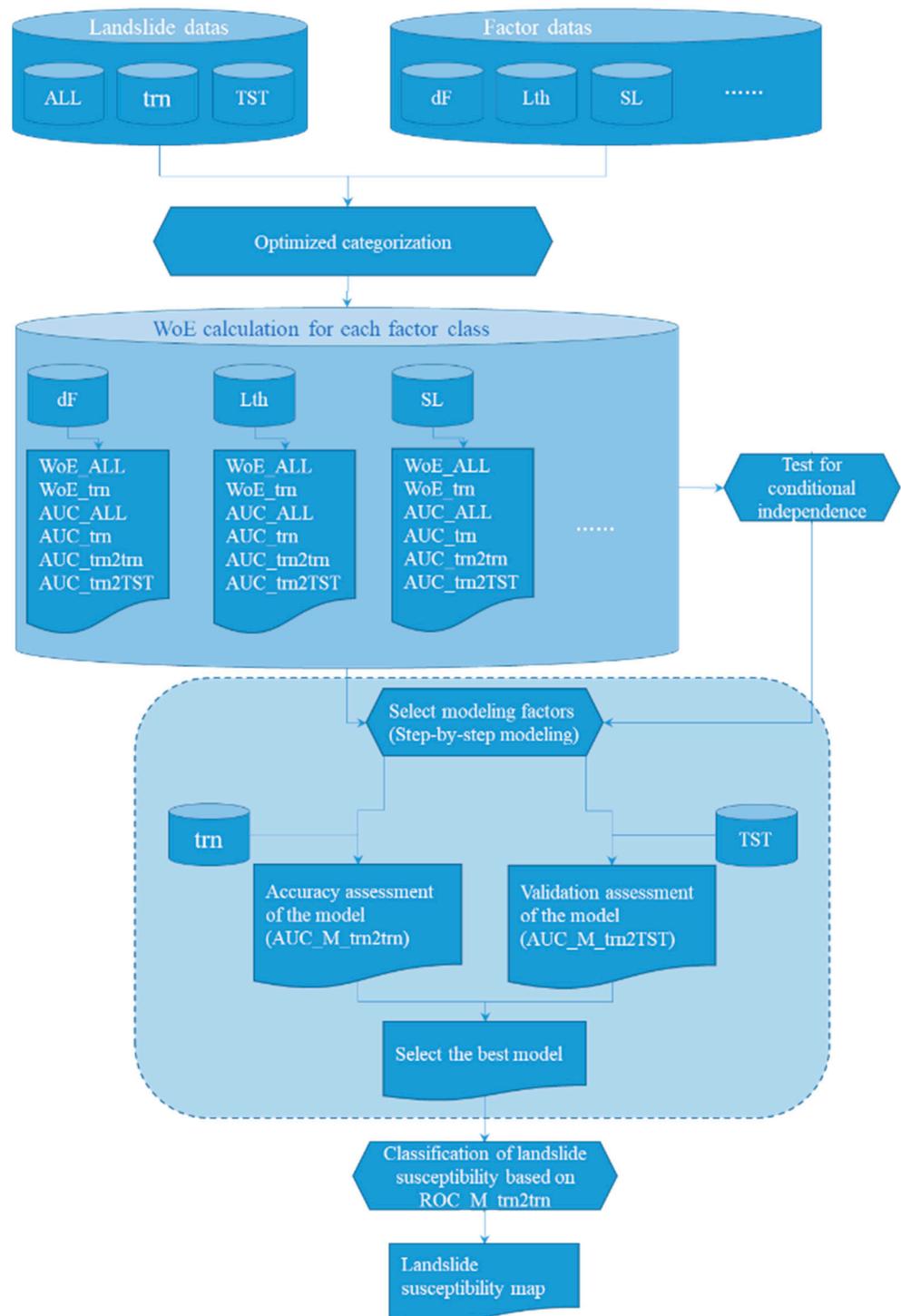


Figure 4. Flowchart of the improved WoE landslide susceptibility assessment.

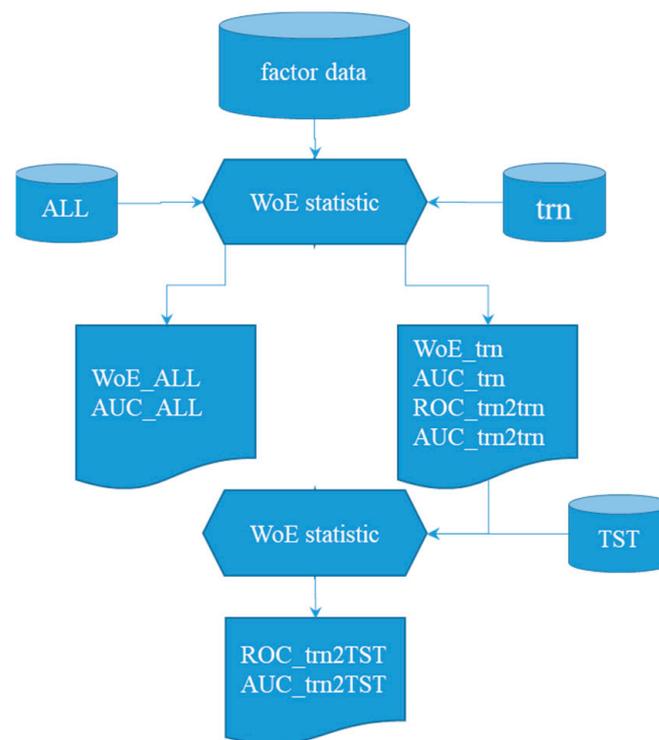


Figure 5. Process flowchart of single-factor WoE statistic. WoE_{ALL} , sC_{ALL} , and AUC_{ALL} are the weight, sC , and AUC calculated on ALL, respectively; WoE_{trn} , sC_{trn} , and AUC_{trn} are the mean weight, sC , and AUC calculated 100 times on trn , respectively; $ROC_{trn2trn}$ and $AUC_{trn2trn}$ are the single-factor accuracy assessment indexes modeled by single-factor weight WoE_{trn} and fit to trn ; and $ROC_{trn2TST}$ and $AUC_{trn2TST}$ are the single-factor validity assessment indexes modeled by single-factor weight WoE_{trn} and fit to TST.

The trn is used to evaluate the accuracy performance of the model, and the TST is used to evaluate the validation performance of the new data prediction model [7,31] (Figure 6). If the ROC curve based on the TST falls within the range of the ROC curve based on the trn (representing MSE), this shows that the accuracy and validation of the model can be good; otherwise, the model may be over-fitted.

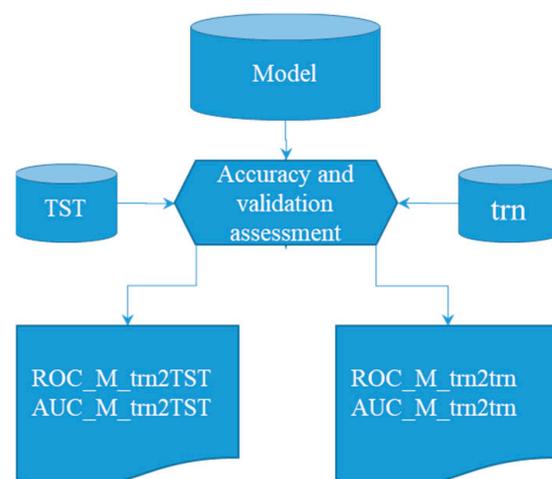


Figure 6. Process flowchart of accuracy and validation assessment of models.

3.4. Optimization Process of Single-Factor Categorization

Because WoE uses discrete data, it is necessary to classify continuous single-factor data discretely, which will lead to a discontinuity of factor weights. The determination of the traditional single-factor discrete classification number and classification threshold is subjective. This paper puts forward a single-factor classification optimization process (Figure 7), and the main steps are as follows.

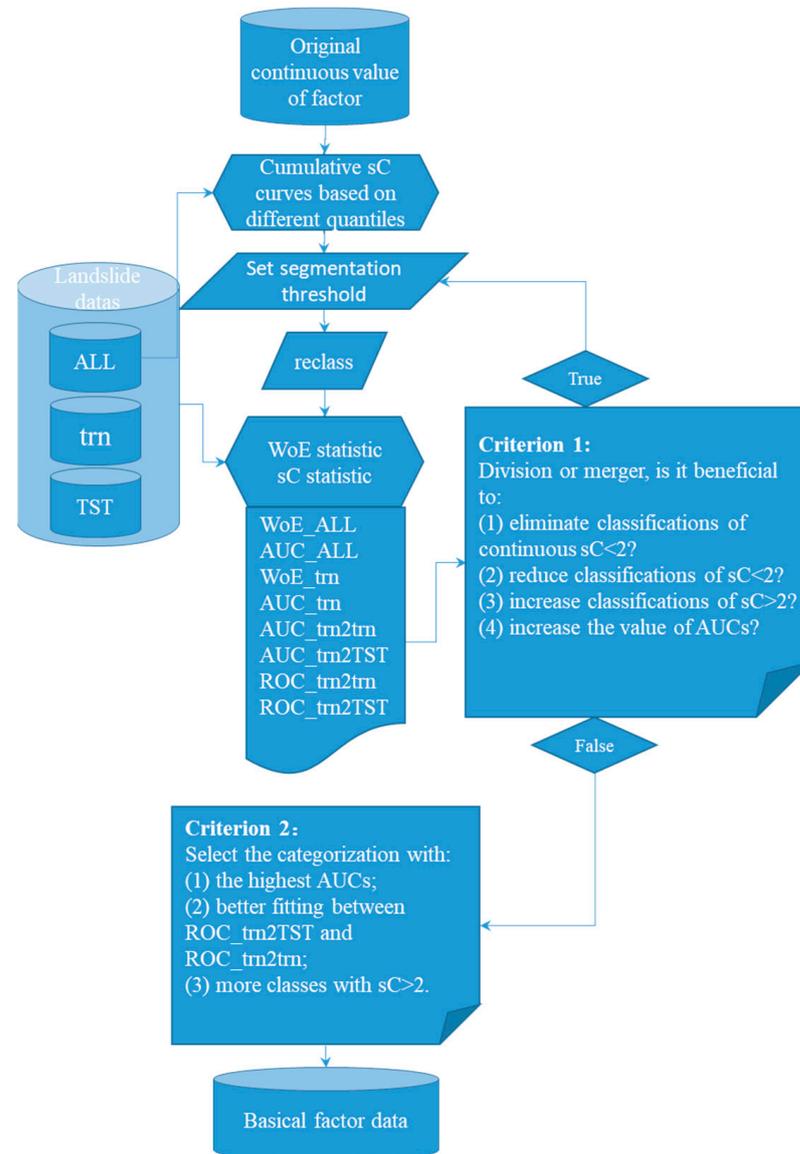


Figure 7. Process flowchart of factor classification optimization strategy based on the cumulative sC curve and WoE statistics.

First, a cumulative sC curve is generated. This method involves subdividing the continuous numerical single-factor raster into multiple classes according to the quantile and calculating the weight and corresponding variance for each class. The difference between the two weights—that is, the comprehensive weight—and the correlation between the quantitative factor and landslides are calculated as follows [26]: $C = W^+ - W^-$. A confidence measure defined via contrast divided by its standard deviation is introduced, which is similar to the Student's comprehensive weight (sC). The sC is relatively large when the standard deviation is small, so the results are more reliable. When the test values of sC are 1.96 and 2.326, confidence levels are 97.5% and 99%, respec-

tively [10,26,55]. $sC = C/\sigma_C = C/\sqrt{\sigma_{W^+}^2 + \sigma_{W^-}^2}$, $\sigma_{W^+}^2 = 1/N(B \cap D) + 1/N(B \cap \bar{D})$, $\sigma_{W^-}^2 = 1/N(\bar{B} \cap D) + 1/N(\bar{B} \cap \bar{D})$, where σ_C , σ_{W^+} , and σ_{W^-} are standard deviations of C , W^+ , and W^- , respectively. A new discrete distance category is defined using the accumulated sC [6]. As long as the weight value is positive, sC should be increased; when the weight is close to zero, it should be flat; when the weight is negative, it should be decreased. Therefore, the shape of the cumulative sC curve shows its maximum value at the position where it is expected to have the greatest influence. If there is more than one maximum value, this indicates the distortion effect of another variable [6].

Based on the cumulative sC curve, the classification and segmentation thresholds are set, the factors are reclassified, and the reclassified factor data are subjected to single-factor WoE statistics (Section 3.3).

Then, set a new trial segmentation threshold is obtained and we repeat the above steps.

Finally, we suggest determining the best classification according to two criteria (Criterion 1 and Criterion 2). Criterion 1: the division or merger beneficial to (1) eliminating classifications of $sC < 2$; (2) reducing classifications of $sC < 2$; (3) increasing classifications of $sC > 2$; or (4) increasing the value of AUCs. After several rounds of trial calculation, the optimal classification is determined according to Criterion 2: select the best categorization with (1) the highest AUCs; (2) the better fitting between ROC_trn2TST and ROC_trn2trn; and (3) the more classes with $sC > 2$.

4. Results

4.1. Cumulative sC Statistical Curve of Continuous Single Factor

ALL and six quantiles, namely, 100, 80, 60, 40, 20, and 10, were used to calculate the continuous numerical factors via the sub-process in Section 3.4. The statistical curve of cumulative sC (Figure 8) reveals the correlation between continuous numerical factors and the spatial distribution of landslides in different quantiles in detail, which not only reflect the changing trend of cumulative sC but also show the details of cumulative sC change.

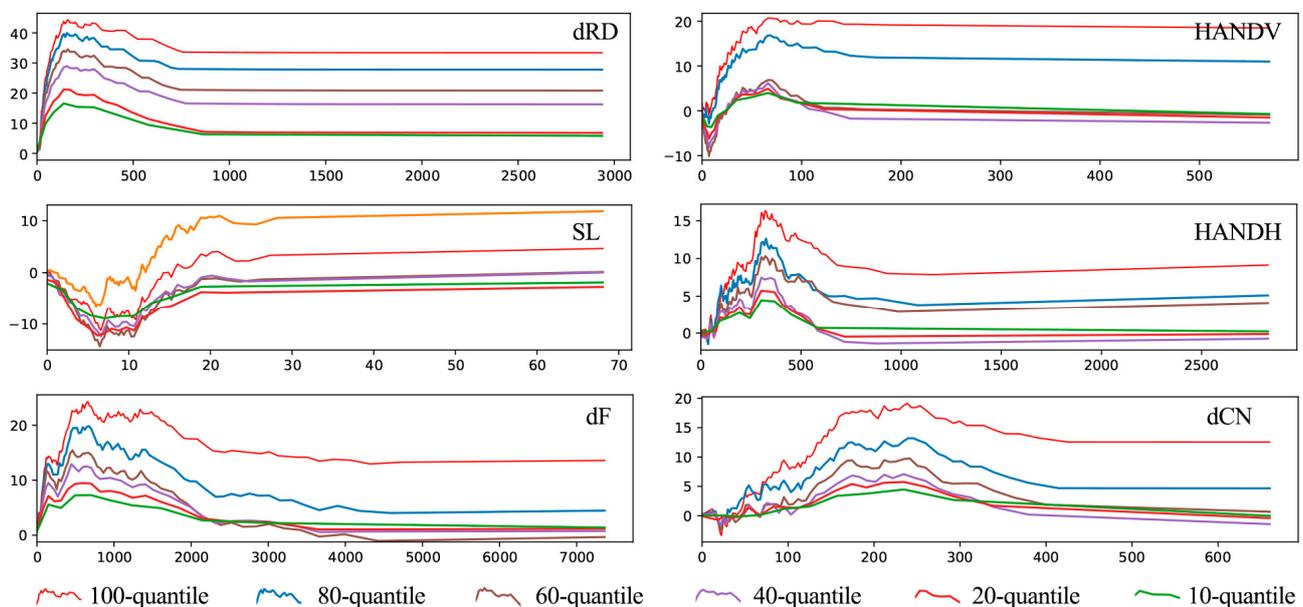


Figure 8. The cumulative sC statistical curve of six factors according to six quantiles of 100, 80, 60, 40, 20, and 10.

In Figure 8, the curves of dRD , SL , and dF on the left are simpler than those of $HANDV$, $HANDH$, and dCN on the right, and the secondary fluctuation is smaller, indicating that dRD , SL , and dF have a strong spatial correlation with landslides and are less affected by other factors. The segmentation value can be further extracted from the cumulative

trend of weight reflected by the slope and continuity of the curve. Specifically, for the dRD, the positive weight at 157.42 m becomes negative; the 0–157.42 m curve has a large rising slope, which indicates that this is the key segment of landslide susceptibility. For the SL, the 10.83–21.10° segment is the key slope gradient that can easily induce landslides. For the dF, the positions of 121 m and 460 m are the key positions with positive and negative weight changes; 0–121 m and 262–460 m are the rising sections of the curve, and the secondary fluctuation is small, which indicates that this is the key area in which to induce landslide. For the HANDV, the weight changes positively and negatively at 6.93 m and 66.60 m, respectively, which are the key dividing points. The section ranging between 6.93–66.60 m is the key section that affects the landslide susceptibility. There are many secondary fluctuations on the HANDH curve, indicating that the spatial correlation between this factor and landslide is affected by other factors. For the dCN, there are also many secondary fluctuations in the curve, but generally speaking, 22.33–174.57 m is a rising section with a large slope, which has a strong correlation with the occurrence of landslides.

4.2. Results of Single-Factor WoE Analysis

Landslides usually do not happen by accident; they are unevenly distributed in different factors and factor categorizations [2,32]. After implementing the technical processes in Sections 3.3 and 3.4, this paper obtained the evidence weight and sensitivity strength analysis results of each factor (Figures 9–18).

(1) Geological Factors

The dF is divided into five categories. Figure 9 shows that many landslides have been found in class 1 and class 3, and the positive weight is very high. As the distance increases, it becomes less and less easy to slide. The third diagram shows the error distribution caused by spatial random effects and also reveals the stability of positive weights for class 1 and class 3. This result reflects that faults make joints and cracks in the nearby rock mass develop, which is conducive to the occurrence of landslides. The position 460 m away from the fault is a key demarcation point, and an area less than 460 m is especially conducive to landslides. The spatial correlation between dF and landslides is moderate, with the mean and the range of the AUC_trn being 0.63 and 0.58–0.68 (the fourth picture).

Rock strata is divided into five categories. Most landslides occurred in the class 24 (mudstone, shale, and siltstone) and class 23 (sandstone, mudstone, and siltstone) categories. The statistical results of cross-validation technology (the third figure) helped us further confirm the distribution of positive weight. The spatial correlation between Ltd. and landslide is moderate, with the mean and the range of the AUC_trn being 0.63 and 0.58–0.70 (the fourth picture).

(2) Land Cover Factors

Landslides do not easily occur in forest areas because the roots of trees reinforce the slopes. The statistical results of NDVIlog and CLCD factors obtained the same understanding. Among these two factors, the areas with low NDVIlog value (<3.81) and grassland are the most prone to landslides. The spatial correlation between NDVIlog and landslide is high, with the mean and the range of the AUC_trn being 0.66 and 0.61–0.71 (the fourth picture).

(3) Anthropogenic Factors

Slope cutting in road construction and vehicle vibration lead to landslides. According to the cumulative comparative weight analysis, dRD was classified into 9 classes (Figure 13). The first and third pictures show that the area of <157.42 m is prone to landslides. The spatial correlation between dRD and landslide is very high, with the mean and the range of the AUC_trn being 0.71 and 0.68–0.75 (the fourth picture).

(4) Morpho-metric Terrain Parameters

The statistical results show that the spatial correlations between SL (Figure 14), RSP (Figure 15), TRI (Figure 16), and Rou (Figure 17) and landslides are moderately high, while the spatial correlation between CProf (Figure 18) and landslides is moderately low. The results of SL show that the weights of class 5 (10.83–11.65°), class 10 (25.60–28.27°), and class 11 (28.27–39.98°) are very high, and landslides are prone to occur in these areas.

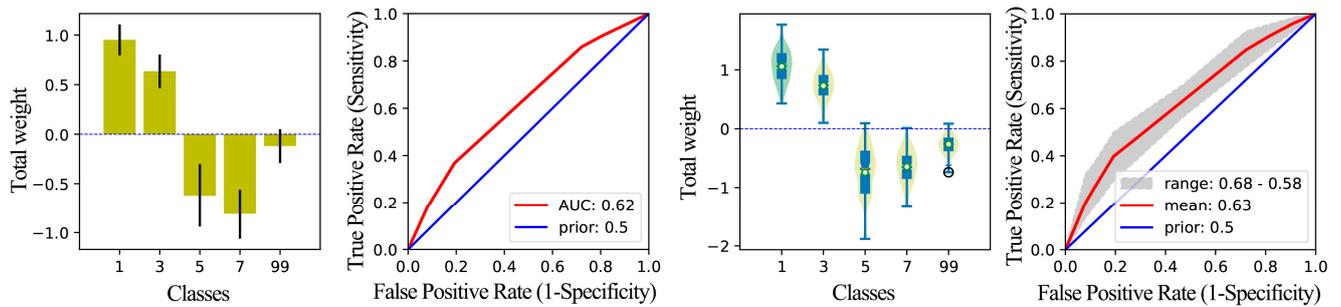


Figure 9. Graphical result of WoE for the factor dF. Class 1 is 0–121 m; class 3 is 262–460 m; class 5 is 657–864 m; class 7 is 1355–2317 m; and class 99 is other ranges. The first picture is the C histogram of factor classification based on statistics of ALL, and the black vertical line is the error bar of C. The second picture presents the ROC_ALL and AUC_ALL based on statistics of ALL. The third picture is a C violin-box diagram of factor classification based on statistics of trn with 100 subsets. The fourth picture presents the ROC and AUC, which have been counted 100 times based on trn with 100 subsets, where the red line is the mean ROC and the gray band is the ROC range.

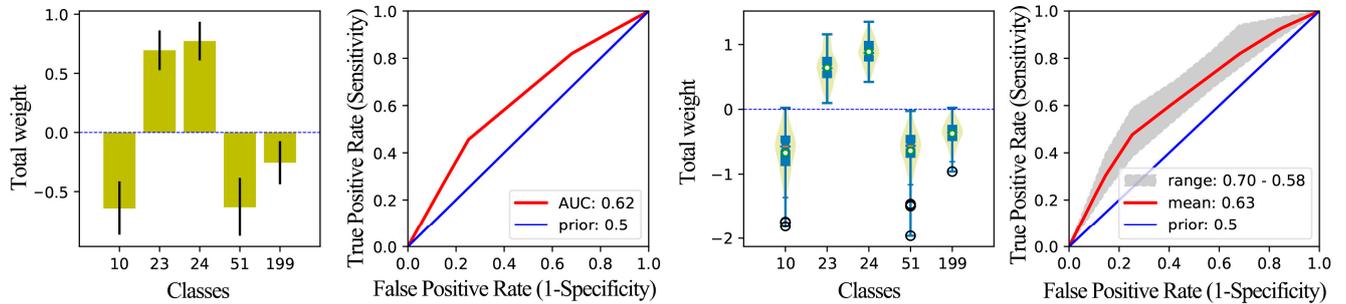


Figure 10. Graphical result of WoE for the factor Lth. Class 10 is loose gravel soil; class 23 is sandstone, mudstone, and shale; class 24 is mudstone, shale, and siltstone; class 51 is basalt; and class 199 is other lithologic strata, including limestone and metamorphic rocks.

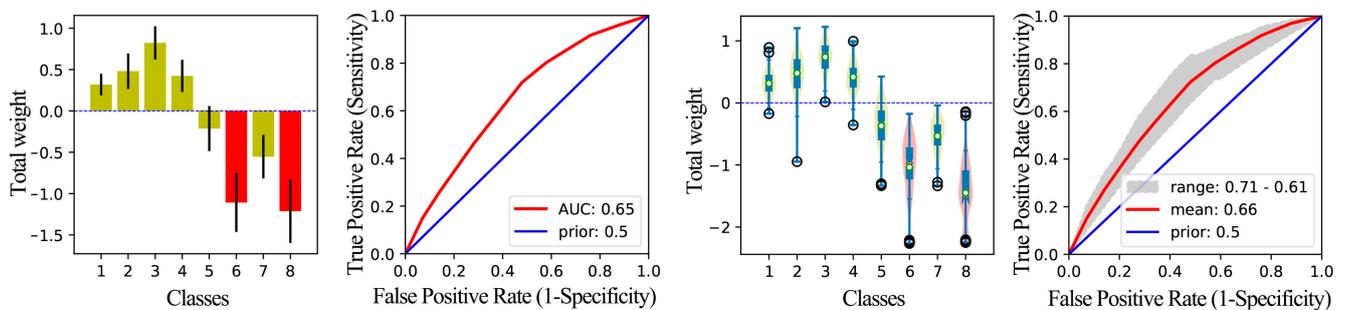


Figure 11. Graphical result of WoE for the factor NDVIIog. Class 1 is 2.79–3.64; class 2 is 3.64–3.71; class 3 is 3.71–3.76; class 4 is 3.76–3.81; class 5 is 3.81–3.84; class 6 is 3.84–3.85; class 7 is 3.85–3.88; and class 8 is 3.88–3.99.

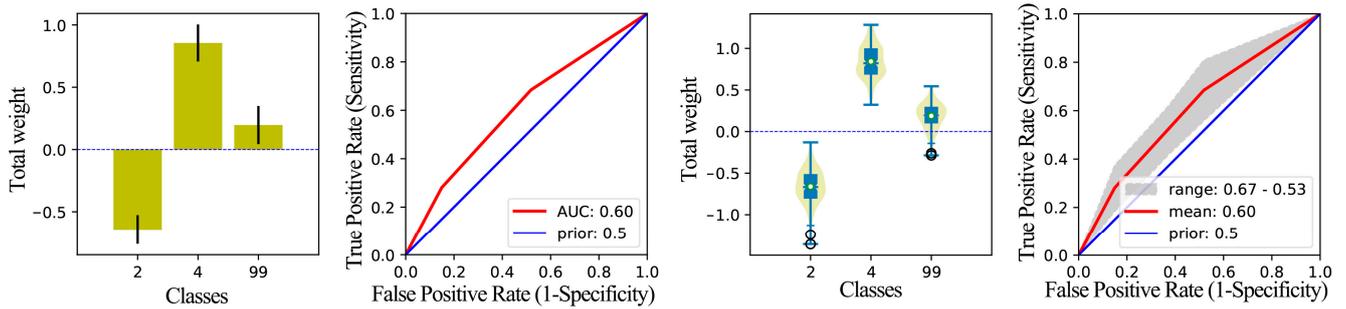


Figure 12. Graphical result of WoE for the factor CLCD. Class 2 is forest; class 4 is grassland; and class 99 is others (cropland, shrub, barren, impervious, wetland).

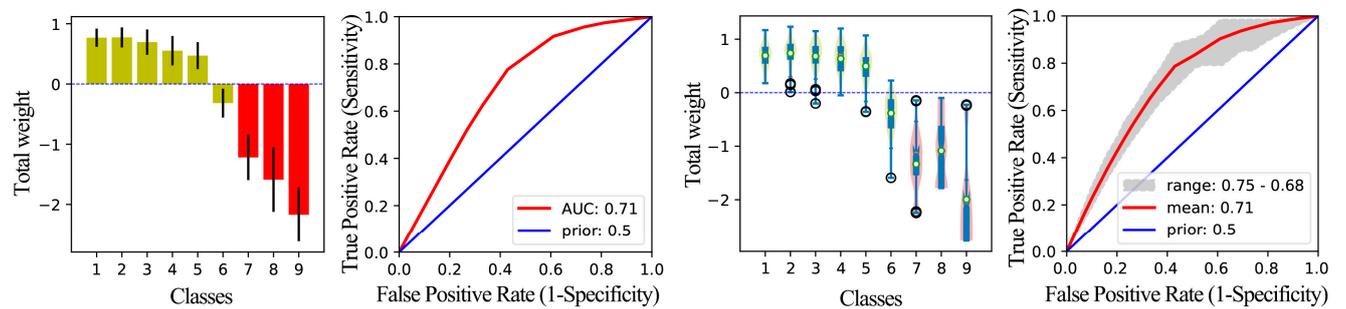


Figure 13. Graphical result of WoE for the factor dRD. Class 1 is 0–22.81 m; class 2 is 22.81–44.56 m; class 3 is 44.56–71.39 m; class 4 is 71.39–99.68 m; class 5 is 99.68–157.42 m; class 6 is 157.42–306.85 m; class 7 is 306.85–458.95 m; class 8 is 458.95–602.39 m; and class 9 is 602.39–2936.07 m.

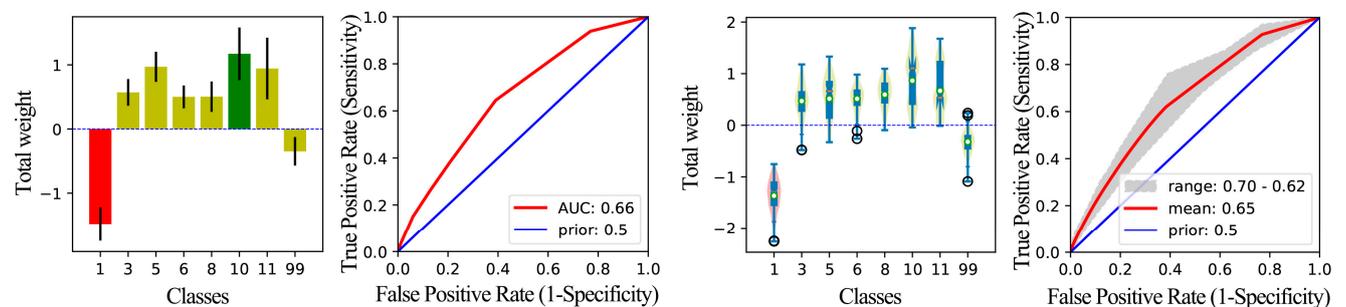


Figure 14. Graphical result of WoE for the factor SL. Class 1 is 0–4.12°; class 3 is 6.44–7.65°; class 5 is 10.83–11.65°; class 6 is 11.65–16.13°; class 8 is 17.12–21.10°; class 10 is 25.60–28.27°; class 11 is 28.27–39.98°; and class 99 is other slopes.

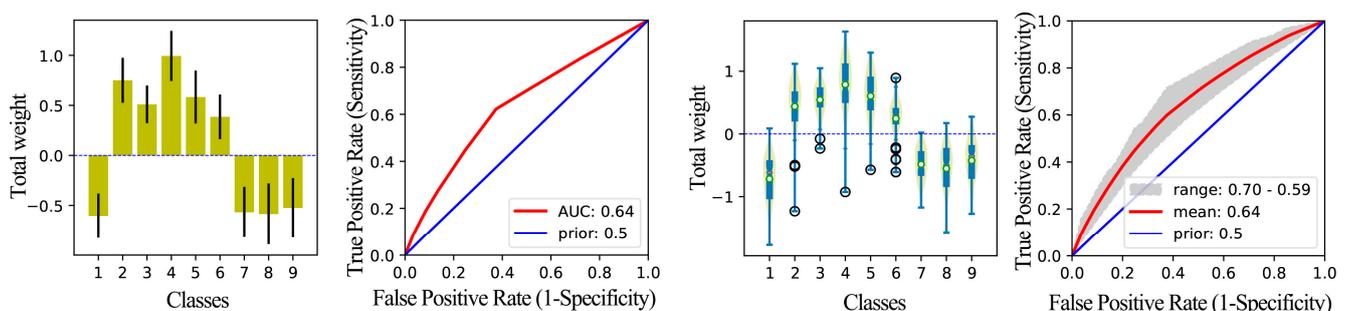


Figure 15. Graphical result of WoE for the factor RSP. Class 1 is 0–0.01; class 2 is 0.01–0.02; class 3 is 0.02–0.05; class 4 is 0.05–0.06; class 5 is 0.06–0.08; class 6 is 0.08–0.14; class 7 is 0.14–0.29; class 8 is 0.29–0.45; and class 9 is 0.45–1.02.

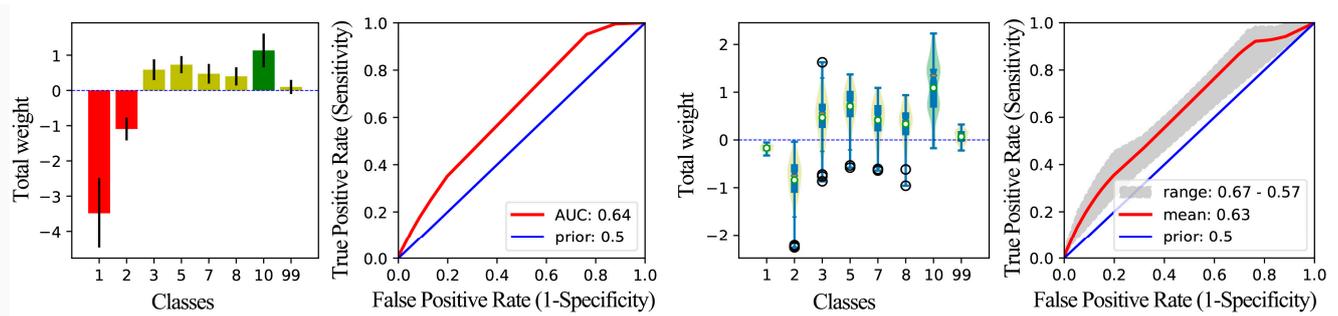


Figure 16. Graphical result of WoE for the factor TRI. Class 1 is 0.00–11.58 m; class 2 is 11.58–20.62 m; class 3 is 20.62–22.98 m; class 5 is 41.98–45.47 m; class 7 is 48.89–52.50 m; class 8 is 52.50–58.39 m; class 10 is 112.52–125.38 m; and class 99 is others in the range of 0–447.60 m.

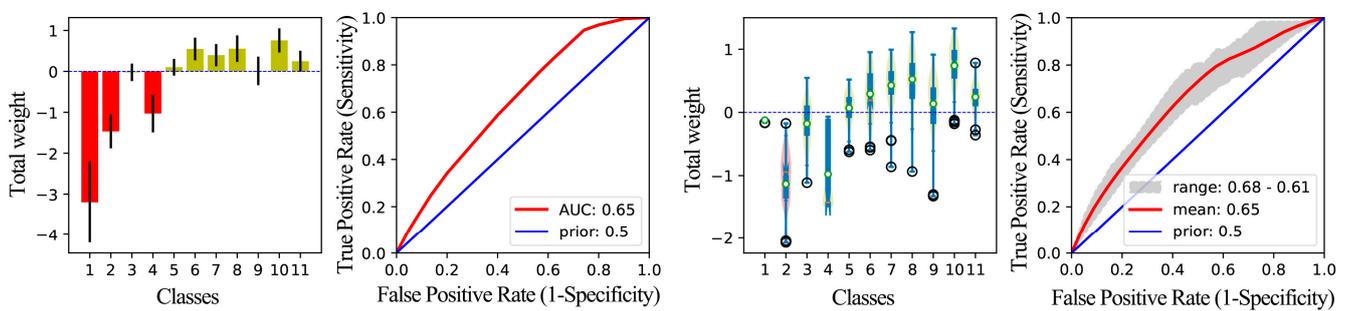


Figure 17. Graphical result of WoE for the factor Rou. Class 1 is 0.00–8.93; class 2 is 8.93–16.53; class 3 is 16.53–24.95; class 4 is 24.95–28.88; class 5 is 28.88–40.73; class 6 is 40.73–44.33; class 7 is 44.33–49.50; class 8 is 49.50–52.52; class 9 is 52.52–57.22; class 10 is 57.22–62.32; and class 11 is 62.32–398.73.

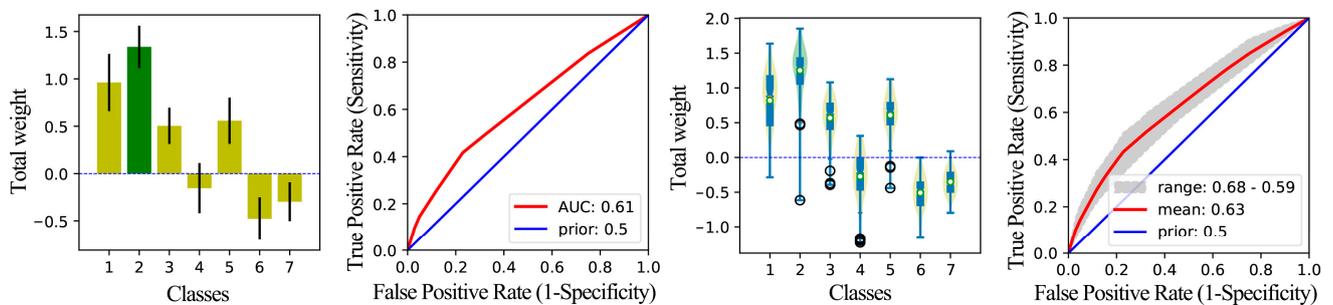


Figure 18. Graphical result of WoE for the factor Cprof. Class 1 is $-12,611.46 \sim -4084.50 (\times 10^{-6})$; class 2 is $-4084.50 \sim -2981.60 (\times 10^{-6})$; class 3 is $-2981.60 \sim -1533.30 (\times 10^{-6})$; class 4 is $-1533.30 \sim -973.62 (\times 10^{-6})$; class 5 is $-973.62 \sim -686.55 (\times 10^{-6})$; class 6 is $-686.55 \sim -37.07 (\times 10^{-6})$; and class 7 is $37.07 \sim 10596.92 (\times 10^{-6})$.

(5) Water-related Factors

The spatial correlation between HANDV and landslide is moderately high (Figure 19), while the spatial correlation between HANDH (Figure 20) and dCN (Figure 21) and landslide is moderate.

The results (Figures 9–21) show 13 factors with $AUC \geq 0.6$ (Figure 22). AUC reflects the strength of spatial correlation between factors and landslides. From the AUC of each factor in the Figure 22, we can see the difference of spatial correlation strength. The order of these factors from high to low is as follows: dRD, HANDV, NDVIIog, SL, RSP, TRI, Rou, Lth, dF, HANDH, Cprof, dCN, and CLCD.

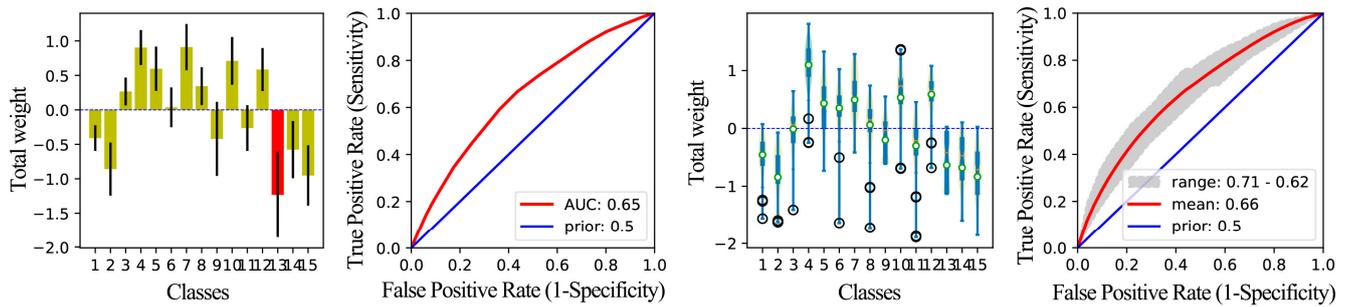


Figure 19. Graphical result of WoE for the factor HANDV. Class 1~class 15 are divided by 0 m, 4.15 m, 6.93 m, 13.03 m, 15.61 m, 17.89 m, 24.11 m, 26.22 m, 34.53 m, 37.77 m, 41.57 m, 55.48 m, 66.60 m, 77.37 m, 101.59 m, and 570.01 m.

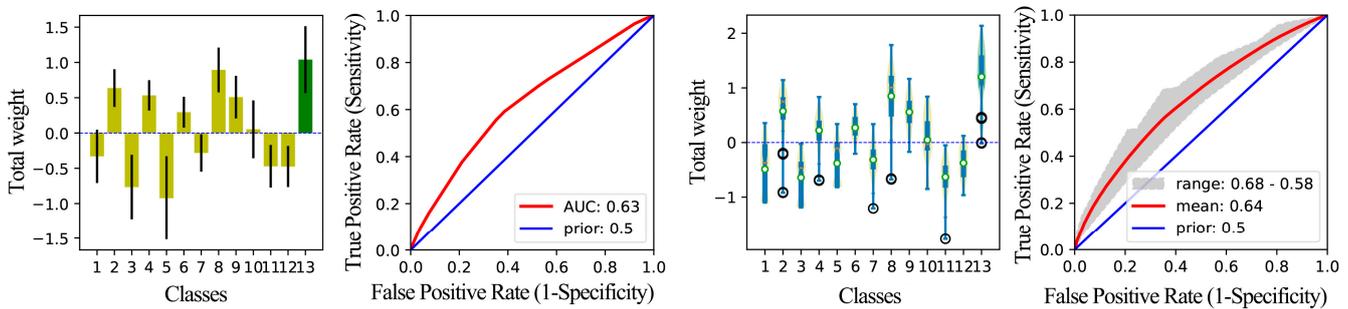


Figure 20. Graphical result of WoE for the factor HANDH. Class 1~class 13 are divided by 0 m, 38.06 m, 49.60 m, 65.22 m, 100.45 m, 115.44 m, 184.98 m, 1255.91 m, 271.86 m, 302.28 m, 323.25 m, 439.08 m, 1176.82 m, and 2831.14 m.

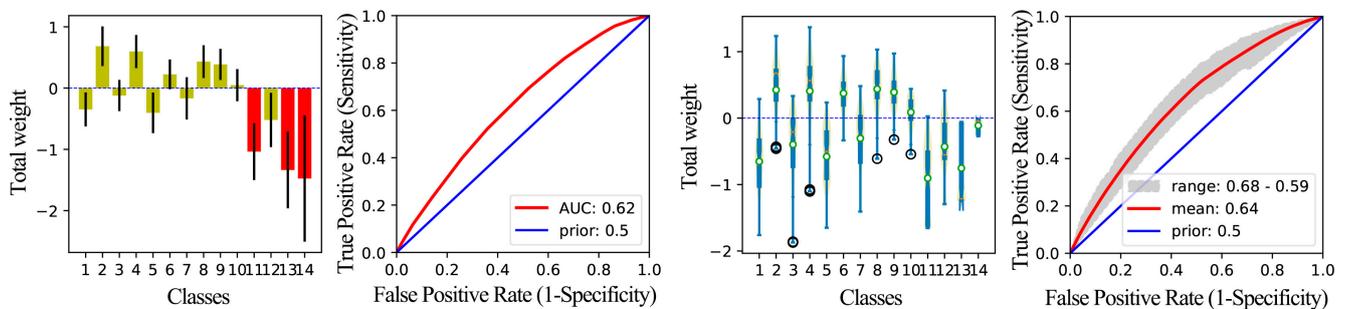


Figure 21. Graphical result of WoE for the factor dCN. Class 1~class 14 are divided by 0 m, 22.33 m, 24.98 m, 40.21 m, 49.85 m, 67.45 m, 94.96 m, 113.16 m, 134.62 m, 174.57 m, 240.09 m, 279.41 m, 320.53 m, 394.72 m, and more than 394.72 m.

We also analyzed the categorizations of factors with $W^+ \geq 0.6$ (Figure 23), which may be the key part of landslide susceptibility. Figure 23 shows the comparison of the spatial correlation between different factor categorizations and landslides.

4.3. Test Results for Conditional Independence

Strong correlation datasets may lead to incorrect estimations of factor contribution and expansion of the estimated probability value [58]. Chi-square-based contingency analysis is performed on the classified data based on the raster [4,11], according to Pearson’s C and Cramer’s V, to measure the correlation between discrete datasets.

Factor	AUC_trn2trn	AUC_trn2TST	AUC_trn
dRD	0.71	0.70	0.71
HANDV	0.65	0.65	0.66
NDVIlog	0.65	0.64	0.66
SL	0.64	0.68	0.65
RSP	0.63	0.66	0.64
TRI	0.63	0.64	0.63
Rou	0.63	0.62	0.65
HANDH	0.62	0.61	0.64
Lth	0.63	0.60	0.63
dF	0.63	0.60	0.63
dCN	0.62	0.60	0.64
CProf	0.62	0.61	0.63
CLCD	0.60	0.61	0.60

Figure 22. Thirteen factors with AUCs ≥ 0.6 and their AUC values.

Factor	classID	class	W+_ALL	W+_trn	sC_trn
dRD	1	0-22.81m	0.671	0.610	430.9
dRD	2	22.81-44.56m	0.673	0.643	29.2
dRD	3	44.56-71.39m	0.639	0.629	51.0
HANDV	4	13.03-15.61m	0.847	1.020	70.7
HANDV	7	24.11-26.22m	0.876	0.480	9.5
HANDV	10	37.77-41.57m	0.679	0.509	72.2
NDVIlog	3	3.71-3.76	0.751	0.681	45.3
SL	5	10.83-11.65°	0.873	0.458	31.0
SL	10	25.60-28.27°	1.115	0.826	32.2
SL	11	28.27-39.98°	0.893	0.636	6.6
RSP	4	0.05-0.06	0.926	0.726	14.7
TRI	5	41.98-45.47m	0.723	0.677	58.2
TRI	10	112.52-125.38m	1.164	1.122	32.4
Rou	10	57.22-62.32m	0.717	0.701	24.8
HANDH	8	255.91-271.86m	0.847	0.801	16.5
HANDH	13	1176.82-2831.14m	1.016	1.173	31.9
Lth	24	Mudstone, shale and siltstone	0.604	0.680	61.5
dF	1	0-121.25m	0.808	0.870	62.7
dCN	2	22.33-24.98m	0.672	0.414	5.2
CProf	1	-12611.46--4084.5 ($\times 10^{-6} \text{ m}^{-1}$)	0.873	0.744	196.1
CProf	2	-4084.50--2981.60 ($\times 10^{-6} \text{ m}^{-1}$)	1.207	1.120	78.9
CLCD	4	Grassland	0.633	0.625	193.8

Figure 23. Factor classification with $W^+ \geq 0.6$.

Figure 24 combines the statistical results of two correlation indexes, Pearson’s C and Cramer’s V, which are located in the upper right half and the lower left half of the heat map, respectively. The results show that according to Pearson’s C index, Rou, and TRI (0.81) and Rou and SL (0.71) are strongly related factor pairs. However, according to Cramer’s V, the correlation among the factors involved in statistics is not strong (<0.60). The correlation between dF, HANDH, and dCN and all other factors is very low. Elevation and its derived TRI, Rou, RSP, and SL have a slight relationship.

	dF	Lth	NDVlog	dRD	TRI	Rou	RSP	SL	HANDH	dCN	HANDV	CLCD	CProf	DEM
dF	1.00	0.22	0.08	0.08	0.12	0.13	0.13	0.10	0.07	0.07	0.09	0.08	0.09	0.14
Lth	0.11	1.00	0.45	0.40	0.54	0.55	0.53	0.48	0.15	0.24	0.48	0.38	0.24	0.52
NDVlog	0.04	0.25	1.00	0.47	0.51	0.55	0.52	0.48	0.13	0.21	0.46	0.53	0.23	0.53
dRD	0.04	0.22	0.20	1.00	0.39	0.42	0.46	0.37	0.14	0.20	0.39	0.31	0.16	0.44
TRI	0.06	0.32	0.22	0.16	1.00	0.81	0.56	0.67	0.16	0.26	0.56	0.34	0.15	0.49
Rou	0.06	0.33	0.25	0.16	0.53	1.00	0.60	0.71	0.20	0.29	0.61	0.39	0.14	0.52
RSP	0.07	0.31	0.23	0.18	0.26	0.26	1.00	0.54	0.32	0.42	0.67	0.37	0.12	0.69
SL	0.05	0.27	0.20	0.15	0.34	0.38	0.24	1.00	0.16	0.26	0.55	0.33	0.11	0.23
HANDH	0.04	0.08	0.05	0.05	0.06	0.06	0.12	0.06	1.00	0.82	0.65	0.09	0.24	0.09
dCN	0.03	0.12	0.08	0.07	0.10	0.10	0.16	0.10	0.41	1.00	0.68	0.14	0.20	0.12
HANDV	0.04	0.27	0.19	0.15	0.25	0.25	0.32	0.25	0.25	0.26	1.00	0.33	0.20	0.26
CLCD	0.05	0.29	0.60	0.41	0.43	0.49	0.46	0.42	0.13	0.19	0.42	1.00	0.19	0.45
CProf	0.04	0.13	0.09	0.07	0.35	0.33	0.28	0.26	0.50	0.45	0.45	0.14	1.00	0.24
DEM	0.07	0.30	0.28	0.22	0.25	0.27	0.43	0.45	0.20	0.26	0.51	0.36	0.11	1.00

Figure 24. Test results for conditional dependence. The upper right half represents the Pearson’s C results, and the factors with a strong correlation indicated by >0.7 are designated by black circles, such as Rou and TRI (0.81), Rou and SL (0.71), and dCN and HANDH (0.82). The lower left presents the Cramer’s V results.

4.4. Step-by-Step Modeling Results of Landslide Susceptibility

According to AUCs and conditional dependencies, factors are sorted and combined. The model M6 is based on the combination of factors with high AUCs. Then, we try to add follow-up factors into the new model in turn, and evaluate the fitting performance, uncertainty, and the prediction performance of the new model via ROC_M and AUC_M. We discard factors that cannot improve the AUC_M or improve the consistency of ROC_M.

As shown in Figures 25 and 26, the success rate of the model M11 is represented by the ROC calculated via trn, and its AUC is ~0.87. The model M11 is the best model. The AUC of the prediction rate calculated via TST was ~0.87 too. Both of them are high, being within the range of excellent classification models. The results also show that the M11 has excellent fitting and prediction performance and has not been over-fitted.

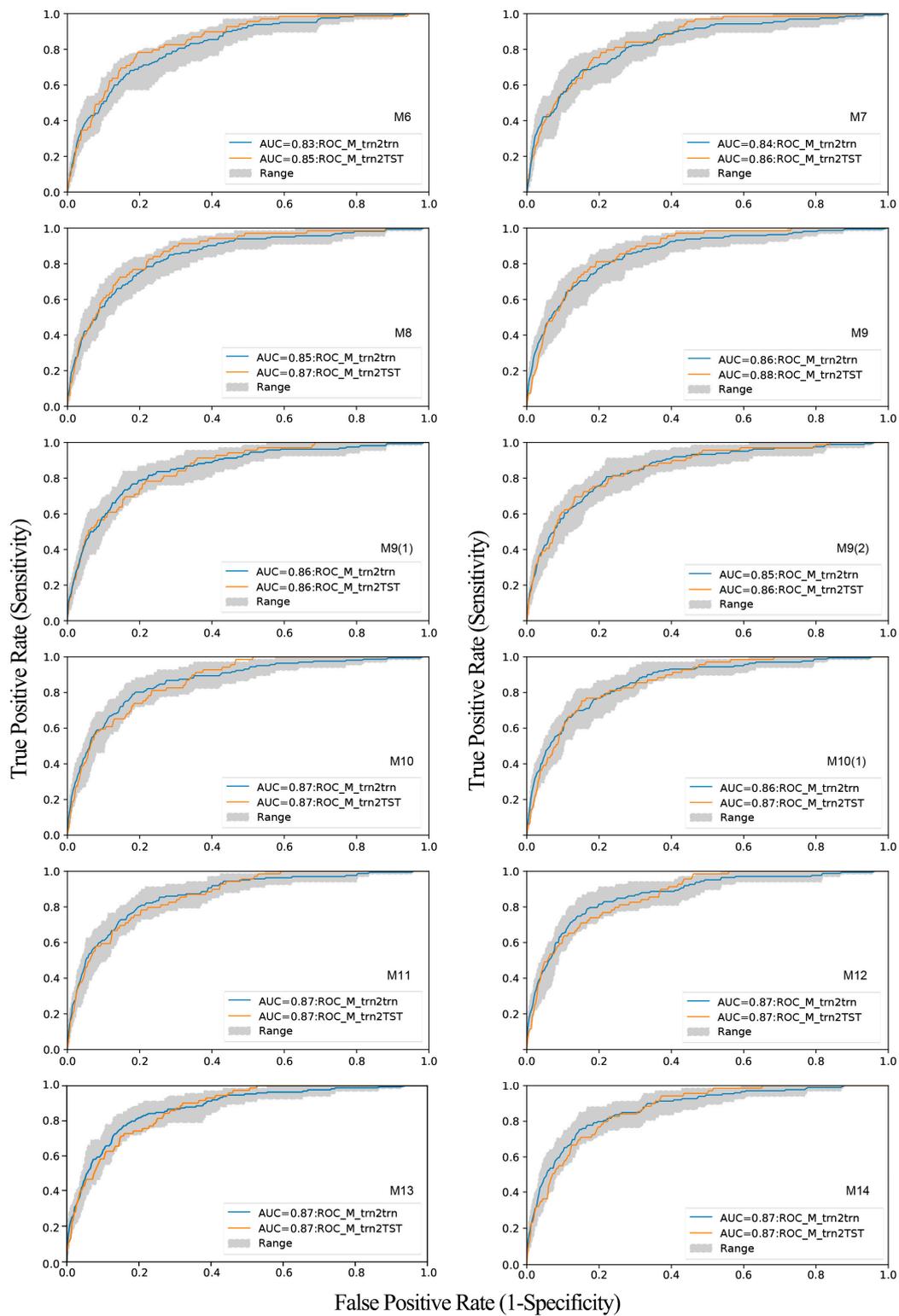


Figure 25. Accuracy and validity assessment of the models. Accuracy assessment of the models of susceptibility to landslides with the ROC_trn2trn of models (the blue line and the grey range). The total weights for the models were based on trn, and the performance of the models was evaluated using trn. One hundred iterations were carried out. The blue line is the mean ROC_M of 100 iterations. The grey range marks the model uncertainty based on the ROCs' MSE for 100 iterations. Test of validity of the models with the ROC_M_trn2TST (the orange line). The total weight maps were based on trn, and the validation was assessed using TST.

Models	AUC_trn2TST	AUC_trn2trn	ROC fitting score	ROC fitting assessment	Fctors
M14	0.87	0.87	10	Good	dF+Lth+CLCD+NDVllog+dRD+CProf+TRI+Rou+RSP+DEM_+SL+HANDH+dCN+HANDV
M13	0.87	0.87	9	Slightly larger in the middle and right	dF+Lth+NDVllog+dRD+CProf+TRI+Rou+RSP+SL+HANDH+dCN+HANDV
M12	0.87	0.87	9	Slightly larger on the left and right	dF+Lth+NDVllog+dRD+CProf+TRI+Rou+RSP+SL+HANDH+dCN+HANDV
M11	0.87	0.87	10	Good	dF+Lth+NDVllog+dRD+TRI+Rou+RSP+SL+HANDH+dCN+HANDV
M10	0.87	0.87	9	Slightly larger in the middle and right	dF+Lth+NDVllog+dRD+TRI+Rou+RSP+SL+HANDH+HANDV
M10(1)	0.87	0.86	10	Good	dF+NDVllog+dRD+TRI+Rou+RSP+SL+HANDH+dCN+HANDV
M9	0.88	0.86	9	Slightly larger on the left	dF+NDVllog+dRD+TRI+Rou+RSP+SL+HANDH+HANDV
M9(1)	0.86	0.86	10	Good	Lth+NDVllog+dRD+TRI+Rou+RSP+SL+HANDH+HANDV
M9(2)	0.86	0.85	10	Good	NDVllog+dRD+TRI+Rou+RSP+SL+HANDH+dCN+HANDV
M8	0.87	0.85	10	Good	NDVllog+dRD+TRI+Rou+RSP+SL+HANDH+HANDV
M7	0.86	0.84	10	Good	NDVllog+dRD+TRI+Rou+RSP+SL+HANDV
M6	0.85	0.83	10	Good	NDVllog+dRD+TRI+RSP+SL+HANDV

Figure 26. Comparison of validity and accuracy (AUCs) of models.

4.5. Landslide Susceptibility Mapping Results

Based on the ROC_M_trn2trn of model M11, we have compiled the landslide susceptibility zoning map (Figure 27). This method uses the success rate to determine that the cumulative landslide area exceeds the cumulative area that is considered vulnerable [59], which can improve the readability of the map. Very-high-susceptibility areas (VHS) comprise only 5.05% of the study area and contain 50% of the landslides. High-susceptibility areas (HS) comprise 14.53% of the study area and contain 30% of the landslides (Figure 27, Table 2). Medium-susceptibility areas (MS), low-susceptibility areas (LS), and very-low-susceptibility areas (VLS) comprise 28.23%, 32.55%, and 19.64% of the study area, respectively, and contain 15%, 4%, and 1% of landslides. Therefore, HS and VHS contain 80% of the landslides and only comprise 19.58% of the study areas. These characteristics of the landslide susceptibility zoning map represent the potential of M11 for the first-order prediction of landslides in this landscape.

Table 2. Statistical table of landslide susceptibility zoning area.

Sub-Regions	Area of Sub-Regions (%)	Total Area of Sub-Regions (%)	Landslides (%)	Total Landslides (%)
VHS	5.05	5.05	50	50
HS	14.53	19.58	30	80
MS	28.23	47.81	15	95
LS	32.55	80.36	4	99
VLS	19.64	100	1	100

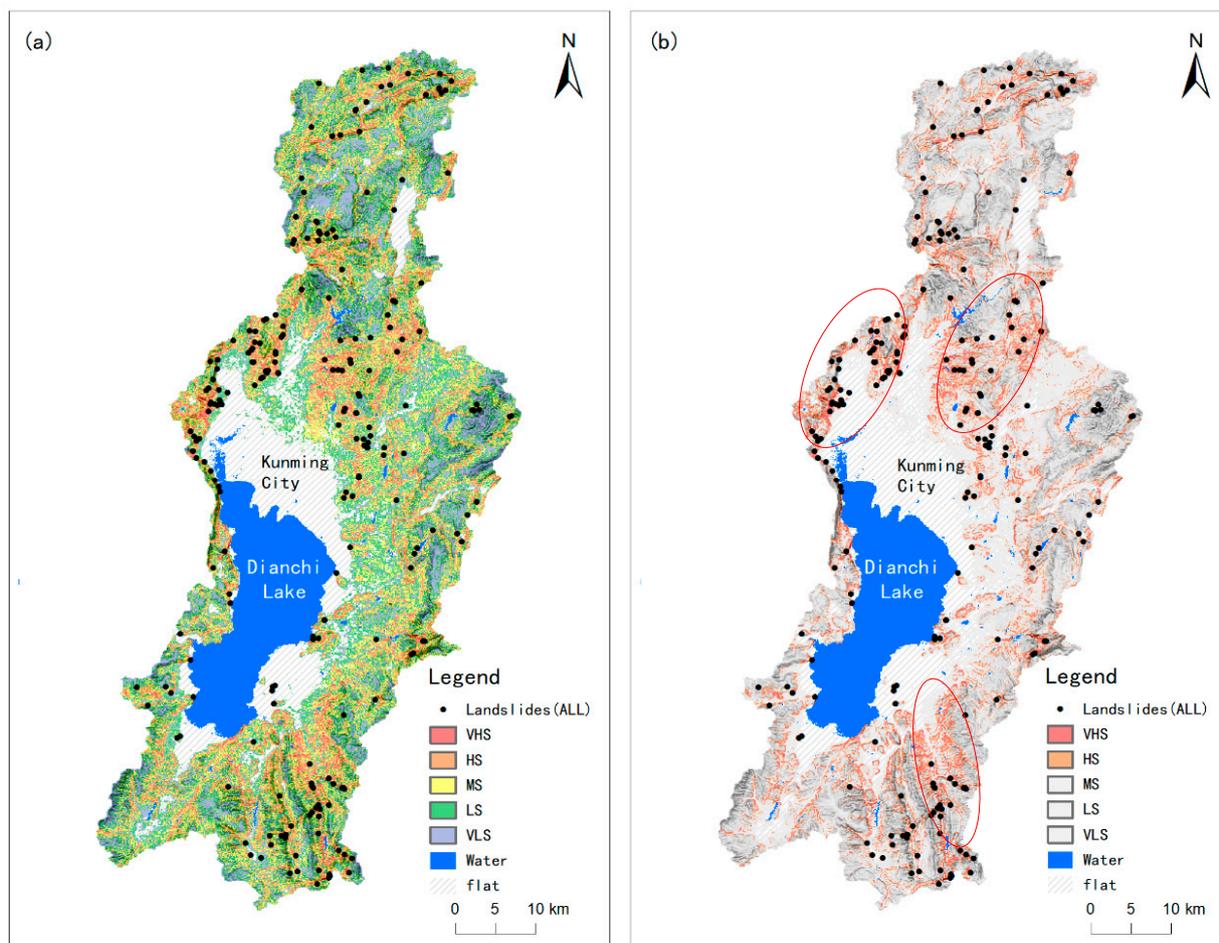


Figure 27. Map of susceptibility to landslides based on model M11 and trn. The model M11 has the highest rate of accuracy and validity; (a,b) are compiled using the same susceptibility partition data. The differences are as follows: (b) MS, LS, and VLS use the same general gray color to highlight VHS and HS; the bottom picture is rendered using elevation and hill shade; the red ellipse roughly delineates the areas of high susceptibility and contiguous distribution.

5. Discussion

5.1. Landslide Susceptibility Zoning and Disaster Prevention Deployment Strategy

Based on the above work, we compiled the landslide susceptibility map of the Dianchi Lake watershed, which has great practical significance. This map provides spatial planners with basic information relating to landslide disasters. It can be used to determine the regional priority for further investigation, support the local planning activities of regional geological disaster prevention and ecological restoration, or create a regional landslide risk exposure assessment. The latter can evaluate the existing elements with landslide risk or those still under planning.

The landslide susceptibility map developed in this paper can effectively predict known and unknown landslides. The fitting accuracy and prediction accuracy of the best model M11 are both ~ 0.87 , and the model coincidence is excellent (Figures 25 and 26). Moreover, ROC_M_trn2TST and the range of ROC_M_trn2trn are closely coincident (Figures 25 and 26), indicating that there is no over-fitting or under-fitting. When only 19.58% of the study area is defined as a high-susceptibility area (VHS + HS), the model can predict 80% of the landslides (Figure 27, Table 2). The above analysis results are satisfactory for the Dianchi Lake watershed.

The map of landslide susceptibility compiled in this paper reveals that the area of high susceptibility (VHS + HS) is relatively large, accounting for about 20% of the study

area (excluding the area with flat and water surface areas), which shows that the natural landslide susceptibility in Dianchi Lake watershed is relatively strong, which poses a great challenge to the comprehensive prevention and control of geological disasters, and this work has a long way to go. In particular, there are large areas of high susceptibility (VHS + HS) in the mountainous area on the edge of the northern basin of the Kunming urban area, and it is almost contiguous. These areas are close to Kunming city and Dianchi Lake, which have a great influence on urban safety and Dianchi Lake water protection and should be regarded as the key areas for landslide prevention and control. Another area with high susceptibility (VHS + HS) is in the southeast of the study area, and mitigation and preventative activities should also be taken in this area.

5.2. Important Factors of Landslide Susceptibility and High Sensitivity and Disaster Prevention Suggestions

AUCs (AUC_ALL, AUC_trn, AUC_trn2trn, AUC_trn2TST) of single factors quantify the sensitivity (spatial correlation) of each factor to the impact of landslide, and the evidence weight of single factors (WoE_ALL, WoE_trn) reveals the impact of each classification on the spatial distribution of landslide, while sC defines the significance of the difference between classifications. AUCs, WoEs, and sCs are meaningful indexes by which to quantify the sensitivity of landslide impact.

We have identified more reliable landslide control factors. The results (Figures 9–21) show thirteen factors with $AUC \geq 0.6$: dRD, HANDV, NDVIlog, SL, RSP, TRI, Rou, Lth, dF, HANDH, Cprof, dCN, and CLCD (Figure 22). The best landslide susceptibility model represents a combination of 11 factors: dRD, HANDV, NDVIlog, SL, RSP, TRI, Rou, Lth, dF, HANDH, and dCN. In the process of step-by-step modeling, according to ROC_M evaluation, Cprof, dCN, and CLCD were rejected because they did not contribute to the explanatory power of the model.

The above results suggest that we should pay attention to the natural conditions and human factors represented by dRD, HANDV, NDVIlog, SL, RSP, TRI, Rou, Lth, dF, HANDH, dCN, CProf, and CLCD, coordinate prevention with planning, construction, and protection, and reduce the induction of landslides.

We also noticed which classification of the above-mentioned important factors is more conducive to the occurrence of landslides. We should pay attention to the slope stability support within 100 m on both sides of the roads and reduce the development in steep slope areas ($25\text{--}40^\circ$), areas where the height difference between the two sides of the stream is 13–67 m, and areas with low vegetation coverage. Attention should also be paid to the preservation and protection of forest vegetation, and the construction planning area should avoid weak rocks such as the affected areas of fault zones (within 121 m on both sides of the faults) and shale siltstone.

5.3. The Landslide Susceptibility Evaluation Based on the WoE Method May Be Improved

The optimized classification process sets the classification value based on the nearly continuous cumulative sC curve of evidence weight distribution. This sub-process can capture the trend of evidence weight distribution, overcome the discontinuity of evidence weight distribution in traditional methods, improve the discrimination of landslide sensitivity of each factor, and reduce the subjectivity of factor classification.

The uncertainty analysis obtained via sub-sampling cross-validation technology enables us to verify the weighted uncertainty sampling process related to the introduced error [6]. trn and TST are spatial random sub-samples of the same size from the same dataset, ALL, which represent the same spatial distribution but have different mean sampling errors (MSE) related to sample size [4]. The model performance evaluation based on TST, which is smaller than TRN, must take this into account in order to correctly interpret the model analysis results [31]. MSE based on trn defines the uncertainty of model performance. If the model is well-summarized and there is no obvious over-fitting, then the ROC curve and AUC value should both fall within the MSE range when the model is evaluated

against corresponding TST [4]. Therefore, compared with the traditional no-sampling process (all landslide data are used for analysis), our process is advantageous because the potential impact of random sub-sampling is considered.

We have compared the accuracy and prediction performance of fourteen models with different combinations of factors. The optimal model M11 contains Rou, TRI, and SL with Pearson's C index > 0.7, but the ROC_M_trn2TST of the model not only has no over-fitting, it also shows excellent coincidence. We think that it is not appropriate to exclude the modeling factors only according to Pearson's C index, and it may be more feasible to determine the Cramer's V index and ROC_M comprehensively.

The comprehensive process proposed in this paper combines many techniques, such as optimized classification, cross-validation, and step-by-step modeling, and obtains the model with high accuracy and predictive performance, which shows that this process has good practical value, may improve landslide susceptibility evaluation based on the WoE method, and is worthy of further promotion and application in similar areas.

6. Conclusions

- (1) The comprehensive process of LSA proposed in this paper has good adaptability, which made a new contribution to the improvement of LSA based on the WoE method. The single-factor categorization optimization sub-process is driven by data, which reduces the subjectivity of factor classification. Cross-validation technology and single-factor WoE statistics reduces the impact of the spatial random effect on factor weight. An effective model was established, and the AUC of fitting and prediction reached 0.8. Cross-validation proves that the model has not been over-fitted.
- (2) Eleven factors, namely, dRD, HANDV, NDVIlog, SL, RSP, TRI, Rou, Lth, dF, HANDH, and dCN, were identified as the key factors sensitive to landslides in the study area, which should be considered emphatically in landslide prevention, monitoring, early warning facility layout, and ecological restoration planning.
- (3) The area of high susceptibility (VHS + HS) in the Dianchi Lake watershed is large, and the comprehensive prevention of landslides have a long way to go. The large-scale and contiguous high-sensitivity areas in the mountainous areas around the basin have caused serious landslide disasters and degraded the urban safety of Kunming and the water source protection of Dianchi Lake, so it is necessary to strengthen the investigation, monitoring, and risk assessment of landslides.

Author Contributions: Conceptualization, G.B., X.Y. and Z.K.; methodology, G.B. and X.Y.; software, G.B. and X.Y.; validation, G.B., X.Y. and Z.K.; formal analysis, G.B. and X.Y.; investigation, G.B., X.Y., Z.K., J.Z., S.Z. and B.S.; resources, G.B., X.Y. and Z.K.; data curation, Z.K. and S.Z.; writing—original draft preparation, G.B., X.Y. and Z.K.; writing—review and editing, G.B., X.Y., Z.K., J.Z., B.S. and S.Z.; visualization, G.B. and X.Y.; supervision, Z.K. and S.Z.; project administration, S.Z. and J.Z.; funding acquisition, S.Z. and J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the science and technology development project of Power China, Sinohydro Foundation Engineering Co., Ltd. (Tianjin, China), the evaluation of rapid excavation of slope cut-off wall in complex geological background area and treatment technology of mud and water inrush in tunnel engineering (Grant No. KKK0202321010); and the scientific and technological development project of Southwest Pipeline Co., Ltd. (Chengdu, China), National Pipe Network Group Research on Hydraulic Protection and Soil and Water Conservation of Oil and Gas Pipelines through Fully Weathered Granite Area (Grant No. KKK0201921153).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets for this study can be obtained by contacting the first author or corresponding author.

Acknowledgments: We are very grateful to our colleagues in the team who supported the implementation of this project. We are also sincerely thankful to the editors and reviewers.

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

References

1. 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](#)]
2. Bai, G.; Yang, X.; Zhu, J.; Zhang, S.; Zhu, C.; Kang, X.; Sun, B.; Zhou, Y. Susceptibility assessment of geological hazards in Wuhua District of Kuming, China using the weight evidence method. *Chin. J. Geol. Hazard Control* **2022**, *33*, 128–138.
3. Guzzetti, F.; Reichenbach, P.; Cardinali, M.; Galli, M.; Ardizzone, F. Probabilistic landslide hazard assessment at the basin scale. *Geomorphology* **2005**, *72*, 272–299. [[CrossRef](#)]
4. Torizin, J.; Schüßler, N.; Fuchs, M. Landslide Susceptibility Assessment Tools v1.0.0b—Project Manager Suite: A new modular toolkit for landslide susceptibility assessment. *Geosci. Model Dev.* **2022**, *15*, 2791–2812. [[CrossRef](#)]
5. Guzzetti, F.; Cardinali, M.; Reichenbach, P.; Carrara, A. Comparing Landslide Maps: A Case Study in the Upper Tiber River Basin, Central Italy. *Environ. Manag.* **2000**, *25*, 247–263. [[CrossRef](#)]
6. Torizin, J.; Fuchs, M.; Awan, A.A.; Ahmad, I.; Akhtar, S.S.; Sadiq, S.; Razzak, A.; Weggenmann, D.; Fawad, F.; Khalid, N.; et al. Statistical landslide susceptibility assessment of the Mansehra and Torghar districts, Khyber Pakhtunkhwa Province, Pakistan. *Nat. Hazards* **2017**, *89*, 757–784. [[CrossRef](#)]
7. Torizin, J.; Wang, L.; Fuchs, M.; Tong, B.; Balzer, D.; Wan, L.; Kuhn, D.; Li, A.; Chen, L. Statistical landslide susceptibility assessment in a dynamic environment: A case study for Lanzhou City, Gansu Province, NW China. *J. Mt. Sci.* **2018**, *15*, 1299–1318. [[CrossRef](#)]
8. Reichenbach, P.; Rossi, M.; Malamud, B.D.; Mihir, M.; Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth-Sci. Rev.* **2018**, *180*, 60–91. [[CrossRef](#)]
9. Torizin, J. Elimination of informational redundancy in the weight of evidence method: An application to landslide susceptibility assessment. *Stoch. Environ. Res. Risk A* **2016**, *30*, 635–651. [[CrossRef](#)]
10. Bonham-Carter, G.; Agterberg, F.P.; Wright, D.F. Weight of evidence modeling: A new approach to mapping mineral potential. *Geol. Surv. Can.* **1989**, *89*, 171–183.
11. Teerarungsikul, S.; Torizin, J.; Fuchs, M.; Kühn, F.; Chonglakmani, C. An integrative approach for regional landslide susceptibility assessment using weight of evidence method: A case study of Yom River Basin, Phrae Province, Northern Thailand. *Landslides* **2016**, *13*, 1151–1165. [[CrossRef](#)]
12. 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](#)]
13. Sun, D.; Xu, J.; Wen, H.; Wang, D. Assessment of landslide susceptibility mapping based on Bayesian hyperparameter optimization: A comparison between logistic regression and random forest. *Eng. Geol.* **2021**, *281*, 105972. [[CrossRef](#)]
14. Yang, J.; Song, C.; Yang, Y.; Xu, C.; Guo, F.; Xie, L. New method for landslide susceptibility mapping supported by spatial logistic regression and GeoDetector: A case study of Duwen Highway Basin, Sichuan Province, China. *Geomorphology* **2019**, *324*, 62–71. [[CrossRef](#)]
15. Den Eeckhaut, M.V.; Marre, A.; Poesen, J. Comparison of two landslide susceptibility assessments in the Champagne–Ardenne region (France). *Geomorphology* **2010**, *115*, 141–155. [[CrossRef](#)]
16. He, S.; Pan, P.; Dai, L.; Wang, H.; Liu, J. Application of kernel-based Fisher discriminant analysis to map landslide susceptibility in the Qinggan River delta, Three Gorges, China. *Geomorphology* **2012**, *171–172*, 30–41. [[CrossRef](#)]
17. Saha, A.; Saha, S. Comparing the efficiency of weight of evidence, support vector machine and their ensemble approaches in landslide susceptibility modelling: A study on Kurseong region of Darjeeling Himalaya, India. *Remote Sens. Appl. Soc. Environ.* **2020**, *19*, 100323. [[CrossRef](#)]
18. Kumar, D.; Thakur, M.; Dubey, C.S.; Shukla, D.P. Landslide susceptibility mapping & prediction using Support Vector Machine for Mandakini River Basin, Garhwal Himalaya, India. *Geomorphology* **2017**, *295*, 115–125.
19. He, Q.; Wang, M.; Liu, K. Rapidly assessing earthquake-induced landslide susceptibility on a global scale using random forest. *Geomorphology* **2021**, *391*, 107889. [[CrossRef](#)]
20. Sun, D.; Wen, H.; Wang, D.; Xu, J. A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. *Geomorphology* **2020**, *362*, 107201. [[CrossRef](#)]
21. Tanyu, B.F.; Abbaspour, A.; Alimohammadlou, Y.; Tecuci, G. Landslide susceptibility analyses using Random Forest, C4.5, and C5.0 with balanced and unbalanced datasets. *Catena* **2021**, *203*, 105355. [[CrossRef](#)]
22. Gameiro, S.; Riffel, E.S.; de Oliveira, G.G.; Guasselli, L.A. Artificial neural networks applied to landslide susceptibility: The effect of sampling areas on model capacity for generalization and extrapolation. *Appl. Geogr.* **2021**, *137*, 102598. [[CrossRef](#)]
23. Amato, G.; Palombi, L.; Raimondi, V. Data-driven classification of landslide types at a national scale by using Artificial Neural Networks. *Int. J. Appl. Earth Obs.* **2021**, *104*, 102549. [[CrossRef](#)]
24. Lucchese, L.V.; de Oliveira, G.G.; Pedrollo, O.C. Investigation of the influence of nonoccurrence sampling on landslide susceptibility assessment using Artificial Neural Networks. *Catena* **2021**, *198*, 105067. [[CrossRef](#)]

25. Ng, A.; Jordan, M. On Discriminative vs. Generative Classifiers: A comparison of logistic regression and naive Bayes. *Adv. Neural Inf. Process. Syst.* **2002**, *2*, 841–848.
26. Agerberg, F.P.; Bonham-Carter, G.F.; Cheng, Q.; Wright, D.F.; Davis, J.C.; Herzfeld, U.C. Weights of evidence modeling and weighted logistic regression for mineral potential mapping. *Comput. Geol.* **1993**, *5*, 13–32.
27. Alsabhan, A.H.; Singh, K.; Sharma, A.; Alam, S.; Pandey, D.D.; Rahman, S.A.S.; Khursheed, A.; Munshi, F.M. Landslide susceptibility assessment in the Himalayan range based along Kasauli—Parwanoo road corridor using weight of evidence, information value, and frequency ratio. *J. King Saud Univ. Sci.* **2022**, *34*, 101759. [\[CrossRef\]](#)
28. Chen, L.; Guo, H.; Gong, P.; Yang, Y.; Zuo, Z.; Gu, M. Landslide susceptibility assessment using weights-of-evidence model and cluster analysis along the highways in the Hubei section of the Three Gorges Reservoir Area. *Comput. Geosci.* **2021**, *156*, 104899. [\[CrossRef\]](#)
29. Mathew, J.; Jha, V.K.; Rawat, G.S. Weights of evidence modelling for landslide hazard zonation mapping in part of Bhagirathi valley, Uttarakhand. *Curr. Sci.* **2007**, *92*, 628–638.
30. Neuhäuser, B.; Terhorst, B. Landslide susceptibility assessment using “weights-of-evidence” applied to a study area at the Jurassic escarpment (SW-Germany). *Geomorphology* **2007**, *86*, 12–24. [\[CrossRef\]](#)
31. Torizin, J.; Fuchs, M.; Kuhn, D.; Balzer, D.; Wang, L. Practical Accounting for Uncertainties in Data-Driven Landslide Susceptibility Models. Examples from the Lanzhou Case Study. In *Understanding and Reducing Landslide Disaster Risk: Volume 2 From Mapping to Hazard and Risk Zonation*; Guzzetti, F., Mihalić Arbanas, S., Reichenbach, P., Sassa, K., Bobrowsky, P.T., Takara, K., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 249–255.
32. Guzzetti, F.; Mondini, A.C.; Cardinali, M.; Fiorucci, F.; Santangelo, M.; Chang, K. Landslide inventory maps: New tools for an old problem. *Earth-Sci. Rev.* **2012**, *112*, 42–66. [\[CrossRef\]](#)
33. Yang, J.; Huang, X. The 30m annual land cover dataset and its dynamics in China from 1990 to 2019. *Earth Syst. Sci. Data* **2021**, *13*, 3907–3925. [\[CrossRef\]](#)
34. Jasiewicz, J.; Stepinski, T.F. Geomorphons—A pattern recognition approach to classification and mapping of landforms. *Geomorphology* **2013**, *182*, 147–156. [\[CrossRef\]](#)
35. Stepinski, T.F.; Jasiewicz, J. Geomorphons—A new approach to classification of landforms. *Proc. Geomorphometry* **2011**, *2011*, 109–112.
36. Luo, Y.; Tang, L.; Yang, K.; Zhou, X.; Liu, J.; Zhang, Y.; Peng, Z. Investigating the warming effect of urban expansion on lake surface water temperature in the Dianchi lake watershed. *J. Hydrol. Reg. Stud.* **2023**, *49*, 101516. [\[CrossRef\]](#)
37. Chung, C.; Fabbri, A.G. Predicting landslides for risk analysis—Spatial models tested by a cross-validation technique. *Geomorphology* **2008**, *94*, 438–452. [\[CrossRef\]](#)
38. Xu, Y.; Goodacre, R. On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. *J. Anal. Test* **2018**, *2*, 249–262. [\[CrossRef\]](#)
39. Lan, H.; Tian, N.; Li, L.; Wu, Y.; Macciotta, R.; Clague, J.J. Kinematic-based landslide risk management for the Sichuan-Tibet Grid Interconnection Project (STGIP) in China. *Eng. Geol.* **2022**, *308*, 106823. [\[CrossRef\]](#)
40. Tanyaş, H.; Görüm, T.; Fadel, I.; Yıldırım, C.; Lombardo, L. An open dataset for landslides triggered by the 2016 Mw 7.8 Kaikōura earthquake, New Zealand. *Landslides* **2022**, *19*, 1405–1420. [\[CrossRef\]](#)
41. Xiong, H.; Ma, C.; Li, M.; Tan, J.; Wang, Y. Landslide susceptibility prediction considering land use change and human activity: A case study under rapid urban expansion and afforestation in China. *Sci. Total Environ.* **2023**, *866*, 161430. [\[CrossRef\]](#)
42. Zhang, Y.; Ayyub, B.M.; Gong, W.; Tang, H. Risk assessment of roadway networks exposed to landslides in mountainous regions—A case study in Fengjie County, China. *Landslides* **2023**, *20*, 1419–1431. [\[CrossRef\]](#)
43. Zanaga, D.; Van De Kerchove, R.; De Keersmaecker, W.; Souverijns, N.; Brockmann, C.; Quast, R.; Wevers, J.; Grosu, A.; Paccini, A.; Vergnaud, S.; et al. ESA WorldCover 10 m 2020 v100 [Data Set]. Available online: <https://zenodo.org/records/5571936> (accessed on 31 October 2021). [\[CrossRef\]](#)
44. Xu, X. China 30m Annual NDVI Maximum Dataset [Data Set]. Resource and Environmental Science Data Registration and Publishing System. 2022. Available online: <https://www.resdc.cn/DOI/DOI.aspx?DOIID=68> (accessed on 23 August 2023).
45. Jpl, N. NASADEM Merged DEM Global 1 arc Second V001. Nasa Eosdis Land Process. Daac. 2020. Available online: https://lpdaac.usgs.gov/products/nasadem_hgtv001/ (accessed on 14 January 2021).
46. Guisan, A.; Weiss, S.B.; Weiss, A.D. GLM versus CCA spatial modeling of plant species distribution. *Plant Ecol.* **1999**, *143*, 107–122. [\[CrossRef\]](#)
47. Riley, S.; Degloria, S.; Elliot, S.D. A Terrain Ruggedness Index that Quantifies Topographic Heterogeneity. *Int. J. Sci.* **1999**, *5*, 23–27.
48. Nobre, A.D.; Cuartas, L.A.; Hodnett, M.; Rennó, C.D.; Rodrigues, G.; Silveira, A.; Waterloo, M.; Saleska, S. Height Above the Nearest Drainage—A hydrologically relevant new terrain model. *J. Hydrol.* **2011**, *404*, 13–29. [\[CrossRef\]](#)
49. Rennó, C.D.; Nobre, A.D.; Cuartas, L.A.; Soares, J.V.; Hodnett, M.G.; Tomasella, J.; Waterloo, M.J. HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. *Remote Sens. Environ.* **2008**, *112*, 3469–3481. [\[CrossRef\]](#)
50. Moore, I.D.; Grayson, R.B.; Ladson, A.R. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrol. Process* **1991**, *5*, 3–30. [\[CrossRef\]](#)
51. Beven, K.; Kirkby, M. A Physically Based, Variable Contributing Area Model of Basin Hydrology. *Hydrol. Sci. Bull.* **1979**, *24*, 43–69. [\[CrossRef\]](#)

52. Böhner, J.; Selige, T. Spatial prediction of soil attributes using terrain analysis and climate regionalisation. *Saga—Anal. Model. Appl.* **2006**, *115*, 13–27.
53. Böhner, J.; Koethe, R.; Conrad, O.; Gross, J.; Ringeler, A.; Selige, T. Soil regionalisation by means of terrain analysis and process parameterisation. *Soil Classif.* **2001**, *2002*, 213–222.
54. Agterberg, F.P. Combining indicator patterns in weights of evidence modeling for resource evaluation. *Nonrenewable Resour.* **1992**, *1*, 39–50. [[CrossRef](#)]
55. Agterberg, F.P.; Bonham-Carter, G.F.; Wright, D.F. Statistical Pattern Integration for Mineral Exploration**Geological Survey of Canada Contribution No. 24088. In *Computer Applications in Resource Estimation*; GAÁL, G., Merriam, D.F., Eds.; Pergamon: Amsterdam, The Netherlands, 1990; pp. 1–21.
56. Bonham-Carter, G.F. *Geographic Information Systems for Geoscientists: Modelling with GIS*; Pergamon: Oxford, UK, 1994; pp. 1–398.
57. Fawcett, T. An introduction to ROC analysis. *Pattern. Recogn. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
58. Agterberg, F.P.; Cheng, Q. Conditional Independence Test for Weights-of-Evidence Modeling. *Nat. Resour. Res.* **2002**, *11*, 249–255. [[CrossRef](#)]
59. Chung, C.F.; Fabbri, A.G. Validation of Spatial Prediction Models for Landslide Hazard Mapping. *Nat. Hazards* **2003**, *30*, 451–472. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.