

Article

Landslide Susceptibility Evaluation Using Hybrid Integration of Evidential Belief Function and Machine Learning Techniques

Yang Li ^{1,2,3} and Wei Chen ^{1,2,3,*} 

¹ College of Geology and Environment, Xi'an University of Science and Technology, Xi'an 710054, China; 19209071008@stu.xust.edu.cn

² Key Laboratory of Coal Resources Exploration and Comprehensive Utilization, Ministry of Natural Resources, Xi'an 710021, China

³ Shaanxi Provincial Key Laboratory of Geological Support for Coal Green Exploitation, Xi'an 710054, China

* Correspondence: chenwei0930@xust.edu.cn

Received: 7 November 2019; Accepted: 26 December 2019; Published: 29 December 2019



Abstract: In this study, Random SubSpace-based classification and regression tree (RSCART) was introduced for landslide susceptibility modeling, and CART model and logistic regression (LR) model were used as benchmark models. 263 landslide locations in the study area were randomly divided into two parts (70/30) for training and validation of models. 14 landslide influencing factors were selected, such as slope angle, elevation, aspect, sediment transport index (STI), topographical wetness index (TWI), stream power index (SPI), profile curvature, plan curvature, distance to rivers, distance to road, soil, normalized difference vegetation index (NDVI), land use, and lithology. Finally, the hybrid RSCART model and two benchmark models were applied for landslide susceptibility modeling and the receiver operating characteristic curve method is used to evaluate the performance of the model. The susceptibility is quantitatively compared based on each pixel to reveal the system spatial pattern between susceptibility maps. At the same time, area under ROC curve (AUC) and landslide density analysis were used to estimate the prediction ability of landslide susceptibility map. The results showed that the RSCART model is the optimal model with the highest AUC values of 0.852 and 0.827, followed by LR and CART models. The results also illustrate that the hybrid model generally improves the prediction ability of a single landslide susceptibility model.

Keywords: landslide susceptibility; hybrid models; random subspace; classification and regression tree

1. Introduction

Landslides are large-scale movements of rocks, mud, and gravel from the top to the bottom of the mountain [1]. According to statistics, five percent of the world's natural disasters were landslides from 1994 to 2013 [2]. Landslides pose serious risks to human life, environment, resources, and property. Therefore, for the sake of decreasing the hazards caused by landslides, landslide susceptibility evaluation is becoming a common topic.

Landslide susceptibility refers to the probability of landslide occurring in an area on account of local geological environmental factors [3,4]. Geographic information system (GIS) has been diffusely applied for evaluating landslide susceptibility in recent decades, and many methods have been proposed. These methods mainly include deterministic methods, traditional statistical methods, and machine learning technologies [4,5].

In deterministic models, the stability of the slope is in the form of a calculated safety factor, which is suitable for small watersheds [6,7]. Statistical models, such as evidential belief function (EBF) [8,9],

weights of evidence [10–12], frequency ratio [13–17], logistic regression [18–21], linear multivariate regression, multivariate adaptive regression spline [22–24], and statistical index [25,26] have been widely used. However, these traditional statistical methods do not provide satisfactory evaluation of the correlation between landslide influencing factors [4,27].

Therefore, machine learning technologies have drawn extensive attention, and many kinds of machine learning methods have been developed and used, such as classification and regression trees [28,29], adaptive neuro-fuzzy inference systems [30,31], fuzzy logic [32,33], alternating decision trees [34–36], support vector machine [37–39], artificial neural networks [40,41], and random forest [4,42–45]. In particular, hybrid models are increasingly used, such as the rotation forest-based decision trees [46,47], frequency ratio-based ANFIS model [48], bagging-based reduced error pruning trees [49], and multiboost-based support vector machines [50].

Spatial prediction of landslide is not only the first important step, but also one of the most difficult tasks [31,51]. Many modeling methods have been used in the construction of landslide susceptibility maps in the past, but the accuracy of these models has not been accepted by all researchers [52]. Youssef et al. used CART model to study landslide susceptibility in Wadi Tayah Basin. The result of post-cart model is not optimal [53]. However, when Felicísimo et al. studied the landslide susceptibility, CART shows a high predictive ability [54]. Hong et al. studied the landslide susceptibility in Wuning area of China after modeling the hybrid model of free subspace and support vector machine. The mixed model RSSVM model also achieved optimal performance (AUC = 0.857) [55]. Therefore, the hybrid models are often considered to be more precise to single models [55].

The Loess Plateau is one of the most vulnerable regions in China. The special geological environment of the area has led to the development of environmental geological problems. Zichang County is in the hinterland of the Loess Plateau. Due to the large area coverage of the Quaternary loess, various geological problems induced by geological processes became more apparent. Therefore, the main purpose of this study is to apply and analyze the hybrid integration method of Random Subspace (RS) and CART, namely RSCART, in landslide susceptibility assessment for the Zichang County. Also, the EBF model was used to evaluate the correlation between landslides and influencing factors, and the single CART model and well known LR model were selected as benchmark models. Finally, the ROC curve and AUC values were used for comparing the performance of models.

2. Description of the Study Area

The study area selected for this study is Zichang County in Shaanxi Province, China (Figure 1). The length of the County is 72 km from east to west, 55.7 km from north to south, and the total area is 2450 km². It is in east longitude 109°11'22"~110°01'22" E, latitude 36°30'59"~37°30'00" N. The altitude ranges from 933 m to 1574 m above sea level. Topographically, the slope less than 10° accounts for about 10.44% of the total area, the slope of 10–20° is about 26.09%, 20–30° is about 35.14%, 30–40° is about 23.9%, and 40–50° is about 4.41%. The slope greater than 50° is only 0.02%. From the perspective of geomorphology, Zichang County belongs to the Loess Plateau gully region in northern Shaanxi Province. The complex geomorphology types in the area were formed after the loess was eroded and cut by the Xiuyan River, the Luanhe River, and other tributaries. The loess from the Middle Pleistocene to the Late Pleistocene formed a lot of valleys and river systems after the Loess Plateau uplift. The penetration of the Yellow River led to the formation of many loess ridges, loess hills, and gullies in the area. According to the landform genesis, the area is divided into loess landforms and river landforms.

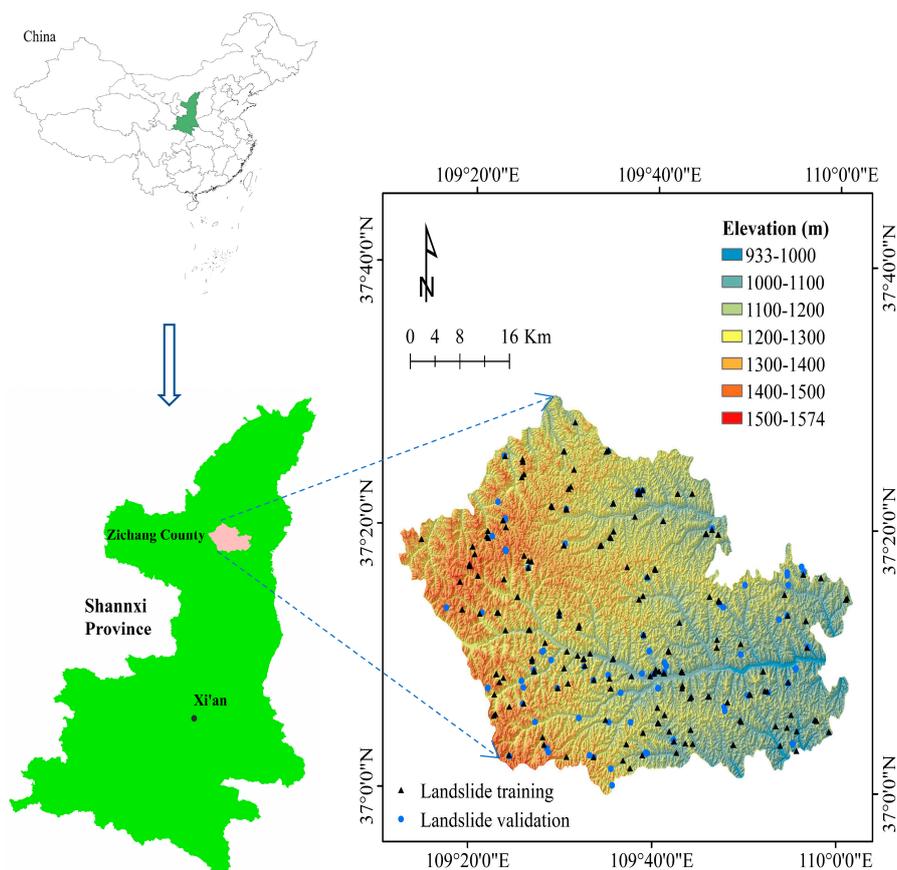


Figure 1. Study area.

3. Methodology

The methodology of this study is shown in Figure 2. There are mainly four steps in the current study: (1) data preparation including preparation of a landslide inventory map and landslide conditioning factors, (2) multicollinearity analysis and consideration of the correlation between landslide locations and conditioning factors, (3) three models are used for landslide susceptibility model (RSCART,CART,LR), and (4) validation of landslide susceptibility maps produced by the ROC curve.

3.1. Data Preparation

The landslides inventory includes the location of past and recent landslides [56]. It also allows people to understand the type and timing of landslides [57]. This study constructed a landslide inventory including 263 landslide locations by consulting aerial photos and collecting historical landslide events, of which most of the landslides were slides (201), the others included 62 falls [58]. According to an analysis in the GIS environment, the largest landslide was more than $1 \times 10^7 \text{ m}^3$, and the smallest landslide was nearly 120 m^3 . Rainfall and human engineering activities, such as urban and rural construction, road engineering construction and water conservancy engineering, are the main triggering factors of landslides. To establish and verify the landslide model, the landslides were randomly divided into two parts: (1) 70% were used to construct the training dataset; (2) the remaining 30% were used to generate validation dataset.

Based on literature review and data availability [42,56,59,60], 14 landslide influencing factors were used in this study. Slope angle, elevation, aspect, sediment transport index (STI), topographical wetness index (TWI), stream power index (SPI), profile curvature, plan curvature, distance to rivers, distance to roads, soil, NDVI, land use, and lithology were considered.

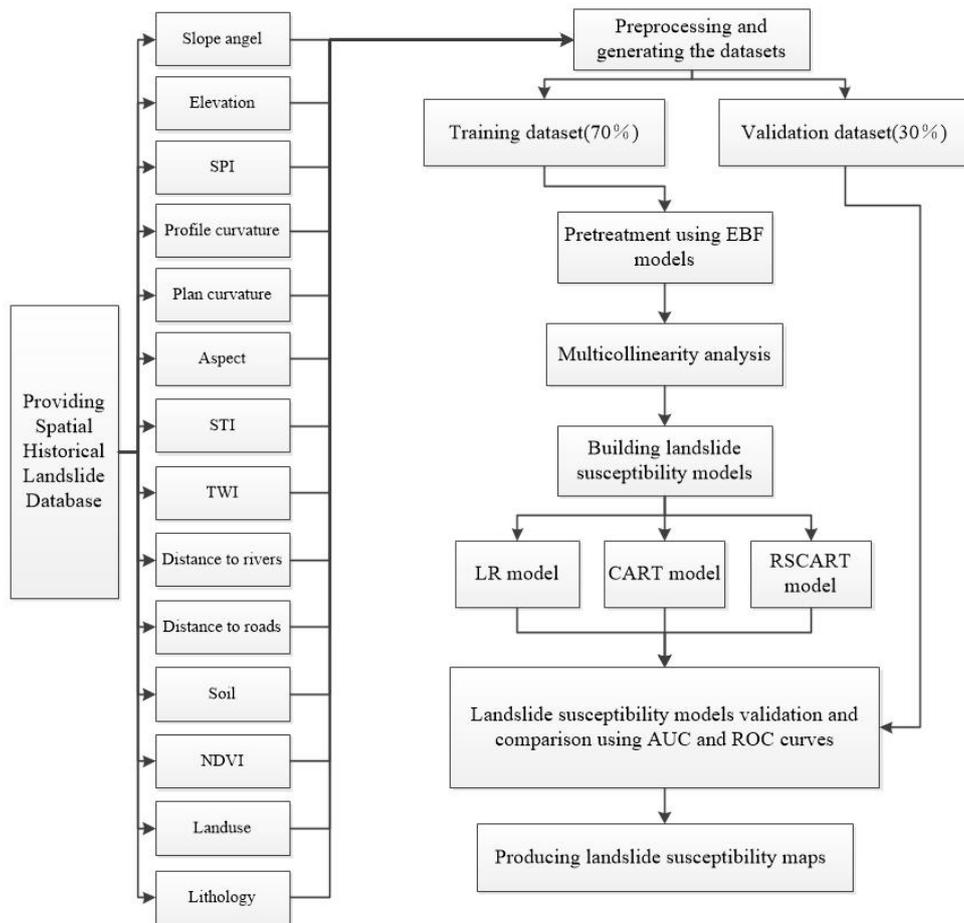


Figure 2. Flowchart of the study.

Slope angle, elevation and slope aspect are often used in landslide susceptibility mapping [61–65]. In this study, slope angles were divided into six classes: $<10^\circ$, $10\text{--}20^\circ$, $20\text{--}30^\circ$, $30\text{--}40^\circ$, $40\text{--}50^\circ$, $>50^\circ$ (Figure 3a, Table 1). The elevation of the study area was reclassified into seven categories with an interval of 100 m, such as <1000 m, $1000\text{--}1100$ m, $1100\text{--}1200$ m, $1200\text{--}1300$ m, $1300\text{--}1400$ m, $1400\text{--}1500$ m, and >1500 m (Figure 3b, Table 1). Slope aspect was divided into nine types as follows: flat, north, northeast, east, southeast, south, southwest, west, and northwest (Figure 3c, Table 1).

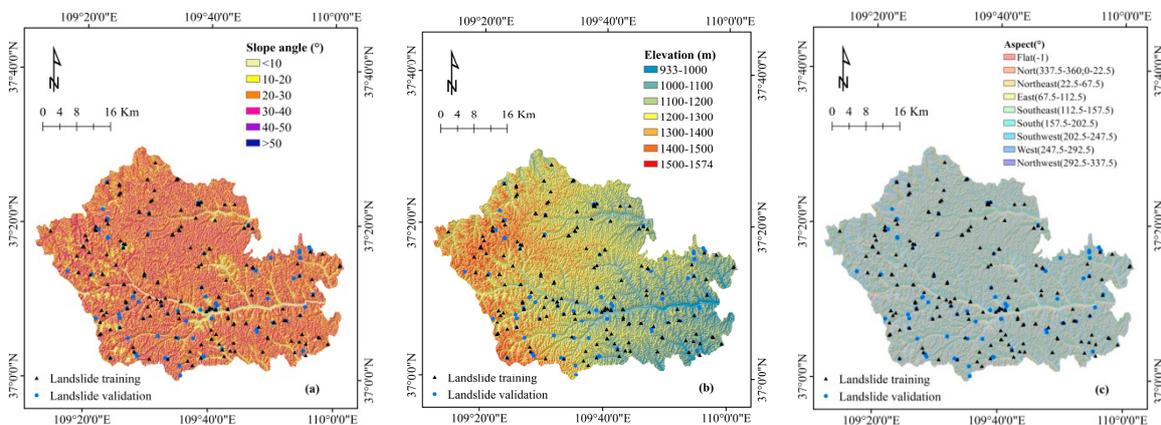


Figure 3. Cont.

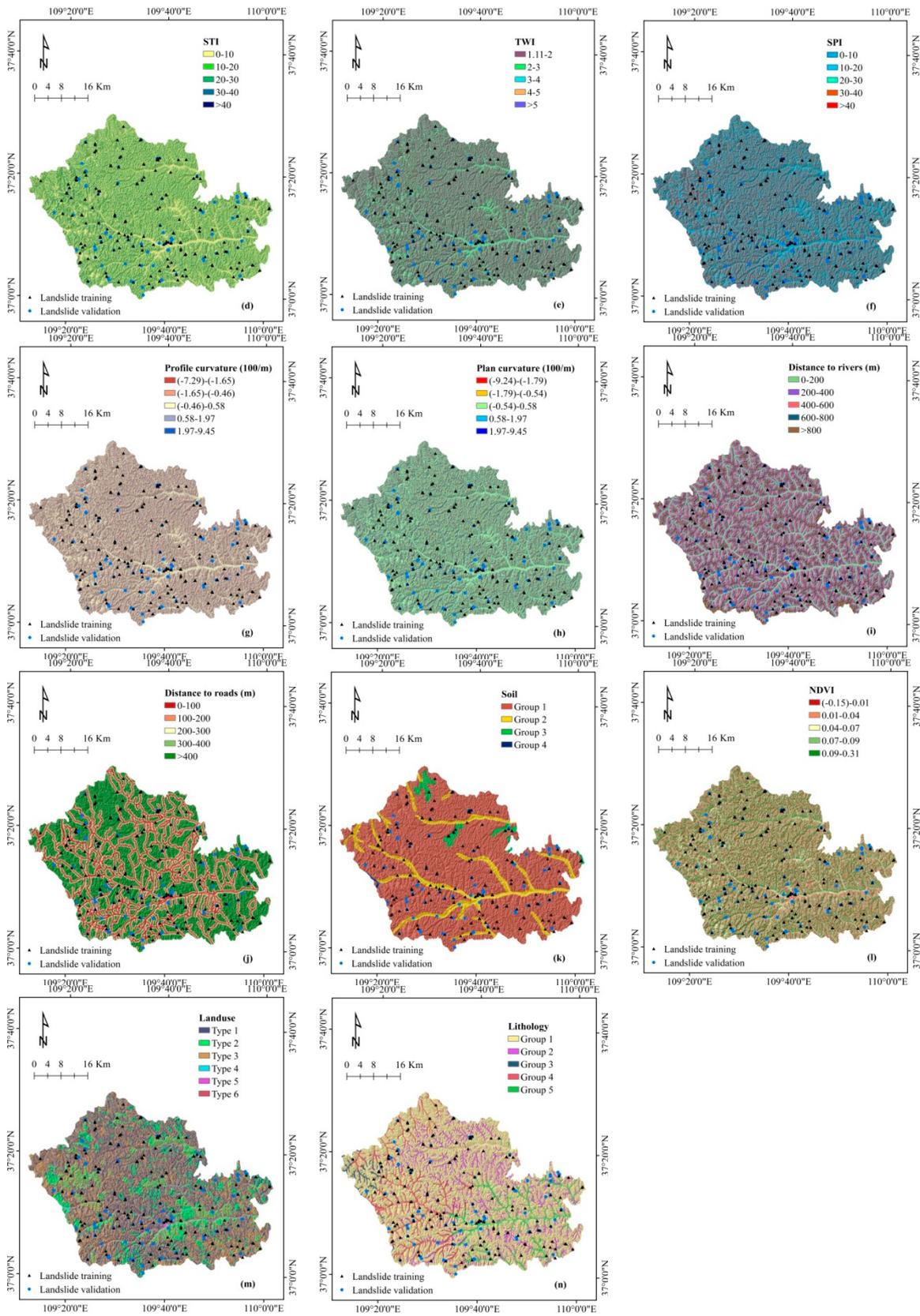


Figure 3. Thematic maps of the study area.

Table 1. Correlation between landslides and influencing factors using EBF model.

Factors	Class	No. of Pixels	No. of Landslide	Bel
Slope angle	<10	278,839	0	0.000
	10–20	696,966	56	0.244
	20–30	938,802	68	0.220
	30–40	638,483	48	0.228
	40–50	117,745	12	0.309
	>50	527	0	0.000
Elevation (m)	933–1000	30,442	4	0.197
	1000–1100	357,423	41	0.172
	1100–1200	753,794	61	0.121
	1200–1300	829,706	54	0.098
	1300–1400	546,264	17	0.047
	1400–1500	148,806	6	0.060
	1500–1574	4927	1	0.305
Aspect	F (−1)	1237	0	0.000
	N (0–22.5; 337.5–360)	247,049	14	0.103
	NE (22.5–67.5)	351,476	12	0.062
	E (67.5–112.5)	436,578	32	0.133
	SE (112.5–157.5)	300,883	25	0.151
	S (157.5–202.5)	270,755	27	0.181
	SW (202.5–247.5)	341,265	31	0.165
	W (247.5–292.5)	412,506	33	0.145
NW (292.5–337.5)	309,613	10	0.059	
STI	(0–10)	1,289,473	81	0.147
	(10–20)	827,143	62	0.176
	(20–30)	299,541	28	0.219
	(30–40)	112,670	6	0.125
	>40	142,535	7	0.115
TWI	(1.11–2)	1,504,887	102	0.319
	(2–3)	885,873	73	0.388
	(3–4)	196,513	7	0.168
	(4–5)	75,716	2	0.124
	>5	8373	0	0.000
SPI	(0–10)	867,208	37	0.113
	(10–20)	526,106	46	0.232
	(20–30)	362,454	32	0.234
	(30–40)	218,983	19	0.230
	>40	696,611	50	0.190
Profile curvature	(−7.29)–(−1.65)	215,910	13	0.182
	(−1.65)–(−0.46)	643,747	29	0.136
	(−0.46)–(0.58)	1,050,656	77	0.222
	(0.58)–(1.97)	567,015	54	0.288
	(1.97)–(9.45)	194,034	11	0.172
Plan curvature	(−9.24)–(−1.79)	143,702	5	0.106
	(−1.79)–(−0.54)	480,381	29	0.185
	(−0.54)–0.38	1,124,169	83	0.226
	0.38–1.44	703,523	47	0.204
	1.44–7.56	219,587	20	0.279
Distance to rivers (m)	0–200	765,053	127	0.590
	200–400	678,212	27	0.141
	400–600	597,921	17	0.101
	600–800	417,041	6	0.051
	>800	213,135	7	0.117

Table 1. Cont.

Factors	Class	No. of Pixels	No. of Landslide	Bel
Distance to roads (m)	0–100	406,132	51	0.313
	100–200	304,978	32	0.262
	200–300	303,291	22	0.181
	300–400	238,548	12	0.126
	>400	1,418,413	67	0.118
Soil	Cultivated loessal soils	2,288,420	141	0.158
	Alluvial soils	316,038	28	0.228
	Red clay soils	62,809	15	0.614
	Water	4095	0	0.000
NDVI	(−0.15–0.01)	372,914	26	0.207
	(0.01–0.04)	452,559	21	0.138
	(0.04–0.07)	599,799	31	0.154
	(0.07–0.09)	733,152	65	0.264
	(0.09–0.31)	512,938	41	0.238
Land use	Farmland	987,416	47	0.142
	Forestland	505,630	37	0.219
	Grassland	1,167,441	99	0.254
	Water bodies	2665	0	0.000
	Residential areas	7769	1	0.385
	Others	441	0	0.000
Lithology	Group 1	2,008,004	111	0.133
	Group 2	330,841	40	0.292
	Group 3	25,061	1	0.096
	Group 4	178,708	23	0.310
	Group 5	128,748	9	0.169

STI reflects the erosion force of surface water flow on the ground [66]. It can be expressed by Equation (1) [67,68]:

$$STI = \left(\frac{\alpha}{22.13} \right)^{0.6} \times \left(\frac{\sin \beta}{0.0896} \right)^{1.3} \quad (1)$$

where α is the specific watershed area, and β is the slope. In this study, STI values were classified into five classes, such as <10, 10–20, 20–30, 30–40, and >40 (Figure 3d, Table 1).

TWI affects slope material by affecting soil moisture and groundwater flow. It can be represented by Equation (2) as follows [68]:

$$TWI = \ln(\alpha / \tan \beta) \quad (2)$$

where α is the area drained per unit contour length at a point, and β is the slope. In the study area, the TWI was also divided into five categories: 1–2, 2–3, 3–4, 4–5, and >5 (Figure 3e, Table 1).

The SPI is an important hydrologic factor, which shows the process of potential flow erosion [68]. It can be represented by Equation (3) as follows [68]:

$$SPI = \alpha \times \tan \beta \quad (3)$$

where α is the specific watershed area, and β is the slope. In the study area, the SPI was also divided into five categories: 1–2, 2–3, 3–4, 4–5 and >5 (Figure 3f, Table 1).

Profile curvature and plane curvature are important topographic factors. Profile curvature is defined as the curvature at any point is perpendicular to the slope [69]. Profile curvature is divided into five types: (−7.92)–(−1.65), (−1.65)–(−0.46), (−0.46)–0.58, 0.58–1.97, 1.97–9.45 (Figure 2g, Table 1). Plan curvature is described as the curvature of the contour formed by the intersection with the plane [70].

Plan curvature was divided into five categories: $(-9.24)-(-1.79)$, $(-1.79)-(-0.54)$, $(-0.46)-0.38$, $0.38-1.44$, $1.44-7.56$ (Figure 3h, Table 1).

The distance to rivers and distance to roads are important factors in landslide susceptibility modeling [71,72]. The river causes the bedrock to produce enough topographic change through erosion. This process makes the slope more prone to block failure [73]. Distance to rivers was divided into five classes with an interval of 200 m (Figure 3i, Table 1). The natural conditions of the slope were destroyed during the construction of transportation facilities, and oil and gas development, and other human engineering activities [74]. In the present study, distance to roads was grouped into five buffering zones using an interval of 100 m (Figure 3j, Table 1).

With the increase of soil depth, the runoff velocity decreases. The unstable nature of shallow is more likely to lead to landslides [75]. There are four types of soil, such as: cultivated loessal soils, alluvial soils, red clay soils and water (Figure 3k, Table 1).

NDVI can quantitatively estimate vegetation growth and biomass by measuring surface reflectance [76]. The NDVI value is calculated based on the following formula:

$$NDVI = (IR - R) / (IR + R) \quad (4)$$

where IR is the infrared band, and R is the red band. The NDVI value was reclassified into five categories: $-0.15-0.11$, $0.01-0.04$, $0.04-0.07$, $0.07-0.09$, $0.09-0.31$ (Figure 3l, Table 1).

Land use is affected by environmental changes and plays an important role in slope stability. In this paper, land use is divided into six categories, including farmland, forestland, grasslands, water bodies, residential areas and others (Figure 3m, Table 1).

Lithological units greatly influence the landslide occurrence [77]. The lithologic units were divided into five groups (Figure 3n, Table 1): the first group is quaternary (loess, silt), the second group is tertiary (mudstone, conglomerate), the third group is cretaceous (arkose), the fourth group is Jurassic (shale, sandstone, mudstone, conglomerate), and the fifth group is Triassic (mudstone, sandstone, conglomerate).

Finally, all the 14 landslide influencing factors were converted to the same resolution of $30\text{ m} \times 30\text{ m}$.

3.2. Evidential Belief Function (EBF)

The EBF is an evidence algorithm which is a mathematical method based on bivariate statistics [78]. The EBF is a flexible model, because it not only accepts uncertainty but also absorbs the belief of many sources. The EBF model is made up of four mathematical functions: the degree of belief (Bel), the degree of disbelief (Dis), the degree of uncertainty (Unc), and the degree of plausibility (Pls). 0 to 1 is the range of these functions [79]. This study uses Bel values to represent the correlation between landslides and factors, and Bel values can be represented by the following expression:

$$Bel = \frac{Bel_1 + Bel_2 + \dots + Bel_n}{1 - \sum_{i=1}^n Bel_{i-1}Bel_i - Bel_{i-1}Bel_i} \quad (5)$$

where Bel_i is the extent of belief of i th influencing factor.

3.3. Classification and Regression Tree (CART)

The CART is an effective method because it has attested to be a technique for dealing with difficult classification problems [80]. CART is a decision tree proposed by Breiman [81]. The data is processed in a recursive form by the CART model [82]. In this process, the value of internal node features is "yes" and "no". Categorical and continuous dependent variables (regression) are predicted by classification and regression trees.

If the dependent variable is a category scale, CART will generate a classification tree, and if the dependent variable is a continuous data, CART will generate a regression tree [83]. The classification tree using CART can be constructed in four main steps: (1) building tree, (2) stopping tree building,

(3) pruning tree, and (4) selecting optimal tree [84]. The information that needs to be processed is represented by CART in an intuitive way, and the metrics between the predictive variables do not affect the model results [53].

CART has many advantages as a classifier that it is not only a method suitable for numeric data types, but also generates an invariant for the transformation of independent variables. In addition, CART does not have to pick a variable first [85].

3.4. Random Subspace (RS)

The RS was proposed by Ho in 1998, which is defined as a classical integration algorithm [86]. The RS can establish the training subset, which is randomly selected from the original training set [87–92]. Also, the RS can modify the training data set in the feature space. RS is also a very effective integration algorithm because it can solve the data sets with redundant features and over-fitting problems. The RS enables the training data to maintain the highest accuracy. However, it cannot be ignored is that the growth of training data increases the complicity of accuracy [93].

The main goal of RS is to collect the feature set of the high-dimensional feature space into the low-dimensional subspace, and then construct a classifier to classify the class based on the subspace. The final result is obtained by the majority voting rule [94]. More specifically, let the n -dimensional vector X_{in} be the n features (landslide influencing factors) of the training sample X_i . In RS, randomly select $r < n$ features from the n -dimensional data set of the original space X to obtain the random subspace of r dimension. The modified training data set X^e contains the r -dimensional training object X_i^e . Finally, the classifier is constructed in the RS X^e and the majority voting rule is adopted, as follows [91]:

$$\alpha(x) = \operatorname{argmax}_{y \in \{0,1\}} \sum_a \delta_{\operatorname{sgn}(C^a(x)), y} \quad (6)$$

where $y \in (0, 1)$ is a decision of the classifier, δ_{ij} is the Kronecker symbol, $C^a(x)$ are the generated classifiers ($a = 1, 2, \dots, A$) [91].

3.5. Logistic Regression (LR)

Logistic regression can analyze a series of problems whose results are affected by one or more factors. The factors influencing the results are called independent variables, which can be continuous, discrete, or a combination of the two types. Logistic regression coefficients can figure out the influence of independent variables on landslide occurrence [95,96]. LR is defined as the follows:

$$A = \beta_0 + \beta_1 Y_1 + \beta_2 Y_2 + \dots + \beta_n Y_n \quad (7)$$

where A is the linear combination, β_0 is the intercept, $\beta_1, \beta_2, \dots, \beta_n$ are the coefficients of logistic regression, and Y_1, Y_2, \dots, Y_n are the independent variables [97]. Using Equation (3), the landslide probability P is expressed as:

$$P = \frac{\exp(A)}{1 + \exp(A)} \quad (8)$$

4. Results

4.1. Correlation Analysis of Influencing Factors

In the present study, Bel value was used to represent the correlation between landslide and various landslide influencing factors. The results (Table 1) showed that Bel value in the south is the highest (0.181), followed by Bel value in the southwest (0.165), and Bel value in the southeast (0.151). At an altitude of more than 1500 m (0.305) greater regional influence landslide occurred right. The results indicate that slopes between 40–50 have a greater impact on the occurrence of landslides (Bel = 0.309). In terms of plan curvature, Bel value is the highest for the class of 1.44–7.56 (0.279).

In terms of profile curvature, the class of 0.58–1.97 has the highest value (0.288), so it has a greater impact on the landslide occurrence. TWI value between 2 and 3 has the highest Bel value, which is 0.388. Areas within 200 m to rivers are more prone to landslides (Bel = 0.590) because rivers provide wetter soil for landslides. Similarly, areas within 100 m to the road networks have a greater impact on landslides occurrence (Bel = 0.313). This is because the soil near the road becomes more loosely structured during human activity. For the NDVI, the area between 0.07 and 0.09 has a higher impact on landslide occurrence. Residential areas have higher impact on landslide events (Bel = 0.385). In terms of lithology, Bel value of the fourth group is 0.31, indicating that Jurassic rocks were more sensitive to landslide occurrences. The area covered by red clay has a greater impact on landslide occurrence (Bel = 0.614). The highest Bel values of SPI and STI are 0.234 and 0.219, respectively, and the corresponding ranges are 20–30 and 30–40, respectively. The regions within these ranges have a greater impact on the landslide occurrence.

The multicollinearity test of landslide influencing factors is very important in landslide susceptibility mapping. The two most widely used indices in multicollinearity analysis are tolerance and variance inflation factors (VIF). If the tolerance value is <0.1 or the VIF value is >10 , it indicates that there is serious multicollinearity among landslide influencing factors [98–100]. Tolerance and VIF values of 14 landslide influencing factors show that the distance to the river has the smallest tolerance value of 0.715 (>0.1), and the largest VIF value is 1.399 (<10) (Table 2). Therefore, there is no multicollinearity among the 14 landslide influencing factors.

Table 2. Multicollinearity analysis.

Factors	Collinearity Statistics	
	Tolerance	VIF
Slope angle	0.873	1.145
Elevation	0.878	1.139
Aspect	0.865	1.156
STI	0.881	1.135
TWI	0.830	1.205
SPI	0.848	1.180
Profile curvature	0.821	1.219
Plan curvature	0.926	1.080
Distance to rivers	0.715	1.399
Distance to roads	0.869	1.150
Soil	0.954	1.048
NDVI	0.830	1.205
Land use	0.954	1.048
Lithology	0.830	1.205

At the same time, this study also evaluated the importance of 14 landslide influencing factors using the 10-fold cross-validation correlation attribute evaluation method (CAE) in Weka software [101]. The evaluation results are sorted in descending order according to the average merit (Table 3). The results show that all factors have a positive effect on the landslide and can be further analyzed. The distance to rivers has the highest average merit among all influencing factors (AM = 0.378), followed by slope angle (AM = 0.213), lithology (AM = 0.181), distance to roads (AM = 0.173), and elevation (AM = 0.172), TWI (AM = 0.171), SPI (AM = 0.154), aspect (AM = 0.143), soil (AM = 0.143), profile curvature (AM = 0.138), NDVI (AM = 0.103), land use (AM = 0.098), plan curvature (0.042) and STI (0.04).

Table 3. Selection of conditioning factors.

Landslide Conditioning Factor	Average Merit (AM)	Standard Deviation (SD)
Distance to rivers	0.378	±0.015
Slope angle	0.213	± 0.008
Lithology	0.181	± 0.012
Distance to roads	0.173	±0.014
Elevation	0.172	± 0.016
TWI	0.171	± 0.014
SPI	0.154	± 0.015
Aspect	0.143	± 0.012
Soil	0.143	± 0.013
Profile curvature	0.138	± 0.019
NDVI	0.103	± 0.024
Land use	0.098	± 0.013
Plan curvature	0.042	± 0.012
STI	0.04	± 0.015

4.2. Application of Hybrid and Benchmark Model

RS integration can not only use random subspace to construct and aggregate the basic classifier, but also make the basic classifier easier to train than in the smaller subspaces. Therefore, the performance of the base classifier in the RS may be better than the original feature space, and the feature-to-instance ratio is significantly improved. The RSCART model is constructed by using Weka software [102] through optimization and classification steps. In the optimization step, RS is trained to obtain the sub-data set, which is randomly divided by the original data set. The training data set is divided by the iterative method. In the classification step, combined with the CART algorithm, the optimal training data set obtained in the previous step is used for spatial prediction of landslide [28]. In the process of RSCART model construction, the optimal number of iterations was 28, the optimal number of execution slots was 1. Finally, the landslide susceptibility map is developed by using the RSCART model and reclassified into five classes using the natural break method [103,104]. The very high class occupies a very small area of 9.27%. The moderate, very low, low, and high areas are accounting for 30.56%, 13.31%, 27.22% and 19.63%, respectively (Figures 4a and 5).

$$\begin{aligned}
 Z = & (10.866 * \text{Slope angle}_{bel}) + (5.226 * \text{Elevation}_{bel}) \\
 & + (6.428 * \text{Slope aspect}_{bel}) + (0.708 * \text{STI}_{bel}) \\
 & + (4.833 * \text{TWI}_{bel}) + (5.437 * \text{SPI}_{bel}) \\
 & + (4.139 * \text{Profile curvature}_{bel}) + (1.150 * \text{Plan curvature}_{bel}) \\
 & + (2.855 * \text{Distancetorivers}_{bel}) + (1.645 * \text{Distancetoroads}_{bel}) \\
 & + (2.285 * \text{Soil}_{bel}) + (2.390 * \text{NDVI}_{bel}) \\
 & + (1.137 * \text{Landuse}_{bel}) + (1.449 * \text{Lithology}_{bel}) - 10.521
 \end{aligned} \tag{9}$$

According to Equation (8), logistic regression coefficients of all landslide influencing factors are positive (Table 4), which indicate that all influencing factors are positively correlated with the landslide occurrence. Finally, the landslide susceptibility map was divided into five categories according to the natural break method. It can be seen that the area percentage with low susceptibility is the largest (29.42%), followed by moderate, very low and high (22.13%, 22%, and 14.38%, respectively). The area proportion of very high is the smallest (12.06%) (Figures 4c and 5)

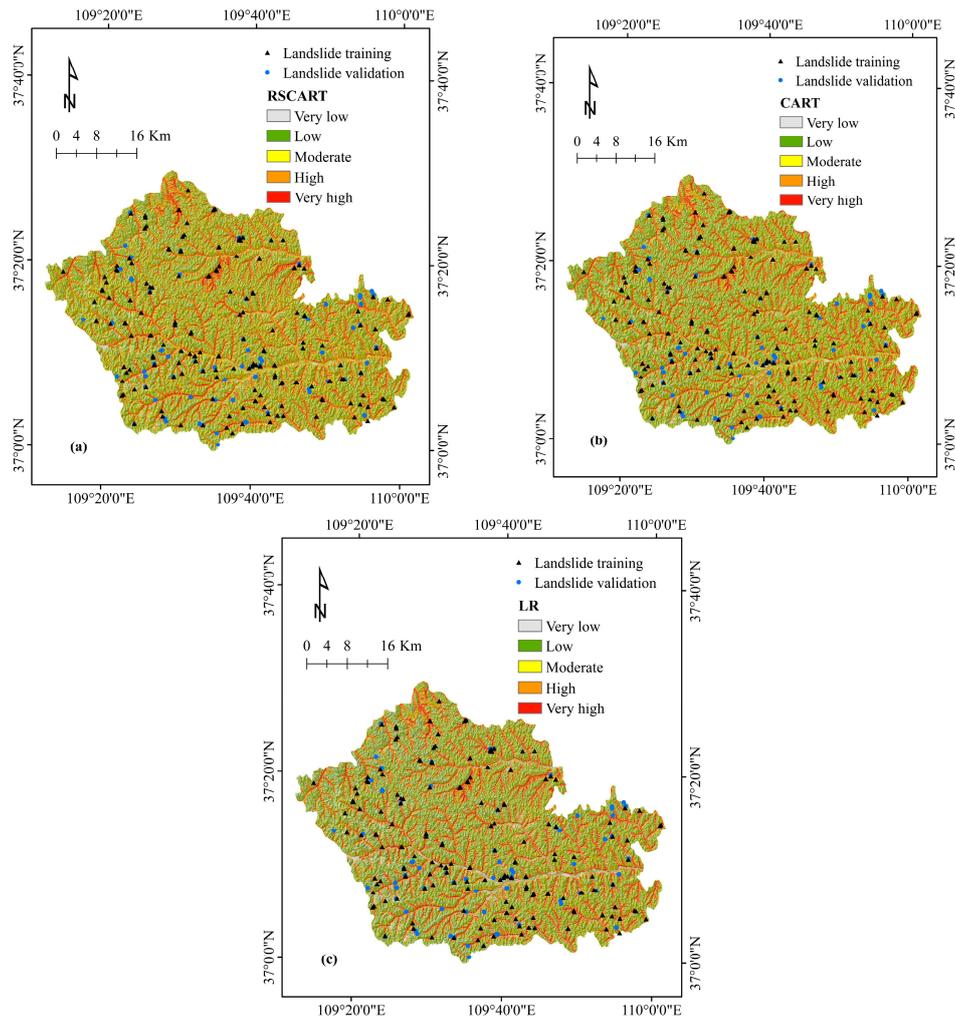


Figure 4. Landslide susceptibility maps: (a) RSCART model; (b) CART model; (c) LR model.

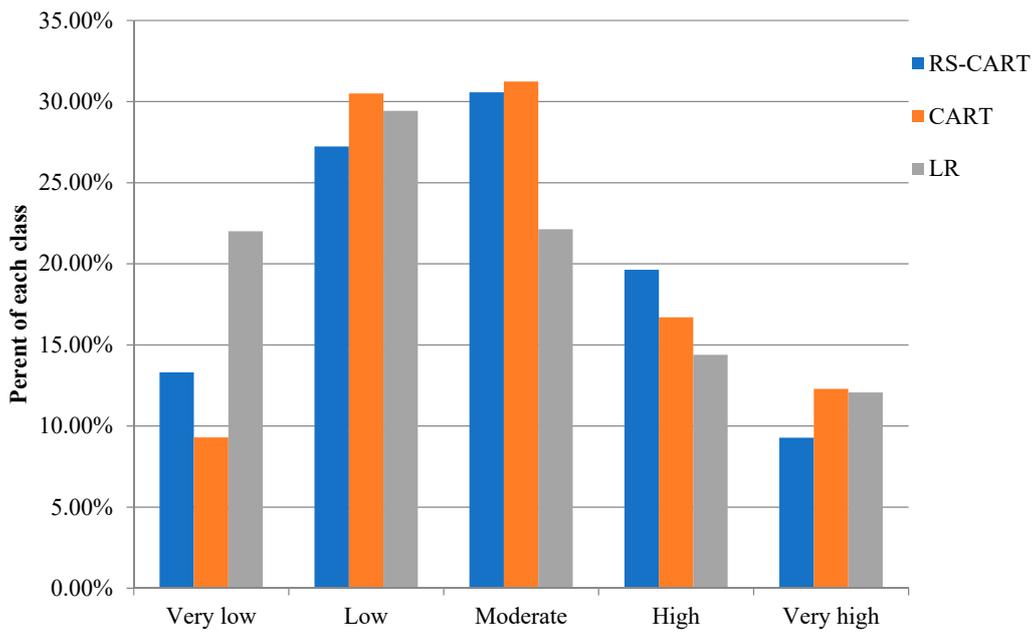


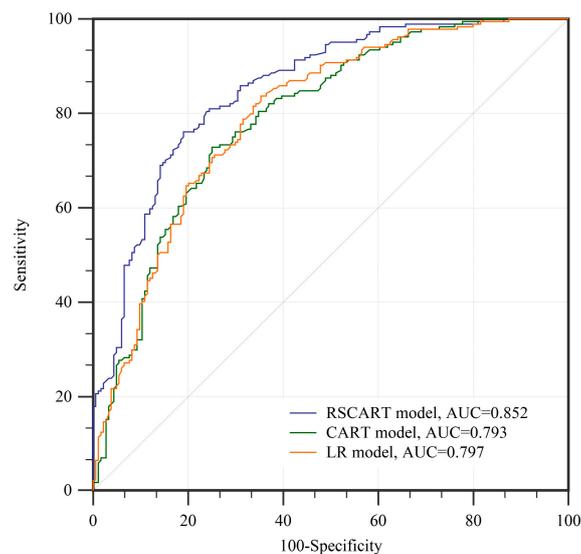
Figure 5. Percentages of landslide susceptibility classes.

Table 4. Coefficients of LR model.

Landslide Influencing Factor	Coefficients
Slope angle	10.866
Elevation	5.226
Aspect	6.428
STI	0.708
TWI	4.833
SPI	5.437
Profile curvature	4.139
Plan curvature	1.150
Distance to rivers	2.855
Distance to roads	1.645
Soil	2.285
NDVI	2.390
Land use	1.137
Lithology	1.449
Intercept	−10.521

4.3. Validation and Comparison of Models

The landslide susceptibility map should have the ability to verify with existing landslide data and predict future landslides [105]. Therefore, in this study, the ROC curve and the AUC are used to assess the prediction capability of models [106–108]. The best models tend to have the highest AUC among the models studied [4,109]. ROC curves and AUC values of the training dataset of the three models are shown in Figure 6. The RSCART model has the highest AUC value (0.852), followed by LR model (0.797) and CART model (0.793). The ROC curves and AUC values of the validation dataset are shown in Figure 7. The prediction accuracy of RSCART model (AUC = 0.827) is higher than that of the LR model (AUC = 0.758) and CART model (AUC = 0.749). Therefore, the comparison of AUC values indicates that the RSCART model is the best of the three models. In addition, the landslide susceptibility map can be verified by calculating the landslide density. LD is defined as the ratio of the percentage of landslide points in each susceptibility classification to the percentage in each susceptibility classification [110]. The landslide density calculation results of the three models are shown in Table 5. RSCART model, CART model, and LR model had the highest values in the very high category, which were 4.264, 3.156 and 3.845, respectively. Then, high (1.743, 2.004, 1.692), moderate (0.634, 0.670, 0.825), low (0.223, 0.212, 0.310) and very low (0.057, 0.041, 0.086).

**Figure 6.** ROC curves using training dataset.

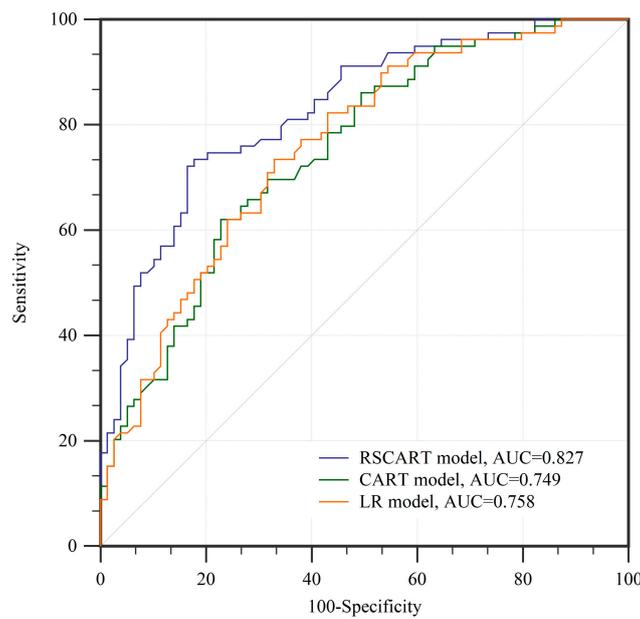


Figure 7. ROC curves using validation dataset.

Table 5. Landslide density analysis on landslide susceptibility maps.

Class	RSCART Model		CART Model		LR Model	
	% Landslides	LD	% Landslides	LD	% Landslides	LD
Very Low	0.760	0.057	0.004	0.041	0.019	0.086
Low	6.084	0.223	0.065	0.212	0.091	0.310
Moderate	19.392	0.634	0.209	0.670	0.183	0.825
High	34.221	1.743	0.335	2.004	0.243	1.692
Very High	39.544	4.264	0.388	3.156	0.464	3.845

4.4. Comparison of Landslide Susceptibility Maps

This study also quantitatively compares the susceptibility values on each pixel to reveal the systematic spatial pattern of the differences between susceptibility maps, following the methodology proposed by [111]. The RSCART model was selected as the benchmark because it has a higher AUC value than the other two models. The susceptibility map of the baseline model is used in a GIS system to pair with the remaining models and subtract their values to define their differences (Figure 8). The values of the comparison map are divided into three levels: “underestimation”, “approximation” and “overestimation”. The values of both comparison maps are broken at -0.2 and 0.2 . The percentage of each grade in the total area is shown in Table 6. At the same time, to explore the key factors influencing the susceptibility difference, overestimation and underestimation statistics were performed for each category of each impact factor (Tables 7 and 8). For each class, we calculate “A” as a percentage of the total area of each class. “B” is the ratio of the underestimated (overestimated) pixels found in this class to the total underestimated (overestimated) pixels, “B–A”, as the difference ratio between the two maps, can be used to identify key class of underestimation (overestimation) anomaly clustering. According to the “B–A” value defined for each class, the class with the highest degree of imbalance was identified (Table 8). To be able to clearly illustrate the relationship between the most imbalanced classification and the underestimated or overestimated area, the visual inspection is required. Underestimations of “RSCART-LR” driven by slope angle (40° – 50°) (Figure 9a), overestimations of “RSCART-LR” driven by slope angle ($<10^{\circ}$) (Figure 10a), Underestimations of “RSCART-CART” driven by distance to rivers (0–200 m) (Figure 9d) and overestimations of “RSCART-LR” driven by slope angle ($<10^{\circ}$) (Figure 10d). As shown in Figure 9a, almost all the underestimated pixels are in the classification with slope of

40–50, and the percentage is 98.55%. In Figure 10a, 99.35% of the overestimation pixels are in the slope of 0°–10°. In Figure 9d, all underestimation pixels are clustered in a distance to rivers of 0–200 m. In Figure 10d, 76.87% of the overestimation pixels are in the classification with a slope less than 10°.

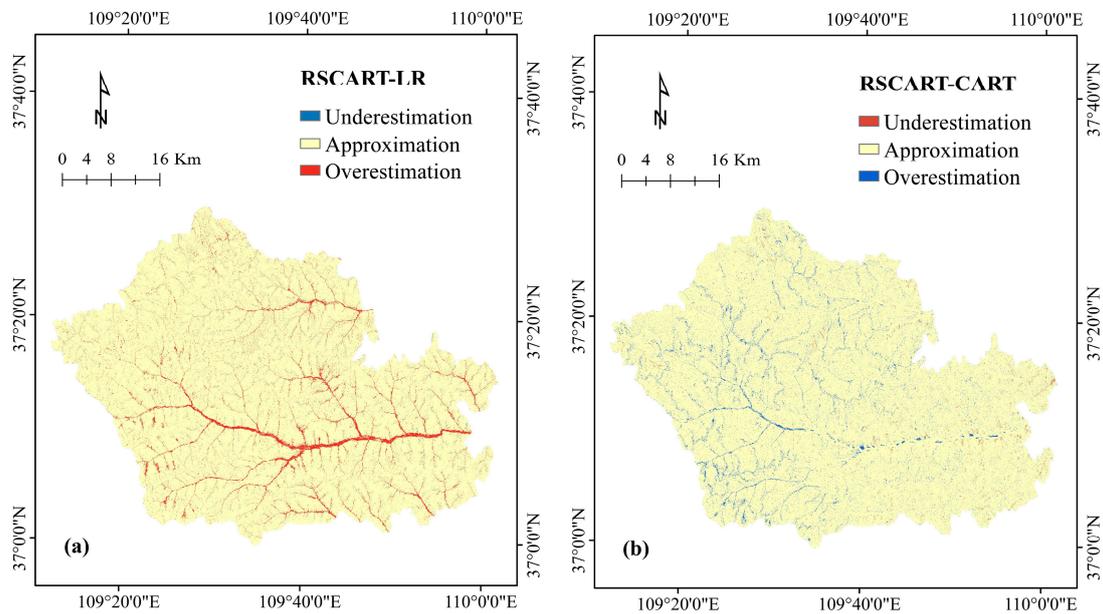


Figure 8. Comparison maps: (a) RSCART-LR; (b) RSCART-CART.

Table 6. Classification of comparison maps.

Comparison	Value	Classification	Percentage
RSCART-LR	−0.27–0.386	Underestimation	−0.27–(−0.2)
		Approximation	−0.2–0.2
		Overestimation	0.2–0.386
RSCART-CART	−0.31–0.42	Underestimation	−0.31–(−0.2)
		Approximation	−0.2–0.2
		Overestimation	0.2–0.42

Table 7. Statistics on underestimation pixels and overestimation pixels of RSCART-LR.

Factors	Class	A (%)	Underestimation RSCART-LRB (%)	B-A (%)	Overestimation RSCART-LRB (%)	B-A (%)
Slope angle	<10	10.44	0.00	−10.44	99.35	88.91
	10–20	26.09	1.44	−24.65	0.00	−26.09
	20–30	35.14	0.00	−35.14	0.38	−34.76
	30–40	23.90	0.01	−23.89	0.05	−23.85
	40–50	4.41	98.55	94.14	0.00	−4.41
	>50	0.02	0.00	−0.02	0.22	0.20
Elevation (m)	933–1000	1.14	1.69	0.55	8.45	7.31
	1000–1100	13.38	23.76	10.38	33.85	20.47
	1100–1200	28.22	32.57	4.35	27.21	−1.01
	1200–1300	31.06	39.37	8.31	19.09	−11.97
	1300–1400	20.45	2.45	−18.00	8.33	−12.12
	1400–1500	5.57	0.11	−5.46	2.70	−2.87
	1500–1574	0.18	0.05	−0.14	0.39	0.20

Table 7. Cont.

Factors	Class	A (%)	Underestimation RSCART-LRB (%)	B-A (%)	Overestimation RSCART-LRB (%)	B-A (%)
Aspect	F	0.05	0.00	−0.05	0.00	−0.05
	N	9.25	7.52	−1.73	10.07	0.82
	NE	13.16	11.01	−2.15	8.06	−5.09
	E	16.34	22.56	6.22	13.25	−3.09
	SE	11.26	7.45	−3.82	14.65	3.39
	S	10.14	6.03	−4.10	19.69	9.55
	SW	12.77	9.48	−3.29	15.88	3.10
	W	15.44	20.99	5.55	12.62	−2.83
	NW	11.59	14.96	3.37	5.78	−5.81
STI	(0–10)	48.27	0.77	−47.50	85.09	36.82
	(10–20)	30.96	59.22	28.25	7.74	−23.22
	(20–30)	11.21	33.37	22.16	3.64	−7.57
	(30–40)	4.22	5.23	1.01	1.80	−2.42
	>40	5.34	1.41	−3.92	1.73	−3.61
TWI	(1.11–2)	56.33	93.66	37.33	0.41	−55.93
	(2–3)	33.16	6.34	−26.82	61.88	28.72
	(3–4)	7.36	0.00	−7.36	21.60	14.24
	(4–5)	2.83	0.00	−2.83	14.79	11.96
	>5	0.31	0.00	−0.31	1.32	1.01
SPI	(0–10)	32.46	0.01	−32.45	61.60	29.13
	(10–20)	19.69	22.99	3.29	9.28	−10.41
	(20–30)	13.57	0.65	−12.92	4.81	−8.76
	(30–40)	8.20	39.35	31.15	3.14	−5.06
	>40	26.08	37.01	10.93	21.17	−4.91
Profile curvature	(−7.29)–(−1.65)	8.08	22.85	14.77	2.80	−5.28
	(−1.65)–(−0.46)	24.10	11.86	−12.24	10.60	−13.50
	(−0.46)–(0.58)	39.33	22.82	−16.51	57.65	18.32
	(0.58)–(1.97)	21.23	29.05	7.82	25.00	3.78
	(1.97)–(9.45)	7.26	13.42	6.16	3.94	−3.33
Plan curvature	(−9.24)–(−1.79)	5.38	1.33	−4.05	3.75	−1.63
	(−1.79)–(−0.54)	17.98	18.25	0.26	12.85	−5.13
	(−0.54)–0.38	42.08	40.14	−1.94	62.01	19.93
	0.38–1.44	26.34	22.84	−3.49	17.30	−9.03
	1.44–7.56	8.22	17.44	9.22	4.09	−4.13
Distance to rivers (m)	0–200	28.64	99.94	71.30	73.01	44.37
	200–400	25.39	0.01	−25.38	11.65	−13.74
	400–600	22.38	0.00	−22.38	7.16	−15.22
	600–800	15.61	0.00	−15.61	4.90	−10.72
	>800	7.98	0.05	−7.93	3.28	−4.69
Distance to roads (m)	0–100	15.20	0.98	−14.23	46.89	31.69
	100–200	11.42	3.96	−7.46	15.81	4.40
	200–300	11.35	7.14	−4.21	8.45	−2.91
	300–400	8.93	9.43	0.50	4.36	−4.57
	>400	53.10	78.49	25.39	24.49	−28.61
Soil	Cultivated loessal soils	85.66	87.94	2.28	59.88	−25.79
	Alluvial soils	11.83	11.31	−0.52	36.29	24.46
	Red clay soils	2.35	0.71	−1.64	3.61	1.26
	Water	0.15	0.04	−0.12	0.22	0.07
NDVI	(−0.15–0.01)	13.96	22.88	8.92	11.12	−2.84
	(0.01–0.04)	16.94	12.85	−4.09	12.55	−4.39
	(0.04–0.07)	22.45	15.18	−7.28	26.68	4.23
	(0.07–0.09)	27.44	32.28	4.83	35.72	8.28
	(0.09–0.31)	19.20	16.82	−2.38	13.94	−5.27

Table 7. Cont.

Factors	Class	A (%)	Underestimation RSCART-LRB (%)	B-A (%)	Overestimation RSCART-LRB (%)	B-A (%)
Land use	Farmland	36.96	16.63	−20.34	35.67	−1.29
	Forestland	18.93	19.87	0.94	16.81	−2.12
	Grassland	43.70	63.37	19.67	44.74	1.04
	Water bodies	0.10	0.09	−0.01	0.36	0.26
	Residential areas	0.29	0.05	−0.24	2.36	2.07
	Others	0.02	0.00	−0.02	0.05	0.04
Lithology	Group 1	75.17	68.36	−6.81	42.36	−32.81
	Group 2	12.38	17.25	4.86	16.10	3.72
	Group 3	0.94	4.90	3.96	0.80	−0.14
	Group 4	6.69	3.19	−3.50	16.22	9.53
	Group 5	4.82	6.30	1.48	24.52	19.70

Table 8. Statistics on underestimation pixels and overestimation pixels of RSCART-CART.

Factors	Class	A (%)	Underestimation RSCART-CART B (%)	B-A (%)	Overestimation RSCART-CART B (%)	B-A (%)
Slope angle	<10	10.44	0.00	−10.44	76.87	66.43
	10–20	26.09	34.83	8.74	6.94	−19.15
	20–30	35.14	11.04	−24.11	14.28	−20.87
	30–40	23.90	8.30	−15.60	1.63	−22.27
	40–50	4.41	45.83	41.42	0.00	−4.41
	>50	0.02	0.00	−0.02	0.29	0.27
Elevation (m)	933–1000	1.14	5.45	4.31	1.26	0.12
	1000–1100	13.38	42.21	28.83	9.86	−3.52
	1100–1200	28.22	22.95	−5.27	23.76	−4.45
	1200–1300	31.06	24.88	−6.18	26.64	−4.42
	1300–1400	20.45	4.37	−16.08	27.01	6.56
	1400–1500	5.57	0.15	−5.42	10.78	5.21
	1500–1574	0.18	0.00	−0.18	0.68	0.50
Aspect	F	0.05	0.00	−0.05	0.00	−0.04
	N	9.25	8.15	−1.10	8.63	−0.62
	NE	13.16	4.51	−8.65	12.34	−0.82
	E	16.34	13.65	−2.69	14.35	−1.99
	SE	11.26	10.49	−0.78	14.01	2.75
	S	10.14	15.24	5.11	14.73	4.60
	SW	12.77	14.38	1.60	14.09	1.32
	W	15.44	27.72	12.28	13.04	−2.40
	NW	11.59	5.86	−5.73	8.80	−2.79
STI	(0–10)	48.27	26.71	−21.56	86.04	37.77
	(10–20)	30.96	46.44	15.48	6.85	−24.11
	(20–30)	11.21	17.54	6.33	3.42	−7.79
	(30–40)	4.22	6.27	2.05	1.50	−2.71
	>40	5.34	3.04	−2.29	2.18	−3.15
TWI	(1.11–2)	56.33	49.57	−6.77	20.92	−35.41
	(2–3)	33.16	49.94	16.78	56.19	23.02
	(3–4)	7.36	0.46	−6.90	12.35	4.99
	(4–5)	2.83	0.04	−2.80	8.99	6.16
	>5	0.31	0.00	−0.31	1.55	1.24
SPI	(0–10)	32.46	2.30	−30.16	73.36	40.90
	(10–20)	19.69	28.85	9.15	5.97	−13.73
	(20–30)	13.57	13.85	0.28	2.93	−10.64
	(30–40)	8.20	25.32	17.12	1.78	−6.41

Table 8. Cont.

Factors	Class	A (%)	Underestimation RSCART-CART B (%)	B-A (%)	Overestimation RSCART-CART B (%)	B-A (%)
	>40	26.08	29.68	3.61	15.96	-10.12
Profile curvature	(-7.29)-(-1.65)	8.08	8.85	0.76	7.63	-0.46
	(-1.65)-(-0.46)	24.10	4.55	-19.55	32.30	8.20
	(-0.46)-(0.58)	39.33	24.42	-14.91	41.48	2.15
	(0.58)-(1.97)	21.23	53.71	32.48	14.45	-6.78
	(1.97)-(9.45)	7.26	8.47	1.21	4.14	-3.12
Plan curvature	(-9.24)-(-1.79)	5.38	4.15	-1.23	3.26	-2.12
	(-1.79)-(-0.54)	17.98	21.91	3.93	11.23	-6.75
	(-0.54)-0.38	42.08	50.18	8.10	43.99	1.91
	0.38-1.44	26.34	18.79	-7.54	31.82	5.48
	1.44-7.56	8.22	4.96	-3.26	9.70	1.48
Distance to rivers (m)	0-200	28.64	100.00	71.36	19.01	-9.63
	200-400	25.39	0.00	-25.39	26.60	1.21
	400-600	22.38	0.00	-22.38	21.30	-1.08
	600-800	15.61	0.00	-15.61	23.98	8.37
	>800	7.98	0.00	-7.98	9.11	1.13
Distance to roads (m)	0-100	15.20	0.14	-15.06	44.63	29.43
	100-200	11.42	1.11	-10.31	13.67	2.26
	200-300	11.35	6.50	-4.85	8.83	-2.52
	300-400	8.93	10.64	1.71	5.58	-3.35
	>400	53.10	81.61	28.52	27.29	-25.81
Soil	Cultivated loessal soils	85.66	75.02	-10.64	82.11	-3.56
	Alluvial soils	11.83	14.70	2.87	14.95	3.12
	Red clay soils	2.35	10.27	7.92	2.61	0.26
	Water	0.15	0.00	-0.15	0.32	0.17
NDVI	(-0.15-0.01)	13.96	14.62	0.66	8.82	-5.14
	(0.01-0.04)	16.94	30.21	13.27	6.53	-10.41
	(0.04-0.07)	22.45	34.02	11.57	14.70	-7.75
	(0.07-0.09)	27.44	12.11	-15.33	43.89	16.44
	(0.09-0.31)	19.20	9.04	-10.16	26.05	6.85
Land use	Farmland	36.96	33.24	-3.72	41.50	4.53
	Forestland	18.93	21.84	2.91	15.58	-3.34
	Grassland	43.70	44.23	0.53	41.66	-2.04
	Water bodies	0.10	0.60	0.50	0.10	0.00
	Residential areas	0.29	0.08	-0.21	1.13	0.84
	Others	0.02	0.00	-0.02	0.03	0.01
Lithology	Group 1	75.17	78.27	3.11	66.03	-9.13
	Group 2	12.38	3.55	-8.83	15.56	3.17
	Group 3	0.94	5.21	4.27	0.26	-0.68
	Group 4	6.69	0.38	-6.31	15.20	8.51
	Group 5	4.82	12.58	7.76	2.95	-1.87

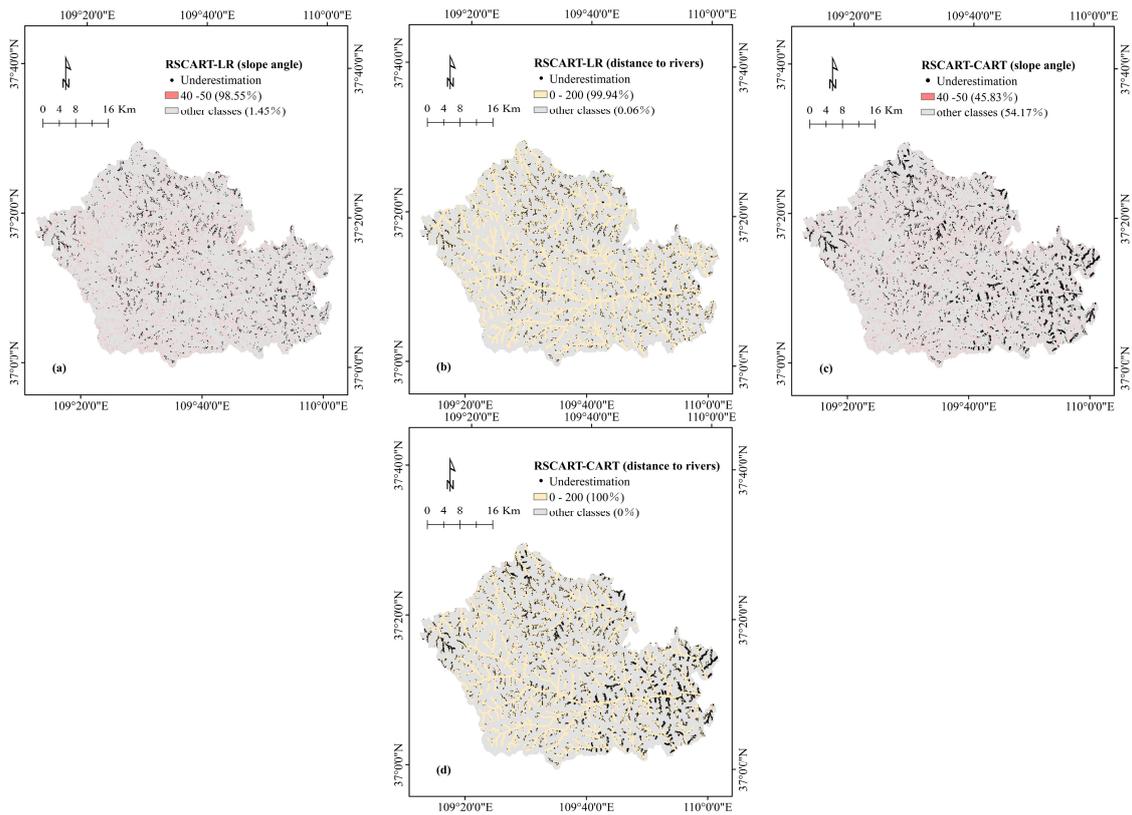


Figure 9. Spatial location of underestimations, in relationship with the imbalanced classes.

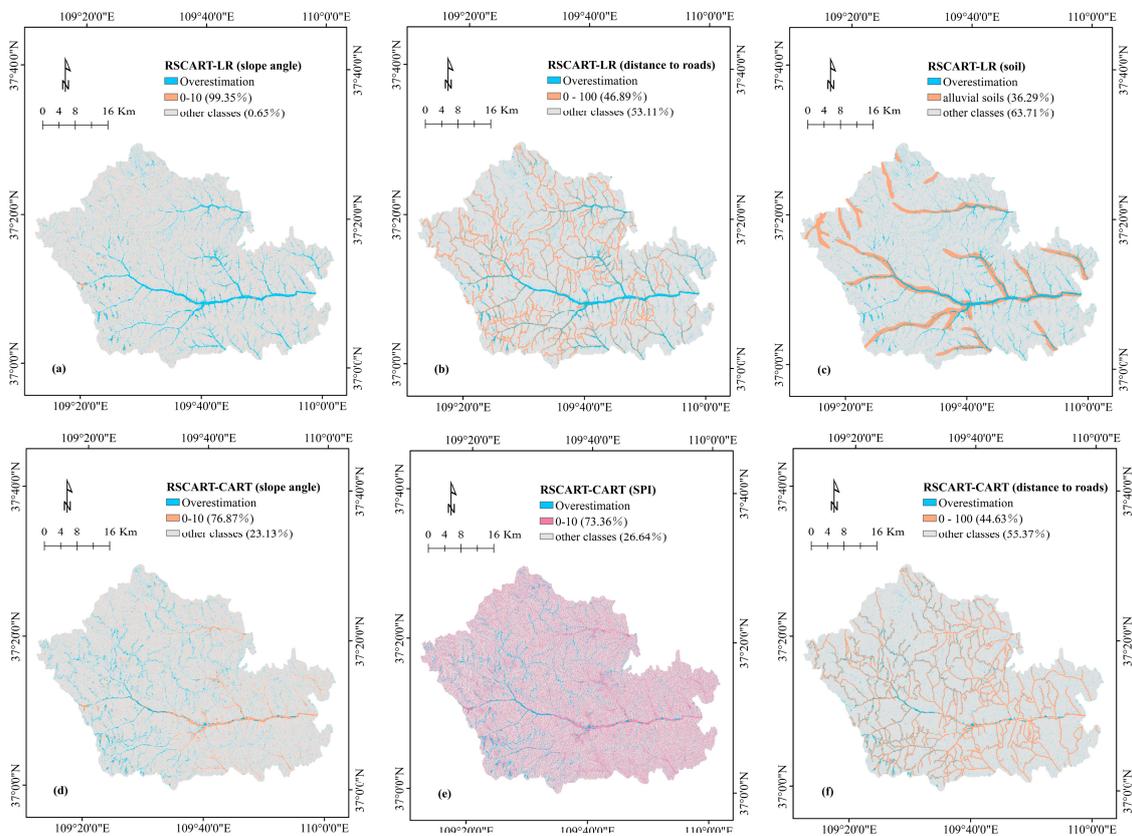


Figure 10. Spatial location of overestimations, in relationship with the imbalanced classes.

5. Discussion

The selection of landslide influencing factors will affect the quality of landslide susceptibility analysis [112,113]. In this study, 14 landslide influencing factors were selected. The EBF model is used to analyze the correlation between the subclasses of landslide influencing factors and landslide, and to reclassify the landslide influence factors. Then, through the multicollinearity analysis, it was found that there was no multicollinearity among all landslide influencing factors (Table 2), and all factors contributed to the model. Finally, the results of the CAE model's importance analysis and Bel values were used to analyze the various landslide influencing factors. According to the results of the importance analysis, the distance to rivers (AM = 0.378) is the most important landslide influencing factor in the study area. The results of the Bel value are similar to those found in previous studies, with a higher probability of landslides near the river area [114]. Different geomorphic parts of the river cause different external forces and stress distributions on the slope, so the failure modes of the slope are also different. The distance to roads (AM = 0.173, Bel = 0.313) has similar results. The closer we get to the road, the higher the probability of a landslide [115]. This is easy to understand because road construction can destabilize the slope by breaking the support of the slope foot [116]. As for the slope (AM = 0.213), the possibility of landslides will increase with the increase of the slope, because the gravity load and stress of the material forming the slope will increase [117]. It can be seen from the Bel value that landslides are most likely to occur in areas with a slope of 40–50. Different types of lithology (AM = 0.181) will result in different slope strengths [118]. According to the Bel value, it can be seen that shale, sandstone, mudstone, and conglomerate have a certain effect on the occurrence of landslides. Elevation (AM = 0.172) has always been a key factor in landslide susceptibility mapping [119,120]. The weathering and shear strength of rocks at different elevations are also different [119,120]. According to the Bel value, it can be concluded that the possibility of landslides of 1500 m–1574 m in the study area is higher. Therefore, it can be judged that landslides are more likely to occur in higher research areas. TWI (AM = 0.171) and SPI (AM = 0.154) are usually indispensable constraints in landslide susceptibility modeling [110,121,122]. Aspect (AM = 0.143, Bel = 0.181) in the south are more likely to cause landslides. This may be due to the more intense solar radiation when the slope is facing south, which may result in different weathering degree. Previous studies have shown that the susceptibility of different types of soil (AM = 0.143) to landslides is also different [101]. According to the Bel value, red clay is the key factor leading to landslides in the study area. The shear strength of red clay soils during rainwater infiltration is reduced, making slopes more prone to landslides [123]. Profile curvature (AM = 0.138) can affect the stress distribution of the slope, and the variations in curvature may be useful to identify depletion and accumulation zones [124]. NDVI (AM = 0.103) is an important influencing factor for landslides. Many studies have shown that plants play an active role in the occurrence of landslides because their root systems can increase soil strength and reduce water infiltration [125–127]. In addition, for those factors whose AM value is less than 1, many scholars have studied their relationship with the occurrence of landslides, and these factors should not be ignored when mapping landslide susceptibility [128–130].

From the results we can observe that the number of overestimation and underestimation is limited (the highest proportion is 0.057). However, it has a great influence on the value of AUC. This is because overestimation and underestimation are not randomly distributed, but there are some spatial patterns [111]. From Figure 2 we can see that the overestimations of the two comparisons have some similarities in spatial distribution. From Table 9 we can find that in the two comparisons, the overestimation has almost the same class of imbalance factors. Slopes less than 10 are classified as the most unbalanced class. From Figure 10, we can find that the classification with a slope of less than 10 has a high degree of overlap with the spatial distribution of the river. Areas with a slope of less than 10 are all closer to the river. For the "RSCART-LR" comparison map, Alluvial soils is also an unbalanced class. In Figure 10c, we can find that the spatial distribution of alluvial soil also overlaps with the river. This is because alluvial soil develops. The classification with the distance to the road less than 100 is also the overestimated imbalanced class of the two comparison maps (Figure 10b,f),

and it has a high spatial overlap with the overestimated pixels. The class with a slope of 40–50 is the most unbalanced class in the “RSCART-LR” comparison, which contains 98.55% of the underestimated pixels (Figure 9a). At the same time, “RSCART-LR” was also significantly affected by the distance to the river less than 200 (“B-A” = 71.3%) (Figure 9b). In the comparison of “RSCART-CART”, the classification with the river distance less than 200 included all the underestimated pixels (Figure 9d) and was also driven by the slope classification of 40–50 (“B-A” = 41.42%). From Figure 9, we can see that the overlap of the spatial distribution of the underestimated pixels with a distance from the river less than 200 is extremely high. The remaining imbalance classes do not significantly dominate the underestimated spatial distribution. In summary, the integrated model RSCART can better use the landform information related to the river and the information closer to the road than the other two models. However, it is easy to ignore the influence of the river itself and the high slope on the susceptibility to landslides.

Table 9. Most imbalanced classes driving the spatial distribution of underestimations and overestimation.

	Comparison Maps	Imbalanced Classes
Underestimation	RSCART-LR	slope, 40–50; STI, 10–20; TWI, 1.11–2; SPI, 30–40; distance to rivers, 0–200; distance to roads, >400
	RSCART-CART	slope, 40–50; elevation, 1000–1100; profile, 0.58–1.97; distance to rivers, 0–200; distance to roads, >400
Overestimation	RSCART-LR	slope, <10; STI, 0–10; TWI, 2–3; SPI, 0–10; distance to roads, 0–100; soil, group2;
	RSCART-CART	slope, <10; STI, 0–10; SPI, 0–10; distance to roads, 0–100

Some studies have proved that LR model is an excellent model for landslide susceptibility research. Polykretis and Chalkias shows that the performance of LR model is better than that of weight of evidence and artificial neural networks in landslide susceptibility mapping in drainage basin of Selinous River [131]. Oh et al. conclude that the LR model has the highest prediction and training accuracy compared with EBF and support vector machine (SVM) in the region surrounding Yongin, South Korea [132]. According to the results of Luc Yen district by Pham et al., the performance of CART model is better than that of LR model [28]. The results also show that the performance of CART model is lower than that of LR model. In addition, some studies have shown that RS is an excellent ensemble algorithm. Hong et al. indicate that the integrated model RSSVM has the optimal performance [55]. In this study, training (goodness of fit) and validation (prediction accuracy) data sets were used to compare the proposed integration model with the benchmark model. The goodness of fit results of the landslide model show that RSCART (AUC = 0.852) is superior to CART (AUC = 0.793) and LR (AUC = 0.797). The validation results show that RSCART model has better prediction accuracy than CART (AUC = 0.749) and LR (AUC = 0.758) models. The RSCART model has the advantages of RS integration and CART classifier. When integrated with the RS model, the performance of the CART model, whose performance is lower than the benchmark LR model, is improved significantly. This is because RS integration constructs the benchmark model by using the random subspace, and the performance of the base classifier in the random subspace is optimized. Therefore, the RSCART model based on machine learning hybrid method is more effective than the single basic model. In addition, the landslide susceptibility figure can be verified by calculating the landslide density (Table 5). In general, higher LD values are associated with higher landslide susceptibility levels. As can be seen from the result of LD, the results of the three models are consistent. The LD value of the very high type is the highest, followed by high, moderate, low, and very low. Therefore, the three base maps of landslide susceptibility are reliable, and the optimal model RSCART also has the highest LD value. Therefore, the model used in this study to generate the landslide susceptibility map has reference significance for the study area. In addition, the selection of factors and the construction of models are also of reference values for similar studies.

6. Conclusions

In this study, the combined CART with RS and two benchmark models (LR and CART) were used to draw three landslide susceptibility maps for Zichang County. The correlation between influencing factors and landslide was evaluated using multicollinearity analysis and EBF method. The AUC value was used to test the performance of the three models. The validating results show that the RSCART model has the highest performance, followed by the LR model and CART model. Finally, this paper also uses a method to quantitatively compare the susceptibility values of each pixel to reveal the systematic spatial pattern of the differences between susceptibility maps. In conclusion, the landslide susceptibility maps compiled in this study are useful for land use and decision-making in landslide-prone areas. In addition, this study also proves the superiority of hybrid model in landslide susceptibility modeling.

Author Contributions: Y.L. and W.C. contributed equally to the work. Y.L. and W.C. collected field data and conducted the landslide susceptibility mapping and analysis. Y.L. and W.C. wrote and revised the manuscript. All the authors discussed the results and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (Grant No. 41807192), Natural Science Basic Research Program of Shaanxi (Program No. 2019JLM-7, ProgramNo. 2019JQ-094), China Postdoctoral Science Foundation (Grant No. 2018T111084, 2017M613168), and Project funded by Shaanxi Province Postdoctoral Science Foundation (Grant No. 2017BSHYDZZ07).

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

References

1. Cruden, D.M. A simple definition of a landslide. *Bull. Eng. Geol. Environ.* **1991**, *43*, 27–29. [[CrossRef](#)]
2. Centre for Research on Epidemiology of Disasters, Université catholique de. *The Human Cost of Natural Disasters 2015: A Global Perspective*; Centre for Research on Epidemiology of Disasters, Université catholique de: Brussels, Belgium, 2015.
3. Van Westen, C.; Van Asch, T.W.; Soeters, R. Landslide hazard and risk zonation—why is it still so difficult? *Bull. Eng. Geol. Environ.* **2006**, *65*, 167–184. [[CrossRef](#)]
4. Chen, W.; Zhang, S.; Li, R.; Shahabi, H. Performance evaluation of the gis-based data mining techniques of best-first decision tree, random forest, and naïve bayes tree for landslide susceptibility modeling. *Sci. Total Environ.* **2018**, *644*, 1006–1018. [[CrossRef](#)] [[PubMed](#)]
5. 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](#)]
6. Canli, E.; Mergili, M.; Thiebes, B.; Glade, T. Probabilistic landslide ensemble prediction systems: Lessons to be learned from hydrology. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 2183–2202. [[CrossRef](#)]
7. Cervi, F.; Berti, M.; Borgatti, L.; Ronchetti, F.; Manenti, F.; Corsini, A. Comparing predictive capability of statistical and deterministic methods for landslide susceptibility mapping: A case study in the northern apennines (reggio emilia province, italy). *Landslides* **2010**, *7*, 433–444. [[CrossRef](#)]
8. Bui, D.T.; Pradhan, B.; Lofman, O.; Revhaug, I.; Dick, Ø.B. Regional prediction of landslide hazard using probability analysis of intense rainfall in the hoa binh province, vietnam. *Nat. Hazards* **2013**, *66*, 707–730.
9. Pradhan, B.; Abokharima, M.H.; Jebur, M.N.; Tehrany, M.S. Land subsidence susceptibility mapping at kinta valley (malaysia) using the evidential belief function model in gis. *Nat. Hazards* **2014**, *73*, 1019–1042. [[CrossRef](#)]
10. Chen, W.; Shahabi, H.; Shirzadi, A.; Hong, H.; Akgun, A.; Tian, Y.; Liu, J.; Zhu, A.X.; Li, S. Novel hybrid artificial intelligence approach of bivariate statistical-methods-based kernel logistic regression classifier for landslide susceptibility modeling. *Bull. Eng. Geol. Environ.* **2019**, *78*, 4397–4419. [[CrossRef](#)]
11. Hong, H.; Naghibi, S.A.; Dashtpajardi, M.M.; Pourghasemi, H.R.; Chen, W. A comparative assessment between linear and quadratic discriminant analyses (lda-qda) with frequency ratio and weights-of-evidence models for forest fire susceptibility mapping in china. *Arab. J. Geosci.* **2017**, *10*, 167. [[CrossRef](#)]

12. Wang, L.-J.; Guo, M.; Sawada, K.; Lin, J.; Zhang, J. A comparative study of landslide susceptibility maps using logistic regression, frequency ratio, decision tree, weights of evidence and artificial neural network. *Geosci. J.* **2016**, *20*, 117–136. [[CrossRef](#)]
13. Lee, S.; Choi, J.; Woo, I. The effect of spatial resolution on the accuracy of landslide susceptibility mapping: A case study in boun, korea. *Geosci. J.* **2004**, *8*, 51. [[CrossRef](#)]
14. Lee, S.; Pradhan, B. Probabilistic landslide hazards and risk mapping on penang island, malaysia. *J. Earth Syst. Sci.* **2006**, *115*, 661–672. [[CrossRef](#)]
15. Dahal, R.K.; Hasegawa, S.; Nonomura, A.; Yamanaka, M.; Masuda, T.; Nishino, K. Gis-based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping. *Environ. Geol.* **2008**, *54*, 311–324. [[CrossRef](#)]
16. Sujatha, E.R.; Rajamanickam, V.; Kumaravel, P.; Saranathan, E. Landslide susceptibility analysis using probabilistic likelihood ratio model—a geospatial-based study. *Arab. J. Geosci.* **2013**, *6*, 429–440. [[CrossRef](#)]
17. Chen, W.; Fan, L.; Li, C.; Pham, B.T. Spatial prediction of landslides using hybrid integration of artificial intelligence algorithms with frequency ratio and index of entropy in nanzheng county, china. *Appl. Sci.* **2020**, *10*, 29. [[CrossRef](#)]
18. Lee, S.; Song, K.-Y.; Oh, H.-J.; Choi, J. Detection of landslides using web-based aerial photographs and landslide susceptibility mapping using geospatial analysis. *Int. J. Remote Sens.* **2012**, *33*, 4937–4966. [[CrossRef](#)]
19. Demir, G.; Aytekin, M.; Akgün, A.; Ikişler, S.B.; Tatar, O. A comparison of landslide susceptibility mapping of the eastern part of the north anatolian fault zone (turkey) by likelihood-frequency ratio and analytic hierarchy process methods. *Nat. Hazards* **2013**, *65*, 1481–1506. [[CrossRef](#)]
20. Devkota, K.C.; Regmi, A.D.; Pourghasemi, H.R.; Yoshida, K.; Pradhan, B.; Ryu, I.C.; Dhital, M.R.; Althuwaynee, O.F. Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in gis and their comparison at mugling–narayanghat road section in nepal himalaya. *Nat. Hazards* **2013**, *65*, 135–165. [[CrossRef](#)]
21. Ozdemir, A.; Altural, T. A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide susceptibility mapping: Sultan mountains, sw turkey. *J. Asian Earth Sci.* **2013**, *64*, 180–197. [[CrossRef](#)]
22. Bui, D.T.; Hoang, N.-D.; Samui, P. Spatial pattern analysis and prediction of forest fire using new machine learning approach of multivariate adaptive regression splines and differential flower pollination optimization: A case study at lao cai province (viet nam). *J. Environ. Manag.* **2019**, *237*, 476–487.
23. Chu, L.; Wang, L.-J.; Jiang, J.; Liu, X.; Sawada, K.; Zhang, J. Comparison of landslide susceptibility maps using random forest and multivariate adaptive regression spline models in combination with catchment map units. *Geosci. J.* **2019**, *23*, 341–355. [[CrossRef](#)]
24. Conoscenti, C.; Agnesi, V.; Cama, M.; Caraballo-Arias, N.A.; Rotigliano, E. Assessment of gully erosion susceptibility using multivariate adaptive regression splines and accounting for terrain connectivity. *Land Degrad. Dev.* **2018**, *29*, 724–736. [[CrossRef](#)]
25. Shafapour Tehrany, M.; Kumar, L.; Neamah Jebur, M.; Shabani, F. Evaluating the application of the statistical index method in flood susceptibility mapping and its comparison with frequency ratio and logistic regression methods. *Geomat. Nat. Hazards Risk* **2019**, *10*, 79–101. [[CrossRef](#)]
26. Nicu, I.C. Application of analytic hierarchy process, frequency ratio, and statistical index to landslide susceptibility: An approach to endangered cultural heritage. *Environ. Earth Sci.* **2018**, *77*, 79. [[CrossRef](#)]
27. Tien Bui, D.; Ho, T.-C.; Pradhan, B.; Pham, B.-T.; Nhu, V.-H.; Revhaug, I. Gis-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with adaboost, bagging, and multiboost ensemble frameworks. *Environ. Earth Sci.* **2016**, *75*, 1–22. [[CrossRef](#)]
28. Pham, B.T.; Prakash, I.; Bui, D.T. Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees. *Geomorphology* **2018**, *303*, 256–270. [[CrossRef](#)]
29. Chen, W.; Xie, X.; Wang, J.; Pradhan, B.; Hong, H.; Tien Bui, D.; Duan, Z.; Ma, J. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena* **2017**, *151*, 147–160. [[CrossRef](#)]

30. Chen, W.; Panahi, M.; Pourghasemi, H.R. Performance evaluation of gis-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (anfis) with genetic algorithm (ga), differential evolution (de), and particle swarm optimization (psa) for landslide spatial modelling. *Catena* **2017**, *157*, 310–324. [[CrossRef](#)]
31. Chen, W.; Pourghasemi, H.R.; Panahi, M.; Kornejady, A.; Wang, J.; Xie, X.; Cao, S. Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. *Geomorphology* **2017**, *297*, 69–85. [[CrossRef](#)]
32. Tsangaratos, P.; Loupasakis, C.; Nikolakopoulos, K.; Angelitsa, V.; Ilia, I. Developing a landslide susceptibility map based on remote sensing, fuzzy logic and expert knowledge of the island of lefkada, greece. *Environ. Earth Sci.* **2018**, *77*, 363. [[CrossRef](#)]
33. Sahana, M.; Sajjad, H. Evaluating effectiveness of frequency ratio, fuzzy logic and logistic regression models in assessing landslide susceptibility: A case from rudraprayag district, india. *J. Mt. Sci.* **2017**, *14*, 2150–2167. [[CrossRef](#)]
34. Chen, W.; Xie, X.; Peng, J.; Wang, J.; Duan, Z.; Hong, H. Gis-based landslide susceptibility modelling: A comparative assessment of kernel logistic regression, naïve-bayes tree, and alternating decision tree models. *Geomat. Nat. Hazards Risk* **2017**, *8*, 950–973. [[CrossRef](#)]
35. Hong, H.; Pradhan, B.; Xu, C.; Tien Bui, D. Spatial prediction of landslide hazard at the yihuang area (china) using two-class kernel logistic regression, alternating decision tree and support vector machines. *Catena* **2015**, *133*, 266–281. [[CrossRef](#)]
36. Wang, Y.; Hong, H.; Chen, W.; Li, S.; Panahi, M.; Khosravi, K.; Shirzadi, A.; Shahabi, H.; Panahi, S.; Costache, R. Flood susceptibility mapping in dingnan county (china) using adaptive neuro-fuzzy inference system with biogeography based optimization and imperialistic competitive algorithm. *J. Environ. Manag.* **2019**, *247*, 712–729. [[CrossRef](#)]
37. Zhang, T.; Han, L.; Chen, W.; Shahabi, H. Hybrid integration approach of entropy with logistic regression and support vector machine for landslide susceptibility modeling. *Entropy* **2018**, *20*, 884. [[CrossRef](#)]
38. Huang, Y.; Zhao, L. Review on landslide susceptibility mapping using support vector machines. *Catena* **2018**, *165*, 520–529. [[CrossRef](#)]
39. Zhang, T.-Y.; Han, L.; Zhang, H.; Zhao, Y.-H.; Li, X.-A.; Zhao, L. Gis-based landslide susceptibility mapping using hybrid integration approaches of fractal dimension with index of entropy and support vector machine. *J. Mt. Sci.* **2019**, *16*, 1275–1288. [[CrossRef](#)]
40. Aditian, A.; Kubota, T.; Shinohara, Y. Comparison of gis-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of ambon, indonesia. *Geomorphology* **2018**, *318*, 101–111. [[CrossRef](#)]
41. Ngadisih; Bhandary, N.P.; Yatabe, R.; Dahal, R.K. Logistic Regression and Artificial Neural Network Models for Mapping of Regional-Scale Landslide Susceptibility in Volcanic Mountains of West Java (Indonesia). *AIP Conf. Proc.* **2016**. [[CrossRef](#)]
42. Hong, H.; Miao, Y.; Liu, J.; Zhu, A.X. Exploring the effects of the design and quantity of absence data on the performance of random forest-based landslide susceptibility mapping. *Catena* **2019**, *176*, 45–64. [[CrossRef](#)]
43. Chen, W.; Xie, X.; Peng, J.; Shahabi, H.; Hong, H.; Bui, D.T.; Duan, Z.; Li, S.; Zhu, A.X. Gis-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method. *Catena* **2018**, *164*, 135–149. [[CrossRef](#)]
44. Lagomarsino, D.; Tofani, V.; Segoni, S.; Catani, F.; Casagli, N. A tool for classification and regression using random forest methodology: Applications to landslide susceptibility mapping and soil thickness modeling. *Environ. Model. Assess.* **2017**, *22*, 201–214. [[CrossRef](#)]
45. Chen, W.; Pourghasemi, H.R.; Naghibi, S.A. Prioritization of landslide conditioning factors and its spatial modeling in shangnan county, china using gis-based data mining algorithms. *Bull. Eng. Geol. Environ.* **2018**, *77*, 611–629. [[CrossRef](#)]
46. Pham, B.T.; Prakash, I.; Dou, J.; Singh, S.K.; Trinh, P.T.; Tran, H.T.; Le, T.M.; Van Phong, T.; Khoi, D.K.; Shirzadi, A. A novel hybrid approach of landslide susceptibility modelling using rotation forest ensemble and different base classifiers. *Geocarto Int.* **2019**. [[CrossRef](#)]

47. Chen, W.; Shirzadi, A.; Shahabi, H.; Ahmad, B.B.; Zhang, S.; Hong, H.; Zhang, N. A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naïve bayes tree classifiers for a landslide susceptibility assessment in langao county, china. *Geomat. Nat. Hazards Risk* **2017**, *8*, 1955–1977. [[CrossRef](#)]
48. Aghdam, I.N.; Pradhan, B.; Panahi, M. Landslide susceptibility assessment using a novel hybrid model of statistical bivariate methods (fr and woe) and adaptive neuro-fuzzy inference system (anfis) at southern zagros mountains in iran. *Environ. Earth Sci.* **2017**, *76*, 237. [[CrossRef](#)]
49. Pham, B.T.; Prakash, I.; Singh, S.K.; Shirzadi, A.; Shahabi, H.; Bui, D.T. Landslide susceptibility modeling using reduced error pruning trees and different ensemble techniques: Hybrid machine learning approaches. *Catena* **2019**, *175*, 203–218. [[CrossRef](#)]
50. Pham, B.T.; Jaafari, A.; Prakash, I.; Bui, D.T. A novel hybrid intelligent model of support vector machines and the multiboost ensemble for landslide susceptibility modeling. *Bull. Eng. Geol. Environ.* **2019**, *78*, 2865–2886. [[CrossRef](#)]
51. Fell, R.; Corominas, J.; Bonnard, C.; Cascini, L.; Leroi, E.; Savage, W.Z. Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. *Eng. Geol.* **2008**, *102*, 99–111. [[CrossRef](#)]
52. Akgun, A. A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: A case study at İzmir, turkey. *Landslides* **2012**, *9*, 93–106. [[CrossRef](#)]
53. Youssef, A.M.; Pourghasemi, H.R.; Pourtaghi, Z.S.; Al-Katheeri, M.M. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at wadi tayyah basin, asir region, saudi arabia. *Landslides* **2016**, *13*, 839–856. [[CrossRef](#)]
54. Felicísimo, Á.M.; Cuartero, A.; Remondo, J.; Quirós, E. Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: A comparative study. *Landslides* **2013**, *10*, 175–189. [[CrossRef](#)]
55. Hong, H.; Liu, J.; Zhu, A.-X.; Shahabi, H.; Pham, B.T.; Chen, W.; Pradhan, B.; Bui, D.T. A novel hybrid integration model using support vector machines and random subspace for weather-triggered landslide susceptibility assessment in the wuning area (china). *Environ. Earth Sci.* **2017**, *76*, 652. [[CrossRef](#)]
56. Mandal, S.; Mandal, K. Bivariate statistical index for landslide susceptibility mapping in the rorachu river basin of eastern sikkim himalaya, india. *Spat. Inf. Res.* **2018**, *26*, 59–75. [[CrossRef](#)]
57. Pradhan, A.; Kim, Y. Evaluation of a combined spatial multi-criteria evaluation model and deterministic model for landslide susceptibility mapping. *Catena* **2016**, *140*, 125–139. [[CrossRef](#)]
58. Varnes, D.J. Slope movement types and processes. In: Schuster RL, Krizek RJ (eds) Landslides, analysis and control, Transportation research board, National Academy of Sciences, Washington, DC. *Spec. Rep.* **1978**, *176*, 11–33.
59. He, Q.; Shahabi, H.; Shirzadi, A.; Li, S.; Chen, W.; Wang, N.; Chai, H.; Bian, H.; Ma, J.; Chen, Y.; et al. Landslide spatial modelling using novel bivariate statistical based naïve bayes, rbf classifier, and rbf network machine learning algorithms. *Sci. Total Environ.* **2019**, *663*, 1–15. [[CrossRef](#)]
60. Chen, W.; Yan, X.; Zhao, Z.; Hong, H.; Bui, D.T.; Pradhan, B. Spatial prediction of landslide susceptibility using data mining-based kernel logistic regression, naive bayes and rbfnetwork models for the long county area (china). *Bull. Eng. Geol. Environ.* **2019**, *78*, 247–266. [[CrossRef](#)]
61. 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](#)]
62. Nefeslioglu, H.A.; Duman, T.Y.; Durmaz, S. Landslide susceptibility mapping for a part of tectonic kelkit valley (eastern black sea region of turkey). *Geomorphology* **2008**, *94*, 401–418. [[CrossRef](#)]
63. 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](#)]
64. Galli, M.; Ardizzone, F.; Cardinali, M.; Guzzetti, F.; Reichenbach, P. Comparing landslide inventory maps. *Geomorphology* **2008**, *94*, 268–289. [[CrossRef](#)]
65. Yalcin, A.; Bulut, F. Landslide susceptibility mapping using gis and digital photogrammetric techniques: A case study from ardesen (ne-turkey). *Nat. Hazards* **2007**, *41*, 201–226. [[CrossRef](#)]
66. Jaafari, A.; Najafi, A.; Pourghasemi, H.; Rezaeian, J.; Sattarian, A. Gis-based frequency ratio and index of entropy models for landslide susceptibility assessment in the caspian forest, northern iran. *Int. J. Environ. Sci. Technol.* **2014**, *11*, 909–926. [[CrossRef](#)]

67. Moore, I.D.; Burch, G.J. Physical basis of the length-slope factor in the universal soil loss equation 1. *Soil Sci. Soc. Am. J.* **1986**, *50*, 1294–1298. [[CrossRef](#)]
68. 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](#)]
69. Yilmaz, C.; Topal, T.; Süzen, M.L. Gis-based landslide susceptibility mapping using bivariate statistical analysis in devrek (zonguldak-turkey). *Environ. Earth Sci.* **2012**, *65*, 2161–2178. [[CrossRef](#)]
70. Pourghasemi, H.R.; Pradhan, B.; Gokceoglu, C. Application of fuzzy logic and analytical hierarchy process (ahp) to landslide susceptibility mapping at haraz watershed, iran. *Nat. Hazards* **2012**, *63*, 965–996. [[CrossRef](#)]
71. Nourani, V.; Pradhan, B.; Ghaffari, H.; Sharifi, S.S. Landslide susceptibility mapping at zonouz plain, iran using genetic programming and comparison with frequency ratio, logistic regression, and artificial neural network models. *Nat. Hazards* **2014**, *71*, 523–547. [[CrossRef](#)]
72. Tang, C.; Zhu, J.; Qi, X.; Ding, J. Landslides induced by the wenchuan earthquake and the subsequent strong rainfall event: A case study in the beichuan area of china. *Eng. Geol.* **2011**, *122*, 22–33. [[CrossRef](#)]
73. Korup, O.; Clague, J.J.; Hermanns, R.L.; Hewitt, K.; Strom, A.L.; Weidinger, J.T. Giant landslides, topography, and erosion. *Earth Planet. Sci. Lett.* **2007**, *261*, 578–589. [[CrossRef](#)]
74. Guo, C.; Qin, Y.; Ma, D.; Xia, Y.; Chen, Y.; Si, Q.; Lu, L. Ionic composition, geological signature and environmental impacts of coalbed methane produced water in china. *Energy Sources Part A Recovery Util. Environ. Effects* **2019**. [[CrossRef](#)]
75. Sharma, L.; Patel, N.; Debnath, P.; Ghose, M. Assessing landslide vulnerability from soil characteristics—A gis-based analysis. *Arab. J. Geosci.* **2012**, *5*, 789–796. [[CrossRef](#)]
76. Hall, F.G.; Townshend, J.R.; Engman, E.T. Status of remote sensing algorithms for estimation of land surface state parameters. *Remote Sens. Environ.* **1995**, *51*, 138–156. [[CrossRef](#)]
77. Restrepo, C.; Vitousek, P.; Neville, P. Landslides significantly alter land cover and the distribution of biomass: An example from the ninole ridges of hawai'i. *Plant Ecol.* **2003**, *166*, 131–143. [[CrossRef](#)]
78. Shafer, G. A mathematical theory of evidence. *Technometrics* **1976**, *20*, 242.
79. Althuwaynee, O.F.; Pradhan, B.; Lee, S. Application of an evidential belief function model in landslide susceptibility mapping. *Comput. Geosci.* **2012**, *44*, 120–135. [[CrossRef](#)]
80. Lee, T.-S.; Chiu, C.-C.; Chou, Y.-C.; Lu, C.-J. Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Comput. Stat. Data Anal.* **2006**, *50*, 1113–1130. [[CrossRef](#)]
81. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. *Classification and Regression Trees*; CRC Press: Boca Raton, FL, USA, 1984.
82. Aertsen, W.; Kint, V.; Van Orshoven, J.; Özkan, K.; Muys, B. Comparison and ranking of different modelling techniques for prediction of site index in mediterranean mountain forests. *Ecol. Model.* **2010**, *221*, 1119–1130. [[CrossRef](#)]
83. Subarkah, P.; Ikhsan, A.N.; Setyanto, A. In The effect of the number of attributes on the selection of study program using classification and regression trees algorithms. In Proceedings of the 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE 2018), Yogyakarta, Indonesia, 15–22 October 2018; pp. 1–5.
84. Pham, B.T.; Bui, D.T.; Prakash, I. Application of Classification and Regression Trees for Spatial Prediction of Rainfall-Induced Shallow Landslides in the Uttarakhand Area (India) Using Gis. In *Climate Change, Extreme Events and Disaster Risk Reduction*; Springer: Cham, Switzerland, 2018; pp. 159–170.
85. Timofeev, R. *Classification and Regression Trees (Cart) Theory and Applications*; Humboldt University: Berlin, Germany, 2004.
86. Barandiaran, I. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 1–22.
87. Kotsiantis, S. Combining bagging, boosting, rotation forest and random subspace methods. *Artif. Intell. Rev.* **2011**, *35*, 223–240. [[CrossRef](#)]
88. Kuncheva, L.I.; Plumpton, C.O. In Choosing parameters for random subspace ensembles for fmri classification. In Proceedings of the International Workshop on Multiple Classifier Systems, Cairo, Egypt, 7–9 April 2010; Springer: Berlin/Heidelberg, Germany, 2010; pp. 54–63.
89. Mielniczuk, J.; Teisseyre, P. Using random subspace method for prediction and variable importance assessment in linear regression. *Comput. Stat. Data Anal.* **2014**, *71*, 725–742. [[CrossRef](#)]

90. Bertoni, A.; Folgieri, R.; Valentini, G. Bio-molecular cancer prediction with random subspace ensembles of support vector machines. *Neurocomputing* **2005**, *63*, 535–539. [[CrossRef](#)]
91. Skurichina, M.; Duin, R.P. Bagging, boosting and the random subspace method for linear classifiers. *Pattern Anal. Appl.* **2002**, *5*, 121–135. [[CrossRef](#)]
92. Chen, W.; Hong, H.; Li, S.; Shahabi, H.; Wang, Y.; Wang, X.; Ahmad, B.B. Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. *J. Hydrol.* **2019**, *575*, 864–873. [[CrossRef](#)]
93. Chandra, S.; Maheshkar, S. Verification of static signature pattern based on random subspace, rep tree and bagging. *Multimed. Tools Appl.* **2017**, *76*, 19139–19171. [[CrossRef](#)]
94. Xia, J.; Dalla Mura, M.; Chanussot, J.; Du, P.; He, X. Random subspace ensembles for hyperspectral image classification with extended morphological attribute profiles. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 4768–4786. [[CrossRef](#)]
95. 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](#)]
96. 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](#)]
97. Chen, W.; Zhao, X.; Shahabi, H.; Shirzadi, A.; Khosravi, K.; Chai, H.; Zhang, S.; Zhang, L.; Ma, J.; Chen, Y.; et al. Spatial prediction of landslide susceptibility by combining evidential belief function, logistic regression and logistic model tree. *Geocarto Int.* **2019**, *34*, 1177–1201. [[CrossRef](#)]
98. Bui, D.T.; Lofman, O.; Revhaug, I.; Dick, O. Landslide susceptibility analysis in the hoa binh province of vietnam using statistical index and logistic regression. *Nat. Hazards* **2011**, *59*, 1413. [[CrossRef](#)]
99. O'Brien, R.M. A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.* **2007**, *41*, 673–690. [[CrossRef](#)]
100. Chen, W.; Tsangaratos, P.; Ilia, I.; Duan, Z.; Chen, X. Groundwater spring potential mapping using population-based evolutionary algorithms and data mining methods. *Sci. Total Environ.* **2019**, *684*, 31–49. [[CrossRef](#)]
101. Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. *Data Mining: Practical Machine Learning Tools and Techniques*; Morgan Kaufmann: Burlington, MA, USA, 2016.
102. Frank, E.; Hall, A.M.; Witten, H.I. *The Weka Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*, 4th ed.; Morgan Kaufmann: Burlington, MA, USA, 2016.
103. Chen, W.; Pradhan, B.; Li, S.; Shahabi, H.; Rizeei, H.M.; Hou, E.; Wang, S. Novel hybrid integration approach of bagging-based fisher's linear discriminant function for groundwater potential analysis. *Nat. Resour. Res.* **2019**, *28*, 1239–1258. [[CrossRef](#)]
104. Chen, W.; Li, H.; Hou, E.; Wang, S.; Wang, G.; Panahi, M.; Li, T.; Peng, T.; Guo, C.; Niu, C.; et al. Gis-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models. *Sci. Total Environ.* **2018**, *634*, 853–867. [[CrossRef](#)]
105. Chen, W.; Shahabi, H.; Zhang, S.; Khosravi, K.; Shirzadi, A.; Chapi, K.; Pham, B.T.; Zhang, T.; Zhang, L.; Chai, H.; et al. Landslide susceptibility modeling based on gis and novel bagging-based kernel logistic regression. *Appl. Sci.* **2018**, *8*, 2540. [[CrossRef](#)]
106. Chen, W.; Hong, H.; Panahi, M.; Shahabi, H.; Wang, Y.; Shirzadi, A.; Pirasteh, S.; Alesheikh, A.A.; Khosravi, K.; Panahi, S.; et al. Spat. Predict. Landslide Susceptibility Using Gis-Based Data Min. Tech. Anfis Whale Optim. Algorithm (Woa) Grey Wolf Optim. (Gwo). *Appl. Sci.* **2019**, *9*, 3755. [[CrossRef](#)]
107. Chen, W.; Panahi, M.; Khosravi, K.; Pourghasemi, H.R.; Rezaie, F.; Parvinnezhad, D. Spatial prediction of groundwater potentiality using anfis ensembled with teaching-learning-based and biogeography-based optimization. *J. Hydrol.* **2019**, *572*, 435–448. [[CrossRef](#)]
108. Zhao, X.; Chen, W. Gis-based evaluation of landslide susceptibility models using certainty factors and functional trees-based ensemble techniques. *Appl. Sci.* **2020**, *10*, 16. [[CrossRef](#)]
109. Chen, W.; Li, Y.; Xue, W.; Shahabi, H.; Li, S.; Hong, H.; Wang, X.; Bian, H.; Zhang, S.; Pradhan, B.; et al. Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating decision tree, and random forest methods. *Sci. Total Environ.* **2020**, *701*, 134979. [[CrossRef](#)] [[PubMed](#)]
110. Pham, B.T.; Bui, D.T.; Prakash, I.; Dholakia, M. Rotation forest fuzzy rule-based classifier ensemble for spatial prediction of landslides using gis. *Nat. Hazards* **2016**, *83*, 97–127. [[CrossRef](#)]

111. Xiao, T.; Segoni, S.; Chen, L.; Yin, K.; Casagli, N. A step beyond landslide susceptibility maps: A simple method to investigate and explain the different outcomes obtained by different approaches. *Landslides* **2019**, 1–14. [[CrossRef](#)]
112. Irigaray, C.; Fernández, T.; El Hamdouni, R.; Chacón, J. Evaluation and validation of landslide-susceptibility maps obtained by a gis matrix method: Examples from the betic cordillera (southern Spain). *Nat. Hazards* **2007**, *41*, 61–79. [[CrossRef](#)]
113. Romer, C.; Ferentinou, M. Shallow landslide susceptibility assessment in a semiarid environment—A quaternary catchment of kwazulu-natal, south africa. *Eng. Geol.* **2016**, *201*, 29–44. [[CrossRef](#)]
114. Pandey, V.K.; Sharma, M.C. Probabilistic landslide susceptibility mapping along tipri to ghuttu highway corridor, garhwal himalaya (india). *Remote Sens. Appl. Soc. Environ.* **2017**, *8*, 1–11. [[CrossRef](#)]
115. Zhou, C.; Yin, K.; Cao, Y.; Ahmed, B.; Li, Y.; Catani, F.; Pourghasemi, H.R. Landslide susceptibility modeling applying machine learning methods: A case study from longju in the three gorges reservoir area, china. *Comput. Geosci.* **2018**, *112*, 23–37. [[CrossRef](#)]
116. Aghdam, I.N.; Varzandeh, M.H.M.; Pradhan, B. Landslide susceptibility mapping using an ensemble statistical index (wi) and adaptive neuro-fuzzy inference system (anfis) model at alborz mountains (iran). *Environ. Earth Sci.* **2016**, *75*, 553. [[CrossRef](#)]
117. Dehnavi, A.; Aghdam, I.N.; Pradhan, B.; Varzandeh, M.H.M. A new hybrid model using step-wise weight assessment ratio analysis (swara) technique and adaptive neuro-fuzzy inference system (anfis) for regional landslide hazard assessment in iran. *Catena* **2015**, *135*, 122–148. [[CrossRef](#)]
118. Gassner, C.; Petschko, H.; Bell, R.; Glade, T. In Effect of lithological data of different scales on modelling landslide susceptibility maps. In Proceedings of the EGU General Assembly Conference Abstracts, Vienna, Austria, 22–27 April 2012; p. 11262.
119. Wu, Y.; Li, W.; Liu, P.; Bai, H.; Wang, Q.; He, J.; Liu, Y.; Sun, S. Application of analytic hierarchy process model for landslide susceptibility mapping in the gangu county, gansu province, china. *Environ. Earth Sci.* **2016**, *75*, 422. [[CrossRef](#)]
120. Hong, H.; Tsangaratos, P.; Ilia, I.; Chen, W.; Xu, C. In Comparing the performance of a logistic regression and a random forest model in landslide susceptibility assessments. The case of Wuyaun Area, China. In *Proceedings of the Workshop on World Landslide Forum*; Ljubljana, Slovenia, 29 May–2 June 2017; Springer: Cham, Switzerland, 2017; pp. 1043–1050.
121. Hue, T.; Duong, T.; Toan, D.; Nghinh, L.; Minh, V.; Pho, N.; Xuan, P.; Hoan, L.; Huyen, N.; Pha, P. *Investigation and Assessment of the Types of Geological Hazard in the Territory of Vietnam and Recommendation of Remedial Measures. Phase II: A Study of the Northern Mountainous Province*; Vietnam Academy of Science and Technology, Institute of Geological Sciences: Hanoi, Vietnam, 2004.
122. Conforti, M.; Pascale, S.; Pepe, M.; Sdao, F.; Sole, A. Denudation processes and landforms map of the camastra river catchment (basilicata–south italy). *J. Maps* **2013**, *9*, 444–455. [[CrossRef](#)]
123. Chen, Y.; Li, B.; Xu, Y.; Zhao, Y.; Xu, J. Field study on the soil water characteristics of shallow layers on red clay slopes and its application in stability analysis. *Arab. J. Sci. Eng.* **2019**, *44*, 5107–5116. [[CrossRef](#)]
124. Pascale, S.; Parisi, S.; Mancini, A.; Schiattarella, M.; Conforti, M.; Sole, A.; Murgante, B.; Sdao, F. In Landslide susceptibility mapping using artificial neural network in the urban area of senise and san costantino albanese (basilicata, southern italy). In Proceedings of the International Conference on Computational Science and Its Applications, Ho Chi Minh City, Vietnam, 24–27 June 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 473–488.
125. Gonzalez-Ollauri, A.; Mickovski, S.B. Plant-soil reinforcement response under different soil hydrological regimes. *Geoderma* **2017**, *285*, 141–150. [[CrossRef](#)]
126. Ordak, M.; Wesolowski, M.; Radecka, I.; Muszynska, E.; Bujalska-Zazdrozny, M. Seasonal variations of mercury levels in selected medicinal plants originating from poland. *Biol. Trace Elem. Res.* **2016**, *173*, 514–524. [[CrossRef](#)]
127. Zhang, C.-B.; Chen, L.-H.; Jiang, J. Why fine tree roots are stronger than thicker roots: The role of cellulose and lignin in relation to slope stability. *Geomorphology* **2014**, *206*, 196–202. [[CrossRef](#)]
128. He, Q.; Xu, Z.; Li, S.; Li, R.; Zhang, S.; Wang, N.; Pham, B.T.; Chen, W. Novel entropy and rotation forest-based credal decision tree classifier for landslide susceptibility modeling. *Entropy* **2019**, *21*, 106. [[CrossRef](#)]
129. Yilmaz, I. A case study from koyulhisar (sivas-turkey) for landslide susceptibility mapping by artificial neural networks. *Bull. Eng. Geol. Environ.* **2009**, *68*, 297–306. [[CrossRef](#)]

130. Truong, X.; Mitamura, M.; Kono, Y.; Raghavan, V.; Yonezawa, G.; Do, T.; Tien Bui, D.; Lee, S. Enhancing prediction performance of landslide susceptibility model using hybrid machine learning approach of bagging ensemble and logistic model tree. *Appl. Sci.* **2018**, *8*, 1046. [[CrossRef](#)]
131. Polykretis, C.; Chalkias, C. Comparison and evaluation of landslide susceptibility maps obtained from weight of evidence, logistic regression, and artificial neural network models. *Nat. Hazards* **2018**, *93*, 249–274. [[CrossRef](#)]
132. Oh, H.-J.; Kadavi, P.R.; Lee, C.-W.; Lee, S. Evaluation of landslide susceptibility mapping by evidential belief function, logistic regression and support vector machine models. *Geomat. Nat. Hazards Risk* **2018**, *9*, 1053–1070. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).