

Article

GIS-Based Evaluation of Landslide Susceptibility Models Using Certainty Factors and Functional Trees-Based Ensemble Techniques

Xia Zhao ^{1,2} and Wei Chen ^{1,2,*} 

¹ College of Geology & Environment, Xi'an University of Science and Technology, Xi'an 710054, China; 19209071022@stu.xust.edu.cn

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

* Correspondence: chenwei0930@xust.edu.cn

Received: 15 October 2019; Accepted: 25 November 2019; Published: 18 December 2019



Abstract: The main purpose of this paper is to use ensembles techniques of functional tree-based bagging, rotation forest, and dagging (functional trees (FT), bagging-functional trees (BFT), rotation forest-functional trees (RFFT), dagging-functional trees (DFT)) for landslide susceptibility modeling in Zichang County, China. Firstly, 263 landslides were identified, and the landslide inventory map was established, and the landslide locations were randomly divided into 70% (training data) and 30% (validation data). Then, 14 landslide conditioning factors were selected. Furthermore, the correlation analysis between conditioning factors and landslides was applied using the certainty factor method. Hereafter, four models were applied for landslide susceptibility modeling and zoning. Finally, the receiver operating characteristic (ROC) curve and statistical parameters were used to evaluate and compare the overall performance of the four models. The results showed that the area under the curve (AUC) for the four models was larger than 0.74. Among them, the BFT model is better than the other three models. In addition, this study also illustrated that the integrated model is not necessarily more effective than a single model. The ensemble data mining technology used in this study can be used as an effective tool for future land planning and monitoring.

Keywords: landslide susceptibility mapping; ensemble techniques; functional trees; bagging; rotation forest; dagging

1. Introduction

A landslide is a complex natural phenomenon [1]. It is influenced by many geological environmental factors, such as topography, landform, geology, land use, and vegetation [2]. A landslide is one of the most familiar and disastrous geological hazards with great destructiveness, which always poses a serious threat to human life, property, and living environment, and restricts human progress and development, especially when geological environments are increasingly affected by human engineering activities [3]. Therefore, landslide prediction is of great significance for landslide prevention and control [4,5]. One of the greatest tasks of landslide disaster and risk mitigation is to prepare landslide susceptibility maps [6].

With the development and progress of the geographic information system (GIS), its application in spatial analysis of landslides is becoming more and more popular. With proper use of GIS, most of the landslide susceptibility mapping methods can realize the automation of evaluation and standardization of data management technology, and enable us to build more efficient and accurate maps [7,8]. This is because these technologies can obtain, query, store, analyze, manipulate,

and display a set of spatial and non-spatial data about landslide conditioning factors [8–10]. Landslide susceptibility zoning mapping technology includes a variety of statistical techniques and statistical methods, including Dempster–Shafer [11–13], entropy [14–16], logistic regression [17–19], certainty factors [20–22], statistical index [23,24], analytic hierarchy process [25–27], frequency ratio [20,28], weight of evidence [29–32], index of entropy [20,33], multivariate adaptive regression spline [34–36], and evidential belief function [37–39].

Landslide susceptibility mapping is a typical complex nonlinear problem in a large area of a landslide research area [5]. Thus, the results obtained by statistical techniques and statistical methods may not be able to achieve satisfactory accuracy [5,40]. Later, many researchers proposed a large number of machine learning techniques for evaluating the susceptibility of landslides, which usually have high prediction accuracy and better performance in data-driven models, such as naive Bayes [41–43], random forests [2,44–46], artificial neural networks [47–50], kernel logistic regression [51,52], support vector machine [53,54], and decision trees [55,56]. However, the performance of machine learning methods is generally influenced by the quality and quantity of training data, and the dependence on modeling parameters is very high [5,57]. So far, it is not clear which method is most suitable for landslide susceptibility mapping [5].

In recent years, hybrid technology is considered to be more effective than single technology [58]. In order to explore more reasonable and perfect research results, a variety of integrated algorithms have been developed for landslide susceptibility modeling [6], such as adaptive neuro-fuzzy inference system [59,60], artificial neural networks-Bayes analysis [61], and Evidential Belief Function-fuzzy logic [62]. The important capability of the integrated model is that the method is more accurate in identification and greatly improves the prediction ability compared with the single machine learning model [6].

The purpose of this study is to propose and validate the ability and effect of ensemble techniques in landslide susceptibility modeling, and functional trees are selected as the base classifier to ensemble with bagging, rotation forest, and dagging models in Zichang County (China). Receiver operating characteristics (ROCs) and statistical parameters were used to evaluate and compare the overall performance of the four models.

2. Study Area and Data Used

2.1. The Study Area

Zichang County is located in the north of Yan'an City, Shaanxi Province, China, between longitudes 109°11'58" E and 110°01'22" E and between latitudes 36°59'30" N and 37°30'00" N, with a total area of 2405 km² (Figure 1). Zichang County is a typical hilly and gully region of the Loess Plateau. The terrain is tilted from northwest to southeast, with an elevation of 933 to 1574 m. Zichang County prevails a warm temperate semi-arid continental monsoon climate, with low temperature and large temperature difference. The annual average temperature within the territory is 9.1 °C, the annual average precipitation is 514.7 mm. The rivers in the territory belong to the Yellow River system, which is divided into three tributaries: Qingjian River, Wuding River, and Yanhe River.

2.2. Data Preparation

The quality of landslide inventory is very significant for landslide susceptibility modeling, and an accurate landslide inventory map is the foundation of landslide susceptibility modeling [63,64]. In this study, three techniques were used to improve the reliability and accuracy of the landslide inventory map: historical report, aerial photo interpretation, and field survey using Global Navigation Satellite Systems (GNSS). According to the landslide inventory map in this area, 263 landslides were identified, and 184 landslide locations (70%) were randomly sampled as the training data and the other 79 landslide locations (30%) were used to validate models.

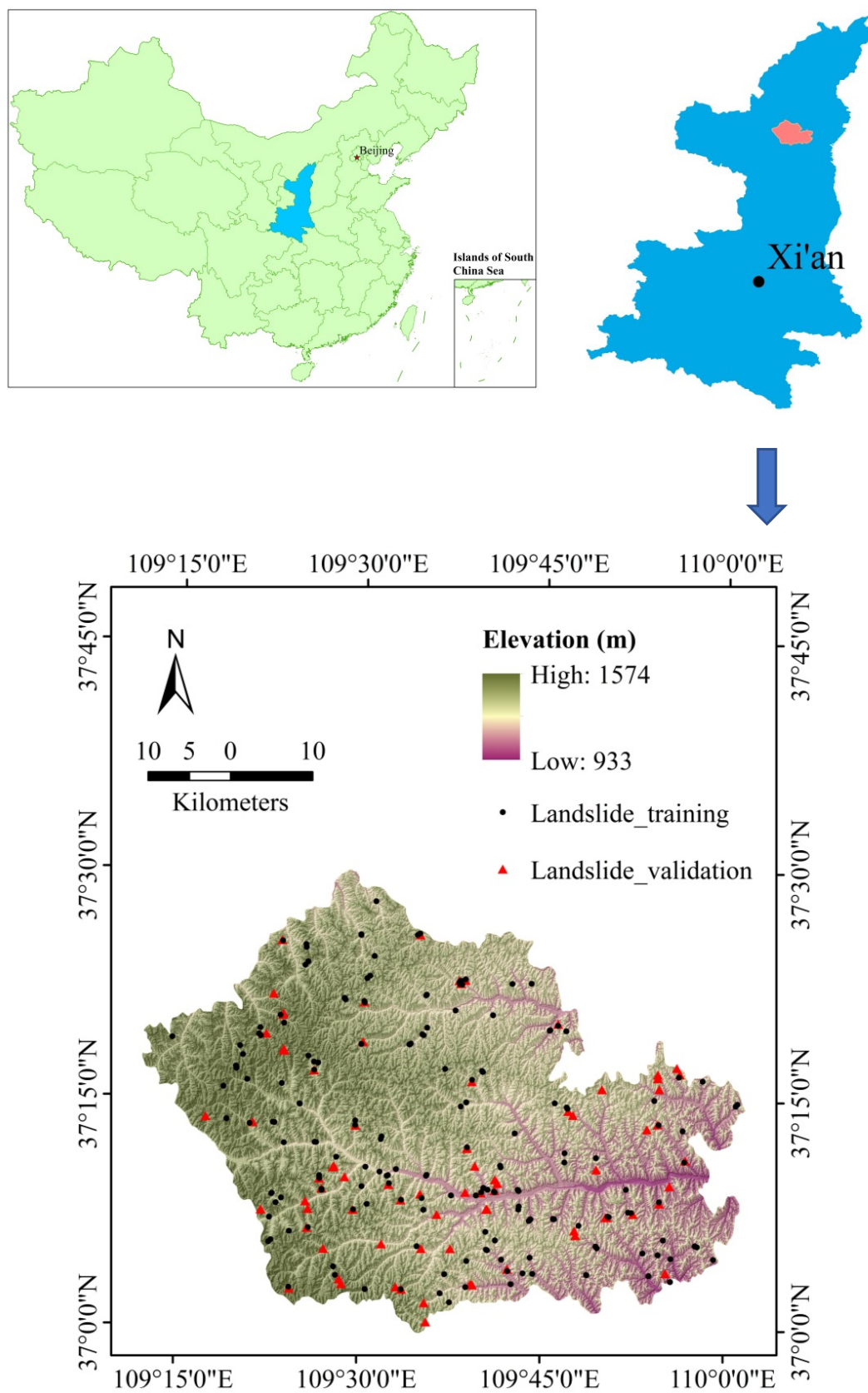


Figure 1. Study area.

After compiling the landslide inventory, it is necessary to choose the landslide conditioning factors to create the landslide susceptibility map [65]. The selection principle is to consider the mechanism and geo-environmental characteristics of landslide occurrence in the study area. Generally, the landslide conditioning factors used to evaluate landslide susceptibility include three categories: topographic factors, geological factors, and environmental factors. In this paper, 14 landslide conditioning factors were selected and transformed into the same resolution (30×30 m), including elevation, slope, aspect, plan curvature, profile curvature, sediment transport index (STI), stream power index (SPI), topographic wetness index (TWI), the normalized difference vegetation index (NDVI), land use, lithology, soil, distance to roads, and distance to rivers (Table 1, Figure 2).

Table 1. Source and scale of conditioning factors.

Factors	Data Source	Format Resolution/Scale
Elevation, slope, aspect, plan curvature, profile curvature, sediment transport index (STI), stream power index (SPI), topographic wetness index (TWI), distance to roads, distance to rivers	ASTER GDEM	Raster, 30 m
Normalized difference vegetation index (NDVI)	Landsat 8 operational land imager	Raster, 30 m
Lithology	Geological maps	Polygon, 1:200,000
Land use	Land use/land cover maps	Polygon, 1:100,000
Soil	Soil type maps	Polygon, 1:1,000,000

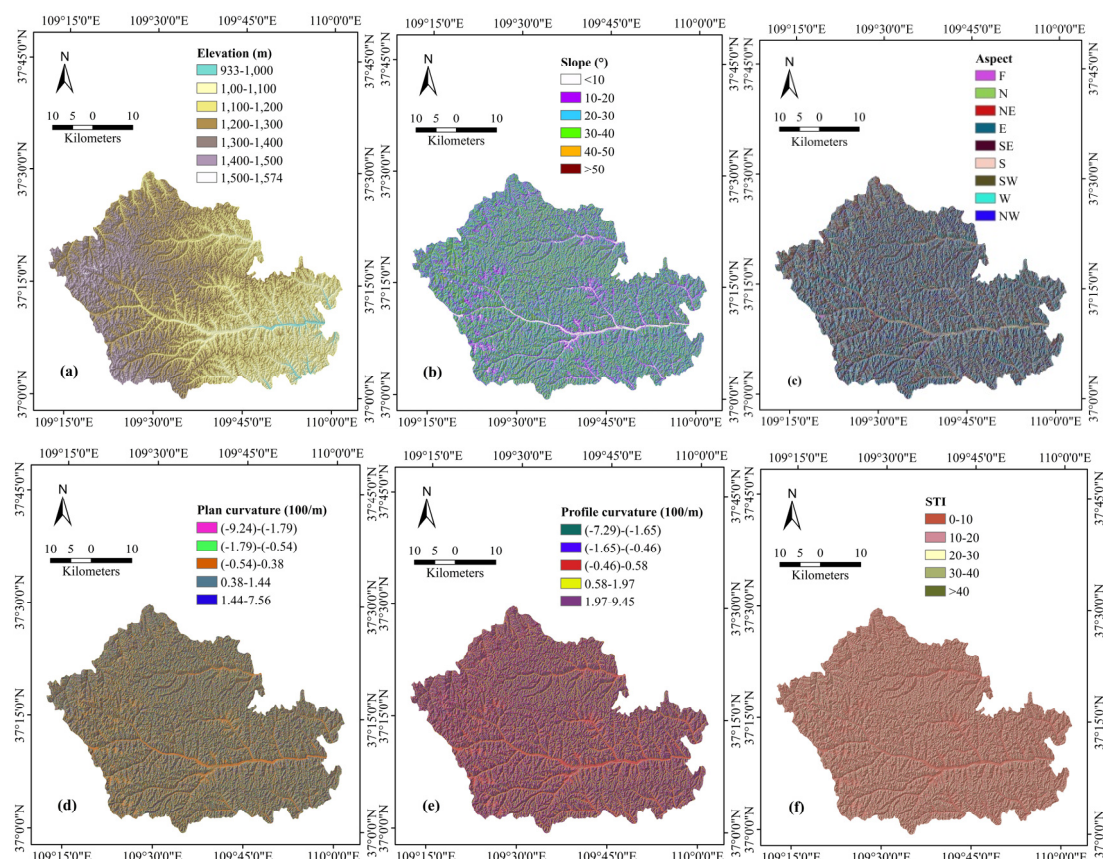


Figure 2. Cont.

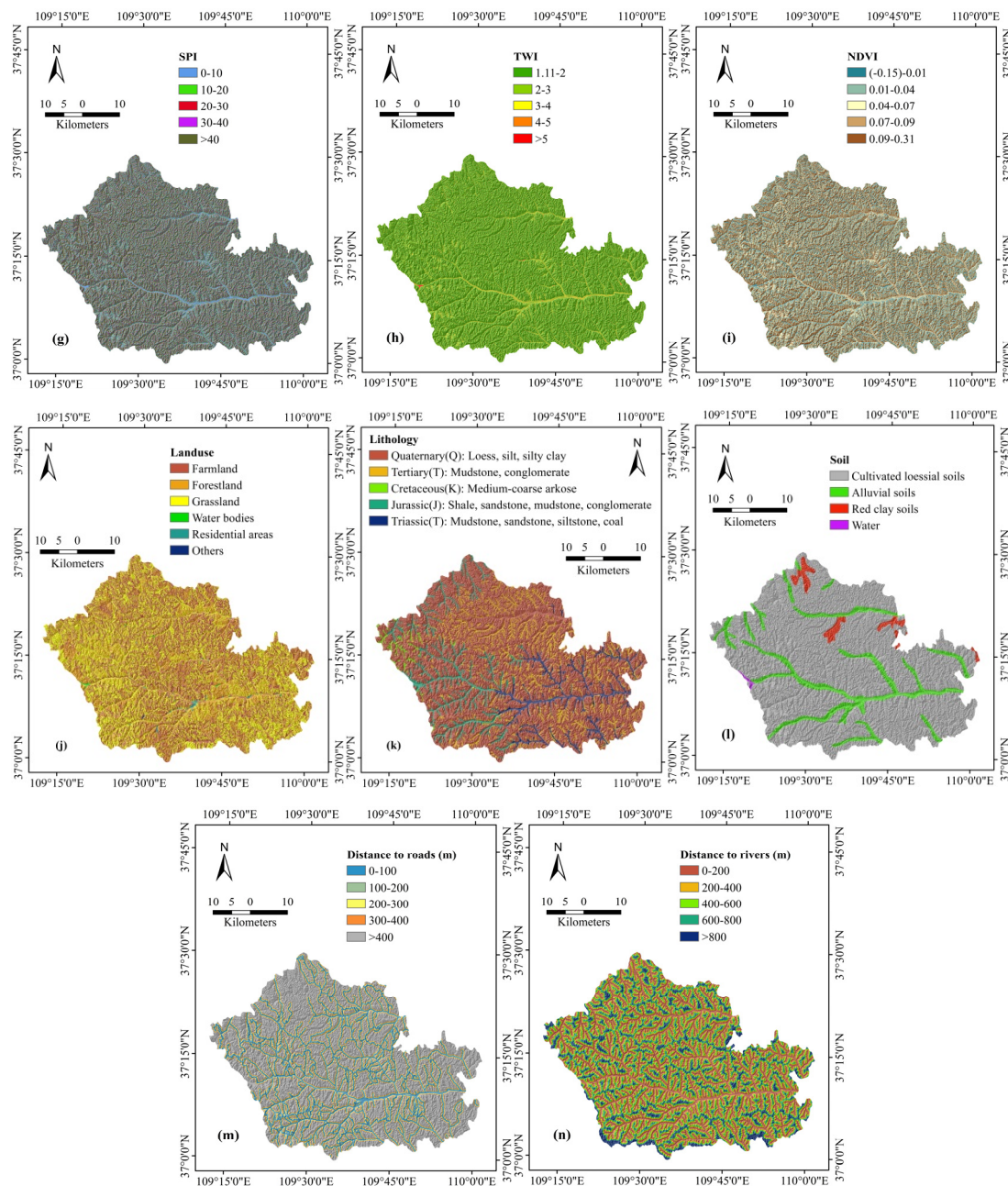


Figure 2. Thematic maps: (a) elevation, (b) slope, (c) aspect, (d) plan curvature, (e) profile curvature, (f) STI, (g) SPI, (h) TWI, (i) NDVI, (j) land use, (k) lithology, (l) soil, (m) distance to roads, (n) distance to rivers.

3. Modeling Approach

The chapter included the illumination of five models, namely certainty factors, functional trees, bagging, rotation forest and dagging. The certainty factors model was used to express the correlation between landslide and conditioning factors, the functional trees model was used as a base classifier, the bagging, rotation forest, and dagging were used as ensemble algorithms.

3.1. Certainty Factors

The certainty factor (CF) belongs to a probability function, which was first proposed in 1990 [66] and modified subsequently [67]. The certainty factor can be expressed as [68]:

$$CF = \begin{cases} \frac{PPa-PPs}{PPa(1-PPs)} & \text{if } PPa \geq PPs \\ \frac{PPa-PPs}{PPs(1-PPa)} & \text{if } PPa < PPs \end{cases} \quad (1)$$

where, PPa is the conditional probability of landslide in class a in study area A , PPs is the prior probability of the total number of landslides in study area A .

The range of CF is -1 to 1 , the positive value indicates that the degree of certainty of landslide occurrence increases, while the negative value indicates that the degree of certainty of landslide occurrence decreases [69–71].

3.2. Functional Trees

Functional trees (FT) are a combination of a discriminant function and multivariable decision tree through constructive induction [72]. Functional trees use logistic regression functions to calculate the splitting of internal nodes (called oblique splitting) and estimation of leaves [73–75]. FT learns the classification tree based on the attributes of leaf nodes, decision nodes or nodes and leaves [38,76]. The decision nodes are built while the trees are growing, while the functional leaves construct when the trees are pruning [76]. Functional trees have the following three usage types: (1) the full functional tree using a regression model for internal nodes and leaves; (2) function tree internal-only uses the regression model for internal nodes; (3) functional tree leaves only use the regression model for leaves [75,76].

In the leaf logic regression function, the logic enhancement (iteration are weighted) of the least-squares function is determined for each output consisting of two classes [77]. Among them, training datasets of D and n samples (A_i, B_i) with $A_i \in R^n$, $B_i \in \{1, 0\}$ [76]. A_i is the input vector containing all landslide condition factors [75,76]; whereas $P(A)$ is the probability prediction value of landslide occurrence; B_i is the coefficient of the i component of the input vector A_i . The posterior probability $P(A)$ of the left ventricle is calculated as follows [78]:

$$f_{B_i}(A) = \sum_{i=1}^{14} B_i A_i + B_0 \quad (2)$$

$$P(A) = \frac{e^{2f_{B_i}(A)}}{1 + e^{2f_{B_i}(A)}} \quad (3)$$

3.3. Bagging

Bagging is based on the concepts of bootstrapping and aggregating, which is used to obtain a more robust and accurate landslide model. Bagging is one of the most popular integration algorithms [79]. The process of a bagging algorithm includes:

Firstly, the bootstrap samples $S(x_i, y_i)$ are randomly resampled from a training set (x_i, y_i) , forming a set of training subsets, where, $x_i \in R$, $y_i \in (\text{landslide, non-landslide})$ [80]. Then, several models based on a classifier are constructed according to each subset, $C_i(x)$ is a classifier constructed from each guiding sample. All models based on classifier (L_i) are aggregated to generate the final model (L') , where, $L_1, L_2, \dots, L_n$ generates a combined classifier (L') . L' predicts the class label of a given instance x by calculating the votes using the following equation [81]:

$$L'(x) = \arg \max_{y \in Y} \sum_{i=1}^t L(C_i(x) = y) \quad (4)$$

3.4. Rotation Forest

Rotation forest (RF) is a popular aggregation technique proposed by Rodriguez et al. [82]. RF is an effective technique for improving weak classifiers [83]. It uses principal component analysis (PCA),

a multivariate technique used for analyzing large multivariate datasets, to reduce its dimensions [84]. In this method, features are extracted from the learning (training) dataset and a base classifier is used to generate learning sub training dataset [82].

For the use of RF: randomly divide the training dataset into D subsets, where D is the parameter of the algorithm, and construct the rotated sparse matrix by performing feature extraction for each subset. The classifier is based on the feature of a repeated matrix projection, and the result is obtained by combining the output of multiple classifiers [84]. RF can be used with any basic classifier, and the feature extraction of each classifier retains all the features that promote variability [84].

In the RF algorithm, $x = (x_1, x_2, \dots, x_n)$ is the training sample set, and Y is the corresponding class label, that is used to consider landslides and non-landslides; $D_1, D_2, \dots, D_L$ are the classifier in the set frame; and P is the set of landslide condition factors. The coefficients of the rotation matrix R_i^a are obtained by transformation and base classifier. Obtain R_i^a by rearranging R_i matrix [84]:

$$R_i = \begin{bmatrix} b_{i,1}^{(1)}, \dots, b_{i,1}^{(M_1)} & 0 & \dots & 0 \\ 0 & b_{i,2}^{(1)}, \dots, b_{i,2}^{(M_2)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & b_{i,K}^{(1)}, \dots, b_{i,K}^{(M_K)} \end{bmatrix} \quad (5)$$

For each sub training dataset extracted by the rotation matrix R_i^a , average grouping method is adopted to obtain the coefficients of each class in a given test sample [85]:

$$\mu_j^{(x)} = \frac{1}{L} \sum_{i=1}^L d_{ij}(xR_i^a), \quad j = 1, \dots, c. \quad (6)$$

where $\mu_j^{(x)}$ is the maximum confidence specified on the class, classifier probability allocation D_i , and the d_{ij} regression $d_{ij}(xR_i^a)$ [85].

3.5. Dagging

Dagging is a well-known resampling integration technique originally proposed by Ting and Witten to generate many disjoint hierarchical folds from a dataset, and each data partition can be sent separately to the basic classifier [86]. The final forecast is based on a majority vote [86]. The main principle is to use a majority vote to combine multiple classifiers to improve the prediction accuracy of the basic classifier [86].

For a given training dataset, which has n samples, the dagging algorithm constructs M datasets (M is a free parameter) from the original training dataset [87]. Each dataset contains n samples [87], and no two datasets have the same sample. A basic classifier is trained for each dataset to build a classification model [87]. Therefore, the M dataset can be summarized into M classification models [86,87].

4. Results

This section consists of the detailed description of the results of the present study, which includes the following four sections: (1) the correlation between landslide and conditioning factors, and then the CF values are used as input to weight the classes of conditioning factors; (2) selection of landslide conditioning factors that are positive to the modeling process; (3) application of four hybrid models and generate landslide susceptibility maps; and (4) validation and comparison of models using ROC and Chi-squared methods.

4.1. Correlation Analysis of Landslide and Conditioning Factors Using the CF Method

The landslide density at each class was calculated by combining each thematic map and landslide inventory map. Meanwhile, this paper summarizes the spatial relationship between the landslides and conditioning factors using the CF method (Table 2). According to the calculation results in Table 2,

the highest CF value (0.661) is found in the elevation category of 1500–1574 m, which indicates that the probability of landslide is the highest. Among the six classes classified by the slope, 40°–50° (0.324) is the highest CF value of the six categories. As far as aspect is concerned, the CF values of slopes facing south (0.309) and southwest (0.242) are the largest. Among the five classes classified by plan curvature, the classes of (−9.24)–(−1.79) have the lowest CF value (−0.495), and the classes of 1.44–7.56 have the highest CF value (0.244). Among the five classes classified by profile curvature, the classes of (−1.65)–(−0.46) have the lowest CF value (−0.346), and the classes of 0.58–1.97 have the highest CF value (0.277). For STI, the frequency of landslide occurrence is the most relevant in 20–30 categories, with the largest CF value (0.220). In TWI, the CF value is the largest in the classes of 2–3 (0.164) and the smallest in the classes of >5 (−1). For NDVI, the lowest CF value (−0.326) was found in the classes of 0.01–0.04, and the highest CF value was found in the categories of 0.07–0.09 (0.223). In terms of land use, landslides mostly occur in residential areas (0.465). Among the five types of lithology, the groups 2 and 4 were relatively more sensitive to landslide occurrence, with CF values of 0.430 and 0.465, respectively. For soil, the majority of landslides occurred in red clay soils with a CF value of 0.712. It can be seen from a distance to roads that the closer the distance is, the more sensitive the landslide. CF value is the largest in the categories of 0–100 m (0.452). For distance to rivers, CF value is the largest in the categories of 0–200 m (0.585).

Table 2. Relationship between landslides and conditioning factors using the certainty factor (CF) method.

Conditioning Factors	Classes	Percentage of Landslide	Percentage of Domain	CF
Elevation (m)	933–1000	2.17	1.14	0.476
	1000–1100	22.28	13.38	0.400
	1100–1200	33.15	28.22	0.149
	1200–1300	29.35	31.06	−0.055
	1300–1400	9.24	20.45	−0.548
	1400–1500	3.26	5.57	−0.415
	1500–1574	0.54	0.18	0.661
Slope (°)	<10	0.00	10.44	−1.000
	10–20	30.43	26.09	0.143
	20–30	36.96	35.14	0.049
	30–40	26.09	23.90	0.084
	40–50	6.52	4.41	0.324
	>50	0.00	0.02	−1.000
Aspect (°)	F (−1)	0.00	0.05	−1.000
	N (0–22.5; 337.5–360)	7.61	9.25	−0.177
	NE (22.5–67.5)	6.52	13.16	−0.504
	E (67.5–112.5)	17.39	16.34	0.060
	SE (112.5–157.5)	13.59	11.26	0.171
	S (157.5–202.5)	14.67	10.14	0.309
	SW (202.5–247.5)	16.85	12.77	0.242
	W (247.5–292.5)	17.93	15.44	0.139
	NW (292.5–337.5)	5.43	11.59	−0.531
Plan curvature (100/m)	(−9.24)–(−1.79)	2.72	5.38	−0.495
	(−1.79)–(−0.54)	15.76	17.98	−0.124
	(−0.54)–0.38	45.11	42.08	0.067
	0.38–1.44	25.54	26.34	−0.030
	1.44–7.56	10.87	8.22	0.244
Profile curvature (100/m)	(−7.29)–(−1.65)	7.07	8.08	−0.126
	(−1.65)–(−0.46)	15.76	24.10	−0.346
	(−0.46)–0.58	41.85	39.33	0.060
	0.58–1.97	29.35	21.23	0.277
	1.97–9.45	5.98	7.26	−0.177
STI	0–10	44.02	48.27	−0.088
	10–20	33.70	30.96	0.081
	20–30	15.22	11.21	0.263
	30–40	3.26	4.22	−0.227
	>40	3.80	5.34	−0.287

Table 2. Cont.

Conditioning Factors	Classes	Percentage of Landslide	Percentage of Domain	CF
SPI	0–10	20.11	32.46	−0.381
	10–20	25.00	19.69	0.212
	20–30	17.39	13.57	0.220
	30–40	10.33	8.20	0.206
	>40	27.17	26.08	0.040
TWI	1.11–2	55.43	56.33	−0.016
	2–3	39.67	33.16	0.164
	3–4	3.80	7.36	−0.483
	4–5	1.09	2.83	−0.617
	>5	0.00	0.31	−1.000
NDVI	(−0.15)–0.01	14.13	13.96	0.012
	0.01–0.04	11.41	16.94	−0.326
	0.04–0.07	16.85	22.45	−0.250
	0.07–0.09	35.33	27.44	0.223
	0.09–0.31	22.28	19.20	0.138
Land use	Farmland	25.54	36.96	−0.309
	Forestland	20.11	18.93	0.059
	Grassland	53.80	43.70	0.188
	Water bodies	0.00	0.10	−1.000
	Residential areas	0.54	0.29	0.465
	Others	0.00	0.02	−1.000
Lithology	1	60.33	75.17	−0.197
	2	21.74	12.38	0.430
	3	0.54	0.94	−0.421
	4	12.50	6.69	0.465
	5	4.89	4.82	0.015
Soil	Cultivated loessial soils	76.63	85.66	−0.105
	Alluvial soils	15.22	11.83	0.223
	Red clay soils	8.15	2.35	0.712
	Water	0.00	0.15	−1.000
Distance to roads (m)	0–100	27.72	15.20	0.452
	100–200	17.39	11.42	0.344
	200–300	11.96	11.35	0.050
	300–400	6.52	8.93	−0.270
	>400	36.41	53.10	−0.314
Distance to rivers (m)	0–200	69.02	28.64	0.585
	200–400	14.67	25.39	−0.422
	400–600	9.24	22.38	−0.587
	600–800	3.26	15.61	−0.791
	>800	3.80	7.98	−0.523

4.2. Selection of Landslide Conditioning Factors

In order to ensure the accuracy of landslide prediction results, it is necessary to remove unimportant or unrelated factors [88,89]. In this study, the Pearson correlation method [90,91] with 10-fold cross-validation was used as an effective feature selection method for evaluating the predictive ability of conditioning factors. The distance to rivers, slope, and lithology has the highest predictive abilities (Table 3). Since a no conditioning factor has a null predictive value, all are included in this analysis.

4.3. Application of Landslide Susceptibility Models

In this study, the training data and CF values were used to construct four models, namely the functional trees (FT) model, bagging-functional trees (BFT) model, rotation forest-functional trees (RFFT) model, and dagging-functional trees (DFT) model, respectively. To get the best performance of the model, the iteration times of the FT model and the minimum number of instances considering the separation of nodes from the training dataset are optimized to 15 and 36, respectively. When building the BFT, RFFT, and DFT models, the two parameters mentioned above were fixed firstly. After completing

the above work, the optimized models were applied to the whole research area to create landslide susceptibility maps. The calculated landslide sensitivity index (LSI) values can be interpreted as the probability in the range of 0 and 1, and all LSI values can be converted to ArcGIS to generate the final landslide susceptibility map.

Table 3. Correlation attribute of landslide conditioning factors.

Factors	Average Merit	Standard Deviation
Distance to rivers	0.382	±0.015
Slope	0.224	±0.008
Lithology	0.180	±0.012
Elevation	0.185	±0.018
Distance to roads	0.172	±0.015
TWI	0.171	±0.014
SPI	0.152	±0.015
Aspect	0.147	±0.011
Soil	0.141	±0.013
Profile curvature	0.136	±0.020
NDVI	0.105	±0.024
Landuse	0.097	±0.013
Plan curvature	0.038	±0.011
STI	0.041	±0.015

Four landslide susceptibility maps generated by FT, BFT, RFFT, and DFT models are shown in Figure 3a–d respectively. The landslide susceptibility maps were reclassified into five classes, namely very low, low, moderate, high, and very high using the natural break method [92]. The comparison of area sizes for each category of the four models is shown in Figure 4. For the FT model, the largest area is the very low class (27.92%), followed by high class (23.47%), very high class (20.21%), low class (17.55%), and the smallest area is the moderate class (10.86%). For the BFT model, the percentages of very low, low, moderate, high, and very high classes are 24.02%, 22.87%, 19.88%, 18.10%, and 15.12%, respectively. The results of landslide susceptibility zoning using the RFFT model show that these percentages are 37.62% (very low), 21.41% (low), 7.79% (moderate), 12.25% (high), and 20.93% (very high), respectively. For the DFT model, the percentages of very low, low, moderate, high, and very high classes are 19.70%, 30.59%, 23.72%, 16.50%, and 9.49%, respectively.

4.4. Model Performances and Comparisons

In this study, the landslide susceptibility models were evaluated by using the areas under the ROC curves (AUC), standard error, 95% confidence interval, and significance level p -value. The ROC curve can be used as a useful tool to indicate the quality of deterministic and probabilistic prediction system [93–95]. The sensitivity (true positive rate) is shown as y-axis and 1-specificity (false positive rate) as x-axis [94,96]. The AUC values are in the range of 0.5 to 1 [97], and the excellent attributes of the model increase with the AUC values [98].

Using the training dataset, the performance of the landslide susceptibility models was evaluated (Table 4). The BFT model has the highest AUC value (0.947), the lowest standard error (0.011), and the narrowest 95% confidence interval (0.925–0.969). It is followed by the RFFT model, the FT model, and the DFT model. For the validation data, the calculation results are shown in Table 5. The BFT model has the highest AUC value (0.804), the lowest standard error (0.035), and the narrowest 95% confidence interval (0.736–0.871). It comes before the DFT model, the FT model, and the RFFT model. These results show that all performance in the validation dataset is slightly worse than those of the training data. These results show that the BFT model is the best model among the four models, and the ensemble model is not necessarily superior to the single model.

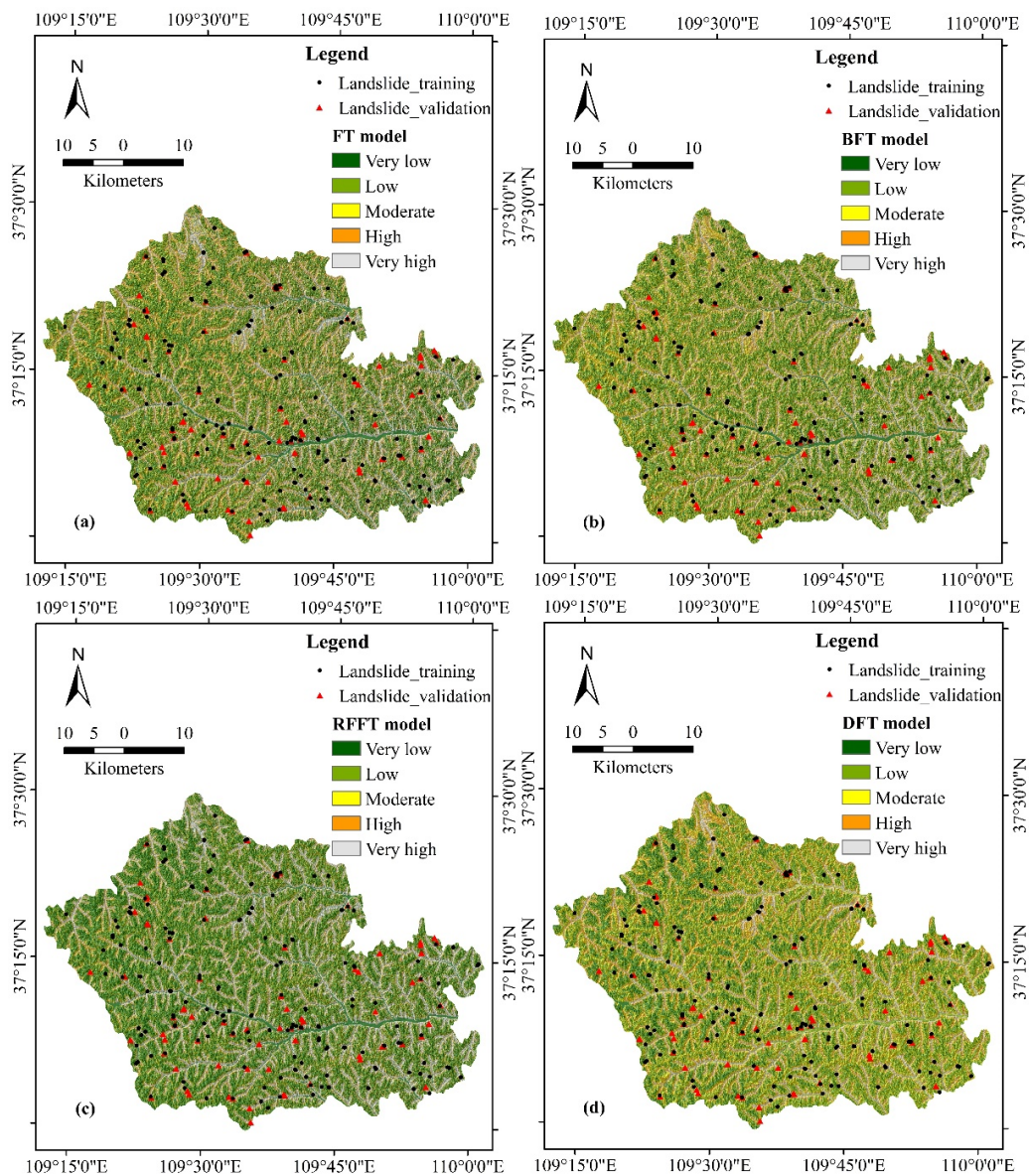


Figure 3. Landslide susceptibility maps: (a) functional trees (FT) model; (b) bagging-functional trees (BFT) model; (c) rotation forest-functional trees (RFFT) model; (d) dagging-functional trees (DFT) model.

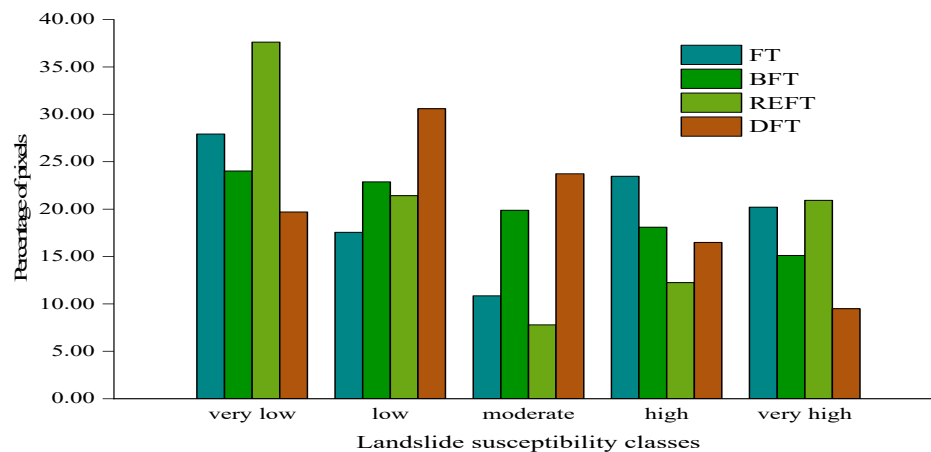


Figure 4. Percentages of landslide susceptibility classes.

Table 4. Parameters of receiver operating characteristic (ROC) curves with the training dataset.

Test Variables	FT Model	BFT Model	RFFT Model	DFT Model
ROC Curve Area	0.836	0.947	0.839	0.797
Standard Error	0.021	0.011	0.020	0.023
95% Confidence Interval	0.795–0.878	0.925–0.969	0.798–0.879	0.752–0.842
<i>p</i> -Value	<0.0001	<0.0001	<0.0001	<0.0001

Table 5. Parameters of ROC curves with the validation dataset.

Test Variables	FT Model	BFT Model	RFFT Model	DFT Model
ROC Curve Area	0.745	0.804	0.740	0.748
Standard Error	0.039	0.035	0.0394	0.039
95% Confidence Interval	0.668–0.822	0.736–0.871	0.663–0.817	0.672–0.824
<i>p</i> -Value	<0.0001	<0.0001	<0.0001	<0.0001

A Chi-squared test was used to analyze the significance of the four models (Table 6). It can be seen that only the comparison of FT and RFFT exhibits lower Chi-squared value (0.044) and higher *p*-value (0.834), which indicate no significant difference between the two models. The other five groups all present larger Chi-squared values and lower *p*-values. The significant differences between the models indicate that the differences between the models are good, which is more conducive to the modeling work and enables this study to obtain the susceptibility results smoothly.

Table 6. Pairwise comparison of four models.

Pair	FT vs. BFT	FT vs. RFFT	FT vs. DFT	BFT vs. RFFT	BFT vs. DFT	RFFT vs. DFT
Chi-squared	40.376	0.044	8.205	44.928	63.681	14.454
<i>p</i> -Value	<0.0001	0.834	0.004	<0.0001	<0.0001	0.000
Significance	Yes	No	Yes	Yes	Yes	Yes

5. Discussions

In this current study, the correlation analysis between conditioning factors and landslides was carried out by the CF method. The probability of landslide occurrence is in inverse correlation with elevation. This may be related to local rainfall and loess and may be related to human engineering activities. With the increase of slope angle, the degree of certainty of landslide occurrence decreases. This may be due to the larger slope angle, the less loose material or more weatherproof material. At the same time, it can be observed that most landslides occur on slopes facing south and southwest with the highest probability. This is mainly because more rain and sunshine are available to the south and landslides are prone to occur. The curvature of plan and profile shows anomalous results. The curvature of the plan (near zero) and convex plan (positive value) are highly sensitive. This anomaly may be related to the overweight effect [28,99,100]. In terms of land use, the probability of landslides in residential areas is the largest, which can explain the impact of human engineering activities on landslides. For the lithology, the second group (Tertiary (T): mudstone, conglomerate) and the fourth group (Triassic (T): mudstone, sandstone, conglomerate) are more sensitive to landslide occurrence. There is groundwater flow in the relatively fractured saturated sandstone and fractured conglomerates, resulting in additional load on the mudstone, resulting in landslides [28,101]. The linear characteristics of the road and river buffers are inversely correlated with landslide susceptibility in the distance. Such an important result has been repeated in many kinds of literature [6,102–104]. However, the remaining five variables make little contributions to the occurrence of landslides.

According to ROC curve analysis (Figures 5 and 6) and statistical index analysis (Tables 4 and 5), it can be concluded that the four machine learning methods selected in the training and testing data assemble a very small *p*-value and significant high performance in the 95% confidence interval. The BFT

model has the highest AUC value (0.947), the lowest standard deviation (0.011), 95% confidence interval (0.925–0.969), and p -value (<0.0001). However, the DFT model has the worst results in this study area. The DFT model has the lowest AUC value (0.797), the highest standard deviation (0.023), 95% confidence interval (0.752–0.842), and p -value (<0.0001). There is no doubt that most ensemble models are superior to single models. However, there is still a phenomenon that the performance of hybrid machine learning methods is not always better than a single model. In order to find more optimal solutions, much more different set models should be applied to the research field.

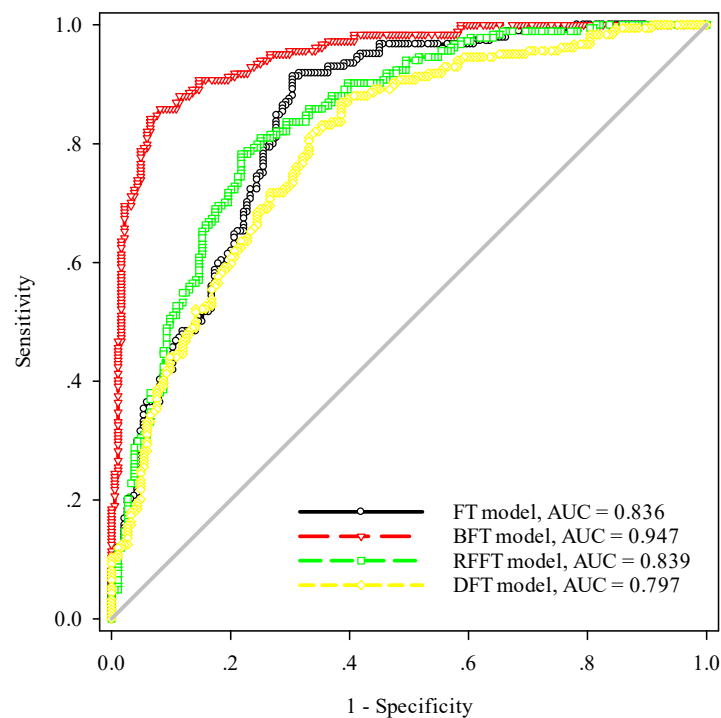


Figure 5. ROC curves of the models using the training dataset. AUC: area under the curve.

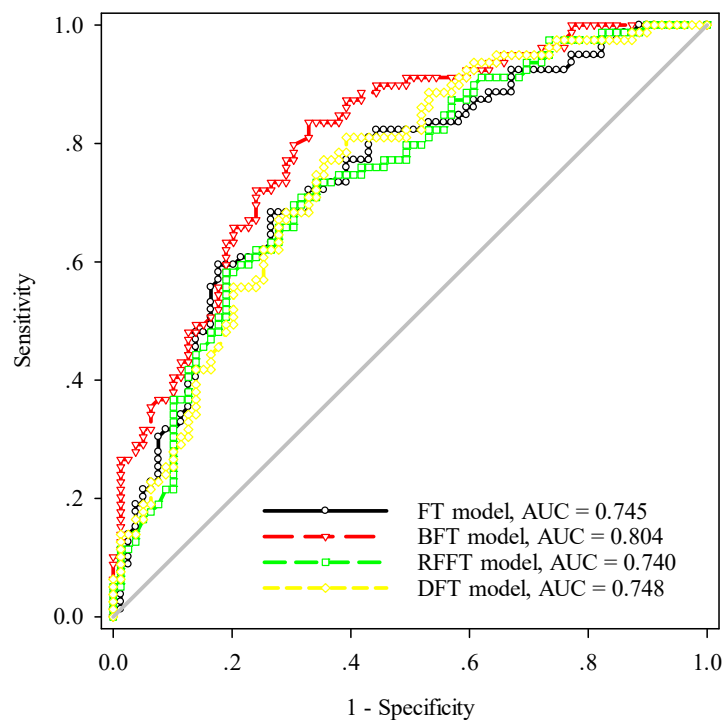


Figure 6. ROC curves of the models using the validation dataset.

According to the paired comparison of the performance of the models (Table 6), the Chi-squared test shows that the Chi-squared values are relatively large. Among them, the Chi-squared value of the FT and RFFT models is smaller, the p -value is larger, and the difference between these two models is not significant. The good results obtained from the other three groups can serve as a powerful basis for modeling in this study. At the same time, the BFT model is compared with the other three models in pairs, and the difference is significant. According to the evaluation results of various evaluation criteria, the performance of the BFT model is better than that of the RFFT model, FT model, and DFT model. As a final recommendation, the obtained results can be useful for policy planning and decision-making in areas prone to landslides. The proposed BFT model, based on performance and prediction accuracy, is suggested in the study area and other regions over the world where they have similar geo-environmental conditions with a logical caution.

6. Conclusions

This study applied functional tree-based ensemble techniques (FT model, BFT model, RFFT model, DFT model) for landslide susceptibility spatial modeling in Zichang County, China. Fourteen conditioning factors and the occurrence of landslides were used to analyze the correlation. Meanwhile, the ROC curve and statistical parameters were used to evaluate and compare the accuracy of the model results. The results showed that the prediction rate of the BFT model is the highest. Therefore, the BFT model is the best optimization ensemble model in this study, and it can be used as an advantageous and promising method for landslide susceptibility modeling. Finally, the landslide susceptibility map generated by this study can be used as an effective tool for future land planning and monitoring by government officials or research experts and scholars.

Author Contributions: X.Z. and W.C. contributed equally to the work. X.Z. and W.C. collected field data and conducted the landslide susceptibility mapping and analysis. X.Z. and W.C. wrote and revised the manuscript. W.C. provided critical comments in planning this paper and edited 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, Program No. 2019JQ-094), China Postdoctoral Science Foundation (Grant No. 2018AT111084, 2017M61318), and project funded by Shaanxi Province Postdoctoral Science Foundation (Grant No. 2017BSHYDZZ07).

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

References

1. Pham, B.T.; Tien Bui, D.; Prakash, I. Bagging based support vector machines for spatial prediction of landslides. *Environ. Earth Sci.* **2018**, *77*, 146. [\[CrossRef\]](#)
2. 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. [\[CrossRef\]](#)
3. 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. Eff.* **2019**, 1–15. [\[CrossRef\]](#)
4. Corominas, J.; Westen, C.V.; Frattini, P.; Cascini, L.; Malet, J.P.; Fotopoulou, S.; Catani, F.; Eeckhaut, M.V.D.; Mavrouli, O.; Agliardi, F. Recommendations for the quantitative analysis of landslide risk. *Bull. Eng. Geol. Environ.* **2014**, *73*, 209–263. [\[CrossRef\]](#)
5. 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. [\[CrossRef\]](#)
6. Chen, W.; Pourghasemi, H.R.; Kornejady, A.; Zhang, N. Landslide spatial modeling: Introducing new ensembles of ann, maxent, and svm machine learning techniques. *Geoderma* **2017**, *305*, 314–327. [\[CrossRef\]](#)
7. He, Y.; Beighley, R.E. Gis-based regional landslide susceptibility mapping: A case study in Southern California. *Earth Surf. Process. Landf.* **2010**, *33*, 380–393. [\[CrossRef\]](#)

8. Hadmoko, D.S.; Lavigne, F.; Samodra, G. Application of a semiquantitative and gis-based statistical model to landslide susceptibility zonation in kayangan catchment, Java, Indonesia. *Nat. Hazards* **2017**, *87*, 437–468. [\[CrossRef\]](#)
9. Saha, A.K.; Gupta, R.P.; Arora, M.K. Gis-based landslide hazard zonation in the bhagirathi (ganga) valley, himalayas. *Int. J. Remote Sens.* **2002**, *23*, 357–369. [\[CrossRef\]](#)
10. Saha, A.K.; Gupta, R.P.; Sarkar, I.; Arora, M.K.; Csaplovics, E. An approach for gis-based statistical landslide susceptibility zonation—with a case study in the Himalayas. *Landslides* **2005**, *2*, 61–69. [\[CrossRef\]](#)
11. Mohammady, M.; Pourghasemi, H.R.; Pradhan, B. Landslide susceptibility mapping at golestan province, iran: A comparison between frequency ratio, dempster–shafer, and weights-of-evidence models. *J. Asian Earth Sci.* **2012**, *61*, 221–236. [\[CrossRef\]](#)
12. Milaghardan, A.H.; Delavar, M.; Chehreghani, A. Uncertainty in landslide occurrence prediction using dempster–shafer theory. *Model. Earth Syst. Environ.* **2016**, *2*, 188. [\[CrossRef\]](#)
13. Mezaal, M.; Pradhan, B.; Rizeei, H. Improving landslide detection from airborne laser scanning data using optimized dempster–shafer. *Remote Sens.* **2018**, *10*, 1029. [\[CrossRef\]](#)
14. Constantin, M.; Bednarik, M.; Jurchescu, M.C.; Vlaicu, M. Landslide susceptibility assessment using the bivariate statistical analysis and the index of entropy in the sibiciu basin (Romania). *Environ. Earth Sci.* **2011**, *63*, 397–406. [\[CrossRef\]](#)
15. Jaafari, A.; Najafi, A.; Pourghasemi, H.R.; Rezaeian, J.; Sattarian, A. Gis-based frequency ratio and index of entropy models for landslide susceptibility assessment in the caspian forest, northern iran. *Int. J. Environ. Sci. Technol.* **2014**, *11*, 909–926. [\[CrossRef\]](#)
16. 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. [\[CrossRef\]](#)
17. Bai, S.-B.; Wang, J.; Lü, G.-N.; Zhou, P.-G.; Hou, S.-S.; Xu, S.-N. Gis-based logistic regression for landslide susceptibility mapping of the zhongxian segment in the three gorges area, China. *Geomorphology* **2010**, *115*, 23–31. [\[CrossRef\]](#)
18. Chen, W.; Pourghasemi, H.R.; Zhao, Z. A gis-based comparative study of dempster–shafer, logistic regression and artificial neural network models for landslide susceptibility mapping. *Geocarto Int.* **2017**, *32*, 367–385. [\[CrossRef\]](#)
19. 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. [\[CrossRef\]](#)
20. Hong, H.; Chen, W.; Xu, C.; Youssef, A.M.; Pradhan, B.; Tien Bui, D. Rainfall-induced landslide susceptibility assessment at the chongren area (China) using frequency ratio, certainty factor, and index of entropy. *Geocarto Int.* **2017**, *32*, 16. [\[CrossRef\]](#)
21. Chen, W.; Li, W.; Chai, H.; Hou, E.; Li, X.; Xiao, D. Gis-based landslide susceptibility mapping using analytical hierarchy process (ahp) and certainty factor (cf) models for the baozhong region of baoji city, China. *Environ. Earth Sci.* **2016**, *75*, 63. [\[CrossRef\]](#)
22. Sujatha, E.R.; Rajamanickam, G.V.; Kumaravel, P. Landslide susceptibility analysis using probabilistic certainty factor approach: A case study on tevakarai stream watershed, India. *J. Earth Syst. Sci.* **2012**, *121*, 1337–1350. [\[CrossRef\]](#)
23. Chen, W.; Chai, H.; Sun, X.; Wang, Q.; Xiao, D.; Hong, H. A gis-based comparative study of frequency ratio, statistical index and weights-of-evidence models in landslide susceptibility mapping. *Arab. J. Geosci.* **2016**, *9*, 204. [\[CrossRef\]](#)
24. Hong, H.; Shahabi, H.; Shirzadi, A.; Chen, W.; Chapi, K.; Ahmad, B.B.; Roodposhti, M.S.; Hesar, A.Y.; Tian, Y.; Tien Bui, D. Landslide susceptibility assessment at the wuning area, China: A comparison between multi-criteria decision making, bivariate statistical and machine learning methods. *Nat. Hazards* **2019**, *96*, 173–212. [\[CrossRef\]](#)
25. Myronidis, D.; Papageorgiou, C.; Theophanous, S. Landslide susceptibility mapping based on landslide history and analytic hierarchy process (ahp). *Nat. Hazards* **2016**, *81*, 1–19. [\[CrossRef\]](#)

26. Gou, C.; Rui, L.; Feng, C. Gis-based landslide susceptibility mapping using analytical hierarchy process in Wenchuan. In Proceedings of the International Conference on Geoinformatics, Kaifeng, China, 20–22 June 2013.
27. Park, S.; Choi, C.; Kim, B.; Kim, J. Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. *Environ. Earth Sci.* **2013**, *68*, 1443–1464. [[CrossRef](#)]
28. 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](#)]
29. Hong, H.; Ilia, I.; Tsangaratos, P.; Wei, C.; Chong, X. A hybrid fuzzy weight of evidence method in landslide susceptibility analysis on the wuyuan area, China. *Geomorphology* **2017**, *290*, 1–16. [[CrossRef](#)]
30. Polykretis, C.; Chalkias, C. Comparison and evaluation of landslide susceptibility maps obtained from weight of evidence, logistic regression, and artificial neural network models. *Nat. Hazards* **2018**, *93*, 1–26. [[CrossRef](#)]
31. 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](#)]
32. Regmi, Deep, A.; Devkota; Chandra, K.; Yoshida; Kohki; Pradhan; Biswajeet; Pourghasemi; Reza, H. Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in central nepal himalaya. *Arab. J. Geosci.* **2014**, *7*, 725–742. [[CrossRef](#)]
33. Chen, W.; Li, W.; Hou, E.; Bai, H.; Chai, H.; Wang, D.; Cui, X.; Wang, Q. Application of frequency ratio, statistical index, and index of entropy models and their comparison in landslide susceptibility mapping for the baozhong region of baoji, China. *Arab. J. Geosci.* **2015**, *8*, 1829–1841. [[CrossRef](#)]
34. Samui, P. Multivariate adaptive regression spline (mars) for prediction of elastic modulus of jointed rock mass. *Geotech. Geol. Eng.* **2013**, *31*, 249–253. [[CrossRef](#)]
35. 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](#)]
36. Alreja, J.; Parab, S.; Mathur, S.; Samui, P. Estimating hysteretic energy demand in steel moment resisting frames using multivariate adaptive regression spline and least square support vector machine. *Ain Shams Eng. J.* **2015**, *6*, 449–455. [[CrossRef](#)]
37. Gayen, A.; Saha, S. Application of weights-of-evidence (woe) and evidential belief function (ebf) models for the delineation of soil erosion vulnerable zones: A study on pathro river basin, jharkhand, India. *Model. Earth Syst. Environ.* **2017**, *3*, 1123–1139. [[CrossRef](#)]
38. Zhang, Z.; Fan, Y.; Han, C.; Wu, Y.; Tao, L.; Li, W.; Wang, Q.; Ping, L. Gis-based landslide susceptibility analysis using frequency ratio and evidential belief function models. *Environ. Earth Sci.* **2016**, *75*, 948. [[CrossRef](#)]
39. 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](#)]
40. Tien Bui, D.; Bui, Q.T.; Nguyen, Q.P.; Pradhan, B.; Nampak, H.; Trinh, P.T. A hybrid artificial intelligence approach using gis-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. *Agric. For. Meteorol.* **2017**, *233*, 32–44. [[CrossRef](#)]
41. Chen, W.; Yan, X.; Zhao, Z.; Hong, H.; Tien Bui, D.; 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](#)]
42. Pham, B.T.; Tien Bui, D.; Pourghasemi, H.R.; Indra, P.; Dholakia, M. Landslide susceptibility assessment in the uttarakhand area (India) using gis: A comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. *Theor. Appl. Climatol.* **2017**, *128*, 255–273. [[CrossRef](#)]
43. Tsangaratos, P.; Ilia, I. Comparison of a logistic regression and naïve bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size. *Catena* **2016**, *145*, 164–179. [[CrossRef](#)]

44. Krkač, M.; Špoljarić, D.; Bernat, S.; Arbanas, S.M.J.L. Method for prediction of landslide movements based on random forests. *Landslides* **2017**, *14*, 947–960. [[CrossRef](#)]
45. Paudel, U.; Oguchi, T.; Hayakawa, Y. Multi-resolution landslide susceptibility analysis using a dem and random forest. *Int. J. Geosci.* **2016**, *7*, 726. [[CrossRef](#)]
46. 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](#)]
47. Arnone, E.; Francipane, A.; Noto, L.V.; Scarbaci, A.; Loggia, G.L. Strategies investigation in using artificial neural network for landslide susceptibility mapping: Application to a sicilian catchment. *J. Hydroinform.* **2014**, *16*, 502. [[CrossRef](#)]
48. Tan, L.; Chen, G.; Zeng, R.Q.; Xiong, M.Q.; Meng, X.M. Application of artificial neural network in landslide susceptibility assessment. *J. Lanzhou Univ.* **2014**, *50*, 15–20.
49. Lee, S.; Ryu, J.H.; Min, K.; Won, J.S. Landslide susceptibility analysis using gis and artificial neural network. *Earth Surf. Process. Landf.* **2010**, *28*, 1361–1376. [[CrossRef](#)]
50. Chen, W.; Hong, H.; Panahi, M.; Shahabi, H.; Wang, Y.; Shirzadi, A.; Pirasteh, S.; Alesheikh, A.A.; Khosravi, K.; Panahi, S.; et al. Spatial prediction of landslide susceptibility using gis-based data mining techniques of anfis with whale optimization algorithm (woa) and grey wolf optimizer (gwo). *Appl. Sci.* **2019**, *9*, 3755. [[CrossRef](#)]
51. 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](#)]
52. Chen, W.; Peng, J.; Hong, H.; Shahabi, H.; Pradhan, B.; Liu, J.; Zhu, A.X.; Pei, X.; Duan, Z. Landslide susceptibility modelling using gis-based machine learning techniques for Chongren County, Jiangxi Province, China. *Sci. Total Environ.* **2018**, *626*, 230. [[CrossRef](#)] [[PubMed](#)]
53. Chen, W.; Pourghasemi, H.R.; Naghibi, S.A. A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy data mining models in China. *Bull. Eng. Geol. Environ.* **2017**, *77*, 1–18. [[CrossRef](#)]
54. 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](#)]
55. Lombardo, L.; Cama, M.; Conoscenti, C.; Marker, M.; Rotigliano, E. Binary logistic regression versus stochastic gradient boosted decision trees in assessing landslide susceptibility for multiple-occurring landslide events: Application to the 2009 storm event in messina (Sicily, Southern Italy). *Nat. Hazards* **2015**, *79*, 1621–1648. [[CrossRef](#)]
56. Tien Bui, D.; Pradhan, B.; Lofman, O.; Revhaug, I. Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and naive bayes models. *Math. Probl. Eng.* **2012**, 974638. [[CrossRef](#)]
57. Ghorbanzadeh, O.; Rostamzadeh, H.; Blaschke, T.; Gholaminia, K.; Aryal, J. A new gis-based data mining technique using an adaptive neuro-fuzzy inference system (anfis) and k-fold cross-validation approach for land subsidence susceptibility mapping. *Nat. Hazards* **2018**, *94*, 497–517. [[CrossRef](#)]
58. 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](#)]
59. Aghdam, I.N.; 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](#)]
60. Polykretis, C.; Chalkias, C.; Ferentinou, M. Adaptive neuro-fuzzy inference system (anfis) modeling for landslide susceptibility assessment in a mediterranean hilly area. *Bull. Eng. Geol. Environ.* **2019**, *78*, 1173–1187. [[CrossRef](#)]
61. Lee, S.; Ryu, J.H.; Won, J.S.; Park, H.J. Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. *Eng. Geol.* **2004**, *71*, 289–302. [[CrossRef](#)]
62. Tien Bui, D.; Pradhan, B.; Revhaug, I.; Nguyen, D.B.; Pham, H.V.; Bui, Q.N. A novel hybrid evidential belief function-based fuzzy logic model in spatial prediction of rainfall-induced shallow landslides in the Lang son city area (Vietnam). *Geomat. Nat. Hazards Risk* **2015**, *6*, 243–271.

63. 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. [\[CrossRef\]](#)
64. Chen, W.; Shahabi, H.; Zhang, S.; Khosravi, K.; Shirzadi, A.; Chapi, K.; Pham, B.; Zhang, T.; Zhang, L.; Chai, H. Landslide susceptibility modeling based on gis and novel bagging-based kernel logistic regression. *Appl. Sci.* **2018**, *8*, 2540. [\[CrossRef\]](#)
65. Süzen, M.L.; Kaya, B.Ş. Evaluation of environmental parameters in logistic regression models for landslide susceptibility mapping. *Int. J. Digit. Earth* **2012**, *5*, 338–355. [\[CrossRef\]](#)
66. Shortliffe, E.H. A model of inexact reasoning in medicine. *Math. Biosci.* **1975**, *23*, 351–379. [\[CrossRef\]](#)
67. Heckerman, D. Probabilistic interpretations for MYCIN's certainty factors. In *Machine Intelligence and Pattern Recognition*; Kanal, L.N., Lemmer, J.F., Eds.; Elsevier: Amsterdam, The Netherlands, 1986; Volume 4, pp. 167–196.
68. Lan, H.X.; Zhou, C.H.; Wang, L.J.; Zhang, H.Y.; Li, R.H. Landslide hazard spatial analysis and prediction using gis in the Xiaojiang watershed, Yunnan, China. *Eng. Geol.* **2004**, *76*, 109–128. [\[CrossRef\]](#)
69. Dou, J.; Tien Bui, D.; Yunus, A.P.; Jia, K.; Song, X.; Revhaug, I.; Xia, H.; Zhu, Z. Optimization of causative factors for landslide susceptibility evaluation using remote sensing and GIS data in parts of Niigata, Japan. *PLoS ONE* **2015**, *10*, e0133262. [\[CrossRef\]](#)
70. Dou, J.; Oguchi, T.; Hayakawa, Y.S.; Uchiyama, S.; Saito, H.; Paudel, U. Gis-based landslide susceptibility mapping using a certainty factor model and its validation in the chuetsu area, central Japan. *Landslide Sci. A Safer Geoenviron.* **2014**, *2*, 419–424.
71. Pourghasemi, H.R.; Pradhan, B.; Gokceoglu, C.; Mohammadi, M.; Moradi, H.R. Application of weights-of-evidence and certainty factor models and their comparison in landslide susceptibility mapping at Haraz watershed, Iran. *Arab. J. Geosci.* **2013**, *6*, 2351–2365. [\[CrossRef\]](#)
72. Gama, J. Functional trees. *Mach. Learn.* **2004**, *55*, 219–250. [\[CrossRef\]](#)
73. Kavzoglu, T.; Sahin, E.K.; Colkesen, I. Landslide susceptibility mapping using gis-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides* **2014**, *11*, 425–439. [\[CrossRef\]](#)
74. Witten, I.H.; Frank, E.; Hall, M.A. *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed.; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 2011; Volume 31, pp. 76–77.
75. 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. [\[CrossRef\]](#)
76. Pham, B.T.; Prakash, I. A novel hybrid intelligent approach of random subspace ensemble and reduced error pruning trees for landslide susceptibility modeling: A case study at Mu Cang Chai District, Yen Bai Province, Vietnam. In Proceedings of the International Conference on Geo-Spatial Technologies & Earth Resources, Hanoi, Vietnam, 5–6 October 2017.
77. Doetsch, P.; Buck, C.; Golik, P.; Hoppe, N.; Kramp, M.; Laudenberg, J.; Steingrube, P.; Forster, J.; Mauser, A. Logistic model trees with AUC split criterion for the KDD cup 2009 small challenge. In Proceedings of the International Conference on KDD-Cup 2009 Competition, New York, NY, USA, 28 June 2009; Volume 7, pp. 77–88.
78. Landwehr, N.; Hall, M.; Frank, E. Logistic model trees. *Mach. Learn.* **2005**, *59*, 161–205. [\[CrossRef\]](#)
79. 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. [\[CrossRef\]](#)
80. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [\[CrossRef\]](#)
81. Bauer, E.; Kohavi, R. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Mach. Learn.* **1999**, *36*, 105–139. [\[CrossRef\]](#)
82. Rodriguez, J.J.; Kuncheva, L.I.; Alonso, C.J. Rotation forest: A new classifier ensemble method. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1619–1630. [\[CrossRef\]](#)
83. Ozcift, A. Svm feature selection based rotation forest ensemble classifiers to improve computer-aided diagnosis of Parkinson disease. *J. Med. Syst.* **2012**, *36*, 2141–2147. [\[CrossRef\]](#)
84. Hong, H.; Liu, J.; Tien Bui, D.; Pradhan, B.; Acharya, T.D.; Pham, B.T.; Zhu, A.-X.; Chen, W.; Ahmad, B.B. Landslide susceptibility mapping using j48 decision tree with adaboost, bagging and rotation forest ensembles in the Guangchang area (China). *Catena* **2018**, *163*, 399–413. [\[CrossRef\]](#)

85. Pham, B.T.; Prakash, I.; Singh, S.K.; Shirzadi, A.; Shahabi, H.; Tien Bui, D. Landslide susceptibility modeling using reduced error pruning trees and different ensemble techniques: Hybrid machine learning approaches. *Catena* **2019**, *175*, 203–218. [\[CrossRef\]](#)
86. Kai, M.T.; Witten, I.H. Stacking bagged and dagged models. In Proceedings of the Fourteenth International Conference on Machine Learning, Nashville, TN, USA, 8–12 July 1997.
87. Li, J.; Huang, T. Predicting and analyzing early wake-up associated gene expressions by integrating gwas and eqtl studies. *Biochim. Biophys. Acta-Mol. Basis Dis.* **2018**, *1864*, 2241–2246. [\[CrossRef\]](#) [\[PubMed\]](#)
88. 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. *Geomatics. Nat. Hazards Risk* **2017**, *8*, 1955–1977. [\[CrossRef\]](#)
89. 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. *Geomatics. Nat. Hazards Risk* **2017**, *8*, 950–973. [\[CrossRef\]](#)
90. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
91. 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. [\[CrossRef\]](#)
92. 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. [\[CrossRef\]](#)
93. Youssef, A.M.; Pradhan, B.; Jebur, M.N.; El-Harbi, H.M. Landslide susceptibility mapping using ensemble bivariate and multivariate statistical models in Fayfa area, Saudi Arabia. *Environ. Earth Sci.* **2015**, *73*, 3745–3761. [\[CrossRef\]](#)
94. 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 (pso) for landslide spatial modelling. *Catena* **2017**, *157*, 310–324. [\[CrossRef\]](#)
95. 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. [\[CrossRef\]](#)
96. 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. [\[CrossRef\]](#)
97. 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. [\[CrossRef\]](#)
98. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
99. Kornejady, A.; Ownegh, M.; Rahmati, O.; Bahremand, A. Landslide susceptibility assessment using three bivariate models considering the new topo-hydrological factor: Hand. *Geocarto Int.* **2018**, *33*, 1155–1185. [\[CrossRef\]](#)
100. Pourghasemi, H.R.; Yousefi, S.; Kornejady, A.; Cerdà, A. Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. *Sci. Total Environ.* **2017**, *609*, 764–775. [\[CrossRef\]](#)
101. Kornejady, A.; Ownegh, M.; Bahremand, A. Landslide susceptibility assessment using maximum entropy model with two different data sampling methods. *Catena* **2017**, *152*, 144–162. [\[CrossRef\]](#)
102. Hong, H.; Pourghasemi, H.R.; Pourtaghi, Z.S. Landslide susceptibility assessment in Lianhua county (China): A comparison between a random forest data mining technique and bivariate and multivariate statistical models. *Geomorphology* **2016**, *259*, 105–118. [\[CrossRef\]](#)

103. 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](#)]
104. Chen, W.; Sun, Z.; Han, J. Landslide susceptibility modeling using integrated ensemble weights of evidence with logistic regression and random forest models. *Appl. Sci.* **2019**, *9*, 171. [[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/>).