

Article

Improved Shallow Landslide Susceptibility Prediction Based on Statistics and Ensemble Learning

Zhu Liang ^{1,2,3}, Wei Liu ^{2,3}, Weiping Peng ^{2,3}, Lingwei Chen ^{2,3} and Changming Wang ^{4,*} 

¹ College of Construction Engineering, South China University of Technology, Guangzhou 510641, China; liangzhu19@mails.jlu.edu.cn

² Guangzhou Urban Planning & Design Survey Research Institute, Guangzhou 510060, China; liuwei@gzpi.com.cn (W.L.); pengweiping@gzpi.com.cn (W.P.); lingweichen@163.com (L.C.)

³ Guangdong Enterprise Key Laboratory for Urban Sensing, Monitoring and Early Warning, Guangzhou 510060, China

⁴ College of Construction Engineering, Jilin University, Changchun 130012, China

* Correspondence: wangcm@jlu.edu.cn; Tel.: +86-135-0441-8751

Abstract: Rainfall-induced landslides bring great damage to human life in mountain areas. Landslide susceptibility assessment (LSA) as an essential step toward landslide prevention has attracted a considerable focus for years. However, defining a reliable or accurate susceptibility model remains a challenge although various methods have been applied. The main purpose of this paper is to explore a comprehensive model with high reliability, accuracy, and intelligibility in LSA by combining statistical methods and ensemble learning techniques. Miyun county in Beijing is selected as the study area. Firstly, the dataset containing 370 landslide locations inventories and 13 conditioning factors were collected and non-landslide samples were prepared by clustering analysis. Secondly, random forest (RF), gradient boosting decision tree (GBDT), and adaptive boosting decision tree (Ada-DT) were selected as base learners for the Stacking ensemble method, and these methods were evaluated using measures like area under the curve (AUC). Finally, the Gini index and frequent ratio (FR) were combined to analyze the major conditioning factors. The results indicated that the performance of the Stacking method was enhanced with an AUC value of 0.944 while the basic classifiers also performed well with 0.906, 0.910, and 0.917 for RF, GBDT, and Ada-DT, respectively. Regions with a distance to a stream less than 2000 m, a distance to a road less than 3000 m, and elevation less than 600 m were susceptible to the landslide hazard. The conclusion demonstrates that the performance of LSA desires enhancement and the reliability and intelligibility of a model can be improved by combining binary and multivariate statistical methods.

Keywords: landslide susceptibility; statistical methods; ensemble techniques; GIS



Citation: Liang, Z.; Liu, W.; Peng, W.; Chen, L.; Wang, C. Improved Shallow Landslide Susceptibility Prediction Based on Statistics and Ensemble Learning. *Sustainability* **2022**, *14*, 6110. <https://doi.org/10.3390/su14106110>

Academic Editors: Stefano Morelli, Veronica Pazzi and Mirko Francioni

Received: 3 May 2022

Accepted: 13 May 2022

Published: 18 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Landslides are a common natural phenomenon and may cause unpredictable damage to human beings and property worldwide, especially in China where geohazards are enormously occurring and widely distributed [1]. Generally, damages can be decreased or mitigated by predicting the area prone to landslides [2,3]. Therefore, landslide susceptibility mapping (LSM), which predicts the spatial distribution of the likelihood of a landslide occurring, is significant and worthwhile for the reduction of hazards.

How to improve the quality of a model is always the focus of attention and discussed by researchers although related studies have been conducted on improving the predictive accuracy [4,5]. The effectiveness of LSM depends greatly on the models adopted [6], which can be roughly divided into knowledge-based and data-driven methods [7]. Conventional knowledge-based methods as a heuristic, are subjective and limited to be applied in small-scale areas. Conventional statistical methods, like logistic regression (LR) and principal component analysis, are popular due to their simplicity. Nevertheless, the mechanism of a

landslide is complicated and usually involves multiple factors. Nevertheless, conventional statistical methods fail to deal with nonlinear problems [8,9]. Geographic information systems (GIS) and computing techniques are increasingly developing so that machine learning techniques (MLT) translated from statistical methods have achieved promising performance for LSM [10]. However, the inductive preference is inevitable for MLT and the performance of different MLT fluctuates as the data change. Besides, the performance of a model will be affected by the purity of the samples. Accordingly, controversy continues over which method is the best and which is feasible for further improvement [11]. A single classifier applied to form an ensemble is called the “base learner”. Base learners are not limited to homogeneous but also can be heterogeneous. Bagging and boosting are two of the most popular ensemble techniques of homogeneous and have been applied to LSM by some researchers [12]. While the stacking ensemble method, which is heterogeneous, has seldom been applied to LSM and needs more exploration [13].

Data-driven methods are binary classification processes and are sensitive to the quality of training data, which require a data set consisting of an equal amount of both disaster presence and absence observations in LSM [14]. In terms of landslide presence data, it is obtained from the landslide inventory, which was achieved through historical records, remote sensing technology, and extra field investigations. While absence samples were not available, they are usually selected randomly or subjectively from the “safe area” based on the acknowledgment and experience of the experts [15]. Nevertheless, it is controversial and difficult to implement especially for a large area. We could not identify the area with low susceptibility based on the historical records because these landslide-free areas may contain locations prone to landslides, which have not been recorded in the past. Unreliable sampling strategies may bring the noise to the data and eventually, a false assessment of the models. This study applies clustering analysis to improve the quality of samples and the performance of models [16]. Two initial landslide susceptibility maps are made by k-means clustering and Fuzzy c-means (FCM) and the reasonability of the initial maps are compared and finally determined the better one. Accordingly, the non-landslide samples are selected from the very low susceptibility area.

Previous studies have emphasized the importance of accuracy and regarded it as the only indicator for evaluating a model. However, emphasizing accuracy is not enough for the requirement of prevention and control. Communication between theory and practice can be improved through a better understanding of major variables. The bivariate statistical method as FR is commonly applied to explore the relationship between conditioning factors and the occurrence of landslides by calculating the FR values of factors in a certain interval [17]. While the bivariate statistical method fails to determine the relative importance among different factors and Gini index (the larger the value indicates the greater the contribution to the occurrence of landslides) makes up for it [18].

The current study aims to explore a model with high reliability, accuracy, and intelligibility on LSM. Three ensemble techniques were evaluated by 5-fold cross-validation according to the Receiver Operating Characteristic (ROC) curve and statistical indexes. RF, GBDT, and Ada-DT are selected as the candidate base learners of the stacking method and LR as the meta-learner. The stacking method is explored as a potential application to LSM and is compared to other ensemble methods. The purity of the samples is improved by selecting the non-landslide samples in a more reliable way. The Gini index and FR were combined to identify and analyze the major conditioning factors to improve intelligibility. Miyun country, China, where shallow landslides occurred frequently, was selected as the study area and a comparison of the ensemble methods above was made.

2. Materials

2.1. Study Area

Miyun country located in Beijing, China, extends from longitudes of 116°39' E to 117°30' E and latitudes of 40°13' N to 40°47' N (Figure 1). It has a population of more than 470,000 and occupies an area of about 2229.45 km², which is composed of mountainous

areas (82.9%), cultivated land (8.3%), and reservoirs, roads, and villages (collectively 9.8%). The average annual precipitation is 663.1 mm (1981–2012) mainly concentrated in summer (76.4%) and it is a continental monsoon semi-arid climate.

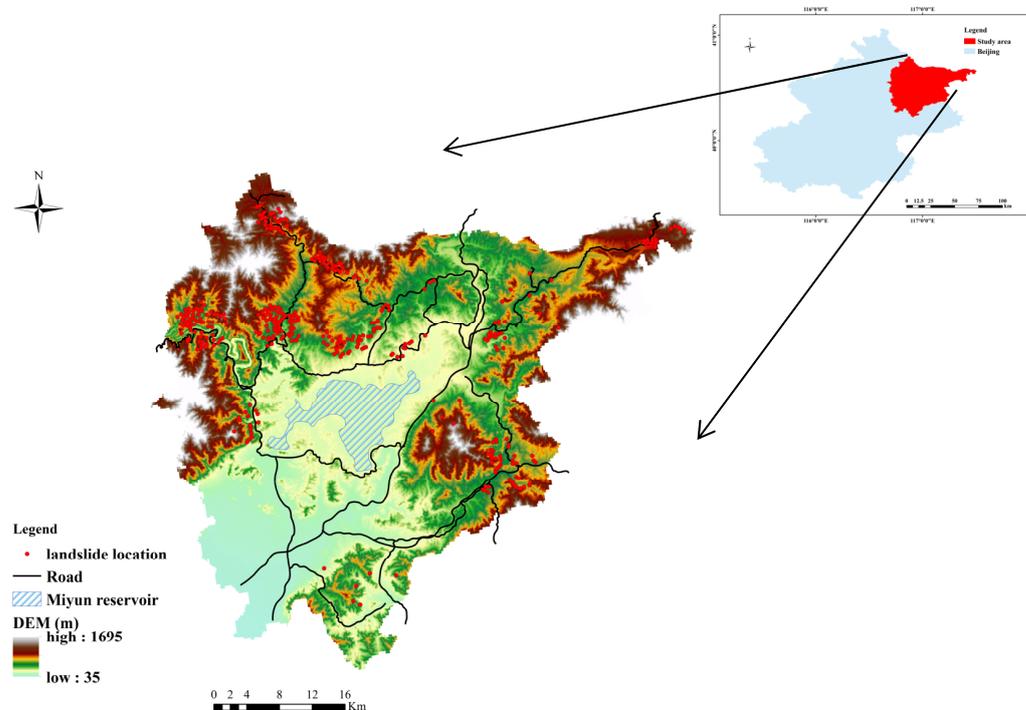


Figure 1. Location map of the study area showing landslide inventory.

The study area is part of the transition zone between the North China Plain and the Yanshan mountains, which leads to a series of large fold and fault structures. The faults are large in scale and widely distributed, mainly in the Northeast and north-south directions. The elevation ranges from 45 m to 1750 m above mean sea level with a slope angle between 10–45°. The strata are mainly composed of Archaean (Ar), Proterozoic (Pt), Mesozoic Jurassic (J), and Quaternary (Q). Three types of lithology are usually exposed in our investigation: gneiss from Middle Archean (ArXdgn), dolomites from Proterozoic (Pt22w), and siltstone from Mesozoic Jurassic (J2z). Magmatic intrusive rocks are widely distributed, accounting for nearly one-third of the total area and are exposed discontinuously in the northeast direction.

Road traffic is developed, and human activities are intensive in the study area, involving mining, reservoir, and power station projects. The disasters are various and frequent, mainly rain-induced landslides, which has affected the normal life of the local villagers.

2.2. Data Preparation

2.2.1. Landslide Inventory

The statistically based models follow a crucial assumption: future landslides have more chances to occur again in the places with the conditions which cause the landslides once and present [18,19]). Accordingly, the landslide inventory map as the initial source is essential and was depicted according to related records (from 1970–2010), field surveys (from 2016–2017) (Figures 2 and 3), and Google Earth satellite images interpretation (May 2018) (Figure 4). Ultimately, 620 landslide locations were identified, including soil slides (370), rockslides (6), and falls (244) [18]. It is accepted that different type of landslides has a different mechanism of occurrence. Soil slides were only considered in our work and were represented as points shown in Figure 1. Landslides occurred during or after heavy rainfall. Based on field investigation, remote sensing interpretation and relevant records, the scale of landslides in the study area is generally small, accounting for about 80%. The area of

landslides ranges from 3.6 km² to 300 m² while the depth of most landslides is less than 4 m, belonging to shallow landslides.



Figure 2. Field investigation photos. (a) shallow landslide in Lama Gate South gully; (b) falls in Lama Gate South gully.



Figure 3. Field investigation photos. (a) early debris-flow deposits in Dawa gully; (b) Partial enlargement.



Figure 4. Stereo remote sensing map of landslides in Duitaizi county (Chen et al., 2016).

2.2.2. Choice of Mapping Units

The selection of mapping units should be determined in advance for LSM [20]. Another piece of literature discussed and compared the difference among mapping units, such as grid cells and slope units [21]. To better predict or identify the locations of landslides, slope units were applied in our work, which describes the topographic and geomorphic conditions of landslides integrally. Finally, the area was divided into 8736 slope units using the hydrologic analysis tool in ArcGIS and indispensable artificial corrections according to remote sensing images. Detailed division steps and discussion can be referred to in other literature [22].

2.2.3. Conditioning Factors

Factors responsible for a landslide are various and there is no consensus on the choice of number and types of factors. It is commonly accepted that landslide is controlled by topographical, geological, and triggering factors. However, data availability, reliability, and accuracy should be given priority [23] and finally, 13 conditioning factors were selected. Detailed information on conditioning factors is shown in Table 1 and Figure 5a–m. A brief description of each controlling factor is given below.

Table 1. Landslide conditioning factors in this study.

Category	Conditioning Factors	Type	Data Source	Values
Topographical	Elevation (m)	Continuous	SRTM	(1) <200; (2) 200–400; (3) 400–600; (4) 600–800; (5) >800
	Plan curvature	Continuous	SRTM	(1) <0; (2) 0–0.01; (3) 0.01–0.02; (4) 0.02–0.03; (5) >0.03
	Profile curvature	Continuous	SRTM	(1) <0; (2) 0–0.01; (3) 0.01–0.02; (4) 0.02–0.03; (5) >0.03
	Slope angle (°)	Continuous	SRTM	(1) <10; (2) 10–20; (3) 20–30; (4) >30
	TWI	Continuous	SRTM	(1) <6.5; (2) 6.5–7; (3) 7–7.5; (4) 7.5–8; (5) 8–8.5; (6) >8.5
	MED (m)	Continuous	SRTM	(1) <100; (2) 100–200; (3) 200–300; (4) 300–400; (5) 400–500; (6) >500
	Slope aspect	Categorical	SRTM	(1) north; (2) northeast; (3) east; (4) southeast; (5) south; (6) southwest; (7) west; (8) northwest
Geological and Geomorphological	Distance to faults (m)	Continuous	Geological map	(1) <1000; (2) 1000–2000; (3) 2000–3000; (4) 3000–4000; (5) >4000
	Distance to streams (m)	Continuous	DNRB	(1) <1000; (2) 1000–2000; (3) 2000–3000; (4) 3000–4000; (5) >4000
	Lithology	Categorical	Geological map	(1) Gneiss; (2) Dolomites; (3) Siltstone (4) Granite; (5) Limestone; (6) Conglomerate
Triggering factors	Maximum 24 h rainfall (mm)	Continuous	BHM	(1) <270; (2) 270–280; (3) 280–290; (4) >290
	Maximum 7 days rainfall (mm)	Continuous	BHM	(1) <320; (2) 320–330; (3) 330–340; (4) >340
	Distance to roads (m)	Continuous	DNRB	(1) <1000; (2) 1000–2000; (3) 2000–3000; (4) 3000–4000; (5) >4000

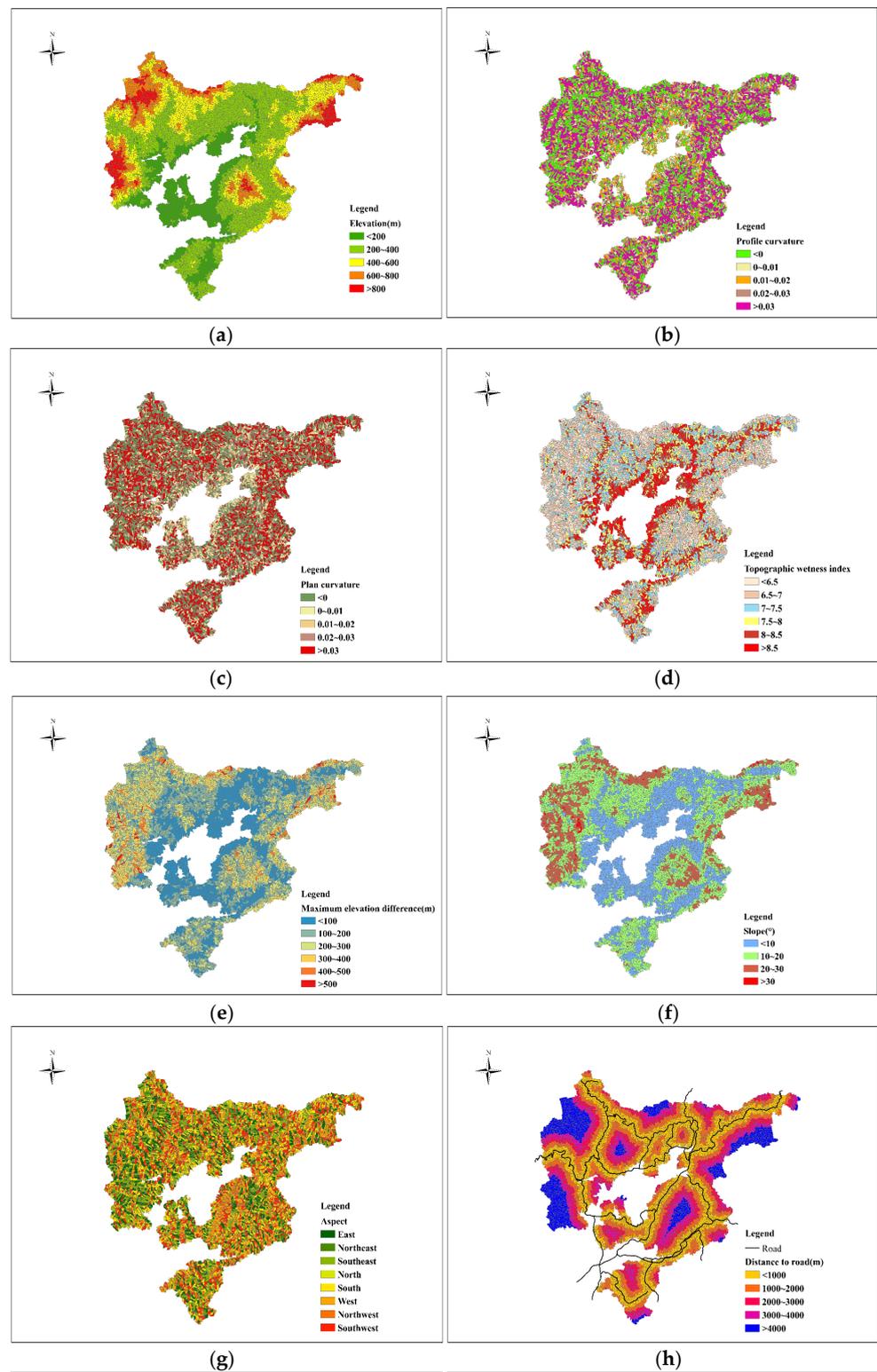


Figure 5. Cont.

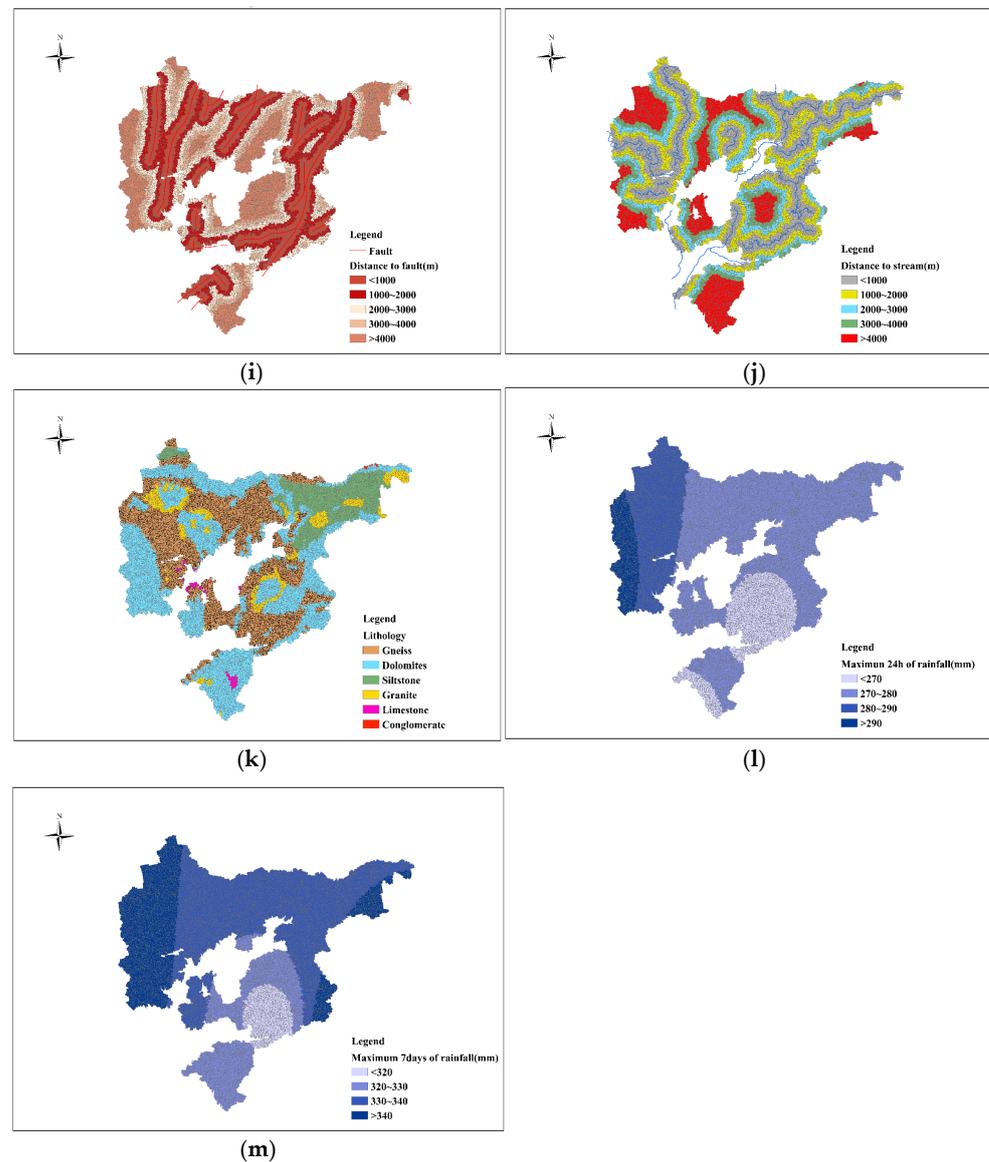


Figure 5. Study area thematic maps: (a) Elevation; (b) Plan curvature; (c) Profile curvature; (d) TWI; (e) MED; (f) Slope; (g) Aspect; (h) DTR; (i) DTF; (j) DTS; (k) Lithology; (l) Maximum 24 h Rainfall; (m) Maximum seven days Rainfall.

Topographic-related factors were derived from the DEM (Digital Elevation Model) with a resolution of 30 m (<http://www.gscloud.cn>, accessed on 4 April 2022) originally sourced from the Shuttle Radar Topography Mission (SRTM) data. Elevation affects slope instability and precipitation properties and was frequently applied to LSM [24,25]. Landslides are likely to occur as slopes become steep and vice versa [26]. Maximum elevation difference (MED) reflects the potential energy of a slope and was calculated in ArcGIS [27]. Topographic wetness index (TWI) and Curvature reflect topographic relief [28]. TWI was reclassified into six classes (Figure 5g) and the related algorithm is as follows:

$$TWI = \ln\left(\frac{A_s}{\tan \beta}\right) \quad (1)$$

where, A_s is the specific catchment area, β is the slope angle.

The plan curvature (Figure 5g) and profile curvature (Figure 5g) are both the most extensively used predisposing factors, which reflect the changes in terrain [29]. The slope aspect map was reclassified into eight classes according to the eight cardinal directions (Figure 5g).

Fault information (Figure 5i) was collected from a geological map of which the ratio was 1:50,000. Faults decrease the rock strength, which acts as potential weak planes in slopes. It was produced by the spatial distance analysis tool in ArcGIS. Similarly, the distance to roads (Figure 5h) and distance to rivers (Figure 5j) were both constructed based on the data from the Department of Natural Resources of Beijing (DNRB).

Shallow landslides are mainly caused by heavy or continuous rainfall [30]. Consequently, both the maximum 24 h rainfall (Figure 5l) and maximum seven days of rainfall (Figure 5m) were selected based on the data (1981–2000) from Beijing Hydrology Manual (BHM) using the kriging interpolation coordinated with elevation in ArcGIS and 11 precipitation stations nearby were taken as reference. Rainfall was regarded as the natural trigger while the distance to the road was the human factor.

Factors were reclassified into four to eight classes and the mean value was regarded as the statistic value of slope units.

3. Methods

3.1. Sampling Strategy

3.1.1. K-Means Clustering

K-means comes out to be a well know clustering method due to its efficiency and feasibility [31]. It is applied to divide n observations into k clusters, where each sample is allocated to the cluster based on the closest Euclidean distance, thus considered as the centroid of the cluster [32]. The procedure is then repeated until the change of the cluster seed from one stage to the next is negligible. The main equation involved in k-means is as follows:

$$\frac{|u_{n+1} - u_n|}{u_{n+1}} \leq \varepsilon \quad (2)$$

where u_{n+1} represents the sum of squares of distances from each point to the cluster center after the n th clustering; ε represents the precision value.

3.1.2. FCM Algorithm

The fuzzy c-means method is a soft clustering method developed by Dunn [33] and it is different from K-means (hard clustering). It has been widely used for statistical analysis of geological problems because of its flexibility and rationality [34]. Its core idea is to assign the objects to the corresponding clusters according to the degree of membership. The function of the FCM clustering is defined by the equation:

$$C_i = \frac{\sum_{j=1}^n \mu_{ij}^m x_j}{\sum_{j=1}^n \mu_{ij}^m} \quad (3)$$

$$J = \sum_{j=1}^n \sum_{i=1}^C \mu_{ij}^m d^2(X_j, V_i) \quad (4)$$

$$\mu_{ij} = 1 / \sum_{k=1}^C \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)} \quad (5)$$

where C_i represents the cluster centers, C represents the number of centers, u_{ij} represents the membership matrix; m represents the degree of fuzziness; J is the objective function and n is the number of objects in the database; d^2 is the Euclidean distance between the i th clustering center and the j th sample [35].

Two parameters as m and C are required to determine in advance. C is determined by the cluster validity function [36] and m is equal to 2 referred to in most applications in this study.

Machine learning methods need both positive and negative datasets. Three-hundred-seventy positive samples (that is, landslide locations) were set as "1" and the same number of negative samples with the value of "0", which were selected based on the result of

K-means and FCM in this study. As the purity of absent samples increases, it is more likely to reflect the characteristics of non-landslide areas. Accordingly, the critical value of the model results distinguishing landslides and non-landslides is 0.5.

3.1.3. Frequency Ratio

The equation for determining the FR value of a certain level of conditioning factor is defined below [16]:

$$FR_i = \frac{\frac{\text{landslide_cell}_{s_i}}{\text{landslide_cell}_{s_{tot}}}}{\frac{\text{total_cell}_{s_i}}{\text{total_cell}_{s_{tot}}}} \quad (6)$$

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

An FR_i greater than 1 manifest that there exists a close relationship between landslide occurring and variable class, and if the values are less than 1 then a weak correlation is reflected. Continuous variables are required to be reclassified into classes before application, as Table 1 showed.

3.2. Modeling Landslide Susceptibility

3.2.1. LR Model

LR establishes a non-linear probability function model, trying to find appropriate regression coefficients to express the correlation between the independent variable and the dependent variable [37]. The LR model is constructed as the equation below:

$$p = \frac{1}{1 + e^{-y}} \quad (7)$$

where p is the probability of a landslide occurring; y is a linear combination function as Equation (7).

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n \quad (8)$$

where b_0 is the constant value, and b_1, b_2, \dots, b_n refer to each significant input variable (x_1, x_2, \dots, x_n) causing the landslide.

The forward7 stepwise method was adopted to screen variables during LR modeling in SPSS software.

3.2.2. RF

RF belongs to a family of ensemble methods based on the decision tree and Bagging technique and it was first introduced by Breiman [17]. The bagging technique, which is also called bootstrap aggregation, is applied to selecting variables and samples randomly as the training data for modeling. Unused observations are applied to calculate the classification error. Consequently, there are two powerful ideas of RF: random feature selection and Bagging [38]. More details about RF can be found in Breiman [17]. RF was modeled in Python 3.7 using the scikit-learn package [39]. The number of trees (k) and the number of predictive variables (n) are required tuning before modeling [40].

3.2.3. GBDT

GBDT forms weak classifiers (DT) iteratively based on Gradient Boosting [41]. The parameter of the weak classifier defaults to the direction of the. The GBDT was applied in Python 3.7 using the GBDT class library of scikit-learn.

3.2.4. AdaBoost-DT

AdaBoost (known as adaptive boosting) is another boosting algorithm, which was invented by Freund and Schapire [42]. Unlike gradient boosting, AdaBoost assigns incorrectly classified samples with modified weights after each iteration. The final classifier is constructed by combining all weak classifiers. AdaBoost-DT is also applied in Python 3.7 using the AdaBoost class library of scikit-learn.

3.2.5. Gini Index

The split method tree-based classifiers adopt is the minimum principle of Gini and thus Gini index is applied to calculate the relative importance of conditioning factors. The relevant formula is as follows:

$$\text{Gini}(T) = 1 - \sum_{j=1}^N P_j^2 \tag{9}$$

where T expresses the training set, N is the number of categories, and P is the probability of a sample that is classified into the k th class.

3.2.6. Stacking

The stacking ensemble consists of base-classifiers and meta-classifier. Stacking takes the results predicted by the base-classifiers as the input attributes and the meta-classifier merges the different predictions into the final prediction. It is believed that stacking performs better than any basic classifiers [43]. Figure 6 shows the structure of the Stacking. The basic classifiers of Stacking were three ensemble learning machines that have been showing great performance in statistical analysis: RF, GBDT, and AdaBoost-DT. LR model was used as the combiner. To avoid over-fitting of the meta-classifier, the dataset is divided into two disjoint subsets: one for training base-classifiers and the other for testing. To train the meta-level classifier, 5-fold cross-validation is applied to construct the meta-levels for all combining methods.

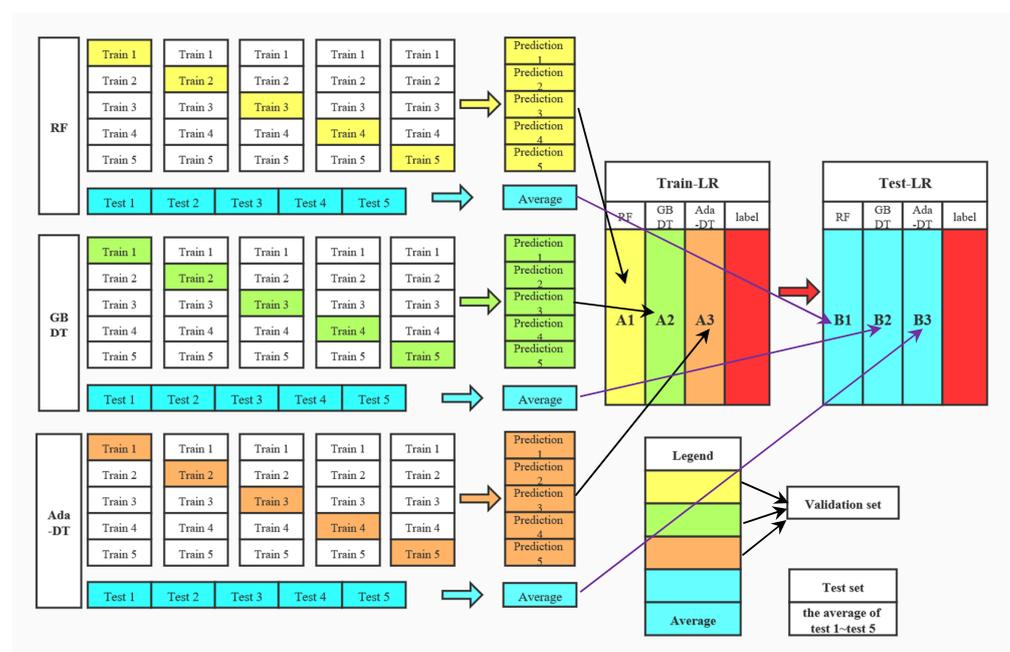


Figure 6. The structure of Stacking.

3.3. Evaluating Model Performance

Models need a reliable evaluation and/or validation process [44]. The capacity of a model to classify was evaluated by a 5-fold cross-validation procedure, where the data is divided into five independent groups, one at a time for testing and the remaining four groups for training [45].

Accuracy, sensitivity, and specificity were three statistical indexes evaluating the performance [13]:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{11}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \tag{12}$$

where True Positive (TP) refers to the number of landslide samples with correct classification, True Negative (TN) refers to the number of non-landslide samples with correct classification, False Positive (FP) refers to the number of landslide samples with incorrect classification and False Negative (FN) refers to the number of non-landslide samples with incorrect classification.

AUC is a metric commonly used to assess the quality of the model and it varies from 0.5 to 1. The higher the AUC value shows the stronger the predictive ability [46].

Non-parametric models need to be optimized by tuning related hyperparameters before application [47]. The involved parameters for modeling utilized in this study were shown in Table 2 and the flowchart of methods involved was shown in Figure 7.

Table 2. The optimized parameters of methods utilized in this study.

Methods	Parameters
DT	Criterion = 'gini'; max_features = None; max_depth = 20; min_samples_split = 2; min_samples_leaf = 1; max_leaf_nodes = None; class_weight = None
RF	n_estimators = 500; criterion = 'gini'; max_depth = None; max_features = 'sqrt';
GBDT	n_estimators = 100; learning_rate = 0.1; max_depth = 2; verbose = 1; subsample = 0.7; max_leaf_nodes = None
AdaBoost-DT	base_estimator = None; n_estimators = 100; learning_rate = 1.0; algorithm = 'SAMME.R'; random_state = None

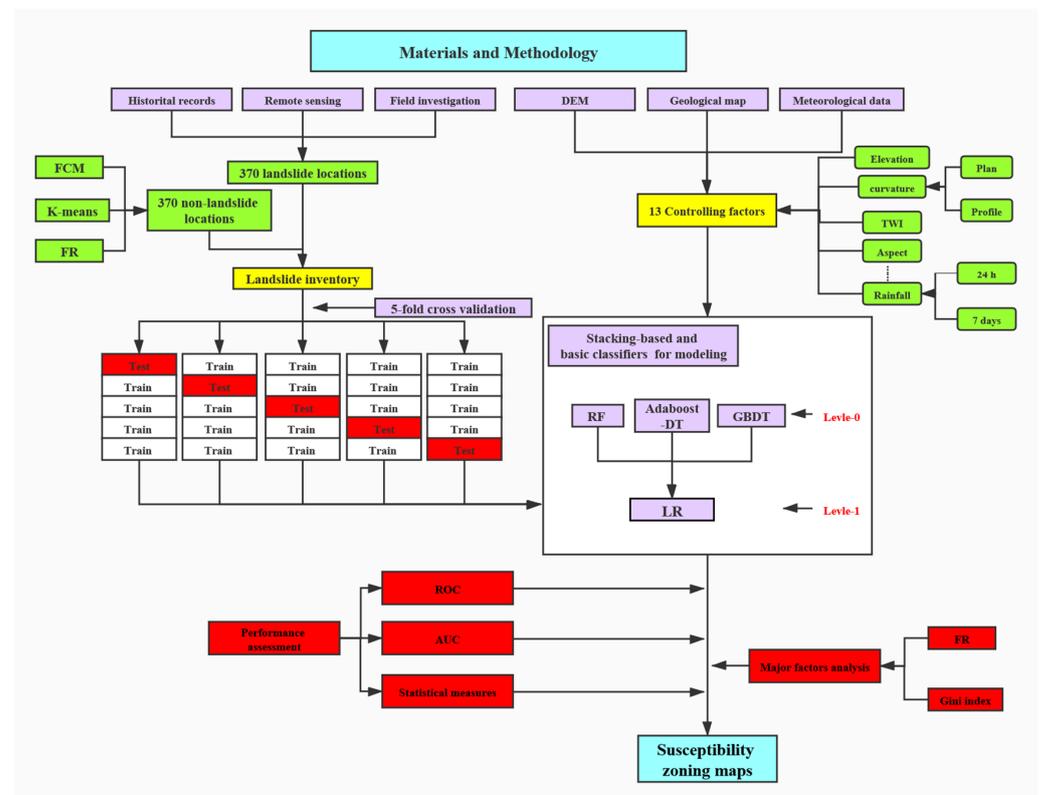


Figure 7. Flowchart of the methodology followed in this study.

4. Results and Verification

4.1. Non-Landslide Samples Selected by FCM and K-Means

LSM generated based on cluster analysis does not need to identify the positive and negative labels of the samples in advance. Based on the curve of the clustering effectiveness

index V_{cs} (Figure 8), the preferred value is five. Consequently, the study area was reclassified into five areas based on the FR values, which were very low, low, moderate, high, and very high. The proportions of each area are: very low (15.97%), low (23.25%), moderate (19.29%), high (33.5%) and very high (8%). Among them, the very-low area accounted for 15.97% of the whole study area with only 3.24% of landslide locations and an FR value of 0.2. Besides, the high or very-high area accounted for 41.5% of the study area with more than 55% of landslide locations and the FR values were both greater than 1.

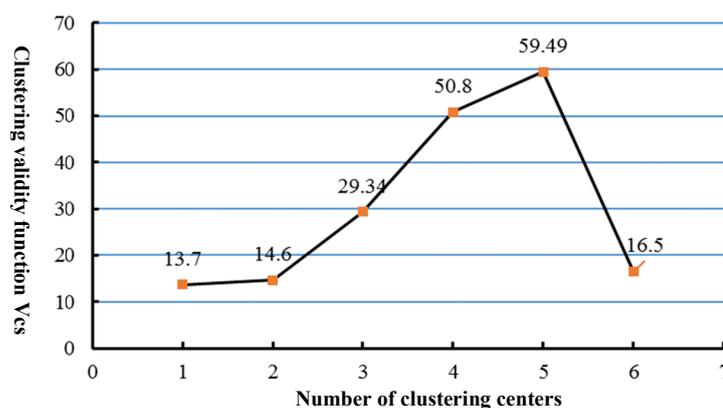


Figure 8. Clustering validity function V_{cs} .

Similarly, the results constructed by K-means were shown in Table 3. The proportions of each area are: very low (11.66%), low (22.30%), moderate (18.71%), high (39.16%), and very high (8.17%). The very-low area accounted for only 1.62% of landslide locations with an FR value of 0.14. The high or very-high area accounted for 47.33% of the study area with more than 55% of landslide locations.

Table 3. Frequency ratios of five susceptibility classes assessed with FCM and K-means.

Method	Class	Landslide Ratio (%)	Area Ratio (%)	FR
FCM	Very low	3.24	15.97	0.20
	Low	19.73	23.25	0.85
	Moderate	21.35	19.29	1.11
	High	40.00	33.50	1.19
	Very high	15.68	8.00	1.96
k-means	Very low	1.62	11.66	0.14
	Low	15.41	22.30	0.69
	Moderate	15.57	18.71	0.83
	High	48.11	39.16	1.22
	Very high	17.30	8.17	2.11

Compared to the results obtained by FCM, the area with low or very low class predicted by K-means occupied a smaller area (5.26%) while a bigger area (5.83) with high or very high class. The zoning maps should follow two rules: (1) the recorded landslides should appear in high-susceptibility areas as many as possible and (2) the high-susceptibility area should occupy a small proportion (Bui et al., 2012). Therefore, the results obtained by FCM were more reasonable. Selecting the non-landslide samples in a more reliable area is the main purpose and it means that the bigger the very-low class area, the easier the sampling will be. Meanwhile, 370 non-landslides samples were collected from the area with very-low susceptibility predicted by FCM.

4.2. Evaluation and Comparison of Different Models

To highlight the performance of the Stacking model, three basic classifiers were also applied for modeling. Analyses of the statistical measures using the training set were shown in Table 4. The Stacking showed the best performance in terms of classifying landslides (sensitivity = 91.89%), followed by the GBDT model (sensitivity = 86.97%), the Ada-DT model (sensitivity = 85.66%) and RF model (sensitivity = 79.93%). In terms of the classification of non-landslides zones, Stacking model also performed best (specificity = 91.84%), followed by the GBDT model (specificity = 85.67%), the Ada-DT model (specificity = 82.26%) and the CART model (specificity = 83.16%). Besides, the Stacking model also had the highest accuracy (91.84%). It was noticed that the Stacking model achieved an AUC of 0.963, while RF was 0.920, GBDT was 0.957 and Ada-DT was 0.959 (Table 5). The standard errors were less than 0.05 and the probability estimation was negligible.

Table 4. Models' performance using training dataset.

Metrics	RF	GBDT	Ada-DT	Stacking
TP (%)	82.46	84.88	81.29	91.22
TN (%)	76.80	87.67	86.44	92.20
FP (%)	17.54	15.12	18.71	8.78
FN (%)	23.2	12.37	13.56	7.80
Sensitivity (%)	79.93	86.97	85.66	91.89
Specificity (%)	83.16	85.67	82.26	91.78
Accuracy (%)	81.56	86.29	83.87	91.84

Table 5. ROC analysis of the four models using training data.

Models	AUC	Standard Error	95% Confidence Interval
RF	0.920	0.011	0.899–0.941
GBDT	0.957	0.008	0.942–0.973
Ada-DT	0.959	0.009	0.942–0.976
Stacking	0.963	0.006	0.950–0.975

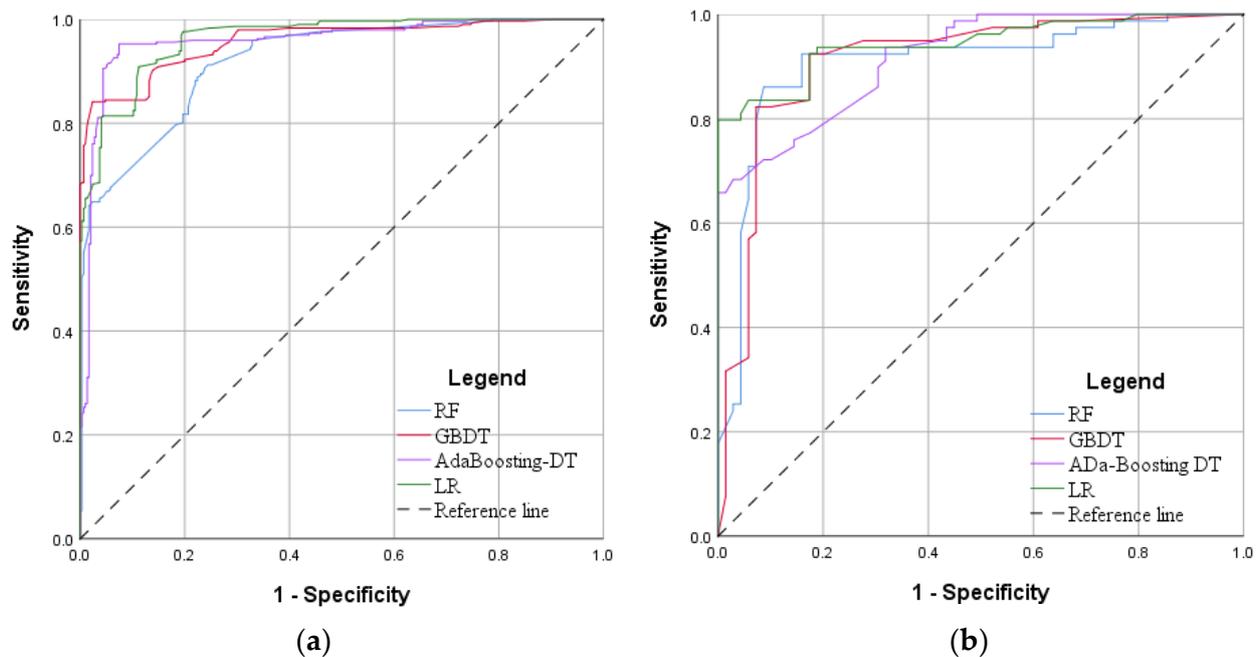
The predictive capacity needs to be evaluated using validation data. The results confirmed that the Stacking model perform the best as the values of sensitivity, specificity, accuracy and AUC were highest (Tables 6 and 7), which was 91.78%, 90.54%, 91.16% and 0.944, respectively, followed by Ada-DT (sensitivity = 86.96%, specificity = 82.19%, accuracy = 85.13% and AUC = 0.917), GBDT (sensitivity = 86.11%, specificity = 84.00%, accuracy = 85.03% and AUC = 0.910), and RF (sensitivity = 81.33%, specificity = 75.34%, accuracy = 78.38 and AUC = 0.906) (Figure 9).

Table 6. Models' performance using verification dataset.

Metrics	RF	GBDT	Ada-DT	Stacking
TP (%)	77.22	86.30	83.54	90.54
TN (%)	79.71	83.78	86.96	91.78
FP (%)	22.78	13.70	16.46	9.46
FN (%)	20.29	16.22	13.04	8.22
Sensitivity (%)	81.33	86.11	86.96	91.78
Specificity (%)	75.34	84.00	82.19	90.54
Accuracy (%)	78.38	85.03	85.13	91.16

Table 7. ROC analysis of the models using validating data.

Models	AUC	Standard Error	95% Confidence Interval
RF	0.906	0.027	0.853–0.959
GBDT	0.910	0.026	0.859–0.962
Ada-DT	0.917	0.021	0.877–0.958
Stacking	0.944	0.018	0.908–0.980

**Figure 9.** Analysis of ROC curve for the landslide susceptibility map: (a) Success rate curve of landslide using the training dataset; (b) Prediction rate curve of landslide using the validation dataset.

The Stacking model exhibited the best both in training and validation data compared to the other three ensemble learning methods, which indicated ideal goodness-of-fit to modeling and generalization capability. The performance of GBDT and Ada-DT was similar, and the RF model performed the worst but was still satisfactory. The gaps in performance between training and validation data were not obvious among the models. Compared to the RF model, the application of the Stacking model enhanced the performance significantly and was regarded as the most suitable model for LSM in this study.

4.3. Application of Stacking Method for LSM

The above analysis proves that the Stacking method has superior ability in LSM compared with the other three models. Therefore, the probability of landslides occurring was calculated for all mapping units in the whole study area. The LSM was also constructed with five susceptible classes, which were very low (0–0.2), low (0.2–0.4), moderate (0.4–0.6), high (0.6–0.8), and very high (0.8–1) (Figure 10). Table 3 showed the distribution ratio of each level. The very low susceptible level occupied 26.04% of the area while low, moderate, high, and very high susceptible levels represented 15.31%, 15.46%, 32.45%, and 10.74%, respectively (Figure 11). It was noticed that LSM has the smallest area percentage in very high susceptibility levels while the largest is in high. Landslide locations were mostly distributed in the red areas. Meanwhile, most of the non-landslide samples screened by FCM clustering appeared in blue areas.

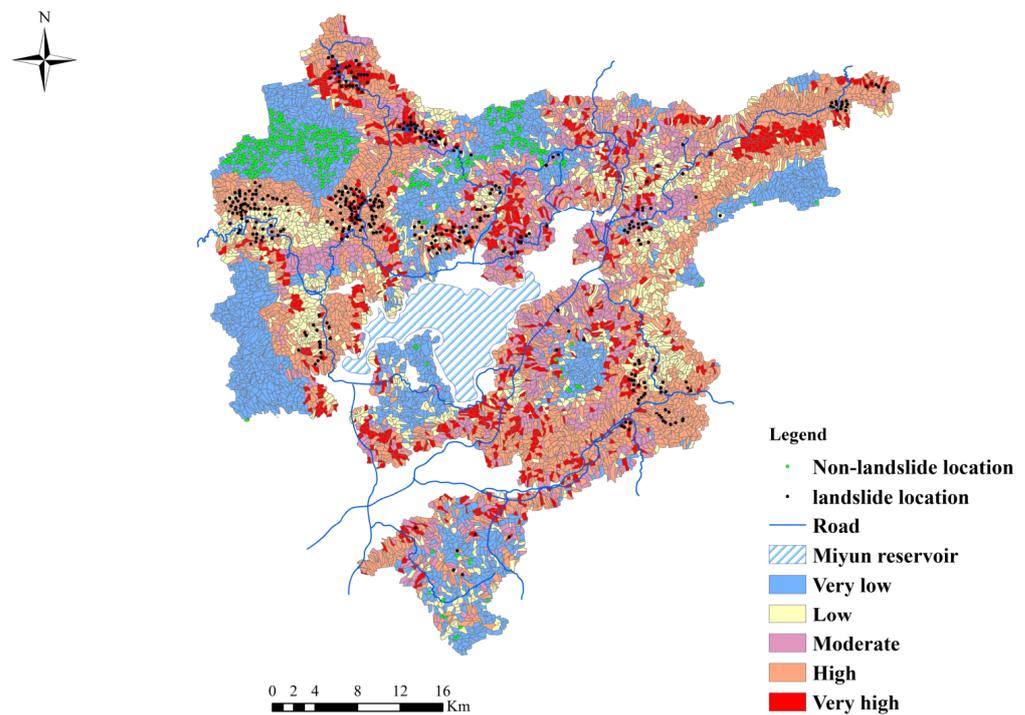


Figure 10. Landslide susceptibility map using the Stacking model.

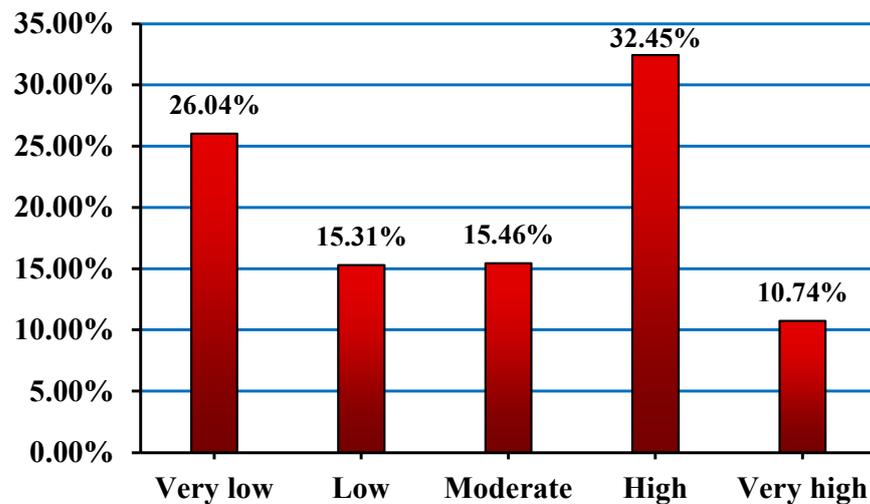


Figure 11. The distribution of susceptible classes on landslide susceptibility maps.

The high or very-high susceptibility areas are mainly distributed closed to streams or provincial highway, which runs through three townships including Fanzipai Town, Sihetang Town, and Fengjiayu Town in the study area. These areas are densely populated.

The landslide susceptibility class ranged from very low to very high around the Miyun reservoir. It is noteworthy that once a landslide occurs in this area, a series of disaster chains may be induced.

4.4. Analysis of Major Conditioning Factors

The stacking method performed the best in terms of accuracy, but the results had a poor analysis of the occurrence of landslides, which was confusing. Understanding the major factors that have a significant contribution to landslides occurring helps in the prevention and treatment of landslides. Based on the Gini index, ten major parameters were selected and normalized as shown in Table 8, including DTS, DTR, elevation, slope

angle, TWI, maximum 24 h rainfall, lithology, MED, maximum seven days of rainfall, and profile curvature. Among them, DTS, DTR, and elevation have a significant impact on the occurrence of landslides (Figure 12), the weight values of which were 0.37, 0.34, and 0.16, respectively. While the weight values of lithology, MED, maximum seven days rainfall, and profile curvature were close to 0.01, which had a limited contribution. The weight values of slope angle, TWI, and maximum 24 h rainfall were close to 0.04, 0.03, and 0.02, respectively.

Table 8. Conditioning factors assigned by the Ada-DT.

Method	DTS	DTR	Elevation	Slope Angel	TWI	Maximum 24 h Rainfall	Lithology	MED	Maximum 7 Days Rainfall	Profile Curvature
GBDT	0.37	0.34	0.16	0.04	0.03	0.02	0.01	0.01	0.01	0.01

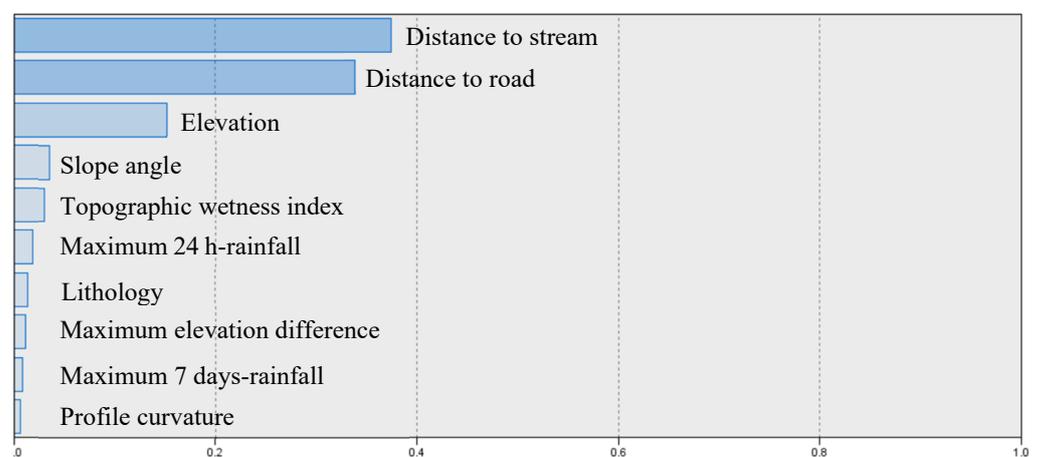


Figure 12. Parametric importance graphics obtained from Ada-DT.

Therefore, three conditioning factors, namely DTS, DTR, and elevation, were considered the major factors responsible for the landslide. Rivers are an important factor affecting the occurrence of landslides. On the slopes closer to the river, the toe of the slope is easily soaked by the river water, which reduces the strength of the rock and makes landslides more likely. Road development and construction are important tasks in mountainous area construction. However, unreasonable road excavation is a common human factor that induces geological disasters. Road construction often produces a large number of slopes, which destroy the stability of the slope and finally, lead to the occurrence of landslides.

The relationship between the major factors and landslides was further explored by calculating the FRi of each parameter (Table 9). As for DTS, the percentages of landslide area of the first two classes (<1000 m and 1000–2000 m) were 46.99% and 24.43% with the FR values of 49.3 and 173.29, accounting for more than 70% of the landslides area. Similarly, DTS showed a positive correlation in the first three classes (<1000 m, 1000–2000 m, and 2000–3000 m) with FR values greater than 1. Regarding elevation also a positive correlation in the first three classes (<200m, 200–400 m, and 400–600 m) with values of FR gradually decreasing with altitude and a negative relationship in the last two classes (>600 m).

The selection and analysis of major factors by combining basic machine learning and bivariate methods made up for the defects of stacking, thereby ensuring the integrity of geological hazard assessment.

Table 9. Spatial relationship between landslide conditioning factors and landslides using frequency ratio.

Conditioning Factor	Zone	Landslide (%)	Non-Landslide (%)	FR
DTS(m)	<1000	46.99%	0.95%	49.30
	1000–2000	24.43%	0.14%	173.29
	2000–3000	14.33%	6.63%	2.16
	3000–4000	5.33%	15.72%	0.34
	>4000	8.91%	76.69%	0.12
DTR(m)	<1000	56.06%	7.13%	7.87
	1000–2000	23.02%	7.13%	3.23
	2000–3000	15.59%	9.29%	1.68
	3000–4000	3.95%	11.51%	0.34
	>4000	1.37%	66.79%	0.02
Elevation(m)	<200	4.36%	2.08%	2.09
	200–400	53.76%	12.29%	4.37
	300–600	30.36%	23.70%	1.28
	400–800	10.06%	34.52%	0.29
	>800	1.46%	27.41%	0.05

5. Discussion

5.1. Ensuring the Reliability of Models

5.1.1. Internal and External Cross-Validation

The basic classifiers used in our work have several hyperparameters that control the behavior and performance. In some cases, reasonable “guesses” are available (e.g., n tree = 500 in RF), in other cases classifiers are very sensitive to the parameters, which means that default hyperparameter settings fail to guarantee optimal performance of machine-learning techniques. Therefore, hyperparameters need to be tuned before application and inner cross-validation should be used for this [48].

On the other hand, external cross-validation was also essential. One can find an “excellent model” using the method “Leave-One-Out” because of the randomness in the sampling scheme, the results of which are unconvincing. Only by implementing a more rigorous k-fold (or other types) cross-validation scheme can one infer the actual capacity of a model to learn the functional relationships between landslides and causative factors as well as the variability that the models and the susceptibility estimates exhibit [45,48].

While various machine-learning algorithms have been recognized in recent years due to their powerful capabilities of data processing and generalization, there are several practical challenges related to bias-reduced assessment of a model’s predictive power and some researchers often ignore them, which leads to an unreliable or uncertain result. Single hold-out model performance measures were popular [49]. However, statistically based landslide susceptibility models desire a more credible validation and assessment before generalization.

5.1.2. The Selection of Non-Landslide Samples

A complete disaster inventory map is emphasized in a multitude of studies, which consists of the locations and number of a certain disaster [9]. The quality of landslide presence samples is more convincing compared to that of landslide absence because non-landslide samples are selected randomly or subjectively although quite a few methods or principles will be adopted. Seldom do studies consider or discuss the noise and influence of the absence of data bring to data-driven models [50]. Non-landslide points need to be selected from low-prone areas as far as possible, which is arduous to implement by selecting randomly. Clustering analysis help solve the problem by combining with the bivariate methods. FR was calculated to judge the area with low susceptibility based on the results of FCM and K-means in this study and the non-landslide samples were generated from it, which improved the quality of non-landslide records and the performance of models logically.

5.2. Increasing the Accuracy of LSM

How to achieve an accurate landslide susceptibility zoning map is always a hot topic and the main concern of researchers. However, determining the most suitable model is challenging because the performance varies according to the study area and methods applied. Actually, related studies have applied various methods and compared their performance based on the value of AUC to obtain the best method for a given region [51,52]. Yet, it is controversial that we claim a model to be better than the other according to the decimal places down the line (AUC). Therefore, it is necessary to explore new methods for significant improvement and ensemble techniques are considered in our study, which have also been proven to be an excellent solution [53–55]. A detailed comparison among three ensemble techniques in LSM, namely bagging, boosting, and stacking was implemented. Bagging and boosting are two algorithms commonly used in LSM while stacking have rarely been applied. The results proved that the ensemble of the GBDT-Adaboost-DT-RF-LR had the ability to enhance the predictive performance and the improvement was obvious. This enhancement originates from reducing both bias and variance and avoiding over-fitting problems [56,57]. It is believed that the stacking technique and its comparison will guarantee a better result for further studies [58].

5.3. Maintain the Integrity of Geological Hazard Assessment

An optimal model should not only focus on accuracy, especially for geological hazard assessment [50]. An outstanding model should also require communication skills, that is, make it easy for researchers to understand, accept and apply, especially for natural disasters [59]. The capacity of communicating model behavior is another valuable quality for LSA, which is arduously achieved by machine learning methods because of the “black box” nature. Stacking performed the best in terms of accuracy while it had a low capacity for recognizing the importance of the variables. Gini index and FR were combined to determine the major conditioning factors and analyze the individual landslide-related factors in each interval and the relative importance among them, which improved the readability of the stacking model.

6. Conclusions

LSM is the basis of supplementary analyses, such as land use and hazard prevention. Meanwhile, this field of geomorphology has become an empty shell with no research question on whether a model can be evaluated in an unreliable sampling strategy and focus only on accuracy. Therefore, a more reliable and accurate landslide susceptibility map is urgently needed through further comparison and application of different methods. In the present study, three ensemble learning machines were compared in terms of the performance of LSM in Miyun County, Beijing, China. Non-landslide samples were determined in a more reliable way with the use of FCM and K-means clustering. Statistical indexes and AUC were combined to assess the accuracy performance of the models. The major conditioning factors were determined and analyzed based on the Gini index and FR. The following conclusions can be drawn from the present study:

1. The performance of different ensemble techniques varies, but achieved satisfactory results as a whole. Stacking was considered the most suitable model with obvious improvement in terms of accuracy compared to the basic classifiers.
2. The combination of the bivariate statistical method and Gini index helps better explore the major conditioning factors and improve the integrity of ensemble techniques.
3. The non-landslide samples selected by FCM are more representative and improved the quality of samples. Overall, improvement of sample quality and selection of advanced methods help improve the practicability of LSM.

Author Contributions: Z.L. writing—original draft, methodology, and software; W.L., W.P. and C.W., review and validation; L.C. reviewing and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Key-Area Research and Development Program of Guangdong Province (Grant No. 2020B0101130009) and Guangdong Enterprise Key Laboratory for Urban Sensing, Monitoring and Early Warning (No. 2020B121202019) and The Science and Technology Foundation of Guangzhou Urban Planning & Design Survey Research Institute (Grant No. RDI2210204140, RDI2210204146).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The related code of machine learning applied in the study is available at <https://github.com/Liangzhu-mz>, accessed on 6 April 2022.

Acknowledgments: The authors would like to thank the Editors and anonymous reviewers for their valuable comments, which improve this paper.

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

References

1. Huang, X.; Guo, F.; Deng, M.; Yi, W.; Huang, H. Understanding the deformation mechanism and threshold reservoir level of the floating weight-reducing landslide in the Three Gorges Reservoir Area, China. *Landslides* **2020**, *17*, 2879–2894. [[CrossRef](#)]
2. Sun, X.; Chen, J.; Li, Y.; Rene, N.N. Landslide Susceptibility mapping along a rapidly uplifting river valley of the Upper Jinsha River, Southeastern Tibetan Plateau, China. *Remote Sens.* **2022**, *14*, 1730. [[CrossRef](#)]
3. Kim, J.C.; Lee, S.; Jung, H.S.; Lee, S. Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. *Geocarto Int.* **2018**, *33*, 1000–1015. [[CrossRef](#)]
4. Safran, E.B.; O'Connor, J.E.; Ely, L.L.; House, P.K.; Grant, G.; Harrity, K.; Jones, E. Plugs or flood-makers? The unstable landslide dams of eastern Oregon. *Geomorphology* **2015**, *248*, 237–251. [[CrossRef](#)]
5. Zhu, A.X.; Miao, Y.; Wang, R.; Zhu, T.; Deng, Y.; Liu, J.; Hong, H. A comparative study of an expert knowledge-based model and two data-driven models for landslide susceptibility mapping. *Catena* **2018**, *166*, 317–327. [[CrossRef](#)]
6. Ayalew, L.; Yamagishi, H. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kaku-da-Yahiko Mountains, Central Japan. *Geomorphology* **2005**, *65*, 15–31. [[CrossRef](#)]
7. Jiao, Y.; Zhao, D.; Ding, Y.; Liu, Y.; Xu, Q.; Qiu, Y.; Liu, C.; Liu, Z.; Zha, Z.; Li, R. Performance evaluation for four GIS-based models purposed to predict and map landslide susceptibility: A case study at a World Heritage site in Southwest China. *Catena* **2019**, *183*, 104221. [[CrossRef](#)]
8. Shi, M.; Chen, J.; Song, Y.; Zhang, W.; Song, S.; Zhang, X. Assessing debris flow susceptibility in Heshigten Banner, Inner Mongolia, China, using principal component analysis and an improved fuzzy C-means algorithm. *Bull. Eng. Geol. Environ.* **2016**, *75*, 909–922. [[CrossRef](#)]
9. Liang, Z.; Wang, C.M.; Zhang, Z.M.; Khan, K.U.J. A comparison of statistical and machine learning methods for debris flow susceptibility mapping. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 1887–1907. [[CrossRef](#)]
10. Lian, C.; Zeng, Z.; Yao, W.; Tang, H. Extreme learning machine for the displacement prediction of landslide under rainfall and reservoir level. *Stoch. Environ. Res. Risk Assess.* **2014**, *28*, 1957–1972. [[CrossRef](#)]
11. Merghadi, A.; Abderrahmane, B.; Tien Bui, D. Landslide susceptibility assessment at Mila Basin (Algeria): A comparative as-sessment of prediction capability of advanced machine learning methods. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 268. [[CrossRef](#)]
12. Tien Bui, D.; Ho, T.C.; Revhaug, I.; Pradhan, B.; Nguyen, D.B. *Landslide Susceptibility Mapping Along the National Road 32 of Vietnam Using GIS-Based J48 Decision Tree Classifier and Its Ensembles[M]//Cartography from Pole to Pole*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 303–317.
13. Hu, X.; Zhang, H.; Mei, H.; Xiao, D.; Li, Y.; Li, M. Landslide susceptibility mapping using the stacking ensemble machine learning method in Lushui, Southwest China. *Appl. Sci.* **2020**, *10*, 4016. [[CrossRef](#)]
14. Bennett, G.L.; Miller, S.R.; Roering, J.J.; Schmidt, D.A. Landslides, threshold slopes, and the survival of relict terrain in the wake of the Mendocino Triple Junction. *Geology* **2016**, *44*, 363–366. [[CrossRef](#)]
15. Du, J.; Glade, T.; Woldai, T.; Chai, B.; Zeng, B. Landslide susceptibility assessment based on an incomplete landslide inventory in the Jilong Valley, Tibet, Chinese Himalayas. *Eng. Geol.* **2020**, *270*, 105572. [[CrossRef](#)]
16. Lee, S.; Min, K. Statistical analysis of landslide susceptibility at Yongin, Korea. *Environ. Earth Sci.* **2001**, *40*, 1095–1113. [[CrossRef](#)]
17. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
18. Varnes, D.J. Landslide types and processes. *Landslides Eng. Pract.* **1958**, *24*, 20–47.
19. Furlani, S.; Ninfo, A. Is the present the key to the future? *Earth-Sci. Rev.* **2015**, *142*, 38–46. [[CrossRef](#)]
20. Guzzetti, F.; Galli, M.; Reichenbach, P.; Ardizzone, F.; Cardinali, M.; Galli, M. Estimating the quality of landslide susceptibility models. *Geomorphology* **2006**, *81*, 166–184. [[CrossRef](#)]
21. Guzzetti, F.; Galli, M.; Reichenbach, P.; Ardizzone, F.; Cardinali, M. Landslide hazard assessment in the Collazzone area, Umbria, Central Italy. *Nat. Hazards Earth Syst. Sci.* **2006**, *6*, 115–131. [[CrossRef](#)]

22. Sun, X.L.; Zhao, Y.G.; Wang, H.L.; Yang, L.; Qin, C.Z.; Zhu, A.X.; Li, B. Sensitivity of digital soil maps based on FCM to the fuzzy exponent and the number of clusters. *Geoderma* **2012**, *171*, 24–34. [[CrossRef](#)]
23. Van Westen, C.J.; Castellanos, E.; Kuriakose, S.L. Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Eng. Geol.* **2008**, *102*, 112–131. [[CrossRef](#)]
24. Feizizadeh, B.; Blaschke, T.; Nazmfar, H. GIS-based ordered weighted averaging and dempster—Shafer methods for landslide susceptibility mapping in the Urmia Lake Basin, Iran. *Int. J. Digit. Earth* **2012**, *7*, 688–708. [[CrossRef](#)]
25. Hong, H.; Pradhan, B.; Xu, C.; Bui, D.T. 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](#)]
26. Magliulo, P.; Di Lisio, A.; Russo, F.; Zelano, A. Geomorphology and landslide susceptibility assessment using GIS and bivariate statistics: A case study in southern Italy. *Nat. Hazards* **2008**, *47*, 411–435. [[CrossRef](#)]
27. Liang, Z.; Wang, C.; Han, S.; Khan, K.U.J.; Liu, Y. Classification and susceptibility assessment of debris flow based on a semi-quantitative method combination of the fuzzy C-means algorithm, factor analysis and efficacy coefficient. *Nat. Hazards Earth Syst. Sci.* **2020**, *20*, 1287–1304. [[CrossRef](#)]
28. Evans, I.S. An integrated system of terrain analysis and slope mapping. *Z. Geomorphol.* **1980**, *36*, 274–295.
29. Camilo, D.C.; Lombardo, L.; Mai, P.M.; Dou, J.; Huser, R. Handling high predictor dimensionality in slope-unit-based landslide susceptibility models through LASSO-penalized generalized linear model. *Environ. Model. Softw.* **2017**, *97*, 145–156. [[CrossRef](#)]
30. Dou, J.; Yamagishi, H.; Xu, Y.; Zhu, Z.; Yunus, A.P. *Characteristics of the Torrential Rainfall-Induced Shallow Landslides by Typhoon Bilis, in July 2006, Using Remote Sensing and GIS[M]/GIS Landslide*; Springer: Tokyo, Japan, 2017; pp. 221–230.
31. Anil, K. Data clustering: 50 years beyond K-Means. *Pattern Recogn. Lett.* **2010**, *31*, 651–666.
32. Hartigan, J.; Wong, M. Algorithm AS 136: A K-means clustering algorithm. *J. R. Stat. Soc. C.* **1979**, *28*, 100–108. [[CrossRef](#)]
33. Dunn, J.C. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *J. Cybern.* **1973**, *3*, 32–57. [[CrossRef](#)]
34. Bezdek, J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
35. Wang, J.; Chen, J.; Yang, J. Application of distance discriminant analysis method in classification of surrounding rock mass in highway tunnel. *J. Jilin Univ.* **2008**, *38*, 999–1004.
36. Chen, J.; Pi, D. A cluster validity index for fuzzy clustering based on non-distance. In Proceedings of the 2013 International Conference on Computational and Information Sciences, Yongzhou, China, 21–23 June 2013; pp. 880–883.
37. Neter, J.; Wasserman, W.; Kutner, M.H. *Applied Linear Statistical Models*; Irwin: Chicago, IL, USA, 1996.
38. Fernández-Delgado, M.; Cernadas, E.; Barro, S.; Amorim, D. Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.* **2014**, *15*, 3133–3181.
39. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
40. 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.* **2014**, *73*, 3745–3761. [[CrossRef](#)]
41. Wang, Y.; Feng, L.; Li, S.; Ren, F.; Du, Q. A hybrid model considering spatial heterogeneity for landslide susceptibility mapping in Zhejiang Province, China. *Catena* **2020**, *188*, 104425. [[CrossRef](#)]
42. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of online learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [[CrossRef](#)]
43. Džeroski, S.; Ženko, B. Is combining classifiers with stacking better than selecting the best one? *Mach. Learn.* **2004**, *54*, 255–273. [[CrossRef](#)]
44. Chung, C.J.F.; Fabbri, A.G. Validation of spatial prediction models for landslide hazard mapping. *Nat. Hazards* **2003**, *30*, 451–472. [[CrossRef](#)]
45. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer: New York, NY, USA, 2013.
46. Green, D.M.; Swets, J.A. *Signal Detection Theory and Psychophysics*; Wiley: New York, NY, USA, 1966.
47. Schratz, P.; Muenchow, J.; Iturrutxa, E.; Richter, J.; Brenning, A. Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecol. Model.* **2019**, *406*, 109–120. [[CrossRef](#)]
48. Duarte, E.; Wainer, J. Empirical comparison of cross-validation and internal metrics for tuning SVM hyperparameters. *Pattern Recognit. Lett.* **2017**, *88*, 6–11. [[CrossRef](#)]
49. Bengio, Y. Gradient-based optimization of hyperparameters. *Neural Comput.* **2000**, *12*, 1889–1900. [[CrossRef](#)] [[PubMed](#)]
50. 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](#)]
51. Ciurleo, M.; Cascini, L.; Calvello, M. A comparison of statistical and deterministic methods for shallow landslide susceptibility zoning in clayey soils. *Eng. Geol.* **2017**, *223*, 71–81. [[CrossRef](#)]
52. Liu, R.; Yang, X.; Xu, C.; Wei, L.; Zeng, X. Comparative study of convolutional neural network and conventional machine learning methods for landslide susceptibility mapping. *Remote Sens.* **2022**, *14*, 321. [[CrossRef](#)]
53. Dou, J.; Yunus, A.P.; Bui, D.T.; Merghadi, A.; Sahana, M.; Zhu, Z.; Chen, C.; Han, Z.; Pham, B.T. Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. *Landslides* **2020**, *17*, 641–658. [[CrossRef](#)]

54. Di Napoli, M.; Carotenuto, F.; Cevasco, A.; Confuorto, P.; Di Martire, D.; Firpo, M.; Pepe, G.; Raso, E.; Calcaterra, D. Machine learning ensemble modelling as a tool to improve landslide susceptibility mapping reliability. *Landslides* **2020**, *17*, 1897–1914. [[CrossRef](#)]
55. Arabameri, A.; Chandra Pal, S.; Rezaie, F.; Chakraborty, R.; Saha, A.; Blaschke, T.; Thi Ngo, P.T. Decision tree based ensemble machine learning approaches for landslide susceptibility mapping. *Geocarto Int.* **2021**, 1–35. [[CrossRef](#)]
56. Li, W.; Fang, Z.; Wang, Y. Stacking ensemble of deep learning methods for landslide susceptibility mapping in the Three Gorges Reservoir area, China. *Stoch. Environ. Res. Risk Assess.* **2021**, 1–22. [[CrossRef](#)]
57. Chen, W.; Xie, X.; Wang, J.; Pradhan, B.; Hong, H.; Bui, D.T.; 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](#)]
58. Dietterich, T.G. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Mach. Learn.* **2000**, *40*, 139–157. [[CrossRef](#)]
59. Youssef, A.M.; Pourghasemi, H.R. Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia. *Geosci. Front.* **2021**, *12*, 639–655. [[CrossRef](#)]