

## Article

# Credal-Decision-Tree-Based Ensembles for Spatial Prediction of Landslides

Jingyun Gui <sup>1,2,\*</sup>, Ignacio Pérez-Rey <sup>2</sup> , Miao Yao <sup>3</sup>, Fasuo Zhao <sup>1</sup> and Wei Chen <sup>4</sup>

<sup>1</sup> Department of Geological Engineering, College of Geological Engineering & Surveying and Mapping, Chang'an University, Xi'an 710064, China

<sup>2</sup> GESSMin Group, CINTECX, Department of Natural Resources and Environmental Engineering, University of Vigo, 36310 Vigo, Spain

<sup>3</sup> China Information Industry Engineering Investigation and Research Institute, Xi'an 710001, China

<sup>4</sup> College of Geology and Environment, Xi'an University of Science and Technology, Xi'an 710054, China

\* Correspondence: guijingyun@chd.edu.cn; Tel.: +34-658-776-729

**Abstract:** Spatial landslide susceptibility assessment is a fundamental part of landslide risk management and land-use planning. The main objective of this study is to apply the Credal Decision Tree (CDT), adaptive boosting Credal Decision Tree (AdaCDT), and random subspace Credal Decision Tree (RSCDT) models to construct landslide susceptibility maps in Zhashui County, China. The observed 169 historical landslides were classified into two groups: 70% (118 landslides) for training and 30% (51 landslides) for validation. To compare and validate the performance of the three models, the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC) were utilized. Specifically, the success rates of the CDT model, AdaCDT model, and RSCDT model were 0.788, 0.821, and 0.847, respectively, while the corresponding prediction rates were 0.771, 0.802, and 0.861, respectively. In sum, the two ensemble models can effectively improve the performance accuracy of an individual CDT model, and the RSCDT model was proven to be superior to the other two models. Therefore, ensemble models are capable of being novel and promising approaches for the spatial prediction and zonation of a certain region's landslide susceptibility.

**Keywords:** Credal Decision Tree; AdaBoost; random subspace; machine learning; landslide



**Citation:** Gui, J.; Pérez-Rey, I.; Yao, M.; Zhao, F.; Chen, W. Credal-Decision-Tree-Based Ensembles for Spatial Prediction of Landslides. *Water* **2023**, *15*, 605. <https://doi.org/10.3390/w15030605>

Academic Editor: Olga Petrucci

Received: 8 January 2023

Revised: 26 January 2023

Accepted: 30 January 2023

Published: 3 February 2023



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

## 1. Introduction

Nowadays, landslides are regarded as one of the most hazardous geological risks in many areas all over the world, especially in mountainous regions. Once a landslide disaster occurs, it may cause huge losses. The occurrence of landslides can be associated with many factors, such as rainfall, earthquakes, human activities, etc. [1]. Therefore, the occurrence of landslides can be varied, which leads to the prediction of this kind of hazard becoming more difficult and it will also leave a heavy burden on economic and societal management. Moreover, due to the fact that China is a country with many mountains, hills, and plateaus, which account for 67% of the total land area, landslides are a quite common geological disaster. This has seriously threatened national security as well as people's safety and property. Over the past few years, there is an imperative and urgent task for disaster risk prevention and management. Hence, landslide sensitivity evaluation has grown to be a serious topic of concern.

Developing a landslide susceptibility map (LSM) with various scales for different specific purposes is a prerequisite for policymakers to assess susceptibility, manage land use and carry out decision-making activities. As dramatic advances have been made in geographic information systems (GIS) and data processing functions over the years, many quantitative techniques and approaches have been applied to make LSMs [2], including weights of evidence (WoE) [3,4], frequency ratios (FR) [5–8], bivariate statistics

(BS) [9,10], multivariate adaptive regression splines (MARS) [11,12], simple additive weighting (SAW) [13,14], analytic hierarchy processes (AHP) [15,16], Fisher's linear discriminant function (FLDA) [17,18], multivariate regression (MR) [11,12], discriminant analysis (DA) [18,19], and Bayesian logistic regression (BLR) [20–22]. Although various GIS-based models have been built to assess landslide susceptibility based on the mentioned quantitative approaches, a consensus on the most effective and accurate model has still not been arrived at since these approaches can be limited by unconsidered factors as well as the model performer's experience. Additionally, the results of quantitative approaches tend to be influenced by low-precision data.

The machine-learning method (MLM) differs from the aforementioned bivariate and multivariate statistical probability approaches. With the help of computer algorithms, the MLM can effectively learn information from training data to analyze and predict the results of spatial landslide sensitivity [23,24]. Additionally, various MLMs contain different distribution functions, which enable the advancement of all kinds of algorithms with various performances. The MLM belongs to artificial intelligence and has been combined with many algorithms, such as reduced error pruning trees (REPT) [25], naive Bayes trees [26], decision trees [27], artificial neural networks [28–31], support vector machine algorithms [7,32,33], long-term policy making algorithms [34], and multi-dimensional models with the constrained recursive least squares algorithms [35]. For different study objectives, these models show their advantages and disadvantages. Therefore, it is necessary to explore ensemble models to produce more accurate predictions for LSMs [36,37].

Recently, integrated models have been developed with the ability to comprehensively evaluate continuous and discrete data. The application of ensemble models contributes to outstanding performance and more reasonable results than individually used classifiers. Many researchers have used various tree-based ensemble learning algorithms, including bagging [38], rotation forest [38], bootstrap aggregation [39], random forest [30,40], gradient boosting [30], and extreme gradient boosting [41]. All of these ensemble models have been successfully applied in LSMs. Meanwhile, deep-learning algorithms have also received increasing attention due to the development of neural network techniques, such as deep convolutional neural networks [42,43], long short-term memory networks [44], and recurrent neural networks [45].

In the process of developing the above-mentioned approaches, though they can produce reasonable results and acceptable accuracy, researchers still try to seek new model ensembles to provide more reliable and accurate spatial predictions of landslide occurrence. Presently, various LSM models have been proposed, including the application of adaptive boosting (AdaBoost) [37,39,46,47] and random subspace (RS) [20,37,48,49] algorithms in several publications; however, there is still a gap in the application and integration of adaptive boosting (AdaBoost) and random subspace (RS).

Consequently, in this study, we introduce two CDT-based ensembles, AdaCDT and RSCDT, to assess landslide susceptibility. The purpose of this study is to estimate and compare the performance of CDT, AdaCDT, and RSCDT for the spatial prediction of landslides on a regional scale. The ROC curve has been applied to measure these MLM models. Data have been collected in Zhashui County (China). The analysis of landslide data and model study has been performed on ArcMap 10.5.

## 2. Study Area

Zhashui County, selected as the study area, is located in the western region of Shangluo (Shaanxi Province, China) and covering an area of nearly 2322 km<sup>2</sup>. The area is between the longitudes 108°50' E and 109°36' E, and the latitudes 33°25' N and 33°56' N (Figure 1). The lowest elevation in the study area is 516 m above sea level (a.s.l) and the highest elevation is 2763 m a.s.l. The slope angle varies from 0° to 74°. The averages of elevation and slope are 1306.96 m and 27.28°, respectively. The climate is a transition zone between subtropical and warm climates, with the characteristics of a monsoon climate due to the barrier function of the Qinling Mountains and their mountain topography. Precipitation is seasonal, with

80.5% of the annual rainfall falling in summer and early autumn, concentrated in July, August, and September. The average annual rainfall is 750 mm. Landslides frequently occur in the southeastern part of Zhashui County as illustrated in Figure 1.

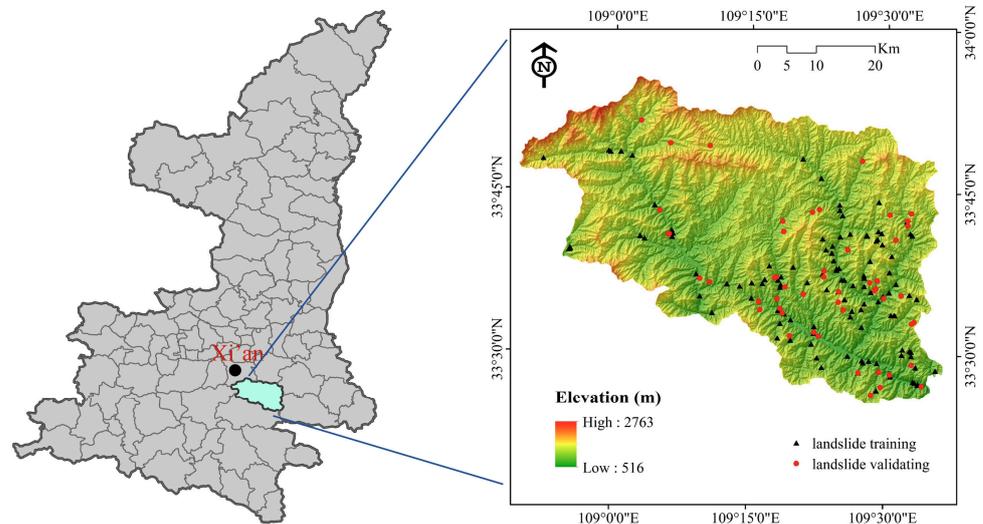


Figure 1. Geographical position of the study area.

### 3. Methodology

In this study, an analysis of landslide susceptibility was carried out using three approaches. The methodological steps are illustrated in detail in Figure 2, including: (1) the preparation of a landslide inventory map; (2) selecting the appropriate landslide conditioning factors; (3) modeling landslide susceptibility using CDT and its two ensembles—AdaCDT and RSCDT; (4) evaluating the success rate and prediction accuracy of the three models; and (5) making the landslide susceptibility map.

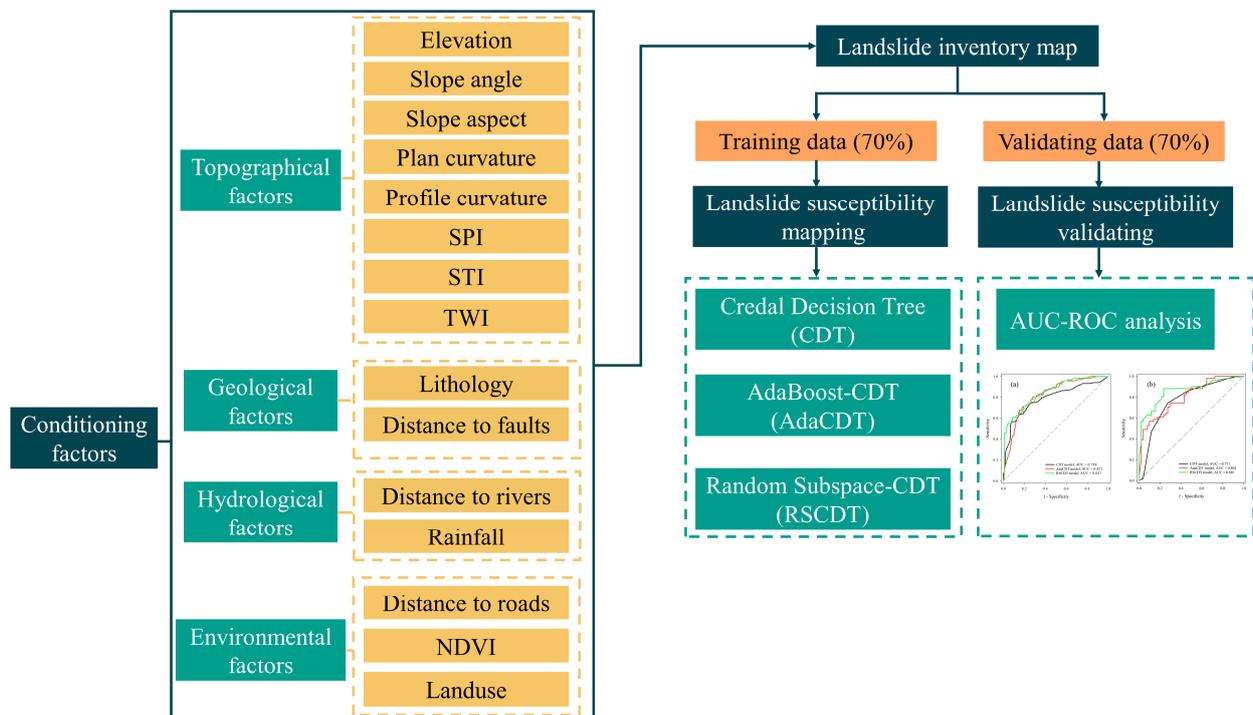


Figure 2. Methodological flowchart of the landslide susceptibility mapping used in the study.

### 3.1. Landslide Inventory Map

The target variable we wanted to model is required to be digitally represented in a landslide susceptibility inventory map, where the spatial distribution of landslides is reflected. The information on past and recent landslides is the foundation for the LSM. After fieldwork and remote mapping, a total of 169 landslides were identified in the present study area, which were then used as the target variable to model susceptibility. Specifically, each of the implemented models was fitted with 118 landslides (70%) as the training data and 51 landslides (30%) as the validating data [30,50,51].

### 3.2. Landslide Conditioning Factors

Generally, there are no universal criteria for the selection of landslide conditioning factors. According to a literature review, nearly 95 factors have been adopted to model LSMs by many researchers [25]. The selected conditioning factors are supposed to be the most important factors causing landslides, which will be vital to successfully obtain the general pattern of the formation of historical landslides. For this study, 15 conditioning factors were chosen for the LSMs depending on the local characteristics of the study area and the previous literature, which can be classified into four groups: topographic, geological, hydrological, and environmental factors. Topographic factors include elevation, slope angle, slope aspect, plan curvature, profile curvature, the stream power index (SPI), sediment transport index (STI), and topographic wetness index (TWI). The geological factors are lithology and distance to faults. Hydrological factors include distance to rivers and rainfall. Environmental factors include distance to roads, the normalized difference vegetation index (NDVI), and land use/cover.

Elevation is highly related to landslide occurrence; therefore, it is a vital conditioning factor when carrying out regional landslide susceptibility mapping [52]. In mountainous areas, the regional microclimate and human activities will be mainly affected by elevation, which will in turn trigger landslides, meaning landslides will be characterized by vertical zoning. In Figure 3a, elevation values were grouped into eight classes with an interval of 200 m.

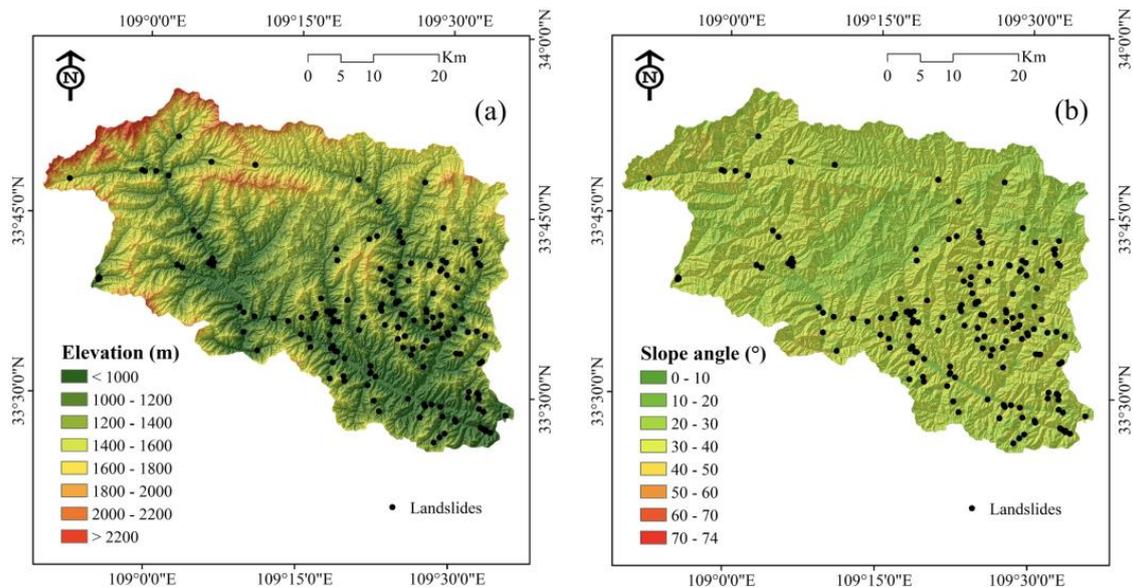


Figure 3. Cont.

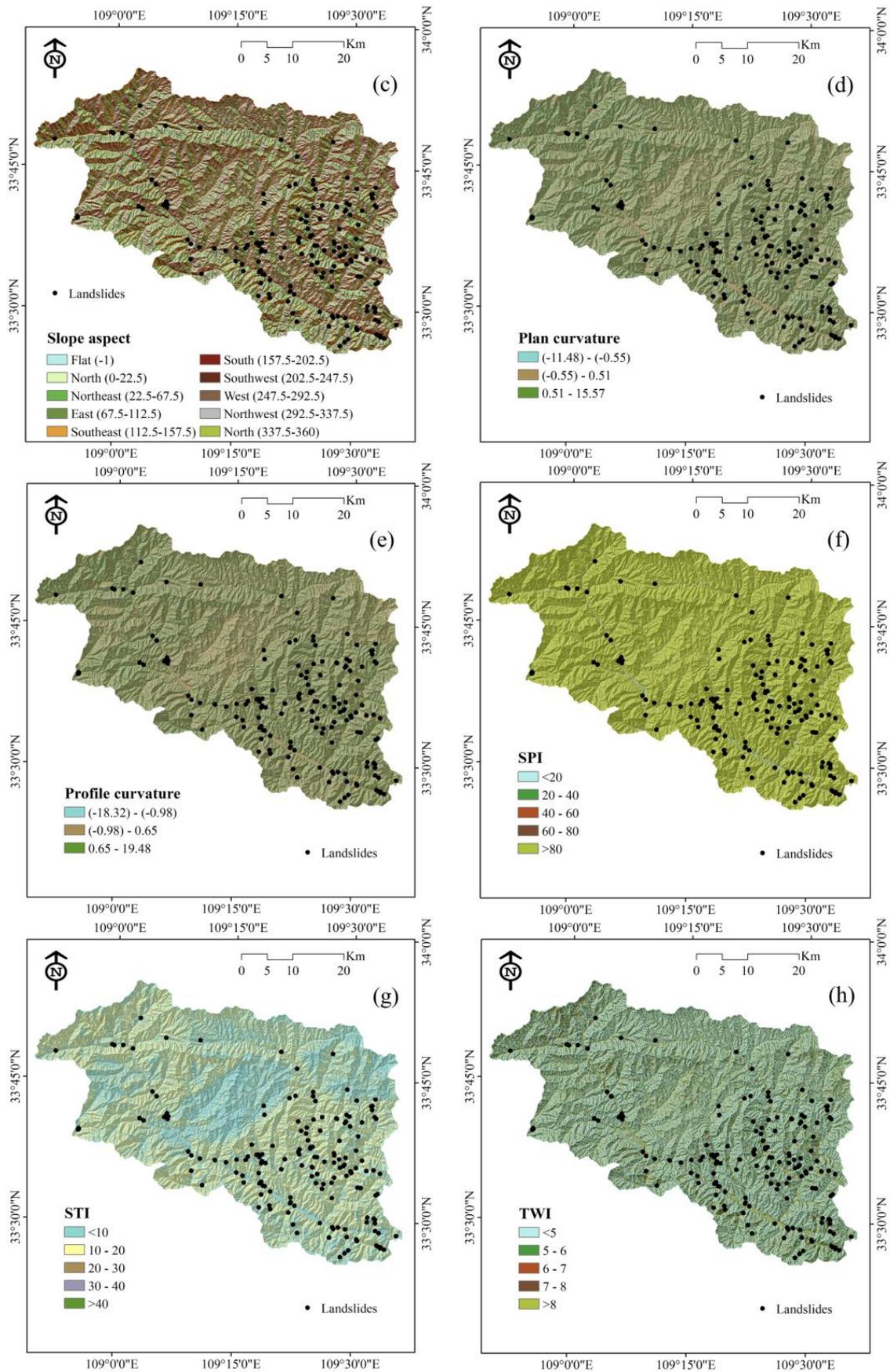


Figure 3. Cont.

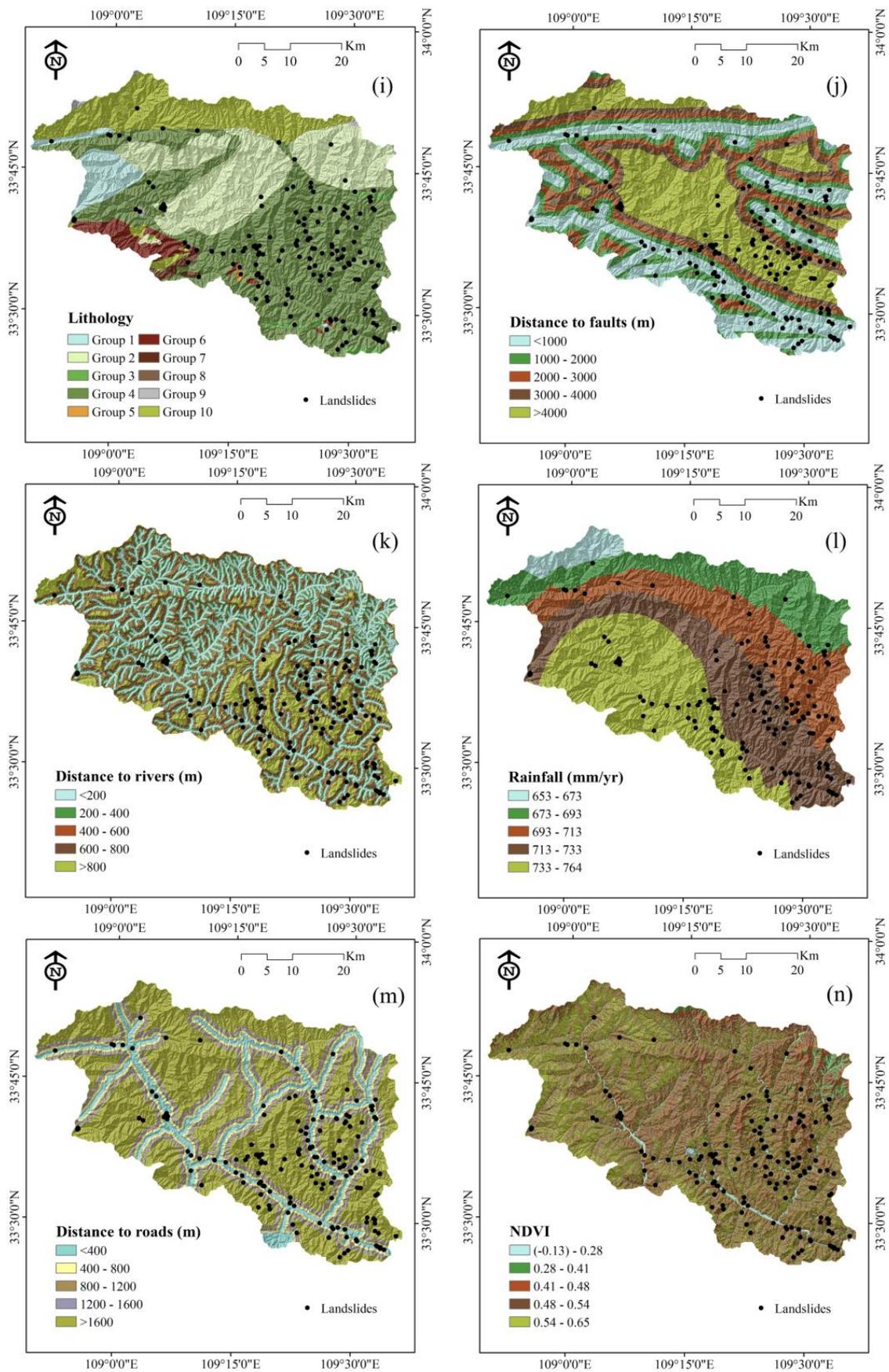
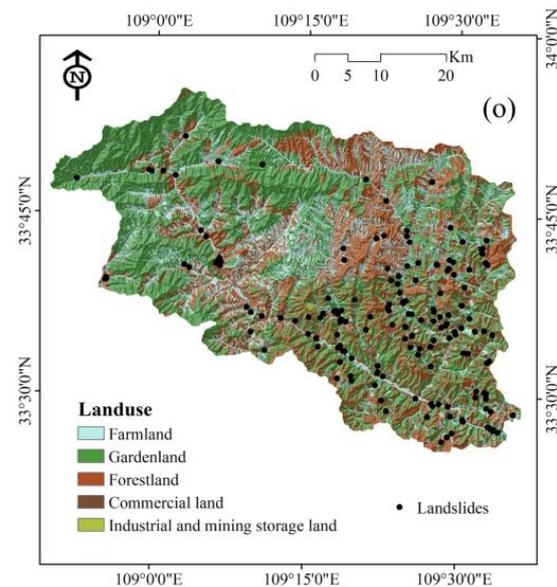


Figure 3. Cont.



**Figure 3.** Thematic maps with landslide dataset of the study area: (a) Elevation; (b) slope angle; (c) slope aspect; (d) plan curvature; (e) profile curvature; (f) SPI; (g) STI; (h) TWI; (i) lithology; (j) distance to faults; (k) distance to rivers; (l) rainfall; (m) distance to roads; (n) NDVI; (o) land use.

Slope angle is a key parameter to describe topography and one of the main controlling factors of landslide formation [16,53,54]. Slope angle will affect the stress distribution and hydraulic condition in the slope. Generally, steeper slopes tend to be more unstable and more prone to slide or topple than gentle ones. Additionally, slope gradient is a dominant factor in the erosion process by controlling the direction of the runoff. In this case, the values of slope angle were divided into eight classes with an interval of  $10^\circ$ , as illustrated in Figure 3b.

Slope aspect has an impact on the hours of sunshine, the intensity of solar radiation, and rainfall, which will affect vegetation coverage and soil moisture [16]. Hence, slope aspect has also been used as a common conditioning factor for a landslide susceptibility inventory. Ultimately, slope aspect within the study area was reclassified as: flat ( $-1$ ), north ( $0-22.5^\circ$ ), northeast ( $22.5^\circ-67.5^\circ$ ), east ( $67.5^\circ-112.5^\circ$ ), southeast ( $112.5^\circ-157.5^\circ$ ), south ( $157.5^\circ-202.5^\circ$ ), southwest ( $202.5^\circ-247.5^\circ$ ), west ( $247.5^\circ-292.5^\circ$ ), northeast ( $292.5^\circ-337.5^\circ$ ), and north ( $337.5^\circ-360^\circ$ ) (seeing in Figure 3c).

Ground curvature is a quantitative measure index of a point's distortion on the terrain surface, including plan curvature (Figure 3d) and profile curvature (Figure 3e) in this paper. Plan curvature is calculated from a contour line produced by the intersection of a horizontal plane and the surface. Plan curvature affects the erosion process by changing the overland flow rate [55]. Profile curvature is described as the corresponding normal section, which is tangential to the streamline. It is positive for an upwardly convex surface and negative for an upwardly concave one [56]. The plan curvature and profile curvature values can be computed by GIS software and then reclassified into three groups.

SPI was initially proposed by Moore et al. [57], indicating the erosion capacity of concentrated flow [58]. Therefore, SPI gives a good prediction of flow detachment risk. High SPI values show a strong erosion ability on the slope surface. SPI can be rigorously calculated according to Equation (1),

$$\text{SPI} = A_s \cdot \tan \beta \quad (1)$$

where  $A_s$  ( $\text{m}^2/\text{m}$ ) is the unit contributing area,  $\beta$  is the slope angle (degrees), and  $\tan \beta$  is slope gradient ( $\text{m}/\text{m}$ ). In this study, STI values were computed and then divided into five groups with an interval of 20, including  $<20$ , 20–40, 40–60, 60–80, and  $>80$  (Figure 3f).

Similar to SPI, STI was used to evaluate the sediment transporting capacity [57]. This index is expressed as a function of the local slope and contributing area:

$$STI = \left[ \frac{A_s}{22.13} \right]^{0.6} \cdot \left[ \frac{\sin \beta}{0.0896} \right]^{1.3} \quad (2)$$

In this case, five categories with an interval of 10 were generated for STI values: <10, 10–20, 20–30, 30–40, and >40 (Figure 3g).

TWI was established by Beven and Kirkby [59] within a runoff model and improved by Moore et al. [60]. It can be used to indicate the quantitative correlation between topographic features and soil wetness. It is relevant to note that the soil covering slopes tend to be unstable with larger moisture content due to the decreases in effective forces. TWI is computed as:

$$TWI = \ln \left( \frac{A_s}{\tan \beta} \right) \quad (3)$$

Generally, high TWI values mean that the soil in the corresponding area has a high moisture content. In this study, TWI values were reclassified into five groups: <5, 5–6, 6–7, 7–8, and >8 (Figure 3h).

Lithology has an effect on slope stability since different lithological units have different physical and mechanical properties. Therefore, different lithological units generate various susceptibilities to landslides [61]. In rock masses, soft and hard interlayers with a high moisture content are more likely to cause landslides. In the study area, ten groups of lithology were generated based on their lithofacies and geological ages, as illustrated in Figure 3i and Table 1.

**Table 1.** Description of the groups of lithology.

Group	Code	Lithology	Geological Age
1	J <sub>2</sub>	Monzonitic granite, quartz monzonite, granodiorite, quartz diorite	Middle Jurassic
2	T <sub>2</sub> , T <sub>3</sub>	Quartz monzonite, monzonitic granite, granodiorite	Middle and late Triassic
3	C <sub>1</sub> , C <sub>2</sub>	Lower: carbonaceous phyllite; middle: siltstone, gray-green phyllite; upper: medium-thin bedded limestone; carbonaceous slate with quartz sandstone, carbonaceous slate, slate-sandwiched sandstone, quartz conglomerate, and limestone, breccia limestone	Early and middle Carboniferous
4	D <sub>1</sub> , D <sub>2</sub> , D <sub>3</sub>	Lower: sandstone sandwiches slate, sandy argillaceous limestone, and local siderite sandwiches; upper: slate and phyllite-sandwiched sandstone, dolomite, limestone, sandstone, siltstone with a small amount of slate, locally intercalated argillaceous limestone, slate mixed with fine sandstone	Devonian
5	S	Granite	Silurian
6	O	Quartz diorite, diorite, gabbro, gabbro-norite, alaskite	Ordovician
7	Є <sub>1</sub>	Lower: black carbonaceous slate and siliceous rock; upper: variegated (dark gray, gray-purple, light gray, gray-white) limestone, dolomitic limestone; dolomite with flint	Cambrian
8	Z <sub>1</sub> , Z <sub>2</sub>	Lower: conglomerate, sandstone, shale with limestone; upper: dolomite, marl with sandstone, shale	Early and middle Sinian
9	Pz <sub>2</sub>	Lower: mainly metamorphic quartz sandstone, meta granulite with mica-quartz schist; upper: sandy conglomerate, meta-sandstone, mica-quartz schist with a few marble layers from bottom to top	Upper Paleozoic
10	Pt <sub>1</sub> , Qn	Biotite schist, graphite marble, clastic rock interbedded with basic lava, volcanic rock with marble, clastic rock with basic lava, volcanic rock with carbonaceous phyllite, marble, and siliceous rock	Lower Proterozoic, Qingbaikouan

The proximity to fault structures is a factor influencing slope stability. It is believed that areas where tectonic activity plays a role will lead to more frequent geological hazards [62,63]. According to relevant studies [56,64], most landslides occur within a distance of 250–1000 m from faults. Hence, the distance to faults was selected as an important conditioning factor for LSMs. Consequently, buffers of the main faults within the study area were computed and grouped into five classes with an interval of 1000 m, namely <1000 m, 1000–2000 m, 2000–3000 m, 3000–4000 m, and >4000 m (Figure 3j).

The stability of slopes is affected by the neighboring rivers since rivers tend to change the degree of saturation. Streams may adversely cause slope instability by eroding the toes of slopes and saturating the slope [65]. Five different buffer ranges were generated with an interval of 200m. The map of the distance to rivers is given in Figure 3k.

Rainfall is a prime triggering factor for landslides. It has been proven that most landslides occur during or after continuous rainfall due to the rising groundwater level and pore water pressure [66]. The rainfall map was obtained by reclassifying the meteorological data into five groups with an interval of 20 mm/yr (Figure 3l): 653–673 mm/yr, 673–693 mm/yr, 693–713 mm/yr, 713–733 mm/yr, and 733–764 mm/yr.

The study area is in a mountainous area, and road construction activities are inevitable which have an adverse effect on slope stability [67]. Mountain excavation and subgrade filling will create a large number of new artificial slopes, destroy the integrity of slope structures, and lead to landslides with a considerable risk to road facilities and human life. Hence, the distance to roads was chosen as a factor to make the LSM map. For the study area, the values of distance to roads were obtained and reclassified into five categories: <400 m, 400–800 m, 800–1200 m, 1200–1600 m, and >1600 m (Figure 3m).

It is a complex task to evaluate the relationship between vegetation coverage and slope stability [68]. In particular, high vegetation coverages of slopes have prominent inhibitory impacts on shallow landslides and surface erosion [69]. NDVI has been utilized by many researchers to quantitatively express the degree of vegetation coverage on slope surfaces [69] and was considered a conditioning factor in this paper. NDVI is defined as,

$$NDVI = \frac{NIR - R}{NIR + R} \tag{4}$$

where NIR is the near-infrared band, and R is the red band of the electromagnetic spectrum. The values of NDVI vary from −1 to 1, where 1 means the corresponding areas are perfectly covered by vegetation. In the present case, NDVI values ranged from −0.13 to 0.65, which were then grouped into five classes based on the natural break method, i.e., −0.13–0.28, 0.28–0.41, 0.41–0.48, 0.48–0.54, and 0.54–0.65 (Figure 3n).

Land cover is an indirect conditioning factor for slope stability. Usually, barren and sparsely vegetated regions tend to be more susceptible to erosion and will bear greater instabilities compared with a thick forest [70]. Generally, agriculture is conducted on gentle and low slopes. However, some farming activities take place on moderately steep slopes, which will lead to rising water levels due to long-term irrigation. Therefore, landslides are more likely to occur on agricultural slopes. The study area was classified into five groups (Figure 3o), including farmland, garden land, forestland, commercial land, and industrial and mining storage land.

The information on the source and scale of the conditioning factors is illustrated in Table 2.

**Table 2.** Source and scale of conditioning factors.

Factors	Data Source	Format Resolution/Scale
Elevation, slope angle, slope aspect, plan curvature, profile curvature, SPI, STI, TWI, distance to faults, distance to roads, distance to rivers	ASTER GDEM	Raster, 30 m
NDVI	Landsat 8 operational land imager	Raster, 30 m
Lithology	Geological maps	Polygon, 1:200,000
Rainfall	National Earth System Science Data Center	Raster, 30 m
Land use/cover	Land use/cover maps	Polygon, 1:100,000

### 3.3. Modeling Approaches

#### 3.3.1. Credal Decision Tree

Abellán and Moral [71] initially built the Credal Decision Tree (CDT) to solve classification problems involving credal sets. This approach applies an original split criterion by considering imprecise probability and uncertainty. In order to reduce the generation of complex decision trees during the CDT construction process, an exclusion criterion was used. The construction process stopped when the summation uncertainty increased due to splitting. An improved approach was recommended based on Dempster–Shafer’s theory [72,73], which has been extensively applied to analyze the uncertainty measures of a credal dataset. During the development of CDT, the following equation was introduced to quantitatively calculate the entire uncertainty (EU) with two parts [71],

$$EU(\varphi) = NG(\varphi) + RG(\varphi) \tag{5}$$

where  $\varphi$  expresses a general credal set on a frame  $X$ , EU represents the entire uncertainty, NG is a general non-specificity function, and RG denotes a general randomness function of credal sets.

Function  $m_\varphi$  is defined as an assignment of masses on  $\varphi$ . For a general credal set on the frame  $X$ , the formula of the non-specificity state can be written as [71],

$$NG(\varphi) = \sum_{A \subset X} m_\varphi(A) \ln(|A|) \tag{6}$$

where  $A$  is the power set of  $X$ .

The function of the randomness of a general credal set can be computed as [71],

$$RG(\varphi) = \text{Max} \left\{ - \sum_{x \in X} p_x \ln p_x \right\} \tag{7}$$

where the maximum is taken over all probability distributions on a credal set  $\varphi$ .

The basic arithmetical function for the CDT model can be expressed as the following computations. Based on the landslide dataset  $D$ , the distribution of probability  $p(LI)$  can be defined as [71]:

$$p(LI) \in \left[ \frac{n(LI)}{N + s}, \frac{n(LI) + s}{N + s} \right] \tag{8}$$

where  $N$  represents the sample dataset size,  $LI$  is a landslide indicator,  $n(LI)$  is the frequency value, and  $s$  denotes the hyper-parameter which ranges from 1 to 2, as stated by Walley [74].

Therefore, a new kind of credal set  $K^D$  can be expressed as follows:

$$K^D = \left\{ p | p(LI) \in \left[ \frac{n(LI)}{N + s}, \frac{n(LI) + s}{N + s} \right] \right\} \tag{9}$$

Based on the new credal set  $K^D$ , the procedure to build the CDT algorithm utilizes the maximum entropy function. This function is a total uncertainty measure in the imprecise Dirichlet Model (IDM). Figure 4 shows the basic learning process. A more detailed procedure of the CDT algorithm can be found in the literature [71].

#### 3.3.2. AdaBoost

AdaBoost or adaptive boosting was initially proposed by Freund and Schapire [75,76]. This kind of approach was derived from an online allocation algorithm, training and assembling multiple weak classifiers to achieve a strong one by means of a boosting process. Every classifier within the ensemble model attempts to classify the training data accurately. The working procedure of AdaBoost can be expressed as follows. First, a weak learner is generated with the original training dataset. Then, the distribution of the training data is adjusted based on the prediction performance for the next iteration of weak learner

training. It is relevant to be noted that the misclassified samples are recognized and will be assigned higher weights while the correctly graded samples weigh the same. Next, a strong classifier is generated based on all the weak classifiers and the corresponding weights. Finally, a good classifier model is built from a weighted sum of all the classifier-based models previously constructed.

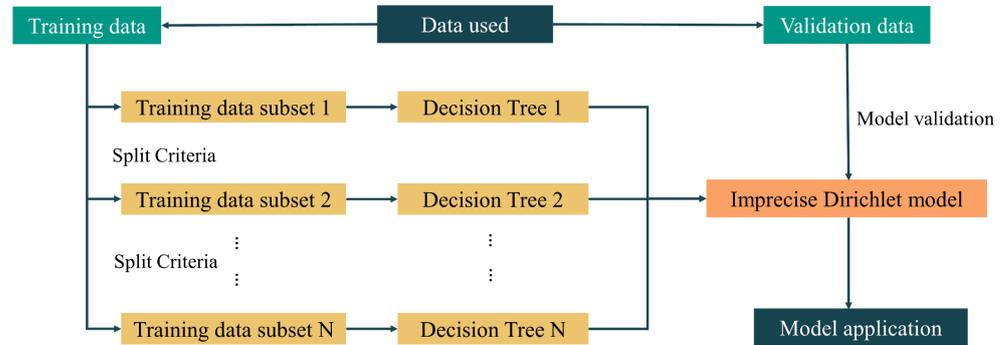


Figure 4. Credal Decision Tree learning process: training and validating.

A weak classifier is defined as  $y_m$  and the corresponding classification error rate is  $\epsilon_m$ .  $\alpha_m$  represents the weight of the weak classifier  $y_m$ .  $Y_M$  is the final strong classifier integrated from all the weak classifiers. The detailed mathematical steps are presented as follows [76].

First, the classification error rate of a weak classifier for the training data is,

$$\epsilon_m = \sum_{n=1}^N w_n^{(m)} I(y_m(x_n) \neq t_n) \tag{10}$$

where  $y_m(x_n)$  is the prediction outcome of the weak classifier,  $t_n$  is the true label,  $I$  represents the weight coefficient optimization function, and  $w_n^{(m)}$  is the weight of the current weak classifier.

Next, the weights of weak classifiers can be obtained,

$$\alpha_m = \frac{1}{2} \ln\left(\frac{1 - \epsilon_m}{\epsilon_m}\right) \tag{11}$$

Finally, the function of a strong classifier  $Y_M$  can be given based on  $\alpha_m$  (Figure 5),

$$Y_M(x) = \text{sign}\left(\sum_{m=1}^M \alpha_m y_m(x)\right) \tag{12}$$

where  $M$  is the number of weak classifiers and  $y_m(x)$  is the prediction result of each weak classifier. Figure 5 gives a basic depiction of AdaBoost.

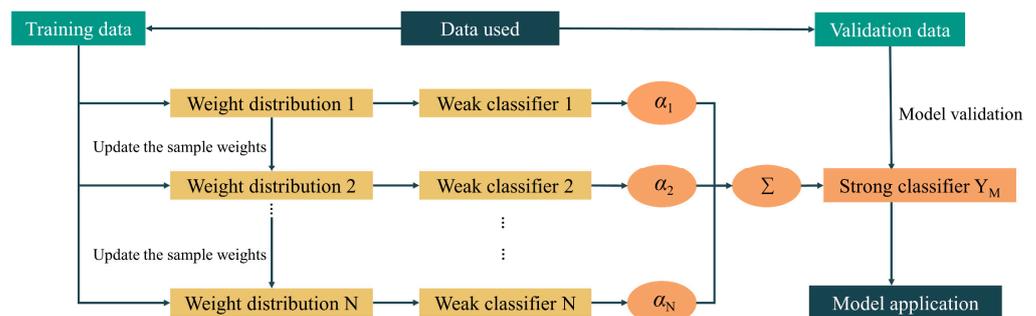


Figure 5. AdaBoost learning process: training and validating.

### 3.3.3. Random Subspace (RS)

Random subspace was originally established by Ho [77], combining and training multiple classifiers on a modified feature space to strengthen weak classifiers. The basic concept of a random subspace ensemble model is the implementation of a pseudorandom process to select components of a feature vector. In the RS classifier, two algorithms are joined. First, low-dimensional subspaces are generated by randomly sampling vectors of the original high-dimensional feature. Then, multiple classifiers are blended into these random subspaces at the end of the predictions. The significant discrepancy between the RS method and other approaches is that RS randomly selects features of the original training data [78]. Figure 6 gives a basic depiction of random subspace.

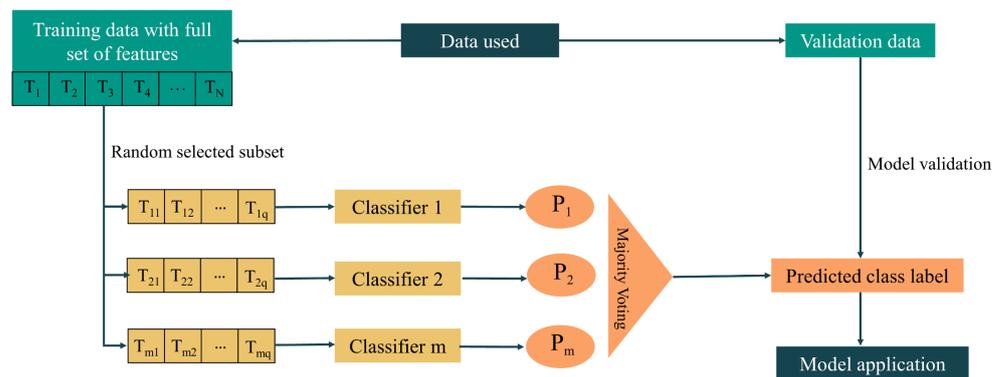


Figure 6. Random subspace learning process: training and validating.

With the predictions of weak classifiers, a decision is made by simple majority voting in the final decision rule. Accounting for the fact that it is easy to train the classifiers based on smaller subspaces using the random subspace approach, a distinct improvement in the features to instance ratio can be obtained. The detailed process involves the following steps.

First, given a training dataset  $X$  of size  $m$ ,

$$X = \{T_1, T_2, \dots, T_m\} \tag{13}$$

For a given  $q$ -dimensional data set, each set of a training sample  $T_i$  is assigned a  $q$ -dimensional feature vector,

$$T_i = \{T_{i1}, T_{i2}, \dots, T_{iq}\}; i = (1, 2, \dots, m) \tag{14}$$

Next, the  $q^*$  dimension feature subset is randomly generated from  $T_i$ . It is relevant to note that  $q^*$  should be smaller than  $q$ .

Then, the training sample of primordial dataset  $X$  turns into  $X^r$ , written as,

$$X^r = \{T_1^r, T_2^r, \dots, T_m^r\} \tag{15}$$

Taking into account that each training sample of  $X^r$  is a  $q^*$ -dimensional feature vector,

$$X^r = \{T_1^r, T_2^r, \dots, T_{q^*}^r\} \tag{16}$$

Randomly assigning the feature element  $X_{ik}^r$  with uniform distribution,  $k$  belongs to 1 to  $q^*$ , and the range of  $i$  is from 1 to  $m$ .

Subsequently, using the random subspace method, the set of  $N$  number of base classifiers is obtained with  $X^r$ , namely,  $C^n(x)$  ( $n = 1, 2, \dots, N$ ).

Ultimately, a decision is made using the simple majority voting combination rule [79],

$$h(X) = \underset{y \in [-1,1]}{\operatorname{argmax}} \sum_{i=1}^n C^n(x), y \tag{17}$$

where  $y \in [1, -1]$  means the class label decision, and  $C^n$  is the ensemble size of classifier.

#### 4. Results

##### 4.1. Correlation Analysis between Landslide and Conditioning Factors Using Frequency Ratio Method

To assess and quantify the spatial relationship between the historical landslides and the selected fifteen conditioning factors, the frequency ratio (FR) approach was used. FR is a straightforward statistical approach for exploring the spatial relationship between a case and its influence factors. This approach has been widely applied [39,46,78]. The larger the FR value, the closer the relationship between the case and a factor. In this study, FR values of each class of the fifteen conditioning factors are illustrated in Table 3. It can be noted that the most landslide-prone areas fall on commercial land of the land use/cover, with an FR value of 9.04, making this the most relevant factor leading to landslide occurrence. The following highly correlated factors are an SPI of 0–20 (FR = 9.00), NDVI of (−0.13)–0.28 with an FR value of 6.66, a slope angle of 0–10 (FR = 3.50), and an elevation of 0–1000 m with an FR value of 3.39. For a specific class, an FR value of 0 denotes the landslide-insusceptible portions of the study area.

**Table 3.** Spatial relationship between conditioning factors and historical landslides using FR method.

Factor	Subclass	No. of Class Pixels	No. of Landslide Pixel	FR Value
Elevation (m)	<1000	434680	66	3.39
	1000–1200	582467	35	1.34
	1200–1400	684338	13	0.42
	1400–1600	497394	4	0.18
	1600–1800	252707	0	0.00
	1800–2000	114189	0	0.00
	2000–2200	47075	0	0.00
	>2200	23672	0	0.00
Slope (°)	0–10	134147	21	3.50
	10–20	488228	36	1.65
	20–30	927360	32	0.77
	30–40	794942	23	0.65
	40–50	269518	3	0.25
	50–60	21531	3	3.11
	60–70	777	0	0.00
7–74	19	0	0.00	
Aspect	Flat	162829	8	1.10
	North	295253	19	1.44
	Northeast	355388	15	0.94
	East	376859	21	1.25
	Southeast	306409	23	1.68
	South	321254	10	0.70
	Southwest	338043	17	1.12
	West	318849	5	0.35
Northwest	161638	5	0.69	
Plan curvature	(−11.48)–(−0.55)	517544	20	0.86
	(−0.55)–0.51	1520401	82	1.21
	0.51–15.57	598577	16	0.60
Profile curvature	(−18.32)–(−0.98)	400907	12	0.67
	(−0.98)–0.65	1632087	74	1.01
	0.65–19.48	603528	32	1.18

Table 3. Cont.

Factor	Subclass	No. of Class Pixels	No. of Landslide Pixel	FR Value
SPI	0–20	29795	12	9.00
	20–40	6605	2	6.77
	40–60	9282	2	4.81
	60–80	8520	3	7.87
	>80	2582320	99	0.86
STI	0–10	1467316	82	1.25
	10–20	1148396	36	0.70
	20–30	19775	0	0.00
	30–40	841	0	0.00
	>40	194	0	0.00
TWI	0–5	1294286	27	0.47
	5–6	714821	28	0.88
	6–7	284237	25	1.97
	7–8	126771	10	1.76
	>8	216407	28	2.89
Lithology	Group 1	123884	1	0.18
	Group 2	639471	3	0.10
	Group 3	31452	2	1.42
	Group 4	1311573	110	1.87
	Group 5	4482	0	0.00
	Group 6	89099	0	0.00
	Group 7	21375	1	1.05
	Group 8	2655	0	0.00
	Group 9	10151	1	2.20
	Group 10	402380	0	0.00
Distance to faults (m)	0–1000	684637	42	1.37
	1000–2000	518613	16	0.69
	2000–3000	416778	19	1.02
	3000–4000	310547	17	1.22
	>4000	705947	24	0.76
Distance to rivers (m)	0–200	803551	67	1.86
	200–400	656896	17	0.58
	400–600	454343	10	0.49
	600–800	279554	6	0.48
	>800	442178	18	0.91
Rainfall (mm/yr)	653–673	93654	0	0.00
	673–693	466030	5	0.24
	693–713	573189	32	1.25
	713–733	724650	47	1.45
	733–764	778999	34	0.98
Distance to roads (m)	0–400	803551	47	1.31
	400–800	656896	8	0.27
	800–1200	454343	6	0.30
	1200–1600	279554	5	0.40
	>1600	442178	52	2.63
NDVI	(−0.13)–0.28	57074	17	6.66
	0.28–0.41	169814	43	5.66
	0.41–0.48	513496	22	0.96
	0.48–0.54	989002	21	0.47
	0.54–0.65	907136	15	0.37
Land use/cover	Farmland	531645	46	1.93
	Garden land	1231021	34	0.62
	Forestland	861886	34	0.88
	Commercial land	9889	4	9.04
	Industrial and mining storage land	2081	0	0.00

#### 4.2. Application of Landslide Susceptibility Models

Landslide susceptibility frameworks were constructed with the training dataset using the CDT model and its two ensembles, namely the AdaCDT model and RSCDT model. Although there are still some discussions about the accuracy of proposed modeling approaches in various publications, it is indeed necessary to carry out some research on assessing the performance of new models on landslide susceptibility mapping in terms of spatial analysis and prediction ability. In the present study, landslide susceptibility indices (LSI) based on the three models for Zhashui County were computed to produce the rasterized LSM in the ArcGIS 10.5 software. Finally, all the LSMs were regrouped into five classes, including very low, low, moderate, high, and very high, based on the natural breaks method [25,50]. The area percentage of each group for every model is illustrated in Table 4.

**Table 4.** Percentages of different landslide susceptibility classes for CDT, AdaCDT, and RSCDT models.

Class	CDT	AdaCDT	RSCDT
Very low	44.92	15.84	35.45
Low	4.26	15.94	24.43
Moderate	15.70	7.80	19.65
High	20.08	11.10	14.88
Very high	15.05	49.33	5.59

For the CDT model (Figure 7), the LSI ranges from 0.000 to 1.000, and the reclassified five groups are as follows: very low (0.000–0.031), low (0.031–0.165), moderate (0.165–0.325), high (0.325–0.706), and very high (0.706–1.000). The corresponding area percentage is 44.92%, 4.26%, 15.70%, 20.08%, and 15.05%, respectively. Regarding the two ensemble models, LSIs of the AdaCDT (Figure 8) and RSCDT (Figure 9) vary from 0.000 to 1.000, and 0.097 to 0.900, respectively. The LSM generated by the AdaCDT model is covered by an area percentage of 15.84% with very low (0.000–0.090), 15.94% with low (0.090–0.267), 7.80% with moderate (0.267–0.502), 11.10% with high (0.502–0.780), and 49.33% with very high susceptibility (0.780–1.000). Additionally, regarding the RSCDT model, the five regrouped classes are very low (0.097–0.289), low (0.289–0.428), moderate (0.428–0.579), high (0.579–0.742), and very high susceptibility (0.742–0.900), accounting for 35.45%, 24.43%, 19.65%, 14.88%, and 5.59%, respectively.

#### 4.3. Model Performance and Validation

To figure out the optimal model for the study area among the three models, the ROC curve was applied, and AUC was computed based on the training data. The results are illustrated in Table 5 and Figure 10a. It can be noted that the AUC of the CDT model is 0.788, showing that the goodness-of-fit to the training data is 78.8%. Apparently, the two ensemble frameworks perform much better on goodness-of-fit to the training data than the CDT model. The RSCDT model outperforms the three models with an AUC of 0.847, followed by the AdaCDT model (0.821). Based on the training data, an AdaBoost algorithm and random subspace algorithm can enhance the success rate of the CDT model, by percentage of 4.02% and 6.97%, respectively.

**Table 5.** ROC analysis of CDT, AdaCDT, and RSCDT models using training data.

Models	AUC	Standard Error	95% Confidence Interval
CDT	0.788	0.0304	0.728–0.847
AdaCDT	0.821	0.0274	0.767–0.875
RSCDT	0.847	0.0245	0.799–0.895

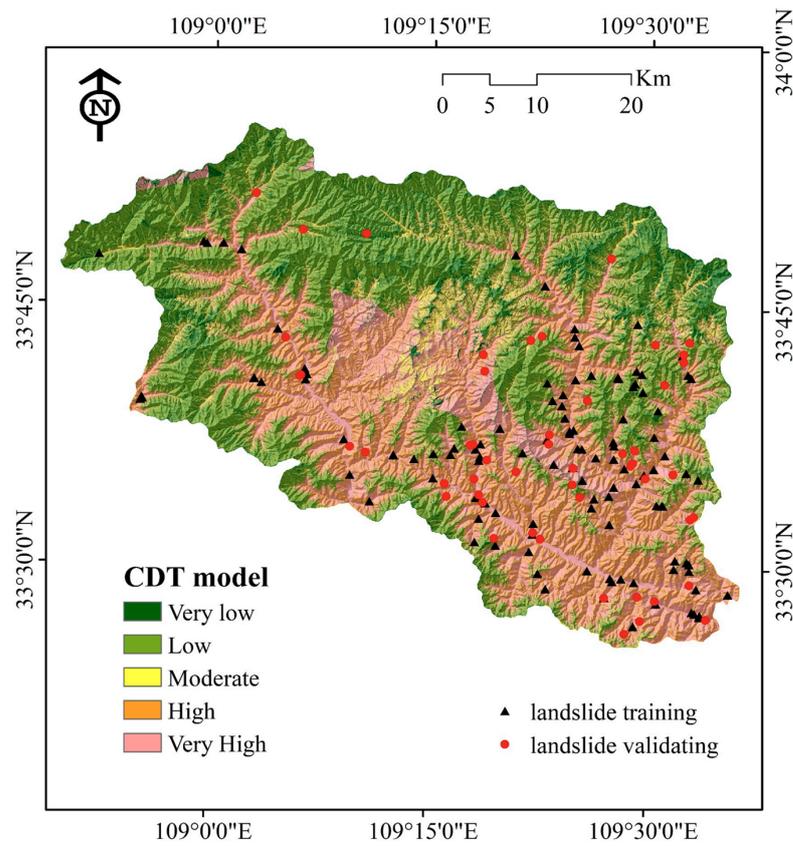


Figure 7. Landslide susceptibility map using the CDT model.

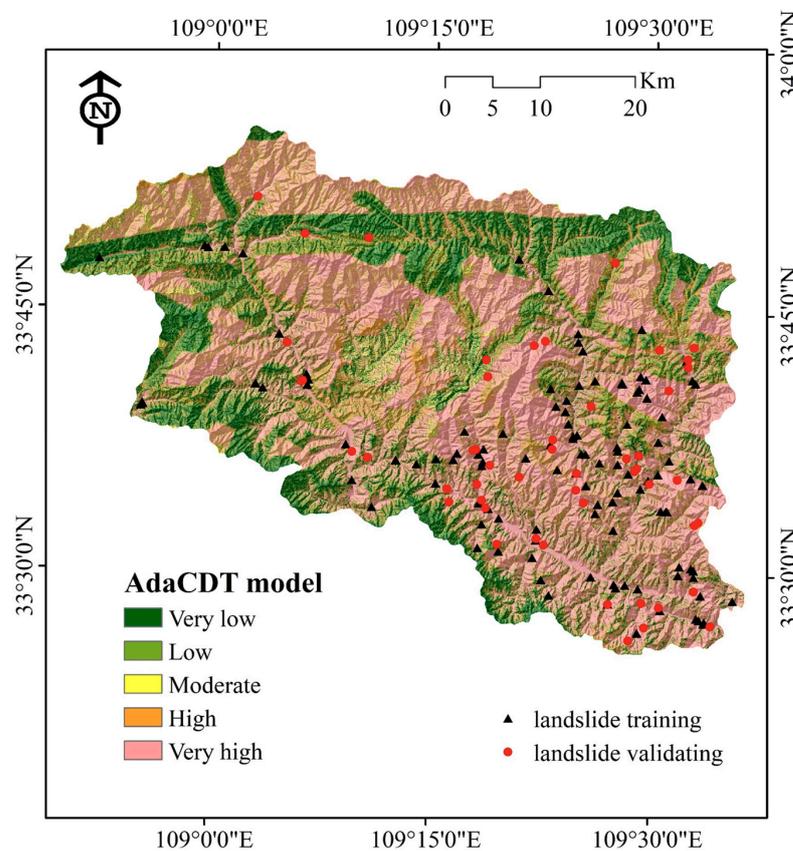


Figure 8. Landslide susceptibility map using the AdaCDT model.

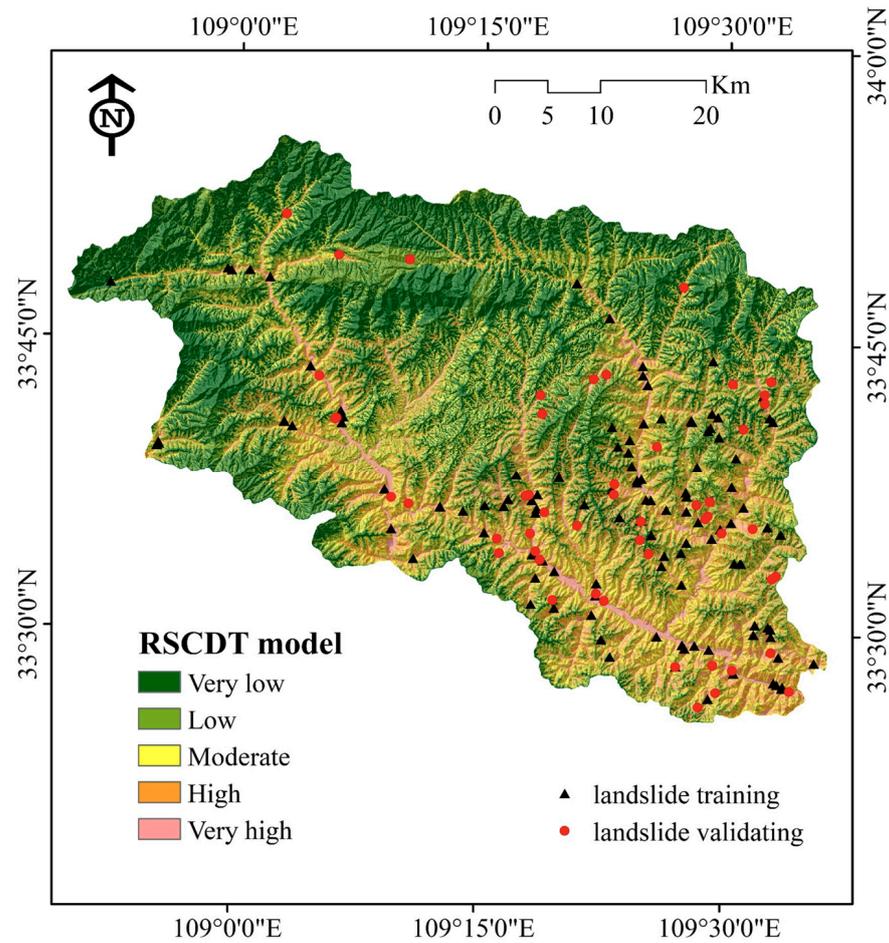


Figure 9. Landslide susceptibility map using the RSCDT model.

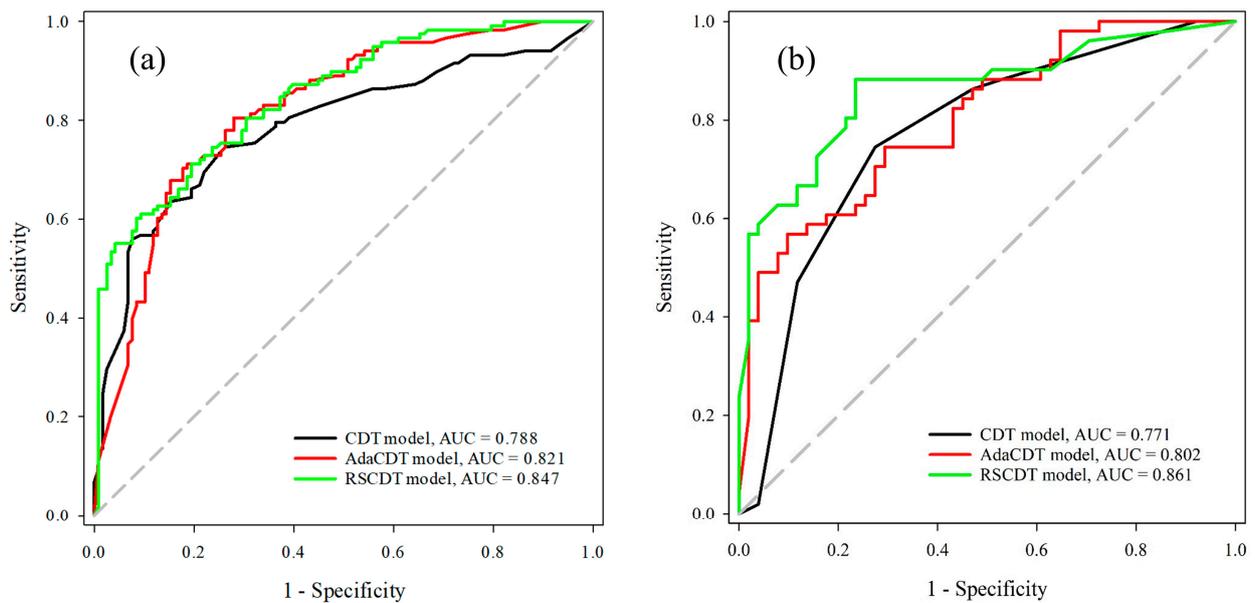


Figure 10. ROC curves and AUC analysis for LSM using the three models: (a) goodness-of-fit from the training data; (b) prediction rates from the validating data.

One more vital step is to evaluate the practicability of the CDT model and its two ensembles. Therefore, based on the validation data (30% of all the data), ROC analysis

was also utilized to estimate the prediction rate. The results are shown in Figure 10b and Table 6. The RSCDT model outperforms with the biggest value of AUC (0.861), followed by AdaCDT (0.802), and CDT (0.771). Consequently, CDT and its two ensemble models all perform well in landslide susceptibility prediction. The distinct discrepancies in the ROC curves of the three models are demonstrated in Table 6. As can be seen, visible differences lie between the CDT model and its ensemble models. In terms of AUC, the RSCDT algorithm improved the prediction accuracy of the CDT model by 10.45%, and the AdaCDT model increased the prediction of the CDT model by 3.87%. Therefore, for the validation data, the RSCDT model performed better than the AdaCDT model. In addition, for the training data, the RSCDT model yielded the most accurate results compared with the other two models.

**Table 6.** ROC analysis of CDT, AdaCDT, and RSCDT models using validation data.

Models	AUC	Standard Error	95% Confidence Interval
CDT	0.771	0.0467	0.680–0.863
AdaCDT	0.802	0.0426	0.719–0.886
RSCDT	0.861	0.0375	0.788–0.935

In previous studies, ensemble models have been widely applied to landslide susceptibility, showing their advantages in enhancing a single classifier's prediction capability and decreasing a single classifier's overfitting problem [80]. Consequently, accounting for the outstanding performance with the validation data, RSCDT is selected as the optimal model for LSMs and provides a more accurate prediction of future landslides in the study area.

## 5. Discussion

Assessing landslide susceptibility of a certain area is a complex task and remains challenging. The key to the assessment relies on the applied models and selected conditioning factors. In various publications, many researchers have tried to improve the performance of models for landslide susceptibility prediction, while the quality and prediction accuracy of these models are determined by the used methods. New machine-learning techniques have been proven to be capable and efficient in boosting prediction performance [48]. Therefore, the authors tried to combine the CDT model with two machine-learning ensemble frameworks (AdaBoost and random subspace) to create an LSM. According to current publications, this kind of investigation has not been performed in Zhashui County; hence, it was selected as a case study.

AdaBoost and random subspace algorithms boost the performance and prediction accuracy of the CDT model in the research area, indicating that ensemble techniques generate higher accuracy than an individually used model. According to the five classes of the LSMs constructed by the proposed three models, the CDT and RSCDT models have the largest area percentage of very low susceptible class, while the largest area percentage of the AdaCDT model is in the very high susceptible class (49.33%). The area percentage of 49.33% in the AdaCDT model indicates that nearly half of the study area is evaluated to be highly landslide prone, which is inconsistent with the distribution of historical landslides in the study area. The abnormal area percentage with a very high susceptible class in AdaCDT ensemble may be related to the way AdaBoost processes. The AdaBoost method dynamically changes the sample weight distribution to make the classifier more focused on the misclassified samples, and these samples are often at the classification boundary of the base classifier, which may cause overfitting and lead to unsatisfactory prediction results. However, the rotation forest ensemble approach can effectively increase the difference between the base classifiers by randomly splitting the sample attribute dataset and adopting the feature transformation strategy to obtain a good integration effect. In summary, for the base classifier in this study, the random subspace method behaves much better than the AdaBoost approach. Therefore, RSCDT is considered to be the optimal model for the

creation of LSMs in Zhashui County thanks to its outstanding performance in both training and validating data. Generally, the reclassified very low and low susceptible areas are distributed in the north and west of Zhashui County, while the very high and high prone areas are concentrated in the southeast. This is related to the fact that people live along the rivers, and the erosion of the rivers and the frequent human activities lead to landslide occurrences. In addition, the concentrated and heavy rainfall in Qinling Mountains will also contribute to landslide occurrence.

Additionally, the CDT, AdaCDT, and RSCDT models assign the lowest area percentage to the low, moderate, and very high susceptible classes, respectively. These distinct discrepancies are related to the fact that AdaBoost and random subspace changed the LSM originally generated by the CDT model used alone, although in different ways. Actually, whether a simple or ensemble model, it processes in a way unique to the approach itself. Hence, every approach generally contributes to a different data processing and modeling result.

Overall, the performance of the two ensembles shows a good consistency between training data and validating data in terms of ROC analysis when compared to the single classifier. AdaBoost and random subspace approaches can enhance the prediction rate of the CDT model, and the RSCDT model has the most outstanding prediction ability. In summary, with the highest success rate and prediction accuracy, the RSCDT model is believed to be the optimal model for mapping landslide susceptibility in this study. It can be noted that the very high susceptible class holds the lowest area percentage in the RSCDT model, which means that the proportion of the area most sensitive to landslide hazards is the smallest. With the help of this LSM, it would be conducive to the development of landslide prevention and control and can prevent the formation of landslides in a targeted manner. However, it will still be a challenging task to figure out the best model for creating LSMs in a certain region from the various models. More case studies need to be carried out to seek the best possibility and to assess the overall performance of AdaBoost and random subspace ensemble techniques.

## 6. Conclusions

Obtaining an accurate landslide susceptibility map is vital for a certain region's sustainable land-use management and planning. Machine-learning ensemble frameworks and ensemble techniques have been widely used for mapping landslide susceptibility. In this study, the CDT model and its two ensembles (AdaCDT and RSCDT) were utilized to map landslide susceptibility in Zhashui County (China). ROC curves were utilized to estimate and compare the performance of each model.

Results of the ROC analysis indicate that the applied landslide models combined with machine-learning ensemble frameworks, namely the AdaCDT and RSCDT models, achieved noteworthy and good results. The RSCDT model has the highest prediction ability, followed by the AdaCDT model. It is also relevant to be noted that machine-learning ensemble frameworks dramatically enhanced the performance of the base model of CDT.

Consequently, it can be concluded that machine-learning ensemble frameworks are effective and promising techniques for the landslide susceptibility prediction of areas suffering from natural hazards, such as landslides. Additionally, the proposed approaches in this study can also be applied to areas with different geo-environmental conditions.

**Author Contributions:** Conceptualization, J.G. and W.C.; methodology, I.P.-R.; software, M.Y.; validation, M.Y., J.G. and W.C.; formal analysis, I.P.-R.; investigation, M.Y.; resources, W.C.; data curation, M.Y.; writing—original draft preparation, J.G.; writing—review and editing, I.P.-R.; visualization, J.G.; supervision, F.Z.; project administration, F.Z.; funding acquisition, F.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (Grant No. 41977228), and Key Research Program of Shaanxi (Program No. 2022SF-335).

**Data Availability Statement:** Not applicable.

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

## References

1. Wang, G.; Lei, X.; Chen, W.; Shahabi, H.; Shirzadi, A. Hybrid Computational Intelligence Methods for Landslide Susceptibility Mapping. *Symmetry* **2020**, *12*, 325. [\[CrossRef\]](#)
2. Kamran, C.; Ramesh, P.R.; Syed, I.A.; Alamgir, A.K.; Bahram, G.; William, T.D.; Pradeep, K.G. Spatial-temporal dynamics of runoff generation areas in a small agricultural watershed in southern Ontario. *J. Water Resour. Prot.* **2015**, *7*, 27. [\[CrossRef\]](#)
3. Bacha, A.S.; Shafique, M.; van der Werff, H. Landslide inventory and susceptibility modelling using geospatial tools, in Hunza-Nagar valley, northern Pakistan. *J. Mt. Sci.* **2018**, *15*, 1354–1370. [\[CrossRef\]](#)
4. Hong, H.; Ilia, I.; Tsangaratos, P.; Chen, W.; Xu, C. A hybrid fuzzy weight of evidence method in landslide susceptibility analysis on the Wuyuan area, China. *Geomorphology* **2017**, *290*, 1–16. [\[CrossRef\]](#)
5. Li, L.; Lan, H.; Guo, C.; Zhang, Y.; Li, Q.; Wu, Y. A modified frequency ratio method for landslide susceptibility assessment. *Landslides* **2017**, *14*, 727–741. [\[CrossRef\]](#)
6. Ramesh, V.; Anbazhagan, S. Landslide susceptibility mapping along Kolli hills Ghat road section (India) using frequency ratio, relative effect and fuzzy logic models. *Environ. Earth Sci.* **2015**, *73*, 8009–8021. [\[CrossRef\]](#)
7. Akinci, H.; Zeybek, M. Comparing classical statistic and machine learning models in landslide susceptibility mapping in Ardanuc (Artvin), Turkey. *Nat. Hazards* **2021**, *108*, 1515–1543. [\[CrossRef\]](#)
8. Lian Zhipeng, X.Y.; Sheng, F. Landslide susceptibility assessment based on multi-model fusion method: A case study in Wufeng County, Hubei Province. *Bull. Geol. Sci. Technol.* **2020**, *39*, 178–186. [\[CrossRef\]](#)
9. Bourenane, H.; Bouhadad, Y.; Guettouche, M.S.; Braham, M. GIS-based landslide susceptibility zonation using bivariate statistical and expert approaches in the city of Constantine (Northeast Algeria). *Bull. Eng. Geol. Environ.* **2015**, *74*, 337–355. [\[CrossRef\]](#)
10. Daniel, M.T.; Ng, T.F.; Kadir, M.F.A.; Pereira, J.J. Landslide Susceptibility Modeling Using a Hybrid Bivariate Statistical and Expert Consultation Approach in Canada Hill, Sarawak, Malaysia. *Front. Earth Sci.* **2021**, *9*, 616225. [\[CrossRef\]](#)
11. Benchelha, S.; Aoudjehane, H.C.; Hakdaoui, M.; El Hamdouni, R.; Mansouri, H.; Benchelha, T.; Layelmam, M.; Alaoui, M. Landslide susceptibility mapping in the commune of Oudka, Taounate Province, North Morocco: A comparative analysis of logistic regression, multivariate adaptive regression spline, and artificial neural network models. *Environ. Eng. Geosci.* **2020**, *26*, 185–200. [\[CrossRef\]](#)
12. 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. [\[CrossRef\]](#)
13. Ashournejad, Q.; Hosseini, A.; Pradhan, B.; Hosseini, S.J. Hazard zoning for spatial planning using GIS-based landslide susceptibility assessment: A new hybrid integrated data-driven and knowledge-based model. *Arab. J. Geosci.* **2019**, *12*, 126. [\[CrossRef\]](#)
14. Sheikh, V.; Kornejady, A.; Ownegh, M. Application of the coupled TOPSIS–Mahalanobis distance for multi-hazard-based management of the target districts of the Golestan Province, Iran. *Nat. Hazards* **2019**, *96*, 1335–1365. [\[CrossRef\]](#)
15. Quan, H.-C.; Lee, B.-G. GIS-based landslide susceptibility mapping using analytic hierarchy process and artificial neural network in Jeju (Korea). *KSCE J. Civ. Eng.* **2012**, *16*, 1258–1266. [\[CrossRef\]](#)
16. Wu, Y.; Li, W.; Liu, P.; Bai, H.; Wang, Q.; He, J.; Liu, Y.; Sun, S. Application of analytic hierarchy process model for landslide susceptibility mapping in the Gangu County, Gansu Province, China. *Environ. Earth Sci.* **2016**, *75*, 422. [\[CrossRef\]](#)
17. Pham, B.T.; Pradhan, B.; Bui, D.T.; Prakash, I.; Dholakia, M.B. A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). *Environ. Model. Softw.* **2016**, *84*, 240–250. [\[CrossRef\]](#)
18. Wang, G.; Chen, X.; Chen, W. Spatial prediction of landslide susceptibility based on GIS and discriminant functions. *ISPRS Int. J. Geo Inf.* **2020**, *9*, 144. [\[CrossRef\]](#)
19. Arabameri, A.; Saha, S.; Roy, J.; Chen, W.; Blaschke, T.; Bui, D.T. Landslide Susceptibility Evaluation and Management Using Different Machine Learning Methods in The Gallicash River Watershed, Iran. *Remote. Sens.* **2020**, *12*, 475. [\[CrossRef\]](#)
20. Abedini, M.; Ghasemian, B.; Shirzadi, A.; Shahabi, H.; Chapi, K.; Pham, B.T.; Bin Ahmad, B.; Tien Bui, D. A novel hybrid approach of Bayesian Logistic Regression and its ensembles for landslide susceptibility assessment. *Geocarto Int.* **2019**, *34*, 1427–1457. [\[CrossRef\]](#)

21. Nhu, V.-H.; Zandi, D.; Shahabi, H.; Chapi, K.; Shirzadi, A.; Al-Ansari, N.; Singh, S.K.; Dou, J.; Nguyen, H. Comparison of support vector machine, Bayesian logistic regression, and alternating decision tree algorithms for shallow landslide susceptibility mapping along a mountainous road in the west of Iran. *Appl. Sci.* **2020**, *10*, 5047. [[CrossRef](#)]
22. Huang Faming, L.J.; Junyu, W.; Daxiong, M.; Mingqiang, S. Modelling rules of landslide susceptibility prediction considering the suitability of linear environmental factors and different machine learning models. *Bull. Geol. Sci. Technol.* **2022**, *41*, 44–59. [[CrossRef](#)]
23. Bui, D.T.; Shahabi, H.; Omidvar, E.; Shirzadi, A.; Geertsema, M.; Clague, J.J.; Khosravi, K.; Pradhan, B.; Pham, B.T.; Chapi, K.; et al. Shallow landslide prediction using a novel hybrid functional machine learning algorithm. *Remote Sens.* **2019**, *11*, 931. [[CrossRef](#)]
24. Zheng Yingkai, C.J.; Chengbin, W. Application of certainty factor and random forests model in landslide susceptibility evaluation in Mangshi City, Yunnan Province. *Bull. Geol. Sci. Technol.* **2022**, *39*, 131–144. [[CrossRef](#)]
25. Pham, B.T.; Prakash, I.; Singh, S.K.; Shirzadi, A.; Shahabi, H.; Tran, T.-T.-T.; Bui, D.T. Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. *Catena* **2019**, *175*, 203–218. [[CrossRef](#)]
26. Pham, B.T.; Prakash, I.; Khosravi, K.; Chapi, K.; Trinh, P.T.; Ngo, T.Q.; Hosseini, S.V.; Bui, D.T. A comparison of Support Vector Machines and Bayesian algorithms for landslide susceptibility modelling. *Geocarto Int.* **2019**, *34*, 1385–1407. [[CrossRef](#)]
27. Althuwaynee, O.F.; Pradhan, B.; Ahmad, N. Landslide susceptibility mapping using decision-tree based CHi-squared automatic interaction detection (CHAID) and Logistic regression (LR) integration. *IOP Conf. Series Earth Environ. Sci.* **2014**, *20*, 12032. [[CrossRef](#)]
28. Huang, C.; Li, F.; Wei, L.; Hu, X.; Yang, Y. Landslide susceptibility modeling using a deep random neural network. *Appl. Sci.* **2022**, *12*, 12887. [[CrossRef](#)]
29. Lian, C.; Zeng, Z.; Yao, W.; Tang, H. Multiple neural networks switched prediction for landslide displacement. *Eng. Geol.* **2015**, *186*, 91–99. [[CrossRef](#)]
30. Akinci, H. Assessment of rainfall-induced landslide susceptibility in Artvin, Turkey using machine learning techniques. *J. Afr. Earth Sci.* **2022**, *191*, 104535. [[CrossRef](#)]
31. Yang Can, L.L.; Yili, Z.; Wenqing, Z.; Shaohe, Z. Machine learning based on landslide susceptibility assessment with Bayesian optimized the hyperparameters. *Bull. Geol. Sci. Technol.* **2022**, *41*, 228–238. (In Chinese) [[CrossRef](#)]
32. Cai, Z.; Xu, W.; Meng, Y.; Shi, C.; Wang, R. Prediction of landslide displacement based on GA-LSSVM with multiple factors. *Bull. Eng. Geol. Environ.* **2016**, *75*, 637–646. [[CrossRef](#)]
33. Cao, Y.; Yin, K.; Zhou, C.; Ahmed, B. Establishment of landslide groundwater level prediction model based on GA-SVM and influencing factor analysis. *Sensors* **2020**, *20*, 845. [[CrossRef](#)] [[PubMed](#)]
34. Tutsoy, O. COVID-19 Epidemic and Opening of the Schools: Artificial Intelligence-Based Long-Term Adaptive Policy Making to Control the Pandemic Diseases. *IEEE Access* **2021**, *9*, 68461–68471. [[CrossRef](#)]
35. Tutsoy, O.; Tanrikulu, M.Y. Priority and age specific vaccination algorithm for the pandemic diseases: A comprehensive parametric prediction model. *BMC Med. Inform. Decis. Mak.* **2022**, *22*, 4. [[CrossRef](#)] [[PubMed](#)]
36. Pham, B.T.; Bui, D.T.; Prakash, I. Bagging based Support Vector Machines for spatial prediction of landslides. *Environ. Earth Sci.* **2018**, *77*, 146. [[CrossRef](#)]
37. Bui, D.T.; Shirzadi, A.; Shahabi, H.; Geertsema, M.; Omidvar, E.; Clague, J.J.; Thai Pham, B.; Dou, J.; Asl, D.T.; Bin Ahmad, B.; et al. New Ensemble Models for Shallow Landslide Susceptibility Modeling in a Semi-Arid Watershed. *Forests* **2019**, *10*, 743. [[CrossRef](#)]
38. Zhao, X.; Chen, W. GIS-Based Evaluation of Landslide Susceptibility Models Using Certainty Factors and Functional Trees-Based Ensemble Techniques. *Appl. Sci.* **2020**, *10*, 16. [[CrossRef](#)]
39. Wu, Y.; Ke, Y.; Chen, Z.; Liang, S.; Zhao, H.; Hong, H. Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. *Catena* **2020**, *187*, 104396. [[CrossRef](#)]
40. Huang Faming, H.S.; Xueya, Y.; Ming, L.; Junyu, W.; Wenbin, L.; Zizheng, G.; Wenyan, F. Landslide susceptibility prediction and identification of its main environmental factors based on machine learning models. *Bull. Geol. Sci. Technol.* **2022**, *41*, 79–90. [[CrossRef](#)]
41. Halil, A.; Mustafa, Z.; Sedat, D. Evaluation of Landslide Susceptibility of Şavşat District of Artvin Province (Turkey) Using Machine Learning Techniques. In *Landslides*; Yuanzhi, Z., Qiuming, C., Eds.; IntechOpen: Rijeka, Croatia, 2021; pp. 69–95.
42. Azarafza, M.; Azarafza, M.; Akgün, H.; Atkinson, P.M.; Derakhshani, R. Deep learning-based landslide susceptibility mapping. *Sci. Rep.* **2021**, *11*, 24112. [[CrossRef](#)] [[PubMed](#)]
43. Sameen, M.I.; Pradhan, B.; Lee, S. Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment. *Catena* **2020**, *186*, 104249. [[CrossRef](#)]
44. Habumugisha, J.M.; Chen, N.; Rahman, M.; Islam, M.; Ahmad, H.; Elbeltagi, A.; Sharma, G.; Liza, S.N.; Dewan, A. Landslide Susceptibility Mapping with Deep Learning Algorithms. *Sustainability* **2022**, *14*, 1734. [[CrossRef](#)]
45. Wang, Y.; Fang, Z.; Wang, M.; Peng, L.; Hong, H. Comparative study of landslide susceptibility mapping with different recurrent neural networks. *Comput. Geosci.* **2020**, *138*, 104445. [[CrossRef](#)]
46. Nhu, V.-H.; Janizadeh, S.; Avand, M.; Chen, W.; Farzin, M.; Omidvar, E.; Shirzadi, A.; Shahabi, H.; Clague, J.J.; Jaafari, A.; et al. GIS-Based Gully Erosion Susceptibility Mapping: A Comparison of Computational Ensemble Data Mining Models. *Appl. Sci.* **2020**, *10*, 2039. [[CrossRef](#)]

47. Bui, D.T.; 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*, 1101. [[CrossRef](#)]
48. Arabameri, A.; Chen, W.; Lombardo, L.; Blaschke, T.; Bui, D.T. Hybrid Computational Intelligence Models for Improvement Gully Erosion Assessment. *Remote. Sens.* **2020**, *12*, 140. [[CrossRef](#)]
49. Pham, B.T.; Prakash, I.; Bui, D.T. Spatial prediction of landslides using a hybrid machine learning approach based on Random Subspace and Classification and Regression Trees. *Geomorphology* **2018**, *303*, 256–270. [[CrossRef](#)]
50. Akinci, H.; Kilicoglu, C.; Dogan, S. Random Forest-Based Landslide Susceptibility Mapping in Coastal Regions of Artvin, Turkey. *ISPRS Int. J. Geo Inf.* **2020**, *9*, 553. [[CrossRef](#)]
51. 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](#)]
52. Riaz, M.T.; Basharat, M.; Hameed, N.; Shafique, M.; Luo, J. A Data-Driven Approach to Landslide-Susceptibility Mapping in Mountainous Terrain: Case Study from the Northwest Himalayas, Pakistan. *Nat. Hazards Rev.* **2018**, *19*, 05018007. [[CrossRef](#)]
53. Katz, O.; Morgan, J.K.; Aharonov, E.; Dugan, B. Controls on the size and geometry of landslides: Insights from discrete element numerical simulations. *Geomorphology* **2014**, *220*, 104–113. [[CrossRef](#)]
54. Chen Qian, Y.E.; Shaoping, H.; Xi, W. Susceptibility evaluation of geological disasters in southern Huanggang based on samples and factor optimization. *Bull. Geol. Sci. Technol.* **2020**, *39*, 175–185. [[CrossRef](#)]
55. Ercanoglu, M.; Gokceoglu, C. Assessment of landslide susceptibility for a landslide-prone area (north of Yenice, NW Turkey) by fuzzy approach. *Environ. Geol.* **2002**, *41*, 720–730. [[CrossRef](#)]
56. Ding, Q.; Chen, W.; Hong, H. Application of frequency ratio, weights of evidence and evidential belief function models in landslide susceptibility mapping. *Geocarto Int.* **2017**, *32*, 619–639. [[CrossRef](#)]
57. Moore, I.D.; Gessler, P.; Nielsen, G.; Peterson, G. Terrain analysis for soil specific crop management. In Proceedings of the Proceedings of Soil Specific Crop Management: A Workshop on Research and Development Issues; Robert, P.C., Rust, R.H., Larson, W.E., Eds.; American Society of Agronomy: Madison, WI, USA, 1993; pp. 27–55.
58. De Roo, A.P.J. Modelling runoff and sediment transport in catchments using GIS. *Hydrol. Process.* **1998**, *12*, 905–922. [[CrossRef](#)]
59. Beven, K.J.; Kirkby, M.J. A physically based, variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.* **1979**, *24*, 43–69. [[CrossRef](#)]
60. Moore, I.D.; Turner, A.K.; Wilson, J.P.; Jenson, S.K.; Band, L.E. GIS and land-surface-subsurface process modeling. *Environ. Model. GIS* **1993**, *20*, 196–230.
61. Mejía-Navarro, M.; Wohl, E.E. Geological hazard and risk evaluation using GIS: Methodology and model applied to Medellin, Colombia. *Environ. Eng. Geosci.* **1994**, *31*, 459–481. [[CrossRef](#)]
62. 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](#)]
63. Fan Yajie, F.X.; Fang, C. County comprehensive geohazard modelling based on the grid maximum method. *Bull. Geol. Sci. Technol.* **2022**, *41*, 197–208. [[CrossRef](#)]
64. Pachauri, A.K.; Gupta, P.V.; Chander, R. Landslide zoning in a part of the Garhwal Himalayas. *Environ. Geol.* **1998**, *36*, 325–334. [[CrossRef](#)]
65. Gökceoglu, C.; Aksoy, H. Landslide susceptibility mapping of the slopes in the residual soils of the Mengen region (Turkey) by deterministic stability analyses and image processing techniques. *Eng. Geol.* **1996**, *44*, 147–161. [[CrossRef](#)]
66. Nam, K.; Wang, F. An extreme rainfall-induced landslide susceptibility assessment using autoencoder combined with random forest in Shimane Prefecture, Japan. *Geoenvirom. Disasters* **2020**, *7*, 6. [[CrossRef](#)]
67. 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](#)]
68. Zhu, H.; Zhang, L.; Xiao, T.; Li, X. Enhancement of slope stability by vegetation considering uncertainties in root distribution. *Comput. Geotech.* **2017**, *85*, 84–89. [[CrossRef](#)]
69. Kim, J.H.; Fourcaud, T.; Jourdan, C.; Maeght, J.-L.; Mao, Z.; Metayer, J.; Meylan, L.; Pierret, A.; Rapidel, B.; Rouspard, O.; et al. Vegetation as a driver of temporal variations in slope stability: The impact of hydrological processes. *Geophys. Res. Lett.* **2017**, *44*, 4897–4907. [[CrossRef](#)]
70. Turrini, M.C.; Visintainer, P. Proposal of a method to define areas of landslide hazard and application to an area of the Dolomites, Italy. *Eng. Geol.* **1998**, *50*, 255–265. [[CrossRef](#)]
71. Abellán, J.; Moral, S. Building classification trees using the total uncertainty criterion. *Int. J. Intell. Syst.* **2003**, *18*, 1215–1225. [[CrossRef](#)]
72. Dempster, A.P. Upper and Lower Probabilities Induced by a Multivalued Mapping. In *Classic Works of the Dempster-Shafer Theory of Belief Functions*; Yager, R.R., Liu, L., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 57–72.
73. Shafer, G. *A Mathematical Theory of Evidence*; Princeton University Press: Princeton, NJ, USA, 1976.
74. Walley, P. Inferences from Multinomial Data: Learning about a Bag of Marbles. *J. R. Stat. Soc. Ser. B Methodolog.* **1996**, *58*, 3–57. [[CrossRef](#)]
75. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [[CrossRef](#)]

76. Freund, Y.; Schapire, R.E. Experiments with a New Boosting Algorithm. In Proceedings of the International Conference on Machine Learning, Bari, Italy, 3–6 July 1996; pp. 148–156.
77. Ho, T.K. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 832–844. [[CrossRef](#)]
78. Peng, T.; Chen, Y.; Chen, W. Landslide Susceptibility Modeling Using Remote Sensing Data and Random SubSpace-Based Functional Tree Classifier. *Remote Sens.* **2022**, *14*, 4803. [[CrossRef](#)]
79. Zhao, C.H.; Zhang, B.L.; Zhang, X.Z.; Zhao, S.Q.; Li, H.X. Recognition of driving postures by combined features and random subspace ensemble of multilayer perceptron classifiers. *Neural Comput. Applic.* **2013**, *22*, 175–184. [[CrossRef](#)]
80. 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](#)]

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