

## Article

# Landslide Susceptibility Mapping Using DIvisive ANAlysis (DIANA) and RObust Clustering Using linKs (ROCK) Algorithms, and Comparison of Their Performance

Deborah Simon Mwakapesa <sup>1</sup>, Yimin Mao <sup>2,3</sup>, Xiaoji Lan <sup>1,\*</sup> and Yaser Ahangari Nanehkaran <sup>4</sup>

<sup>1</sup> School of Civil, and Surveying, & Mapping, Jiangxi University of Science and Technology, Ganzhou 341000, China

<sup>2</sup> School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou 341000, China

<sup>3</sup> School of Information Engineering, Shaoguan University, Shaoguan 512005, China

<sup>4</sup> School of Information Engineering, Yancheng Teachers University, Yancheng 224002, China

\* Correspondence: debrah45@126.com

**Abstract:** Landslide susceptibility mapping (LSM) studies provide essential information that helps various authorities in managing landslide-susceptible areas. This study aimed at applying and comparing the performance of DIvisive ANAlysis (DIANA) and RObust Clustering using linKs (ROCK) algorithms for LSM in the Baota District, China. These methods can be applied when the data has no labels and when there is insufficient inventory data. First, based on historical records, survey reports, and previous studies, 293 landslides were mapped in the study area and 7 landslide-influencing attributes were selected for modeling. Second, the methods were clustered in the study area mapping units into 469 and 476 subsets, respectively; for mapping landslide susceptibility, the subsets were classified into 5 susceptibility levels through the K-means method using landslide densities and attribute values. Then, their performances were assessed and compared using statistical metrics and the receiver operating curve (ROC). The outcomes indicated that similarity measures influenced the accuracy and the predictive power of these clustering models. In particular, when using a link-based similarity measure, the ROCK performed better with overall performance accuracy of 0.8933 and an area under the curve (AUC) of 0.875. The maps constructed from the models can be useful in landslide assessment, prevention, and mitigation strategies in the study area, especially for areas classified with higher susceptibility levels. Moreover, this comparison provides a new perspective in the selection of a considerable model for LSM in the Baota District.



**Citation:** Mwakapesa, D.S.; Mao, Y.; Lan, X.; Nanehkaran, Y.A. Landslide Susceptibility Mapping Using DIvisive ANAlysis (DIANA) and RObust Clustering Using linKs (ROCK) Algorithms, and Comparison of Their Performance. *Sustainability* **2023**, *15*, 4218. <https://doi.org/10.3390/su15054218>

Academic Editors: Jiankun Huang, Yunqi Wang, Liqun Lyu and Jun Li

Received: 28 January 2023

Revised: 17 February 2023

Accepted: 23 February 2023

Published: 26 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

Landslides are among the vast, frequent, and common natural calamities in China [1–5]. One-third of landslide events occurring in China every year occur in the Loess Plateau (which covers about 400,000 square km of land in China), which has a complex geo-environmental nature [6–9]. These landslides always result in the loss of lives and massive damages, as well as the destruction of roads and railways [10–12] (examples are shown in Figure 1). The risk (expected losses or damages) of landslide events will elevate as a result of global climate changes and increasing urbanization caused by the growing population. For this reason, it is crucial to find urgent assessment strategies to prevent and manage landslides and mitigate their consequences.

Landslide susceptibility mapping (LSM) is essential and a common procedure for landslide assessment. Landslide susceptibility refers to the probability of a landslide occurring in an area based on local environmental conditions [13,14]. Typically, LSM is used to predict and map where future landslides may occur [15,16]. Generally, it involves these

stages: A collection of spatial data and selection of landslide-related attributes (factors); application of various methods to develop the LSM model, which processes and analyzes the influence of landslide-related attributes on the spatial distribution of landslides; the construction of a landslide susceptibility map; model evaluation. The maps developed in the process can provide different authorities with insights for landslide-susceptible areas for land assessment and management decisions [17].



**Figure 1.** Landslide events in China: (A,B) Zaolin village, in Shanxi Province-2018; (C) Xiangning county, Shanxi province- 2019; and (D) Huaihua city in Hunan province-2019.

Several methods have been applied for LSM modeling, including direct mapping, deterministic methods, heuristic methods, probabilistic methods, and machine learning methods (MLM) [18–21]. Among them, MLM has received increased popularity over the past decade. MLM can generally be categorized as supervised learning and unsupervised learning [22,23]. Supervised learning takes a known set of input datasets (the training dataset) and known class labels of the dataset (the output) and forms a model to predict the class labels of the new inputs. Supervised learning methods are widely applied in LSM modeling in different regions around the world [15,24–28]. Some of the popular supervised learning methods include Decision Tree (DT) [15,29], Logistic Regression (LR) [30,31], Support Vector Machine (SMV) [25,32], Random Forest (RF) [24,33], Artificial Neural Networks (ANN) [34,35], and Naïve Bayes (NB) [26,36]. Despite their popularity, their application becomes limited when there is insufficient data or when the dataset has no labels. To make use of the available and unlabeled dataset, unsupervised learning can be applied [22].

Clustering is a type of unsupervised learning that groups a dataset into subsets called clusters without any training, which makes overall susceptibility mapping possible even when the available dataset is insufficient or when the dataset is large and difficult or expensive to label [37–40]. The clustering for susceptibility mapping is based on the assumption that mapping units with the same susceptibility have similar values of landslide-influencing attributes. Generally, clustering can be categorized into partition-based methods, hierarchical methods, density-based methods, and model-based clustering methods. Based on these categories, various clustering methods have been proposed for conducting LSM modeling [41–47]. However, from the literature review, it can be noted that these methods are rare in the field of LSM compared to supervised learning-based methods; at present,

there is no agreement on the most suitable method for LSM [48–50]. Thus, on this basis and in light of the advantages of clustering methods, there is still a need for more applications and comparative studies among different clustering methods to assess the effectiveness of the models and improve landslide susceptibility assessments.

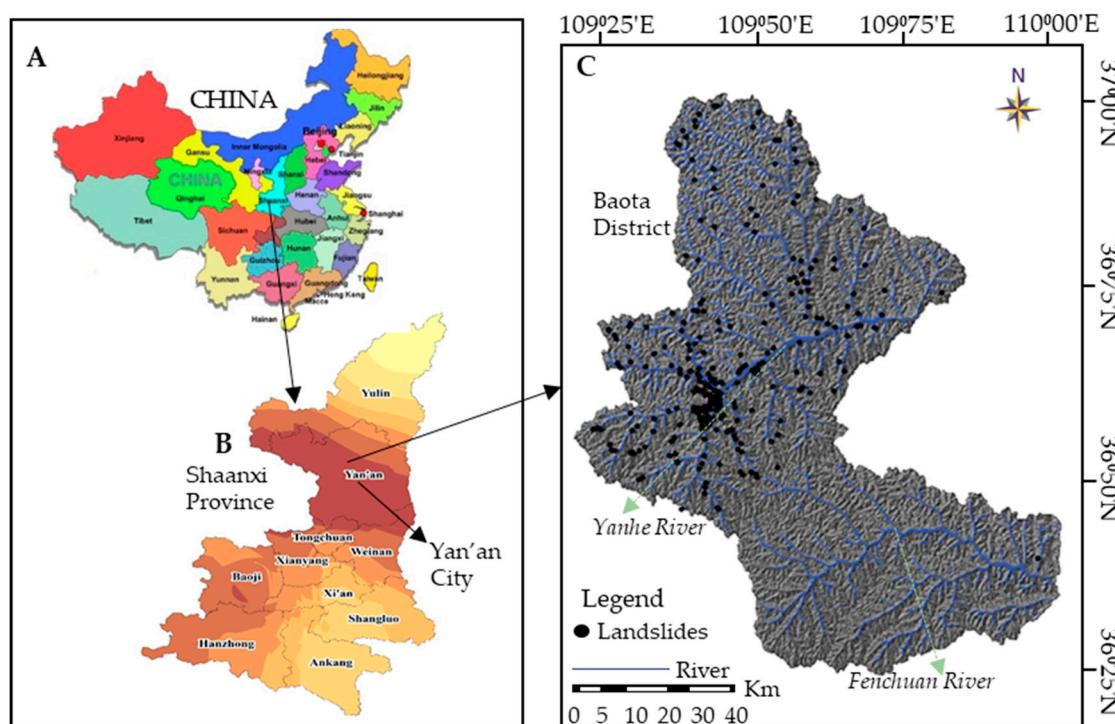
For this reason, this study presented a comprehensive evaluation of the performance of two hierarchical clustering methods, namely the Divisive ANAlysis (DIANA) and Robust Clustering using linKs (ROCK) [51–54] methods for LSM in the Baota District, Shaanxi Province, China. These methods have been frequently applied in other fields, including networking, bioinformatics, GPS data, road network construction, and gene expression [55–58]. The study explored the potential prediction abilities of these clustering algorithms and provided some valuable insight for future research and assessments in LSM, as well as for scientific literature as a whole. The performance evaluation was carried out based on some statistical metrics, namely: sensitivity, specificity, accuracy, kappa, and the receiver operating curve (ROC). Additionally, to show the advantageous capability of clustering algorithms over supervised learning algorithms, the Decision Tree and Random Forest classification methods were employed for comparison based on performance accuracy and ROC.

Moreover, the susceptibility maps generated in this study can recognize and allocate landslide-susceptible places so that land engineers and concerned authorities can decide on favorable places (such as mining areas, new economic zones, and urban areas) for ongoing and forthcoming development arrangements.

## 2. Overview of the Study Area

The Baota District (Figure 2) is the selected study area. It is found in Yan'an City of Shaanxi Province, which is located in the middle of the Loess Plateau in China. Geographically, it occupies an area of approximately 3556 km<sup>2</sup> (0.55% of the Loess Plateau coverage) between longitudes 109°14' E–110°07' E and longitudes and latitudes 36°11' N–37°02' N. Its elevation above sea level ranges between 800 and 1800 m. This area is also recognized as a typical valley area and a fragile environmental area for severe soil erosion and landslides [59]. There are two rivers in the northern and southern parts of the area, the Yanhe River and Fenchuan River, respectively. These features describe the topography of the area. Geomorphological characteristics of the area include gorges and heaved slopes. Additionally, there are deposited rocks and extensive quaternary loess deposits that cover the largest part of the area, explaining the geology of the area [60]. The annual average rainfall is 550 mm, and heavy rainfalls are recorded between June and October, ranging between 58 and 117 mm, which have triggered most of the landslides in the area [44,45,61].

The Baota District serves as Yan'an's administrative center, where various city government offices are located. It is a significant site for various economic activities such as tourism. However, because of the nature of the environment, prolonged heavy rainfalls, and the growing population leading to the expansion of habitation and the development of social and economic activities in the area, various disasters including landslides tend to occur frequently in the area. In response to this, the assessment and management of these disasters are considerably important [59,62]. Thus, we developed this LSM study, particularly for this area, as a helpful tool in assessing and managing landslides and their impacts.



**Figure 2.** The location of the study area and landslide distribution; (A) the location of Shaanxi Province in China; (B) the location of Baota District in Yan'an City, Shaanxi Province; (C) the landslide inventory of Baota District.

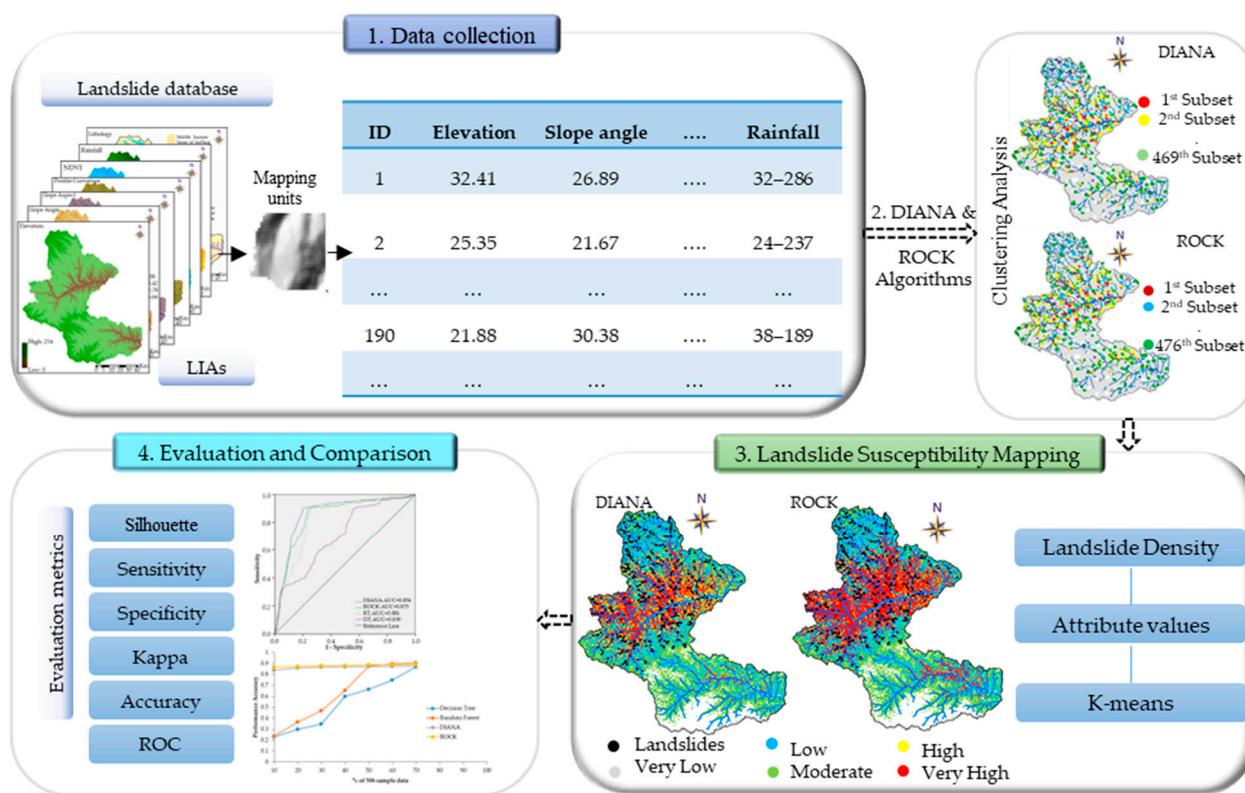
### 3. Methodology

The landslide susceptibility mapping is determined by the characteristics of the area, the availability of the data, and the methods used to produce the susceptibility map. It is essential to develop a proper methodological structure that aids in determining susceptibility mapping. For this purpose, this study adopted the following steps: (1) data collection: preparation of landslide database and selection of the influencing attributes based on an extensive literature survey and local geo-environmental conditions, as well as the description of their relationship to landslide occurrence; (2) description of the research methods and their implementation in LSM; (3) a susceptibility map that was constructed based on the obtained results; (4) the evaluation and comparison of the methods' performance based on internal and external evaluation metrics. The study methodology structure is presented in Figure 3.

#### 3.1. Data Collection

##### 3.1.1. Landslides Database

In LSM studies, the landslide geospatial database is very essential for providing details on past landslides, as well as the relationship between landslides and landslide-influencing attributes [63]. The database used in this study was prepared based on data provided by Xi'an Center for Geological Survey (XCGS). This data was collected through interpretation of the post-disaster aerial photographs and field investigations from 1081 locations in the area, from which 293 historical landslides were recorded (as presented in Figure 2C). The landslides were of two types: rotational and translational, with most of them being rotational landslides. In terms of size, the volume of the sliding mass was classified as: less than  $10^1 \times 10^4 \text{ m}^3$  (small scale, 30.7% of the recorded landslides); between  $10^1 \times 10^4$  and  $10^2 \times 10^4 \text{ m}^3$ , 52.6% of the landslides; between  $10^2 \times 10^4$  and  $10^3 \times 10^4 \text{ m}^3$  (large scale, 16.7% of the recorded landslides) [61]. These 293 landslide sites, along with 213 randomly selected non-landslide sites (making 506 samples in total), were applied to evaluate the LSM models in this study. In addition, the database comprised a set of landslide-influencing attributes.



**Figure 3.** Research methodology structure.

### 3.1.2. Landslide Influencing Attributes (LIAs)

Selecting LIAs for the development and evaluation of models is a basic phase in LSM [64]. Landslide occurrence can be influenced by various attributes. By referring to previous studies [44–46,61] and the availability of data, 7 LIAs were selected: elevation, slope angle, slope aspect, profile curvature, rainfall, lithology, and Normalized Difference Vegetation Index (NDVI). Landslides occur as a result of slope instability [65,66]. Elevation and slope angle play an important role in influencing landslide occurrences because slope instability increases with an increase in elevation and slope angle [67,68]. Landslide events in the Baota District happened in slopes with an elevation range of 20 to 30 m and an angle higher than 60°. The slope aspect describes slope orientation (direction); this regulates the slope's exposure to sunlight, hydrological processes, and wind direction, which have an impact on the physical characteristics of rock and soil masses and ultimately on slope stability [69,70]. In the study area, landslides are more likely to occur in the northeast direction (between 0°–45° and 315°–360°) [47,61]. Profile curvature describes the slope curving in the downward direction, and its values represent the topography structure. The positive and negative values are indicated as upwardly concave and upwardly convex, respectively, while zero represents a flat surface. In addition, the more strongly negative or strongly positive the value, the higher the possibility of a landslide occurring [71,72]. Lithology is an important attribute that describes the rock and soil structure and properties in an area [73]. NDVI was selected to represent and quantify the vegetation cover of the area. Vegetation cover helps to resist slope movements, reduces soil, and improves slope stability. Thus, barren or sparsely vegetated areas are frequently exposed to landslides [74]. In the study area, most of the landslide events were recorded during and after rainfall periods, indicating that rainfall attribute is directly associated with landslide occurrence [43,45,61]. In addition, several researchers have verified that intense and continuous rainfall affects the slope stability, which consequently results in landslide occurrence [15,29,75–79]. Based on their significant relationship and impact on slope stability, the selected attributes are of great importance in landslide susceptibility assessment, not only in this study area but also

in other areas around the world [19,20,80–82]. The category, data types, scale/resolutions, class, and sources of the influencing attribute data are shown in Table 1.

**Table 1.** Information on the Landslide Influencing Attributes used in this study.

Category	Attribute Name	Data Type	Scale	Class	Data Source
Topography	Elevation (m)	Continuous		0~254 0–6.54, 6.54–13.08, 13.08–18.42, 18.42–22.78, 22.78–26.66, 26.66–30.54, 30.54–34.41, 34.41–39.02, 39.02–61.80	Xi'an Center of Geological Survey
	Slope angle (°)	Continuous		Flat, North (N), North-East (NE), North-West (NW), East (E), West (W), South-East (SE), South (S), and South-West (SW)	
	Slope aspect	Discrete		(E), West (W), South-East (SE), South (S), and South-West (SW)	
	Profile curvature	Discrete	1:50,000	<−0.05, −0.05 to 0.05, >0.05	
Geology	Lithology	Discrete		I: loess + nearly horizontal paleo-soil, II: loess + inclined paleo-soil, III: loess + paleo-soil layers + bedrock, IV: loess + paleo-soil layers + the Neogene clay	Xi'an Center of Geological Survey
Underlying surface	NDVI	Continuous		−0.54~0.99	Xi'an Center of Geological Survey
Triggering attribute	Rainfall (mm)	Uncertain		0~200	Baota Weather Bureau

Using ArcGIS 10.2 software, the study area was divided into 5,672,922 mapping units (grids units) with 25 m × 25 m spacing. Each mapping unit was described by the 7 LIAs values. The Digital Elevation Model of the study area was acquired from XCGS, from which the thematic maps for topography and geology attributes were extracted at a scale of 1: 50,000. The NDVI thematic map was constructed using Enhanced Thematic Mapper Plus (ETM+) remote sensing image processing software. Rainfall data was collected based on the meteorological rainfall graphs from Baota Weather Bureau [14]. The thematic maps are presented in Figure 4.

### 3.2. Research Methods

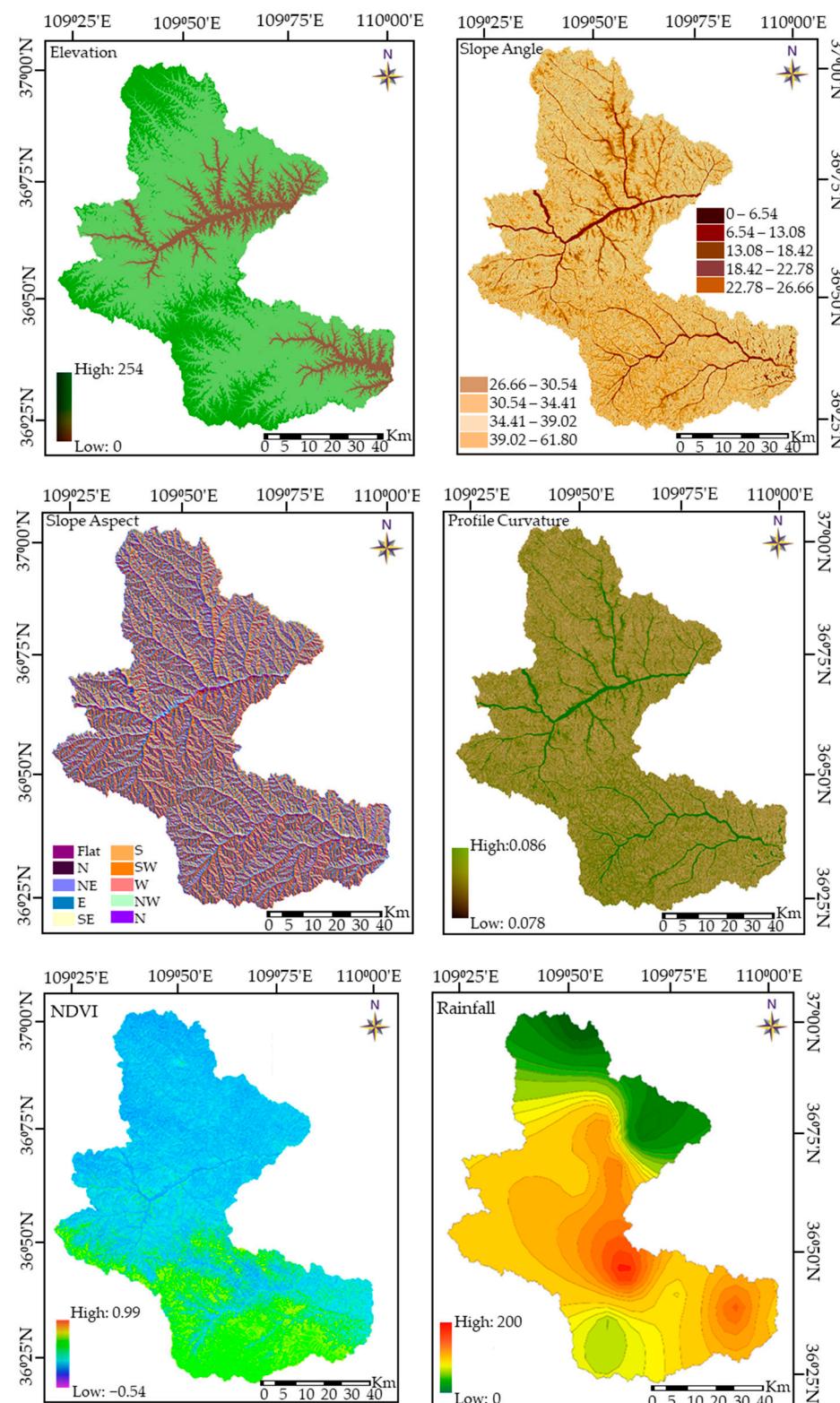
#### 3.2.1. DIANA Algorithm

DIANA is a hierarchical clustering algorithm that constructs a hierarchy of clusters in the dataset [51,52,83,84]. It works in a top-down manner, meaning that it begins with all points in a single cluster and then splits the cluster into smaller and least similar subclusters. To compute the similarity among the points to be clustered, the algorithm employs the basic Euclidean distance function.

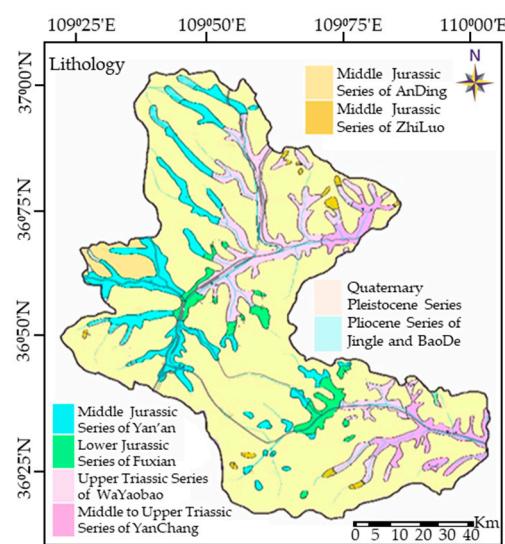
Let  $p = (p_1, p_2)$  and  $q = (q_1, q_2)$  be n-dimensional data points, the Euclidean distance  $d_{Euc}(p, q)$  between the points is given by:

$$d_{Euc}(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} \quad (1)$$

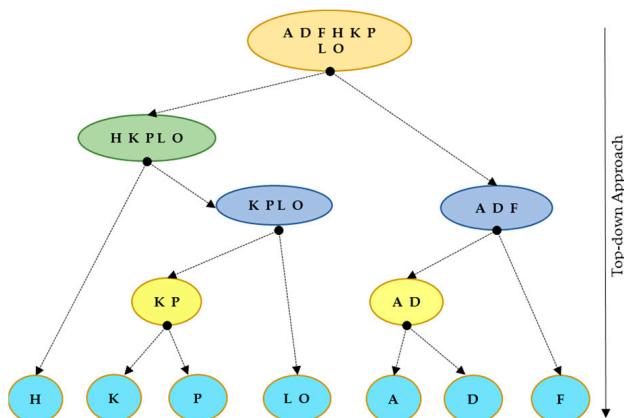
The clusters are split based on the maximum  $d_{Euc}$  (distance between the closest neighboring points in the cluster). The DIANA algorithm is explained in the following steps and illustrated in Figure 5.



**Figure 4.** Cont.



**Figure 4.** Thematic maps for Landslide influencing attributes.



**Figure 5.** Illustration of DIANA algorithm.

- 1: Load into the algorithm the dataset containing  $n$  data points
- 2: Compute  $d_{Euc}$  between the points and find the maximum  $d_{Euc}$
- 3: Form a distance matrix using the  $d_{Euc}$  values acquired from the previous step
- 4: Split the cluster based on the distance matrix into the least similar subclusters
- 5: From new subclusters, calculate new  $d_{Euc}$  and update the distance matrix
- 6: Repeat the process until the points in a cluster are comparatively similar

To its advantage, the DIANA algorithm does not require a pre-defined number of clusters and can detect arbitrarily shaped clusters. However, its main limitation is the use of distance similarity measures, which expose the algorithm to errors and misleading results. This limitation can be avoided using the links approach in the ROCK algorithm.

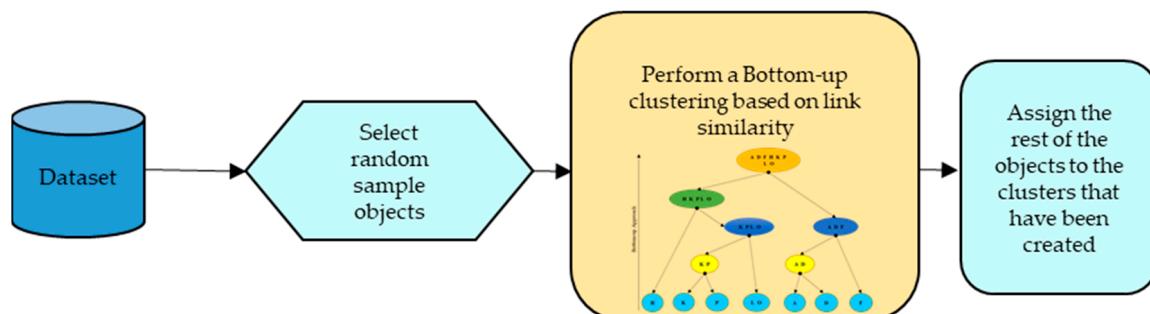
### 3.2.2. ROCK Algorithm

ROCK is a hierarchical clustering algorithm that constructs a hierarchy of clusters in the dataset through a bottom-up approach [51,85–87]. Initially, it treats every object as a solo cluster and then merges the more similar clusters to form a new cluster. To compute similarity, ROCK uses the links approach instead of distance similarity measures. For a pair of objects, a link is described by the number of common neighbors of the objects. Objects that belong to the same cluster will generally contain a large number of common neighbors, and thus more links, and in case two links do not have any neighbors in common, then their link similarity will be equal to 0. In addition, the objects with no or with few neighbors are

regarded as noise and are eliminated. In that manner, merging clusters with the maximum number of links first will lead to the construction of better and more significant clusters.

The implementation of the ROCK algorithm is described in the following steps and illustrated in Figure 6:

1. Load the dataset containing  $n$  objects
2. Select a random sample of objects from the dataset
3. Compute the link value for each pair of objects—that is, the number of shared neighbors between objects
4. Perform a bottom-up hierarchical clustering on the data based on the link's similarity measure
5. Compute and employ a goodness measure (Equation (2)) to identify the pair of objects to be merged at each step.



**Figure 6.** Illustration of ROCK Algorithm.

$$G(l_i, l_j) = \frac{\text{link}(l_i, l_j)}{(n_i + n_j)^{1+2f(\alpha)} - n_i^{1+2f(\alpha)} - n_j^{1+2f(\alpha)}} \quad (2)$$

where:  $\text{link}(l_i, l_j)$  is the number of links between two clusters (say  $c_i$  and  $c_j$ );  $n_i$  and  $n_j$  stand for the number of objects in the two clusters, respectively. Each object in the links has approximately  $n^{f(\alpha)}$  neighbors in the clusters.

6. Repeat the same procedures and assign the rest of the objects to the clusters that have been created.

Note:

- The pair of clusters with maximum  $G(l_i, l_j)$  is considered to be the best pair to be merged.

Unlike many other clustering algorithms, ROCK does not require the user to specify the number of clusters in advance, and can find clusters with varying shapes and sizes. In addition, the link approach supports the successful elimination of noise, and thus can identify better clusters.

### 3.2.3. Implementation of DIANA and ROCK Clustering Methods in LSM

The implementation of these methods in LSM is based on the assumption that mapping units with the same susceptibility have similar values of influencing attributes [88]. In this study, the values of the 7 LIAs were represented by a band of mapping units. The values of all LIAs at a given unit form a vector  $(A_1, A_2, \dots, A_7)$ , which was treated as a point/object (as used in the DIANA and ROCK methods, respectively) in the 7-dimension property space. Then, the mapping units correspond to a point/object set in the space. Therefore, as the mapping unit data were prepared, each unit was normalized by its maximum value so that the value of every attribute is between 0 and 1. Then, the data were used as inputs into the DIANA and ROCK clustering methods. The methods find the units with similar characteristic values and categorize them into various clusters/subsets. All of these

operations, including conversion, calculation, and clustering analysis, were facilitated by ArcGIS 10.2 under its effective spatial analyst toolbox and Python programming language.

### 3.3. Methods for Landslide Susceptibility Classification

Based on the existing LSM literature in the study area, landslide susceptibility is usually classified into 5 levels: very high susceptibility level (VHSL), high susceptibility level (HSL), moderate susceptibility level (MSL), low susceptibility level (LSL), and very low susceptibility level (VSL) [43,44,46,47,60,89,90]. However, the DIANA and ROCK methods categorized the mapping units (points/objects as used in the methods) to their respective subclasses, but did not identify the susceptibility levels in the subsets. Thus, to classify the obtained subsets into the 5 susceptibility levels, a k-means clustering algorithm and landslide density were applied in this study.

#### 3.3.1. K-Means Algorithm

K-means is a simple clustering algorithm that finds k-groups of points from the unlabeled dataset [41,91,92]. Given a dataset containing several points, the algorithm is described below:

1. Define the value of “k”.
2. Randomly select k-initial centroids.
3. Assign each point to its nearest centroid to form a group.
4. Calculate and update the centroid (mean) of each group.
5. Repeat steps 3–4 until no point changes the group.

To implement this algorithm, the subsets obtained from the DIANA and ROCK clustering methods are treated as points in the K-means algorithm, and the value of “k” is set to 5, denoting 5 susceptibility levels. However, the results still present the statistical information about the 5 groups (susceptibility levels) but do not clarify their meanings (as which group belongs to VHSL, HSL, MSL, LSL, or VSL).

#### 3.3.2. Landslide Density

Landslide density ( $LDen$ ) was applied to clarify the meaning of the obtained susceptibility levels. It was calculated in ArcGIS 10.2 platform by computing the number of landslides (from the landslide database) per  $1 \text{ km}^2$  of the mapping unit in each subset. The general principle “the higher the landslide density, the higher the susceptibility level” was applied to clarify the susceptibility levels [45,88,93]. Moreover, for  $LDen = 0$ , the susceptibility levels were identified by examining the values of LIAs.

### 3.4. Performance Evaluation and Comparison Methods

#### 3.4.1. Performance Evaluation

Performance evaluation is the process of applying various evaluation metrics to understand how well or how badly the method has performed on the given data [67,94]. Evaluating the performance of a clustering method is still a problematic and controversial issue because there is no universal criterion for evaluation. However, several criteria have been developed in the literature [95–97]. These criteria are generally categorized into two: (i) internal criteria, which usually measures the compactness of the clusters using some similarity measure, and do not use any external information other than the data itself; (ii) external criteria, which is useful for examining whether the clustering results match some external information about the data (inventory data) [44,98,99].

To evaluate the performance of the models based on internal criteria, the Silhouette value is selected as the clusters’ evaluation criterion [100,101]. Silhouette value is calculated by the closeness of the data samples and it is expressed using the following equation:

Silhouette value for a point  $x$  within a cluster  $C_i$  is given by:

$$\text{Sil}(x) = \frac{v(x) - u(x)}{\max\{v(x), u(x)\}} \quad (3)$$

where  $u(x)$  represents the average distance between sample  $x$  and other samples in the cluster ( $x \in C_i$ ), and  $v(x) = \max \{dis(x, C_i)\}$  [44]. The value of the Silhouette value ranges between  $-1$  and  $1$ . The closer the value is to  $1$ , the better the clustering effect [44,102,103].

However, to assess performance in case of comparison, the external criterion is preferred. Here, the general concept that landslide susceptibility modeling is a binary problem is applied [70]. In addition, the prepared dataset containing 506 samples (293 landslide samples and 213 non-landslide samples) is used for performance evaluation. To this end, the confusion matrix is used, and each sample is given a prediction of either positive (landslide) or negative (non-landslide) [44,104,105]. From the matrix, 4 results are produced: true positive ( $tp$ ), false negative ( $fn$ ), false positives ( $fp$ ), and true negatives ( $tn$ ). Based on these results, sensitivity ( $st$ ), specificity ( $sp$ ), accuracy, and kappa were selected as the model evaluation metrics (Equations (4)–(7), respectively). These metrics are common and widely used in LSM studies [20,28,73,74,106].

$$st = tp / (tp + fn) \quad (4)$$

$$sp = tn / (fp + tn) \quad (5)$$

$$\text{Accuracy} = (tp + tn) / (tp + tn + fp + fn) \quad (6)$$

$$\kappa = (P_a - P_{\text{exp}}) / (1 - P_{\text{exp}}) \quad (7)$$

whereby:  $P_a = (tp + tn) / (tp + tn + fp + fn)$  and  $P_{\text{exp}} = (((tp + fn)(tp + fp) + (tn + fp)) / (tn + fn)) / (\sqrt{tp + tn + fp + fn})$ .

In addition, the ROC (receiver operating characteristic) curve—a quantitative research method—was used to evaluate the models. It is a graph plotted by sensitivity (on the y-axis) that indicates the proportion of correctly predicted landslide samples against 1-specificity (x-axis), which indicates the proportion of incorrectly predicted non-landslide samples [107]. The areas under the ROC curves (AUCs) were applied to judge the prediction performance. The AUC values range between  $0.5$  to  $1$ , and a higher value indicates greater performance [67,68,70].

### 3.4.2. Comparison Methods

To further assess their effectiveness, ROCK and DIANA methods were compared with popular supervised learning methods: DT and RF based on performance accuracy and ROC.

DT is applied in classification tasks to classify labeled data and make predictions [29,108,109]. It aims to construct a model that predicts the class of a target sample by learning some simple decision rules drawn from data structures. DT is based on a tree structure that is composed of a root node, a set of internal nodes, and a set of terminal nodes (leaves). Each node makes a binary decision separating one or more classes from the remaining classes; it is executed by moving down the tree until the terminal node is obtained. The DT model is easy to construct and interpret, which makes it easy for decision makers to use, and thus is commonly adopted to assess landslide predictions [29,109,110].

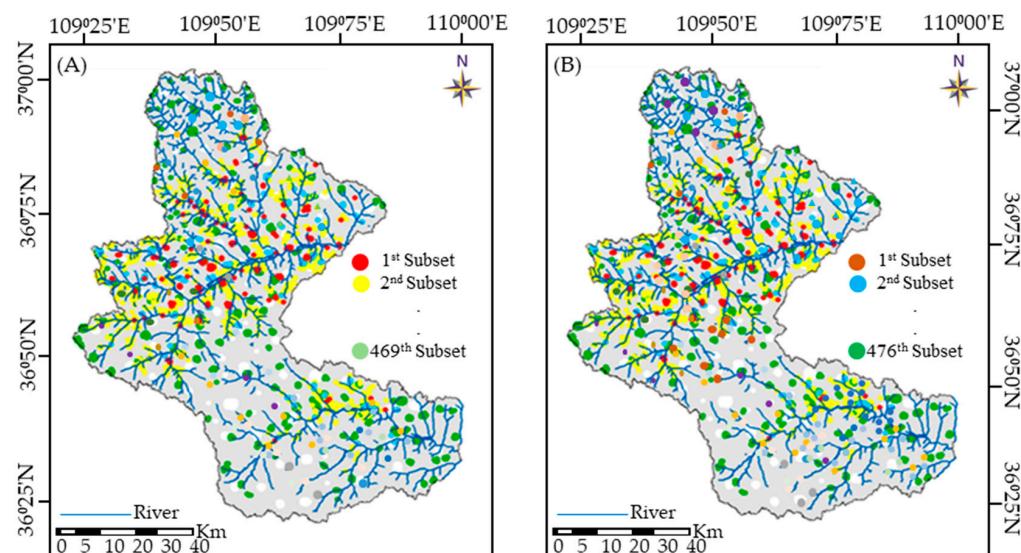
Random Forest is a supervised learning method that consists of multiple decision trees [24,33]. While classifying a sample, the final result is obtained through a voting mechanism of many decision trees. This means that each tree provides a predicted result, and the result with the most votes (selected by most trees) is taken as the optimal output of the RF method. RF can classify a large amount of higher-dimensional data and has a high tolerance for noise. Thus, it is one of the most commonly used classification methods with high prediction accuracy and is very popular in LSM studies [111,112].

To implement the methods in LSM modeling, the dataset containing 506 samples was divided into training and verification sets: 30% of the samples for training and 70% for verification. The procedure was conducted repeatedly by adding 10% of the data from the verification to the training set until the training set had 70% of the total dataset.

## 4. Results

### 4.1. Clustering Analysis

Clustering for mapping susceptibility is based on the assumption that mapping units with the same susceptibility features (geology and geomorphology features) have similar values of LIAs. Based on the aforementioned procedures, 469 and 476 distinct and arbitrary subsets were obtained from the DIANA and ROCK clustering models, respectively. Figure 7 portrays the distribution of those subsets in the study area, whereby different subsets are indicated by different colors in the figure. The ability of the models to identify distinguishing features of the mapping units and cluster them into their respective subsets without prior knowledge of the mapping units belonging indicates that the models have good and effective performance capability [46,113].



**Figure 7.** Distribution of subsets obtained from (A) DIANA and (B) ROCK algorithms.

### 4.2. Landslide Susceptibility Mapping

In this study, to map landslide susceptibility in the obtained subsets, the k-means clustering algorithm, *LDen*, and attribute values were applied to classify the subsets into 5 susceptibility levels. The *LDen* values obtained in both models ranged between 0 and 1.8. Based on the two models, the attribute values, their corresponding *LDen* values, and susceptibility levels for some subsets are shown in Table 2. The distribution of the susceptibility levels obtained in both models is shown in Figure 8. As for the DIANA model, the very high level occupied the largest proportion, at 29% of the study area, while the very low level occupied the smallest proportion, at 11%. High and low levels accounted for 14% and 15%, respectively, and moderate levels accounted for 28%. Compared with the DIANA model, the ROCK model classified the very high and high levels as 33% and 16% of the study area, respectively, which were all more than those of the DIANA model. The moderate, low, and very low levels accounted for 26%, 15%, and 10%, respectively.

### 4.3. Evaluation and Comparison Results

The performance of the two models was assessed and compared using the Silhouette as an internal evaluation criterion, and statistical metrics and ROC for external evaluation criteria. ROCK and DIANA models obtained Silhouette values of 0.8677 and 0.8543, respectively. For the statistical metrics, Table 3 was prepared to show the performance and comparative results of the models. In grouping the landslide samples, the DIANA and ROCK models obtained st values of 0.8805 and 0.8874, respectively, similarly for the prediction of non-landslide samples; they also obtained sp values of 0.8732 and 0.8732, respectively. Moreover, the two models obtained kappa of 0.7518 and 0.7828, respectively, and accuracy of 0.8775 and 0.8933, respectively. In addition, for the case of ROC (Figure 9),

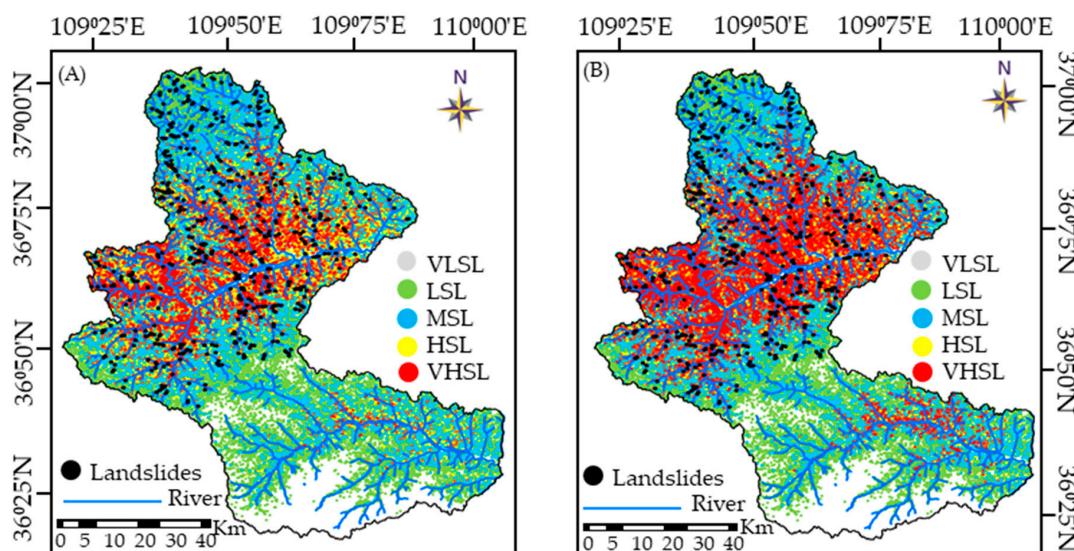
the DIANA and ROCK landslide susceptibility models showed AUC values of 0.854 and 0.875, respectively.

**Table 2.** Description of some subsets obtained based on (A) DIANA (B) ROCK models.

Subset Id.	(A)										Susceptibility Level	
	Attribute Values (Before Normalization)						Landslide Density					
	Elevation	Slope Angle	Profile Curvature	Slope Aspect	Lithology	NDVI	Rainfall	Area (km <sup>2</sup> )	Landslides	LD (/km <sup>2</sup> )		
1	32.41	26.89	0.026	S	II	0.67	32–286	9.54	8	0.84	HSL	
2	25.35	21.67	0.041	SE	IV	0.56	24–237	8.92	0	0	Based on expertise	
...	...	...	...	...	...	...	...	...	...	...	...	
190	21.88	30.38	0.61	S	III	0.69	38–189	12.34	9	0.73	MSL	
...	...	...	...	...	...	...	...	...	...	...	...	

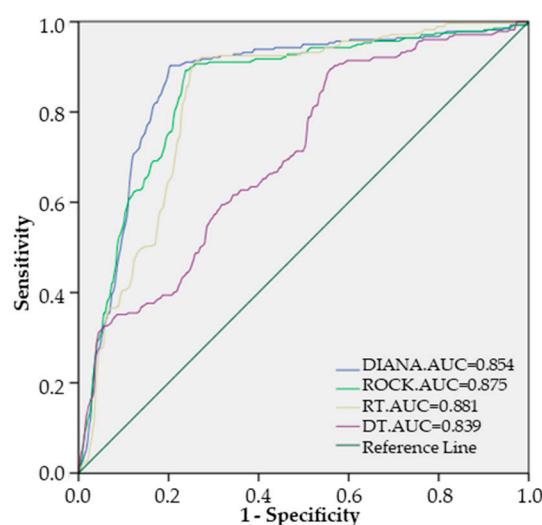
Subset Id.	(B)										Susceptibility Level	
	Attribute Values (Before Normalization)						Landslide Density					
	Elevation	Slope Angle	Profile Curvature	Slope Aspect	Lithology	NDVI	Rainfall	Area (km <sup>2</sup> )	Landslides	LD (/km <sup>2</sup> )		
1	29.89	24.82	0.032	S	II	0.77	30–283	9.53	7	0.73	MSL	
2	21.99	19.19	0.043	N	III	0.64	26–232	6.67	5	0.74	MSL	
...	...	...	...	...	...	...	...	...	...	...	...	
235	14.89	39.43	0.61	NE	II	0.71	20–150	15.32	0	0	Based on expertise	
...	...	...	...	...	...	...	...	...	...	...	...	



**Figure 8.** Maps showing landslide susceptibility levels produced by (A) DIANA and (B) ROCK models.

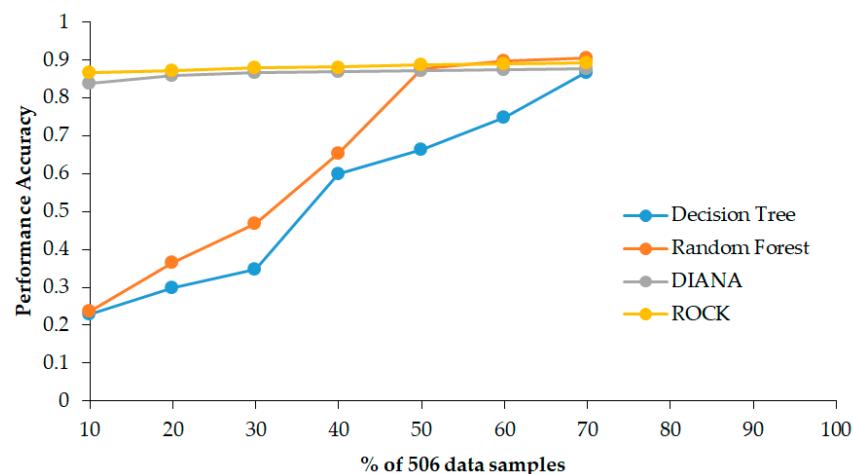
**Table 3.** Model assessments and comparative results.

Models	DIANA	ROCK
tp	258	280
tn	186	192
fp	27	21
fn	35	33
St	0.8805	0.8874
sp	0.8732	0.9014
Kappa	0.7518	0.7828
Silhouette	0.8543	0.8677
Accuracy	0.8775	0.8933



**Figure 9.** ROC results for DIANA, ROCK models, DT and RF.

Moreover, the performances of DIANA and ROCK models were compared with RF and DT-supervised learning models. Under ROC, the AUC of RF was a little higher (0.879) than ROCK (0.875), followed by DIANA (0.854), while DT showed the lowest AUC (0.839) (Figure 9). In addition, as shown in Figure 10, the performance of RF and DT kept on increasing with the increase in data samples. For instance, with 30–70% of the sample data, the accuracy values of DT were 0.348, 0.599, 0.664, 0.747, and 0.868, respectively; for RF: 0.468, 0.654, 0.878, 0.898, and 0.906, respectively. With 30–70% of the sample data, ROCK and DIANA showed steady accuracy values of (0.867, 0.869, 0.872, 0.875, and 0.877) and (0.878, 0.881, 0.887, 0.890, and 0.893), respectively.



**Figure 10.** Comparison analysis of performance accuracy among the DIANA and ROCK clustering algorithms against the Random Forest and Decision Tree classification algorithms.

## 5. Discussion

Landslides are complicated and disastrous natural disasters that require immediate and appropriate assessment measures. Conducting landslide susceptibility mapping is a primary step in assessing landslides, by which the susceptible areas and non-susceptible areas can be located and assessment measures can be employed. Therefore, this study applied and compared two clustering-unsupervised learning methods, namely DIANA and ROCK, in mapping landslide susceptibility in the Baota District of Yan'an city in Shaanxi, China. Silhouette internal evaluation metrics, as well as sensitivity, specificity, accuracy, and kappa and ROC external evaluation metrics were employed to evaluate and compare

their performances. Moreover, to evaluate the efficiency and effectiveness of these methods, DT and RF-supervised methods were applied for comparison.

From the analysis, both DIANA and ROCK models performed very well because their Silhouette values were more than 0.85. This implies that they both have good clustering capability; thus, they can be effective in assessing landslide susceptibility. However, the results indicate that the ROCK model has comparatively better performance than the DIANA model in terms of sensitivity, specificity, accuracy, and kappa metrics. In addition, the overall performance based on the area under the ROC (AUC) was higher for the ROCK model than the DIANA model. The outstanding performance of ROCK was mainly supported by the application of the link similarity measure, which introduces a global approach to the clustering procedure and facilitates the capturing of global information (knowledge) of neighboring mapping units into the relationship between individual pairs of mapping units. Applying this approach in deciding which mapping units to be merged will lead to the construction of better and more significant subsets. In addition, through this approach, the algorithm can detect significant subsets and eliminate noise. These features add value to the ROCK model, thus providing better and more robust performance. The DIANA model is based on the distance similarity measure, which does not reflect the property of the mapping units' neighborhood. It is a local approach to clustering tasks and is vulnerable to errors because two different subsets may have a few mapping units (which are then considered noise) that could be very close, and distance similarity could merge the two subsets, causing the situation to worsen as clustering continues. This problem is successfully handled using the link approach. With these limitations, the DIANA model's performance is less than the ROCK model.

Moreover, in comparison with supervised learning models, Random Forest showed the highest accuracy and AUC, followed by ROCK and DIANA. However, like other supervised learning methods, to obtain higher performance, the models demand a large amount of training data and with less data, performance will always be low. This is a limitation because it is a fact that some study areas may not always have sufficient data to make the algorithms perform highly, and the process of getting information from landslide databases and sites is very exhaustive and expensive. In addition, the supervised learning models are sensitive to changes in the training dataset, meaning that slight changes in the training dataset can cause big variations in the process. These limitations suggest that these supervised learning models cannot be reliable in landslide susceptibility mapping. Fortunately, these limitations can be successively avoided using the ROCK and DIANA models unsupervised learning models. Therefore, from this analysis and discussion, it can be concluded that the ROCK and DIANA unsupervised learning models are more advantageous and more resourceful than the supervised learning models. They can be successfully used in mapping landslide susceptibility and other real-world problems.

It is generally known that the predicted landslide samples should appear in very high or high susceptibility areas as much as possible, while non-landslide samples should appear in a safe area with low or very low susceptibility. Moreover, the constructed landslide susceptibility maps indicate areas that are very high and highly susceptible to landslide occurrences. The map constructed based on the ROCK model showed that a large area (33% coverage) along the Yanhe River is very highly susceptible (indicated by red dots in the figures) to landslides, which is more than the prediction (29%) from the DIANA model. Amounts of 16% and 14% of the subsets from the ROCK and DIANA models, respectively, fell in high susceptible levels (indicated by yellow dots), mostly in the upper part of the study area. In addition, the models predicted that 10% and 11% of the subsets, respectively, were very low, while 15% and 18% of the subsets, respectively, were in the low susceptible level, mostly in the southern part of the area. Upon observing and comparing these results and the inventory map shown in Figure 2, it can be observed that: though the ROCK model had better predictions than the DIANA model, predictions from both models were consistent or in close agreement with the landslide database applied for model construction. This information can be very useful to concerned authorities, decision makers,

and residents regarding activities going on in the susceptible areas and to be aware of the landslide risk and its consequences; thus, appropriate measures can be taken. Meanwhile, areas predicted at lower susceptibility levels can continue to be protected by following environmental guidelines for safety assurance.

## 6. Conclusions

The main purpose of this research was to apply and compare the performance of DIANA and ROCK hierarchical clustering methods for mapping landslide susceptibility in the Baota District, one of the landslide-susceptible areas in China. The study also delivers an evaluation regarding the impact of similarity measures in these clustering methods, while pinpointing the most accurate and reliable clustering method. The models were evaluated and compared based on the Silhouette measure, specificity, sensitivity, accuracy, kappa, and ROC. Both models obtained Silhouette values very close to one, implying that they both have good clustering capability, and can be effective in assessing landslide susceptibility. However, the ROCK model performed better than the DIANA model, with accuracy = 0.8933, kappa = 0.7828, and AUC = 0.875. This is because ROCK uses the link similarity measure, which facilitated the construction of better and more significant clustering results, hence obtaining higher and more robust performance, whereas DIANA uses a distance similarity measure that is prone to errors, which then led to the development of low-quality and less significant clusters, hence lower performance. Moreover, it was noted that the similarity measures have a great impact on the clustering results, and it should be a point of concern while applying these models in LSM studies. In addition, the LSM maps constructed based on these models and the knowledge extracted from them may be significant for the determination and implementation of landslide mitigation schemes. However, the rainfall attribute is of an uncertain data type (because it is recorded in intervals), but, like other clustering methods that rely on distance similarity measurements, the DIANA method did not take into consideration the uncertainty of this attribute. Thus, in the future, this study can be improved by using appropriate methods to process the uncertain data. In addition, more internal criteria should be explored to further evaluate the methods' performance, and data from other study areas should be applied to validate the methods' robustness, reproducibility, reliability, and objectivity.

**Author Contributions:** Conceptualization: D.S.M., Y.M. and X.L.; Methodology: D.S.M. and Y.M.; Software: Y.M., D.S.M. and Y.A.N.; Writing—original draft preparation: D.S.M. and Y.M.; Writing—review and editing: D.S.M., Y.M. and X.L.; Visualization: D.S.M. and Y.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by the Key Promotion Project of Guangdong Province, China (2022ZDJS048) and National Natural Science Foundation of China (Grant No: 41562019, 42250410321).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are available upon request from the corresponding author.

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

## References

1. Huang, R. Mechanisms of Large-Scale Landslides in China. *Bull. Eng. Geol. Environ.* **2012**, *71*, 161–170. [[CrossRef](#)]
2. Hungr, O.; Leroueil, S.; Picarelli, L. The Varnes classification of landslide types, an update. *Landslides* **2014**, *11*, 167–194. [[CrossRef](#)]
3. Fang, K.; Tang, H.; Li, C.; Su, X.; An, P.; Sun, S. Centrifuge modelling of landslides and landslide hazard mitigation: A review. *Geosci. Front.* **2022**, *14*, 101493. [[CrossRef](#)]
4. Peng, J.; Wang, S.; Wang, Q.; Zhuang, J.; Huang, W.; Zhu, X.; Leng, Y.; Ma, P. Distribution and genetic types of loess landslides in China. *J. Asian Earth Sci.* **2019**, *170*, 329–350. [[CrossRef](#)]
5. Ma, S.; Qiu, H.; Zhu, Y.; Yang, D.; Tang, B.; Wang, D.; Wang, L.; Cao, M. Topographic Changes, Surface Deformation and Movement Process before, during and after a Rotational Landslide. *Remote Sens.* **2023**, *15*, 662. [[CrossRef](#)]

6. Feng, L.; Lin, H.; Zhang, M.; Guo, L.; Jin, Z.; Liu, X. Development and evolution of Loess vertical joints on the Chinese Loess Plateau at different spatiotemporal scales. *Eng. Geol.* **2020**, *265*, 105372. [[CrossRef](#)]
7. Guo, Z.; Tian, B.; Li, G.; Huang, D.; Zeng, T.; He, J.; Song, D. Using and comparing three data-driven techniques to generate effective regional landslide susceptibility maps in the Loess Plateau of Northwest China. *Front. Earth Sci.* **1979**. [[CrossRef](#)]
8. Li, Z.; Zhang, Y.; Zhu, Q.; Yang, S.; Li, H.; Ma, H. A gully erosion assessment model for the Chinese Loess Plateau based on changes in gully length and area. *Catena* **2017**, *148*, 195–203. [[CrossRef](#)]
9. Wang, L.; Shao, M.a.; Wang, Q.; Gale, W.J. Historical changes in the environment of the Chinese Loess Plateau. *Environ. Sci. Policy* **2006**, *9*, 675–684. [[CrossRef](#)]
10. Zhou, X.; Wen, H.; Zhang, Y.; Xu, J.; Zhang, W. Landslide susceptibility mapping using hybrid random forest with GeoDetector and RFE for factor optimization. *Geosci. Front.* **2021**, *12*, 101211. [[CrossRef](#)]
11. Dai, K.; Li, Z.; Xu, Q.; Bürgmann, R.; Milledge, D.G.; Tomas, R.; Fan, X.; Zhao, C.; Liu, X.; Peng, J. Entering the era of earth observation-based landslide warning systems: A novel and exciting framework. *IEEE Geosci. Remote Sens. Mag.* **2020**, *8*, 136–153. [[CrossRef](#)]
12. Pei, Y.; Qiu, H.; Yang, D.; Liu, Z.; Ma, S.; Li, J.; Cao, M.; Wufuer, W. Increasing landslide activity in the Taxkorgan River Basin (eastern Pamirs Plateau, China) driven by climate change. *Catena* **2023**, *223*, 106911. [[CrossRef](#)]
13. Wubalem, A. Landslide susceptibility mapping using statistical methods in Uatzau catchment area, northwestern Ethiopia. *Geoenviron. Disasters* **2021**, *8*, 1. [[CrossRef](#)]
14. 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](#)]
15. Rong, G.; Alu, S.; Li, K.; Su, Y.; Zhang, J.; Zhang, Y.; Li, T. Rainfall induced landslide susceptibility mapping based on bayesian optimized random forest and gradient boosting decision tree models—A case study of shuicheng county, China. *Water* **2020**, *12*, 3066. [[CrossRef](#)]
16. Vakhshoori, V.; Pourghasemi, H.R.; Zare, M.; Blaschke, T. Landslide susceptibility mapping using GIS-based data mining algorithms. *Water* **2019**, *11*, 2292. [[CrossRef](#)]
17. Yi, Y.; Zhang, Z.; Zhang, W.; Xu, C. Comparison of different machine learning models for landslide susceptibility mapping. In Proceedings of the IGARSS 2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 9318–9321.
18. 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](#)]
19. Broeckx, J.; Vanmaercke, M.; Duchateau, R.; Poesen, J. A data-based landslide susceptibility map of Africa. *Earth-Sci. Rev.* **2018**, *185*, 102–121. [[CrossRef](#)]
20. Dias, H.C.; Hölbling, D.; Grohmann, C.H. Landslide susceptibility mapping in Brazil: A review. *Geosciences* **2021**, *11*, 425. [[CrossRef](#)]
21. Pokharel, B.; Althuwaynee, O.F.; Aydda, A.; Kim, S.-W.; Lim, S.; Park, H.-J. Spatial clustering and modelling for landslide susceptibility mapping in the north of the Kathmandu Valley, Nepal. *Landslides* **2021**, *18*, 1403–1419. [[CrossRef](#)]
22. Su, C.; Wang, B.; Lv, Y.; Zhang, M.; Peng, D.; Bate, B.; Zhang, S. Improved landslide susceptibility mapping using unsupervised and supervised collaborative machine learning models. *Georisk Assess. Manag. Risk Eng. Syst. Geohazards* **2022**, *1*–19. [[CrossRef](#)]
23. Huang, F.; Cao, Z.; Jiang, S.-H.; Zhou, C.; Huang, J.; Guo, Z. Landslide susceptibility prediction based on a semi-supervised multiple-layer perceptron model. *Landslides* **2020**, *17*, 2919–2930. [[CrossRef](#)]
24. Tanyu, B.F.; Abbaspour, A.; Alimohammadalou, Y.; Tecuci, G. Landslide susceptibility analyses using Random Forest, C4. 5, and C5. 0 with balanced and unbalanced datasets. *Catena* **2021**, *203*, 105355. [[CrossRef](#)]
25. Su, Q.; Tao, W.; Mei, S.; Zhang, X.; Guo, J.; Yang, Y. Landslide Susceptibility Zoning Using C5. 0 Decision Tree, Random Forest, Support Vector Machine and Comparison of Their Performance in a Coal Mine Area. *Front. Earth Sci.* **2021**, *9*, 781472. [[CrossRef](#)]
26. Nhu, V.-H.; Shirzadi, A.; Shahabi, H.; Singh, S.K.; Al-Ansari, N.; Clague, J.J.; Jaafari, A.; Chen, W.; Miraki, S.; Dou, J. Shallow landslide susceptibility mapping: A comparison between logistic model tree, logistic regression, naïve bayes tree, artificial neural network, and support vector machine algorithms. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2749. [[CrossRef](#)] [[PubMed](#)]
27. Lima, P.; Steger, S.; Glade, T.; Murillo-García, F.G. Literature review and bibliometric analysis on data-driven assessment of landslide susceptibility. *J. Mt. Sci.* **2022**, *19*, 1670–1698. [[CrossRef](#)]
28. Xia, D.; Tang, H.; Sun, S.; Tang, C.; Zhang, B. Landslide susceptibility mapping based on the germinal center optimization algorithm and support vector classification. *Remote Sens.* **2022**, *14*, 2707. [[CrossRef](#)]
29. Dou, J.; Yunus, A.P.; Bui, D.T.; Merghadi, A.; Sahana, M.; Zhu, Z.; Chen, C.-W.; Khosravi, K.; Yang, Y.; Pham, B.T. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. *Sci. Total Environ.* **2019**, *662*, 332–346. [[CrossRef](#)]
30. Alqadhi, S.; Mallick, J.; Talukdar, S.; Bindajam, A.A.; Saha, T.K.; Ahmed, M.; Khan, R.A. Combining logistic regression-based hybrid optimized machine learning algorithms with sensitivity analysis to achieve robust landslide susceptibility mapping. *Geocarto Int.* **2021**, *37*, 9518–9543. [[CrossRef](#)]
31. Korma, T.C. GIS-based landslide susceptibility zonation mapping using frequency ratio and logistics regression models in the Dessie area, South Wello, Ethiopia. 2022. Available online: <https://assets.researchsquare.com/files/rs-1633474/v1/4a28b5b9-2bc7-4c3c-9aa8-4c90803720ec.pdf?c=1652373343> (accessed on 19 December 2022).

32. Saha, A.; Saha, S. Comparing the efficiency of weight of evidence, support vector machine and their ensemble approaches in landslide susceptibility modelling: A study on Kurseong region of Darjeeling Himalaya, India. *Remote Sens. Appl. Soc. Environ.* **2020**, *19*, 100323. [[CrossRef](#)]
33. Wang, Y.; Sun, D.; Wen, H.; Zhang, H.; Zhang, F. Comparison of random forest model and frequency ratio model for landslide susceptibility mapping (LSM) in Yunyang County (Chongqing, China). *Int. J. Environ. Res. Public Health* **2020**, *17*, 4206. [[CrossRef](#)] [[PubMed](#)]
34. Bragagnolo, L.; da Silva, R.; Grzybowski, J. Artificial neural network ensembles applied to the mapping of landslide susceptibility. *Catena* **2020**, *184*, 104240. [[CrossRef](#)]
35. Mehrabi, M.; Moayedi, H. Landslide susceptibility mapping using artificial neural network tuned by metaheuristic algorithms. *Environ. Earth Sci.* **2021**, *80*, 1–20. [[CrossRef](#)]
36. Lee, S.; Lee, M.-J.; Jung, H.-S.; Lee, S. Landslide susceptibility mapping using Naïve Bayes and Bayesian network models in Umyeonsan, Korea. *Geocarto Int.* **2020**, *35*, 1665–1679. [[CrossRef](#)]
37. Bindra, K.; Mishra, A. A detailed study of clustering algorithms. In Proceedings of the 2017 6th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 20–22 September 2017; pp. 371–376.
38. Qiu, W.; Wu, B.; Pan, X.; Tang, Y. Application of several cluster-optimization-based machine learning methods in evaluation of landslide susceptibility in Lingtai County. *Northwest. Geol.* **2020**, *53*, 222–233.
39. Guo, Z.; Shi, Y.; Huang, F.; Fan, X.; Huang, J. Landslide susceptibility zonation method based on C5. 0 decision tree and K-means cluster algorithms to improve the efficiency of risk management. *Geosci. Front.* **2021**, *12*, 101249. [[CrossRef](#)]
40. Sinaga, K.P.; Yang, M.-S. Unsupervised K-means clustering algorithm. *IEEE Access* **2020**, *8*, 80716–80727. [[CrossRef](#)]
41. Wang, Q.; Wang, Y.; Niu, R.; Peng, L. Integration of information theory, K-means cluster analysis and the logistic regression model for landslide susceptibility mapping in the Three Gorges Area, China. *Remote Sens.* **2017**, *9*, 938. [[CrossRef](#)]
42. Wan, S.; Yen, J.Y.; Lin, C.Y.; Chou, T.Y. Construction of knowledge-based spatial decision support system for landslide mapping using fuzzy clustering and KPSO analysis. *Arab. J. Geosci.* **2015**, *8*, 1041–1055. [[CrossRef](#)]
43. Hu, J.; Xu, K.; Wang, G.; Liu, Y.; Khan, M.A.; Mao, Y.; Zhang, M. A novel landslide susceptibility mapping portrayed by OA-HD and K-medoids clustering algorithms. *Bull. Eng. Geol. Environ.* **2021**, *80*, 765–779. [[CrossRef](#)]
44. Hu, J.; Zhu, H.; Mao, Y.; Zhang, C.; Liang, T.; Mao, D. Using Uncertain DM-Chameleon Clustering Algorithm Based on Machine Learning to Predict Landslide Hazards. *J. Robot. Mechatron.* **2019**, *31*, 329–338. [[CrossRef](#)]
45. Mao, Y.; Mwakapesa, D.S.; Wang, G.; Nanehkaran, Y.; Zhang, M. Landslide susceptibility modelling based on AHC-OLID clustering algorithm. *Adv. Space Res.* **2021**, *68*, 301–316. [[CrossRef](#)]
46. Mao, Y.; Mwakapesa, D.S.; Xu, K.; Lei, C.; Liu, Y.; Zhang, M. Comparison of wave-cluster and DBSCAN algorithms for landslide susceptibility assessment. *Environ. Earth Sci.* **2021**, *80*, 1–14. [[CrossRef](#)]
47. Mao, Y.-M.; Mwakapesa, D.S.; Li, Y.-C.; Xu, K.-B.; Nanehkaran, Y.A.; Zhang, M.-s. Assessment of landslide susceptibility using DBSCAN-AHD and LD-EV methods. *J. Mt. Sci.* **2022**, *19*, 184–197. [[CrossRef](#)]
48. Ada, M.; San, B.T. Comparison of machine-learning techniques for landslide susceptibility mapping using two-level random sampling (2LRS) in Alakir catchment area, Antalya, Turkey. *Nat. Hazards* **2018**, *90*, 237–263. [[CrossRef](#)]
49. Pourghasemi, H.R.; Rahmati, O. Prediction of the landslide susceptibility: Which algorithm, which precision? *Catena* **2018**, *162*, 177–192. [[CrossRef](#)]
50. Chen, W.; Hong, H.; Panahi, M.; Shahabi, H.; Wang, Y.; Shirzadi, A.; Pirasteh, S.; Alesheikh, A.A.; Khosravi, K.; Panahi, S. Spatial prediction of landslide susceptibility using gis-based data mining techniques of anfis with whale optimization algorithm (woa) and grey wolf optimizer (gwo). *Appl. Sci.* **2019**, *9*, 3755. [[CrossRef](#)]
51. Murtagh, F.; Contreras, P. Algorithms for hierarchical clustering: An overview, II. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2017**, *7*, e1219. [[CrossRef](#)]
52. Han, J.; Pei, J.; Tong, H. *Data Mining: Concepts and Techniques*; Morgan Kaufmann: Burlington, MA, USA, 2022.
53. Han, J.; Kamber, M. *Data Mining: Concepts and Techniques*, 2nd ed.; University of Illinois at Urbana Champaign: Champaign, IL, USA; Morgan Kaufmann: Burlington, MA, USA, 2006.
54. Hastie, T.; Tibshirani, R.; Friedman, J.H.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: Berlin/Heidelberg, Germany, 2009; Volume 2.
55. Yuruk, N.; Mete, M.; Xu, X.; Schweiger, T.A. A divisive hierarchical structural clustering algorithm for networks. In Proceedings of the Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007), Omaha, NE, USA, 28–31 October 2007; pp. 441–448.
56. Umam, K.; Bustamam, A.; Lestari, D. Application of hybrid clustering using parallel k-means algorithm and DIANA algorithm. In Proceedings of the AIP Conference Proceedings 2017, Yogyakarta, Indonesia, 9–10 November 2017; p. 020024.
57. Patnaik, A.K.; Bhuyan, P.K.; Rao, K.K. Divisive Analysis (DIANA) of hierarchical clustering and GPS data for level of service criteria of urban streets. *Alex. Eng. J.* **2016**, *55*, 407–418. [[CrossRef](#)]
58. Oyelade, J.; Isewon, I.; Oladipupo, F.; Aromolaran, O.; Uwoghiren, E.; Ameh, F.; Achas, M.; Adebiyi, E. Clustering algorithms: Their application to gene expression data. *Bioinform. Biol. Insights* **2016**, *10*, BBI-S38316. [[CrossRef](#)]
59. Chen, Z.; Liu, X.; Lu, Z.; Li, Y. The expansion mechanism of rural residential land and implications for sustainable regional development: Evidence from the Baota district in China's Loess Plateau. *Land* **2021**, *10*, 172. [[CrossRef](#)]

60. Mao, Y.; Zhang, M.; Sun, P.; Wang, G. Landslide susceptibility assessment using uncertain decision tree model in loess areas. *Environ. Earth Sci.* **2017**, *76*, 1–15. [[CrossRef](#)]
61. Zhang, M.-S.; Liu, J. Controlling factors of loess landslides in western China. *Environ. Earth Sci.* **2010**, *59*, 1671–1680. [[CrossRef](#)]
62. Bai, Y.; Liu, Y.; Li, Y.; Wang, Y.; Yuan, X. Land consolidation and eco-environmental sustainability in Loess Plateau: A study of Baota district, Shaanxi province, China. *J. Geogr. Sci.* **2022**, *32*, 1724–1744. [[CrossRef](#)]
63. Jaafari, A.; Rezaeian, J.; Omrani, M.S.O. Spatial prediction of slope failures in support of forestry operations safety. *Croat. J. For. Eng. J. Theory Appl. For. Eng.* **2017**, *38*, 107–118.
64. Efiong, J.; Eni, D.I.; Obiefuna, J.N.; Etu, S.J. Geospatial modelling of landslide susceptibility in Cross River State of Nigeria. *Sci. Afr.* **2021**, *14*, e01032. [[CrossRef](#)]
65. Zou, Q.; Jiang, H.; Cui, P.; Zhou, B.; Jiang, Y.; Qin, M.; Liu, Y.; Li, C. A new approach to assess landslide susceptibility based on slope failure mechanisms. *Catena* **2021**, *204*, 105388. [[CrossRef](#)]
66. Intrieri, E.; Carlà, T.; Gigli, G. Forecasting the time of failure of landslides at slope-scale: A literature review. *Earth-Sci. Rev.* **2019**, *193*, 333–349. [[CrossRef](#)]
67. Gao, J.; Shi, X.; Li, L.; Zhou, Z.; Wang, J. Assessment of Landslide Susceptibility Using Different Machine Learning Methods in Longnan City, China. *Sustainability* **2022**, *14*, 16716. [[CrossRef](#)]
68. Feng, L.; Guo, M.; Wang, W.; Chen, Y.; Shi, Q.; Guo, W.; Lou, Y.; Kang, H.; Chen, Z.; Zhu, Y. Comparative Analysis of Machine Learning Methods and a Physical Model for Shallow Landslide Risk Modeling. *Sustainability* **2023**, *15*, 6. [[CrossRef](#)]
69. Mersha, T.; Meten, M. GIS-based landslide susceptibility mapping and assessment using bivariate statistical methods in Simada area, northwestern Ethiopia. *GeoenvIRON. Disasters* **2020**, *7*, 1–22. [[CrossRef](#)]
70. Zhang, Q.; Liang, Z.; Liu, W.; Peng, W.; Huang, H.; Zhang, S.; Chen, L.; Jiang, K.; Liu, L. Landslide Susceptibility Prediction: Improving the Quality of Landslide Samples by Isolation Forests. *Sustainability* **2022**, *14*, 16692. [[CrossRef](#)]
71. Taşoğlu, E.; Abujayyab, S.K. Comparison of the frequency ratio, index of entropy, and artificial neural networks methods for landslide susceptibility mapping: A case study in Pınarbaşı/Kastamonu (North of Turkey). In *Computers in Earth and Environmental Sciences*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 491–508.
72. Selamat, S.N.; Abd Majid, N.; Mohd Taib, A. A Comparative Assessment of Sampling Ratios Using Artificial Neural Network (ANN) for Landslide Predictive Model in Langat River Basin, Selangor, Malaysia. *Sustainability* **2023**, *15*, 861. [[CrossRef](#)]
73. Liu, S.; Zhu, J.; Yang, D.; Ma, B. Comparative Study of Geological Hazard Evaluation Systems Using Grid Units and Slope Units under Different Rainfall Conditions. *Sustainability* **2022**, *14*, 16153. [[CrossRef](#)]
74. Yu, X.; Xia, Y.; Zhou, J.; Jiang, W. Landslide Susceptibility Mapping Based on Multitemporal Remote Sensing Image Change Detection and Multiexponential Band Math. *Sustainability* **2023**, *15*, 2226. [[CrossRef](#)]
75. Ngandam Mfondoum, A.H.; Wokwenmendam Nguet, P.; Mefire Mfondoum, J.V.; Tchindjang, M.; Hakdaoui, S.; Cooper, R.; Gbetkom, P.G.; Penaye, J.; Bekoa, A.; Moudioh, C. Adapting sudden landslide identification product (SLIP) and detecting real-time increased precipitation (DRIP) algorithms to map rainfall-triggered landslides in Western Cameroon highlands (Central-Africa). *GeoenvIRON. Disasters* **2021**, *8*, 1–26. [[CrossRef](#)]
76. Ismail, E.H.; Rogers, J.D.; Ahmed, M.F.; Usery, E.L.; Abdelsalam, M.G. Landslide susceptibility mapping of Blue Nile and Tekeze River Basins using oblique rainfall-aspect rasters. *Bull. Eng. Geol. Environ.* **2018**, *77*, 1311–1329. [[CrossRef](#)]
77. Kuradusenge, M.; Kumaran, S.; Zennaro, M. Rainfall-induced landslide prediction using machine learning models: The case of Ngororero District, Rwanda. *Int. J. Environ. Res. Public Health* **2020**, *17*, 4147. [[CrossRef](#)]
78. Hong, H.; Chen, W.; Xu, C.; Youssef, A.M.; Pradhan, B.; Tien Bui, D. Rainfall-induced landslide susceptibility assessment at the Chongren area (China) using frequency ratio, certainty factor, and index of entropy. *Geocarto Int.* **2017**, *32*, 139–154. [[CrossRef](#)]
79. Chowdhuri, I.; Pal, S.C.; Chakrabortty, R.; Malik, S.; Das, B.; Roy, P. Torrential rainfall-induced landslide susceptibility assessment using machine learning and statistical methods of eastern Himalaya. *Nat. Hazards* **2021**, *107*, 697–722. [[CrossRef](#)]
80. Das, S.; Sarkar, S.; Kanungo, D.P. A critical review on landslide susceptibility zonation: Recent trends, techniques, and practices in Indian Himalaya. *Nat. Hazards* **2022**, *115*, 23–72. [[CrossRef](#)]
81. Lee, S. Current and future status of GIS-based landslide susceptibility mapping: A literature review. *Korean J. Remote Sens.* **2019**, *35*, 179–193.
82. Dias, H.C.; Hölbling, D.; Grohmann, C.H. Landslide inventory mapping in Brazil: Status and challenges. In Proceedings of the XIII International Symposium on Landslides 2021, Cartagena, Colombia, 22–26 February 2021.
83. Roux, M. A comparative study of divisive hierarchical clustering algorithms. *arXiv* **2015**, arXiv:1506.08977. [[CrossRef](#)]
84. Qin, H.; Ma, X.; Herawan, T.; Zain, J.M. MGR: An information theory based hierarchical divisive clustering algorithm for categorical data. *Knowl. Based Syst.* **2014**, *67*, 401–411. [[CrossRef](#)]
85. Tyagi, A.; Sharma, S. Implementation of ROCK clustering algorithm for the optimization of query searching time. *Int. J. Comput. Sci. Eng.* **2012**, *4*, 809.
86. Altameem, A.; Poonia, R.C.; Kumar, A.; Raja, L.; Jilani Saudagar, A.K. P-ROCK: A Sustainable Clustering Algorithm for Large Categorical Datasets. *Intell. Autom. Soft Comput.* **2023**, *35*, 553–566. [[CrossRef](#)]
87. Guha, S.; Rastogi, R.; Shim, K. ROCK: A robust clustering algorithm for categorical attributes. *Inf. Syst.* **2000**, *25*, 345–366. [[CrossRef](#)]
88. Ding, M.; Hu, K. Susceptibility mapping of landslides in Beichuan County using cluster and MLC methods. *Nat. Hazards* **2014**, *70*, 755–766. [[CrossRef](#)]

89. Mao, Y.-M.; Zhang, M.-S.; Wang, G.-L.; Sun, P.-P. Landslide hazards mapping using uncertain Naïve Bayesian classification method. *J. Cent. South Univ.* **2015**, *22*, 3512–3520. [[CrossRef](#)]
90. Yimin, M.; Yican, L.; Simon Mwakapesa, D.; Genglong, W.; Ahangari Nanehkaran, Y.; Asim Khan, M.; Maosheng, Z. Innovative Landslide Susceptibility Mapping Portrayed by CA-AQD and K-Means Clustering Algorithms. *Adv. Civ. Eng.* **2021**, *2021*, 8846779. [[CrossRef](#)]
91. MacQueen, J. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1967*; University of California: Berkeley, CA, USA, 1967; pp. 281–297.
92. Ahmed, M.; Seraj, R.; Islam, S.M.S. The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics* **2020**, *9*, 1295. [[CrossRef](#)]
93. Rosi, A.; Tofani, V.; Tanteri, L.; Tacconi Stefanelli, C.; Agostini, A.; Catani, F.; Casagli, N. The new landslide inventory of Tuscany (Italy) updated with PS-InSAR: Geomorphological features and landslide distribution. *Landslides* **2018**, *15*, 5–19. [[CrossRef](#)]
94. Sun, H.; Burton, H.V.; Huang, H. Machine learning applications for building structural design and performance assessment: State-of-the-art review. *J. Build. Eng.* **2021**, *33*, 101816. [[CrossRef](#)]
95. Jones, P.J.; Catt, M.; Davies, M.J.; Edwardson, C.L.; Mirkes, E.M.; Khunti, K.; Yates, T.; Rowlands, A.V. Feature selection for unsupervised machine learning of accelerometer data physical activity clusters—A systematic review. *Gait Posture* **2021**, *90*, 120–128. [[CrossRef](#)] [[PubMed](#)]
96. Zimmermann, A. Method evaluation, parameterization, and result validation in unsupervised data mining: A critical survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2020**, *10*, e1330. [[CrossRef](#)]
97. Celebi, M.E.; Aydin, K. *Unsupervised Learning Algorithms*; Springer: Berlin/Heidelberg, Germany, 2016.
98. Rokach, L.; Maimon, O. Clustering Methods. In *Data Mining and Knowledge Discovery Handbook*; Springer: Berlin/Heidelberg, Germany, 2005.
99. Palacio-Niño, J.-O.; Berzal, F. Evaluation metrics for unsupervised learning algorithms. *arXiv* **2019**, arXiv:1905.05667.
100. Belyadi, H.; Haghhighat, A. Unsupervised machine learning: Clustering algorithms. *Mach. Learn. Guide Oil Gas Using Python* **2021**.
101. Mao, X.; Zhu, W.; Wu, L.; Zhou, B. Comparative study on methods for computing electrical distance. *Int. J. Electr. Power Energy Syst.* **2021**, *130*, 106923. [[CrossRef](#)]
102. Shahapure, K.R.; Nicholas, C. Cluster quality analysis using silhouette score. In Proceedings of the 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), Sydney, Australia, 6–9 October 2020; pp. 747–748.
103. Thinsungnoena, T.; Kaoungkub, N.; Durongdumronchaib, P.; Kerdprasopb, K.; Kerdprasopb, N. The clustering validity with silhouette and sum of squared errors. *Learning* **2015**, *3*. [[CrossRef](#)]
104. Xie, W.; Nie, W.; Saffari, P.; Robledo, L.F.; Descote, P.-Y.; Jian, W. Landslide hazard assessment based on Bayesian optimization-support vector machine in Nanping City, China. *Nat. Hazards* **2021**, *109*, 931–948. [[CrossRef](#)]
105. Sadighi, M.; Motamedvaziri, B.; Ahmadi, H.; Moeini, A. Assessing landslide susceptibility using machine learning models: A comparison between ANN, ANFIS, and ANFIS-ICA. *Environ. Earth Sci.* **2020**, *79*, 1–14. [[CrossRef](#)]
106. Darminto, M.R.; Widodob, A.; Alfatinahc, A.; Chuc, H.-J. High-resolution landslide susceptibility map generation using machine learning (Case Study in Pacitan, Indonesia). *Int. J. Adv. Sci. Eng. Inf. Technol.* **2021**, *11*, 369–379. [[CrossRef](#)]
107. Yu, X.; Xiong, T.; Jiang, W.; Zhou, J. Comparative Assessment of the Efficacy of the Five Kinds of Models in Landslide Susceptibility Map for Factor Screening: A Case Study at Zigui-Badong in the Three Gorges Reservoir Area, China. *Sustainability* **2023**, *15*, 800. [[CrossRef](#)]
108. Zhang, K.; Wu, X.; Niu, R.; Yang, K.; Zhao, L. The assessment of landslide susceptibility mapping using random forest and decision tree methods in the Three Gorges Reservoir area, China. *Environ. Earth Sci.* **2017**, *76*, 1–20. [[CrossRef](#)]
109. Park, S.-J.; Lee, C.-W.; Lee, S.; Lee, M.-J. Landslide susceptibility mapping and comparison using decision tree models: A Case Study of Jumunjin Area, Korea. *Remote Sens.* **2018**, *10*, 1545. [[CrossRef](#)]
110. Kadavi, P.R.; Lee, C.-W.; Lee, S. Landslide-susceptibility mapping in Gangwon-do, South Korea, using logistic regression and decision tree models. *Environ. Earth Sci.* **2019**, *78*, 1–17. [[CrossRef](#)]
111. Deng, H.; Wu, X.; Zhang, W.; Liu, Y.; Li, W.; Li, X.; Zhou, P.; Zhuo, W. Slope-Unit Scale Landslide Susceptibility Mapping Based on the Random Forest Model in Deep Valley Areas. *Remote Sens.* **2022**, *14*, 4245. [[CrossRef](#)]
112. Yu, C.; Chen, J. Application of a GIS-based slope unit method for landslide susceptibility mapping in Helong City: Comparative assessment of ICM, AHP, and RF model. *Symmetry* **2020**, *12*, 1848. [[CrossRef](#)]
113. Barbará, D. Requirements for clustering data streams. *ACM Sigkdd Explor. Newsl.* **2002**, *3*, 23–27. [[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.