Hindawi
Advances in Civil Engineering
Volume 2021, Article ID 8846779, 17 pages
https://doi.org/10.1155/2021/8846779

Research Article

Innovative Landslide Susceptibility Mapping Portrayed by CA-AQD and K-Means Clustering Algorithms

Mao Yimin[1,2] **Li Yican**[3] **Deborah Simon Mwakapesa**[1] **Wang Genglong**[2] **Yaser Ahangari Nanehkaran**[1] **Muhammad Asim Khan**[1] **and Zhang Maosheng**[2]

[1]*School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou, Jiangxi, China*
[2]*Key Laboratory for Geo-Hazards in Loess Area, MLR, Xi'an, Shaanxi, China*
[3]*Northwest Nonferrous Geological Mining Group Limited Company, Xi'an, Shaanxi, China*

Correspondence should be addressed to Zhang Maosheng; mymlyc2021@163.com

Received 20 September 2020; Revised 21 April 2021; Accepted 19 May 2021; Published 1 July 2021

Academic Editor: Lei Weng

This study aims at proposing and designing an improved clustering algorithm for assessing landslide susceptibility using an integration of a Chameleon algorithm and an adaptive quadratic distance (CA-AQD algorithm). It targets improving the prediction capacity of clustering algorithms in landslide susceptibility modelling by overcoming the limitations found in present clustering models, including strong dependence on the initial partition, noise, and outliers as well as difficulties in quantifying the triggering factors (such as rainfall/precipitation). The model was implemented in Baota District, Shaanxi province, China. The CA-AQD algorithm was adopted to split all grids in the study area into many groups with more similar characteristic values, which also owed to efficiently quantifying the uncertain (rainfall) value by using AQD. The K-means algorithm divides these groups into five susceptibility classes according to the values of landslide density in each group. The model was then evaluated using statistical metrics and the performance was validated and compared to that of the traditional Chameleon algorithm and KPSO algorithm. The results show that the CA-AQD algorithm attained the best performance in assessing landslide susceptibility in the study area. Thus, this work adds to the literature by introducing the first empirical integration and application of the CA-AQD algorithm to the assessment of landslides in the study area, which then is a new insight to the field. Also, the method can be helpful for dealing with landslides for better social and economic development.

1. Introduction

Landslides are one of the world's threatening natural hazards, which are regarded as the part of the masses of rock, compound, or soil that falls down a steep slope [1]. Determining the probability of landslide may be classified into two: (1) the intrinsic causes that accelerate slope failure, such as geological and morphological properties and (2) the extrinsic causes that change the slope from being in a marginally stable state to an unstable state, such as rainfall, earthquake shaking, and human activity [2].

In China, according to related reports [3], nearly all landslides are directly triggered by or related to rainfall. Being regional, aggregate, abrupt, and disastrous characteristics, rainfall-triggered, these landslides have imposed tremendous threats to lives of the people as well as economic activities [4] To diminish the likelihood of damage caused by landslides, it is significantly important to come up with accurate and effective methods for mapping landslide susceptibility to ensure safe environment and steadiness of economic activities and support reliable hazard prevention and reduction [5].

With the technological advancement of remote-sensing (RS) images and geographic information system (GIS) data processing, it has become easier to obtain significant information to analyze landslide susceptibility. Considerable effort has been posed to develop the assessment of landslide susceptibility using RS and GIS technology. Initially, specialists created susceptibility maps to provide an inventory of landslides using qualitative overlays of topographical and geological attributes [6]. Later, more landslide susceptibility assessment methods were developed in specific areas by

applying deterministic approaches, statistical analyses, and computation intelligence method [7, 8]. Given that appropriate soils and rocks engineering data, slope geometry, discontinuity features, and hydrological factors are required to compute the resisting and driving forces association, deterministic models have been limited to small study areas [9, 10]. Statistical models, such as linear and logistic regression [11–13], bivariate statistical models [14–16], frequency ratio [17–21], and weight of evidence models [22–24], have been applied widely to the field of constructing assessment models for landslide susceptibility. These models, however, cannot easily determine the relationship between significant landslide-influencing factors and complicated landslide systems [25].

Prediction models of landslides based on classification algorithms in data mining can overcome such difficulties. Specifically, being interested in coming up with ideal learning methods to determine the nonlinear relationship among landslides and the environmental factors [26], many researchers have successfully adopted them, for example, support vector machine [27–29], decision tree [30–35], naïve Bayesian [36, 37], artificial neural networks [38–41], random forest models [42, 43], and others, to construct landslide susceptibility map. These models, however, depend on a big training data set to improve prediction accuracy. Training data sets need geoscientists or engineers to survey landslide sites, which, in reality, are not easy to be, in particular, to capture rainfall information.

Clustering analysis algorithms classify sets of objects (grids) into groups that are more similar to each other than they are to objects in other clusters (groups) [44]. This process is conducted by primitive observation with little or no prior knowledge; that is, it is unsupervised learning. Because of its advantages, several researchers have been interested in applying K-means [45, 46], fuzzy C-means (FCM) [47], and K-means particle swarm optimization (KPSO) [47, 48] to assess landslide susceptibility. K-means and FCM algorithms can be effective if the choice of initial partitions in the prediction model is correct [49]. In fact, these parameter thresholds (every clustering center) are not easy to be set in large data sets, precisely, large study areas [50]. KPSO can break away from initial partitions dependence using iterations to identify the best cluster partitions, but it is sensitive to data clusters that have diverse shapes, densities, and sizes (called outliers and noise) [51], which restricts the advantages of using KPSO to assess landslide susceptibility in large study areas. Fortunately, the Chameleon algorithm [51] separates itself from initial partition, noise, and outlier dependence, by merging the clusters using a dynamic model to find natural and homogeneous clusters. The Chameleon algorithm, however, regards rainfall as the average value or discrete value, which leads to a distortion of value [33] and influences clustering results. The advantage of the adaptive quadratic distance [52] is changing at every iteration of the algorithm being either similar for all clusters or may change from one cluster to another; thus, we integrated AQD and Chameleon algorithm to construct the spatial prediction model that could be available to any of the study areas and would classify more similar groups with topography and geology from all objects. Finally, the landslide density of each group can be calculated by the sorting tool in ArcGIS. Then, the K-means algorithm [53] was adopted to assign these groups to five susceptibility classes (very high, high, moderate, low, and very low) with the values of landslide density in each group. The algorithms will be incorporated into landslide susceptibility mapping in the study area of Baota District, China. More details of this study will be described in the following sections.

2. Study Area and Materials

2.1. Study Area. The study area is part of the Loess Plateau, sited in the northern part of Shaanxi province and situated between latitude 36°11′ and 37°02′N, and longitude 109°14′ and 110°07′E. It encompasses 3,556 km^2 and is often prone to landslides that are usually triggered by rainfall. Yanhe River is bounded to the north and Fenchuan River extends to the south of the study area (Figure 1).

The geomorphic of the study area is characterized as undulate slopes and ravines, with elevation values ranging between 800 and 1,800 m. The annual mean temperature is 10°C. Historical data shows that the highest amount of rainfall varies from 114 mm to 460 mm between June and October with an annual average of 550 mm. The landslides survey data indicated that 71.4% of the total rainfall and 84.6% of the landslides occur in this area between June and October. Figure 2 depicts how landslides relate to rainfall in the study area [54].

2.2. Landslide Inventory Map. The landslide inventory is an important factor in portraying landslide susceptibility mapping. The Xi'an Center for Geological Survey (CGS) has been done through landslide surveys in Baota District, which includes interpretation of SPOT-5 satellite images for the whole area and of QuickBird satellite images which covered 225 km^2 of the urban area. CGS constructed a landslide inventory of the study area using aerial photos. From the study area, landslides were portrayed and analyzed at 1,081 locations, and 428 landslides were surveyed. Most of the landslides were prompted by rain, and 293 landslides (see the right-hand side of Figure 1) had recorded precipitation information. Most of the landslides were scattered along the sides of the Yanhe and the Fenchuan Rivers. Nearly the entire area was covered by Quaternary loess and thick, loose loess deposits; thus, almost all the landslides were soil landslides. Most of the landslides were medium-scale landslides with a sliding body volume between $10^1 \times 10^4$ and $10^2 \times 10^4$ m^3. The number of small-scale landslides accounted for 30.7% of the landslides, with a sliding body volume of less than $10^1 \times 10^4$ m^3. Large-scale landslides had a sliding body volume between $10^2 \times 10^4$ and $10^3 \times 10^4$ m^3 and accounted for only 16.7% of the landslides [54]. The dominion of landslides is shown in Table 1.

2.3. Data Preparation. Previous studies [54] have classified landslide conditioning factors in the study area into four groups: topography, geology, underlying surface, and triggering factor.

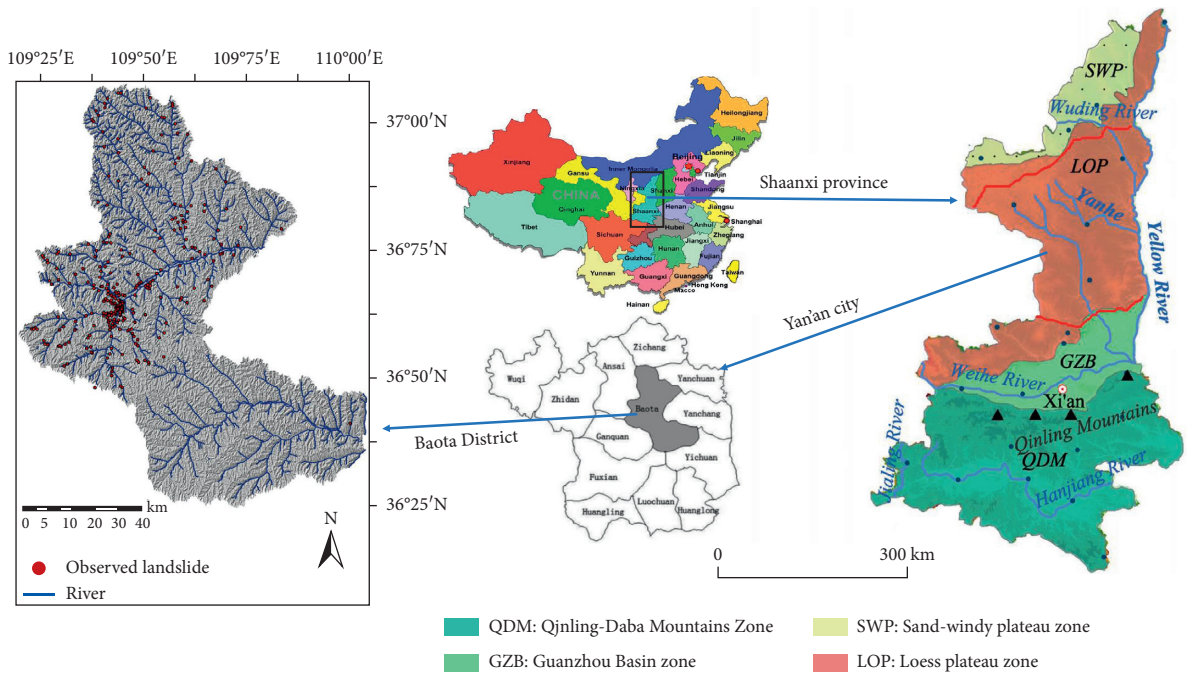


Figure 1: Location of the study area: Baota District, Yan'an city.

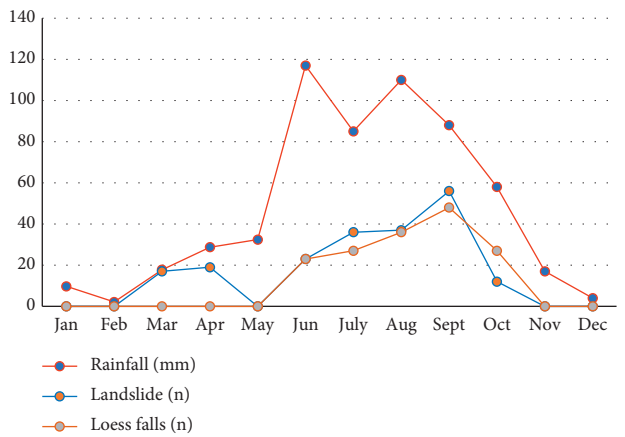


Figure 2: The relation between the landslide occurrence and rainfall.

Table 1: Landslide's dominion.

Number	Name	Location		Length (m)	Width (m)	Thickness (m)	Volume ($\times 10^4$ m^3)
		Coordinates					
		Longitude	Latitude				
1	Liangcun Guojiashi	109°32′12″	36°52′04″	150	200	15	45
2	Urban Market Gap 2	109°28′24″	36°35′01″	30	150	4	1.8
3	Urban market ditch 1	109°28′18″	36°34′52″	60	100	7	4.2
4	Yuyuan Houjiagou	109°25′03″	36°37′13″	150	35	8	42
5	Yao shop Sujiagou	109°41′48″	36°38′25″	112	200	5	10.9
6	Yaoshop ZhaoJiagou	109°41′17″	36°35′54″	220	280	10	17
7	Xichuankou, Summer Palace	109°19′49″	36°38′10″	150	250	10	37.5
8	Wanhua Xiaojialing South	109°24′56″	36°34′12″	100	60	10	24
9	Wanhua Gaojiagou	109°25′07″	36°32′39″	250	200	14	266.4
10	Li Que Yang Xinzhuang	109°34′50″	36°39′35″	240	480	15	172.8
...	...	...	...	...	...	...	...

Topography factors, which depict the geomorphologic and topographic characters in the study area [5], comprise elevation, slope angle, slope aspect, and profile curvature. We derived these data layers (Figures 3(a)–3(d)) from a digital elevation model (DEM) with a resolution of 25 m, which was constructed from the topographic maps at a scale of 1 : 50,000. By computing and analyzing these data layers, previous research results [54] showed that the stability of slope angle and elevation ranged from 25° to 55°, 20 m to 120 m, respectively. The probability of landslides to occur was greater along the shaded slope.

The geological data layer was gained by digitizing a geological map at a scale of 1 : 50,000 (Figure 4) supported by CGS. The geomorphological layer includes Jurassic, Triassic, Neogene, as well as Quaternary strata, whereby the Quaternary loess and the Neogene red clay are exposed to landslides [54]. Because of thick, loose loess deposits in the Baota District, landslides and mudflows occur frequently, making the district more prone to landslides. Therefore, we selected the rock-soil structure to evaluate the geologic condition.

Because of being considered as a factor of underlying surface in the previous study [54], vegetation cover data layer (Figure 3(e)) was extracted using Enhanced Thematic Mapper Plus (ETM+) RS images. The vegetation coverage was more than 60% in the southern area. According to a field survey, landslides were scarce; conversely, poor vegetation and extensive landslides were found in the north [54].

Rain penetrates into rock and soil and erodes them into fractures as a result that the average rainfall has the erosive capability for them. In 19 rainfall stations of the study area, maximal rainfall and minimal rainfall for every month were recorded by CGS. The rainfall map (Figure 3(f)) was constructed by obtaining the maximal average month rainfall in July during 2017 to 2018. The rainfall classes are defined in Table 2.

3. Research Methods

3.1. Chameleon Algorithm. Chameleon is a clustering algorithm that explores dynamic modelling in hierarchical clustering. In its clustering process, two clusters are merged if the interconnectivity and closeness (proximity) between two clusters are highly related to the internal interconnectivity and closeness of objects within the clusters. The merging process based on a dynamic model facilitates the discovery of natural and homogeneous clusters and applies to all types of data as long as a similarity function is specified [51]. To its advantage, the Chameleon algorithm is a user-supplied model, static independent, as well as adapting to the internal characteristics of the clusters of being independent of the initial partitions, as well as insensitive to noise and outliers [44]; thus, we used Chameleon algorithm to assess the landslide susceptibility. In general, for landslide susceptibility assessment using the clustering algorithm, an object is regarded as a grid. In the Chameleon algorithm, an object is considered as a node of the weighted graph.

The main concept of the Chameleon algorithm is presented in Figure 5. Firstly, to cluster the data nodes into a big number of relatively small subgroups, the algorithm applies a graph partitioning algorithm. Then, by combining or merging the subgroups in an iterative process, the algorithm uses an agglomerative hierarchical clustering algorithm to identify the genuine clusters. It then takes into consideration especially the internal characteristics of the groups themselves (both the interconnectivity as well as the closeness of the clusters) in discovering the pairs of most similar subgroups. Thus, to this end, the algorithm is not a static, user-supplied model dependent and can automatically conform to the internal characteristics of the merged groups.

Definition 1. Suppose i, j are nodes in a weighed graph with n-dimensional data; then, the Euclidean distance between them is

$$d(i,j) = \sqrt{\left(x_{i1} - x_{j1}\right)^2 + \left(x_{i2} - x_{j2}\right)^2 + \cdots + \left(x_{in} - x_{jn}\right)^2}, \tag{1}$$

where $i = (x_{i1}, x_{i2}, \ldots, x_{in})$, $j = (x_{j1}, x_{j2}, \ldots, x_{jn})$.

Definition 2. Suppose $G = (V, E, W)$ is a weighted graph composed of a finite nonempty node set V, edge set E, and edge weight function W. Disordered pairs (V_i, V_j) indicating the edge that connects nodes $V_i \in V$ and $V_j \in V$, W_{ij} is the weight of edge (V_i, V_j). If $(V_i, V_j) \in E$; then, $W_{ij} = 1/d(i,j)$; otherwise, $W_{ij} = 0$.

Definition 3. Suppose graph $G = (V, E)$ has been divided into k clusters in one clustering result, which means $G = (G_1', G_2', \ldots, G_k')$, $G_i' = (V_i', E_i')$. We defined a one-dimensional vector $V = (V_i)$, $V_i = |E_i'|/|E|$ and a trade-off standard $Q = \sum_{i=1}^{k} (V_i - V_i^2)$ that is used to define modularity to evaluate clustering results.

Definition 4. If two nodes have exactly the same adjacent nodes, then the two nodes are called structural equivalents. The equivalence similarity of the structure is calculated as follows:

$$S_{ij} = \sqrt{\sum_{k \neq i,j} \left(W_{ik} - W_{jk}\right)^2}. \tag{2}$$

The main processes of the Chameleon algorithm are as follows:

(1) Build a weighted graph, the number of initialized clusters is n; that is, each node is a cluster

(2) Calculate the structural equivalence, merge clusters, according to its value, from small to large

(3) Calculate modularity Q

(4) Repeat 2 and 3, continue to merge clusters until Q drops and stop, and output all clusters

3.2. Uncertain Data and CA-AQD

3.2.1. Uncertain Data. Uncertain numerical data are ubiquitous in real life. For example, someone's annual salary cannot be fixed in a specific value and can be determined

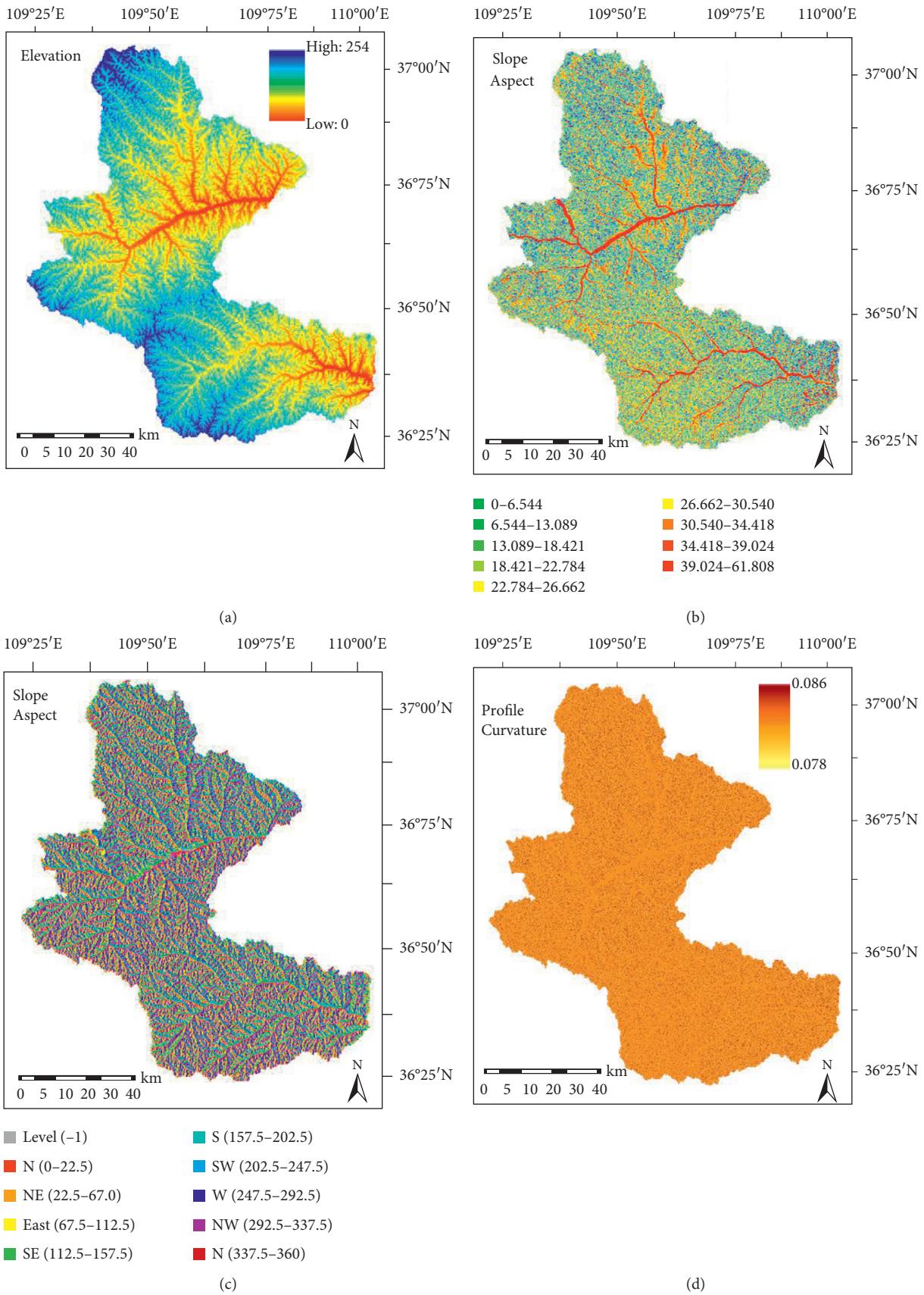


Figure 3: Continued.

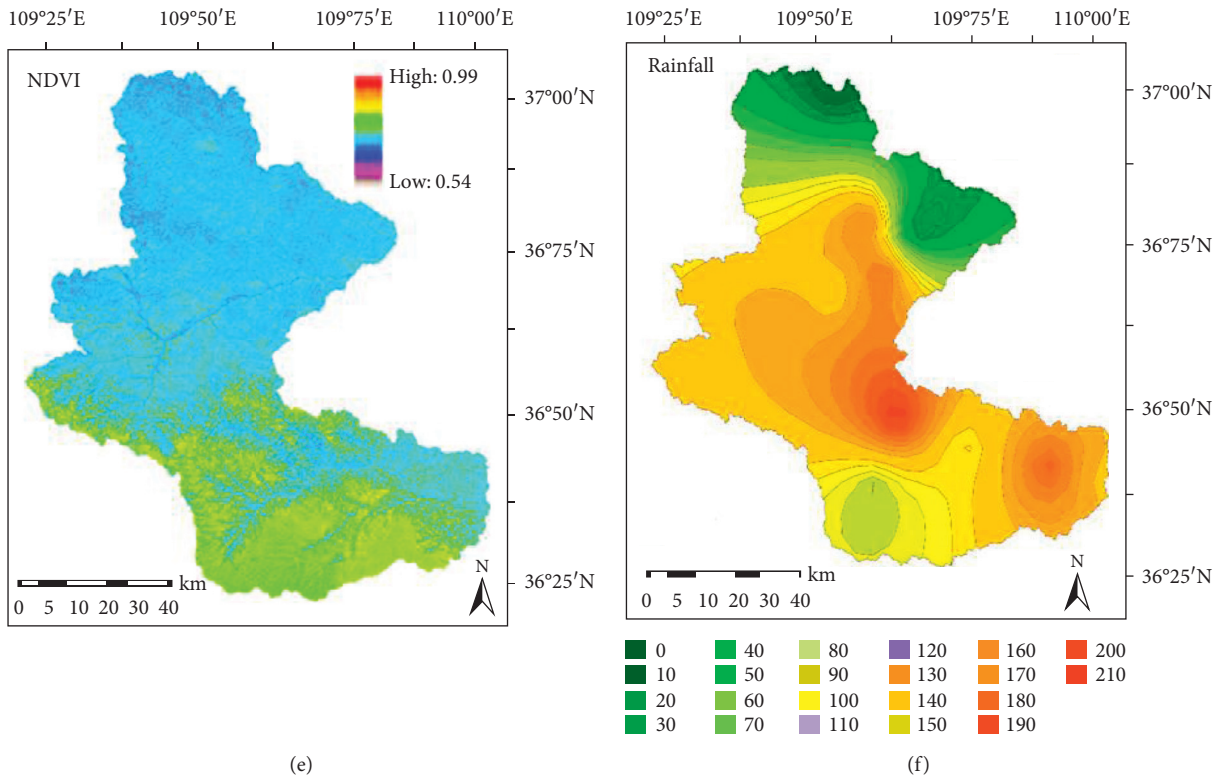


Figure 3: The thematic maps for factors: (a) elevation, (b) slope angle, (c) slope aspect, (d) profile curvature of the slope, (e) vegetation, and (f) rainfall.

only in a range of \$160,000–200,000. Similarly, a rainfall value in one or more days cannot be determined as a specific value and can be determined only in a range say, of 10 mm to 20 mm. For an uncertain data model (interval-value datum), the uncertain data is fixed in a two-dimensional array with the lower and upper bounds of the interval ($\underline{x}_{ij}, \overline{x}_{ij}$, resp.).

3.2.2. CA-AQD Algorithm. In the Chameleon algorithm, the similarity of two nodes depends on their Euclidean distance values. The traditional Euclidean distance formula can process only those nodes whose values are continuous and discrete [50]. In the rainfall-induced regional landslide hazard assessment model, the data types of the node's attributes (landslide conditioning factors) include discrete (slope aspect), continuous (slope height), and uncertain (rainfall) values [37]. The traditional Euclidean distance formula, however, cannot describe nodes with uncertain data (rainfall value). To remedy the week point, the AQD distance between two nodes x, y is calculated in this paper. The Unabridged formula can be found in the literature [52]. Its basic definition follows.

We assume the clustering model $p_k\,(k = 1, \ldots, k)$ can be regarded as a vector of intervals $y_k\,(y_k^1, y_k^2, \ldots, y_k^p)$. Additionally, the vector of intervals can be disposed of a two-dimensional matrix. For example, there is a vector space Ω with n objects $\{1, \ldots, n\}$ (dimensionality $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, n\}$) such that $y\,(i)\,(j) = [a, b] \in \mathfrak{F}$, whereby $\mathfrak{F}$ indicates a set of closed intervals defined from R if we have a vector of intervals $x_i = \left\{x_i^1, x_i^2, \ldots, x_i^p\right\}$, where $x_i^j = [a_i^j, b_i^j] \in \mathfrak{F} = \{[a, b]\colon a, b \in R, a \le b\}$ $(j = 1 \sim p)$. In a similar way, the prototype of a cluster p_k can be disposed by a vector of intervals $y_k = (y_k^1, y_k^2, \ldots, y_k^p)$ where $y_k^j = [\alpha_k^j, \beta_k^j] \in \mathfrak{F}$. Then, we use x_{iL}, x_{iL} as the lower and upper boundaries of the intervals, respectively, to describe x_i. We also consider y_{kL}, y_{kU} indicating two vectors, respectively, of the lower and upper boundaries of the intervals of y_k. This means we also solve the prototype p_k. The equation of adaptive quadratic distance is follows:

$$d_M^2(x_i, y_k) = (x_{iL} - y_{KL})^T M (x_{iL} - y_{KL}) + (x_{iU} - y_{KU})^T M (x_{iU} - y_{KU}), \tag{3}$$

where M is a full positive definite symmetric matrix, $M_k = M\,(k = 1, \ldots, k)$.

Equation (3) can be used to calculate the distance between two nodes whose values are uncertain. We replaced the Euclidean distance formula by equation (3) to obtain the weight and structure equivalent similarity between nodes in the Chameleon algorithm, which called the Chameleon adopted adaptive quadratic distance (CA-AQD) algorithm.

3.3. K-Means Algorithm. K-means algorithm yields better performance in the case of inputting the k of clusters and randomly choosing the center of clusters in advance; thus, we adopted it to classify the landslide density of each group into five clusters (susceptibility classes). Its main steps are as follows:

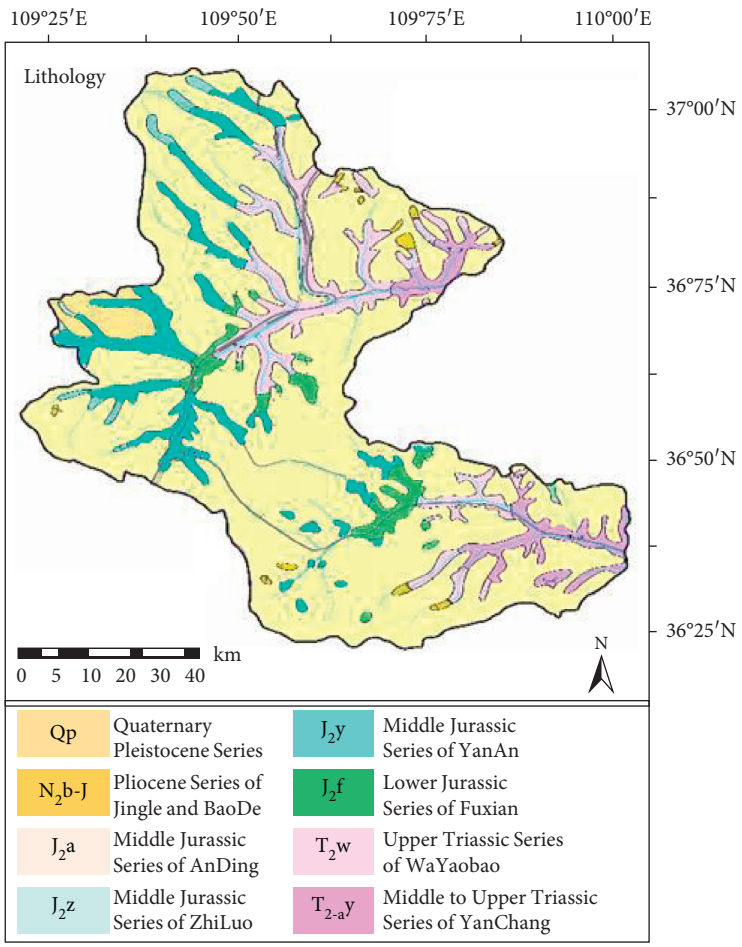


Figure 4: The strata in the study area: Baota District, Yan'an city. Figure from the Department of Survey, Government of Xi'an.

Table 2: The attribute type in different algorithms for constructing and evaluating landslide susceptibility mapping in the study area.

Factors	Attribute name	Attribute type in DTU and NBU	Attribute type in CA-AQD	Attribute type in Chameleon and KPSO	Classes of discrete attribute
Elevation	Elevation	Continuous	Continuous	Continuous	
Slope angle	Angle	Continuous	Continuous	Continuous	
Slope aspect	Aspect	Discrete	Discrete	Discrete	Flat, N, NE, E, SE, S, SW, W, NW
Profile curvature	Curvature	Discrete	Discrete	Discrete	<−0.05, −0.05 to 0.05, >0.05
Rock-soil structure	Rock-soil	Discrete	Discrete	Discrete	1: loess + nearly horizontal paleo-soil, 2: loess + inclined paleo-soil, 3: loess + paleo-soil layers + bedrock, 4: loess + paleo-soil layers + the Neogene clay
Vegetation	NDVI	Continuous	Continuous	Continuous	
Rainfall	Rainfall	Uncertain	Uncertain	Discrete	0~60 mm, 60~80 mm, 80~100 mm, 100~120 mm, 120 mm above

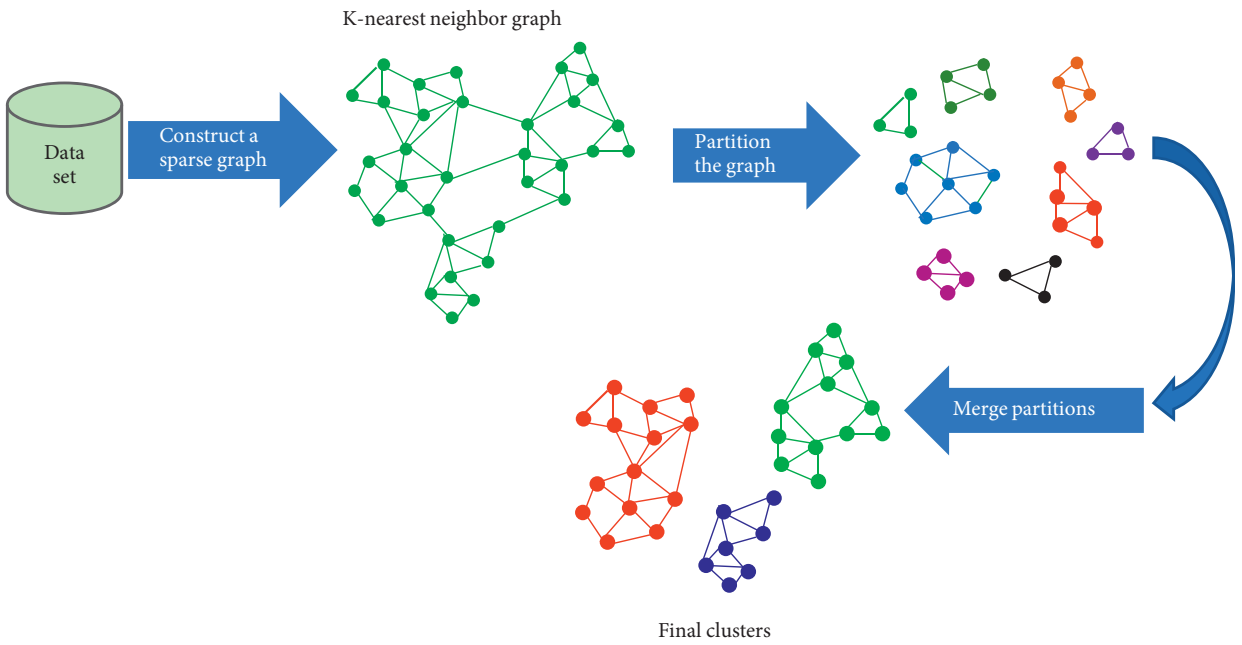


Figure 5: The main approach of Chameleon.

(1) Arbitrarily set k objects as the initial cluster centers

(2) Classify every object to the cluster to which the object is the closest to the cluster centers

(3) Calculate the new cluster centers and update them for new cluster centers

(4) Repeat 2 and 3 until almost not changing the cluster centers

Following the above steps, and having the landslide density of each group as an object, the K-means algorithm classifies these objects to k clusters, where each cluster is regarded as a susceptibility class.

3.4. Evaluation Methods

3.4.1. Model Performance Evaluation. To evaluate the performance of this study and compare with the others studies, some common statistical measure such as Cohen Kappa index (k), sensitivity, specificity, accuracy, and $F1$-measure were used, which are elaborated below.

The Cohen Kappa index (k) [55] was used to analyze and compare the reliability of the model classification results of the landslide susceptibility models:

$$k = \frac{p_c - p_{\exp}}{1 - p_{\exp}}, \tag{4}$$

whereby

$$p_c = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}},$$
$$P_{\exp} = \frac{((\mathrm{TP} + \mathrm{FN}) * (\mathrm{TP} + \mathrm{FP}) + (\mathrm{FP} + \mathrm{TN}) * (\mathrm{FN} + \mathrm{TN}))}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})^2}. \tag{5}$$

The value of the **k** is between 0 and 1 and is defined in the different groups such that the value close to 1 implies observed agreement between the landslide model and the actual data while a value close to 0 implies disagreement [56]. A negative value of k implies low agreement. K value from 0.8 to 1 implies closer complete agreement while the substantial agreement is between 0.60 and 0.80. On the contrary, 0.40 to 0.60 indicates moderate agreement, whilst the value from 0.20 to 0.40 indicates better than fair and slight agreement, respectively:

$$\begin{aligned} \text{sensitivity} &= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \\ \text{specificity} &= \frac{\mathrm{TN}}{\mathrm{FP} + \mathrm{TN}}, \\ \text{accuracy} &= \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}}, \\ F1 - \text{measure} &= \frac{2PR}{P + R}, \end{aligned} \tag{6}$$

whereby

$$\begin{aligned} p &= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \\ R &= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \end{aligned} \tag{7}$$

From the above equations, see the following.

True positive (TP): if the outcome of the prediction model is correctly classified as a landslide. False positive (FP): the outcome of the prediction model is incorrectly classified as a landslide. True negative (TN): the outcome of the prediction model is correctly classified as nonlandslide. False negative (FN): the outcome of the prediction model is nonlandslide while it is landslide.

Moreover, receiver operating characteristics (ROC) was adopted, which is also an essential evaluation metric for assessing the landslide models' performance. The receiver operating characteristics (ROC) curve is a popular model performance metric used to measure the overall performance of a landslide susceptibility model. In the ROC curve, the model can be described by drafting the ratio of the zonation identified as the error in the landslide-prone zonation (1 – specificity) in the x-axis against the ratio of the landslide zonation identified as the correct landslide-prone zonation (Sensitivity) in the y-axis [45]. The area under the ROC curve (AUC) is a model performance and quality measurement, whereby an AUC value of 1 implies an excellent model. The AUC value between 0.8 and 0.9 implies a very good model, while 0.6–0.7 implies an average performance. A poor-quality model has an AUC value between 0.5 and 0.6.

3.4.2. Models Comparison. To compare the performance of the proposed landslide susceptibility models, some performance should be compared among CA-AQD, Chameleon, and KPSO (found in literature [47]) algorithms. KPSO clustering algorithm was chosen due to its remarkable computational efficiency and it is easy to implement. On account of initial parameter thresholds dependence, k-means and FCM algorithms were not taken into consideration. Furthermore, uncertain decision tree (found in literature DTU) [33] and uncertain naïve Bayesian (found in literature NBU) [37] classification algorithms were applied against the proposed model in quantifying the value of rainfall and attaining better prediction accuracy. At the same time, the state-of-the-art benchmark models such as SVM, ANN, and RF [5] can also be chosen for checking the ability of the proposed model.

4. Results

The CA-AQD algorithms outlined in this study were used to construct a landslide susceptibility map and validate and compare its performance with other methods, in different steps (as shown in Figure 6). We used the CA-AQD algorithms outlined in this study to construct a landslide susceptibility map and validate its performances. The study workflow is shown in Figure 6. The process includes four phases: data collection, clustering analysis, describing the landslide susceptibility map, and model validation. To collect data, we extracted the value of landslide conditioning factors from thematic graphs in ArcGIS. To conduct the cluster analysis, we partitioned whole grids into many groups with similar and geomorphology characteristics. To describe the landslide susceptibility map, we evaluated the landslide's susceptibility classes for each group according to the K-means algorithm or according to the characteristics of landslide conditioning factors. To validate the model, validate the performances of the proposed method and compare the performances with the others methods.

4.1. Data Collection. After classifying landslide conditioning factors, the study regions were identified as polygons, which were converted into a raster map. The map featured 25×25 m grid spacing, and the study area had 5,672,922 grids, which included landslides and nonlandslides. Each grid can be extracted from the thematic graphs of the indicator variables (which were treated as attributes of grids). The attributes for each grid included discrete, continuous, and uncertain data types based on different algorithms, as shown in Table 2.

4.2. Clustering Analysis. We used the CA-AQD algorithm to classify all grids into groups with similar geology and geomorphology. The main process was as follows.

Compared with the K-means and FCM algorithms, CA-AQD does not need to set the initial values of parameters. It justly calculated the distance of two grids using AQD distance resulting in the range of values. Thus, we imported the normalized attributes values of each grid into the CA-AQD algorithm and divided 5,672,922 grids in the study area into 465 groups in July, which did not need to manually set the initial parameter of the cluster. Some groups of CA-AQD algorithms in July are shown in Figure 7.

As shown in Figure 7, grids of the groups of CA-AQD algorithms were distributed intricately within the study area. For example, the grids of red color in the second group mainly were located along the side of Yanhe River and sporadically scattered along the sides of Fenchuan Rivers. Additionally, the left side of Figure 7 shows the attribute values for all grids of the first and second groups. As in Figure 7, the characteristics of geology and geomorphology which can indicate the attribute values between the first and second groups were different rather than being very similar in the same group. For example, in Figure 7(a), the values of attributes in the first group were almost similar. But the values of their grid ID were different, which means grids of the same group were distributed intricately within the study area. These results showed that the clustering algorithm based on the CA-AQD algorithm could be used to effectively divide the space grids in the study area into groups.

4.3. Describing Landslide Susceptibility Map. The CA-AQD algorithm concentrated grids in the study area that had similar geological and geomorphic environments in the same group. Being similar in all attributes of one group, we used the mean value of the attribute to reflect the eigenvalues of all attributes for each cluster in the study area, as shown in the left-hand column of Table 3. However, the susceptibility classes cannot be labeled in each group of Table 3. Based on the general principle that the higher the landslide density is, the higher the susceptibility ought to be, we used the K-means algorithm to solve this problem. The steps are as follows:

At first, the landslide density of each group, regarded as only one attribute for all objects in K-means, should be calculated. In the experiment, we interpreted RS of SPOT-5 satellite images for 1,081 locations to compute the landslide density of the entire study area and each group by using the

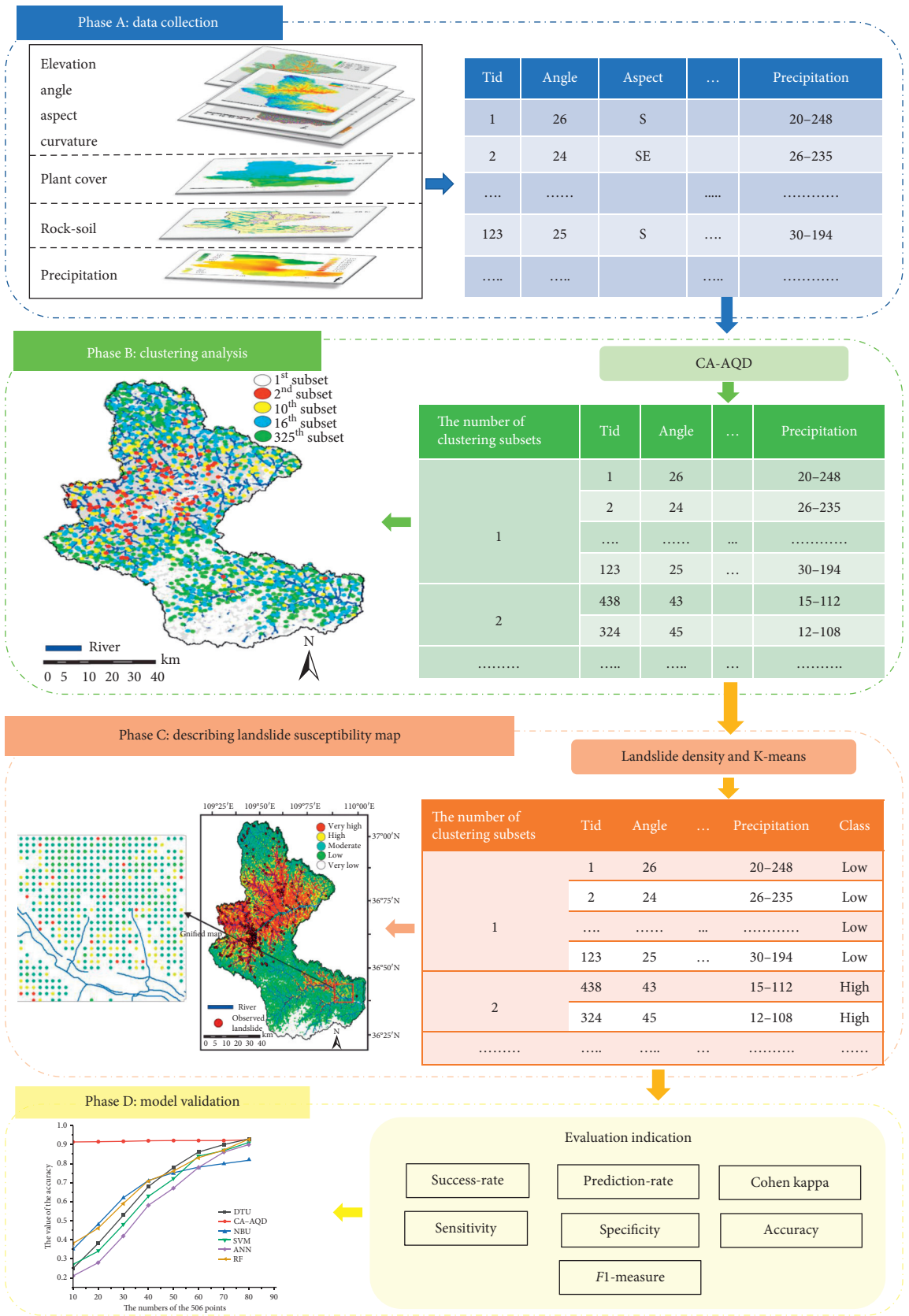
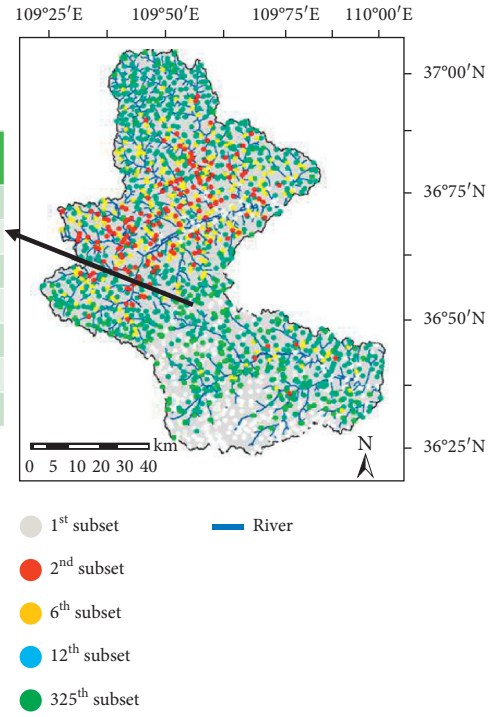


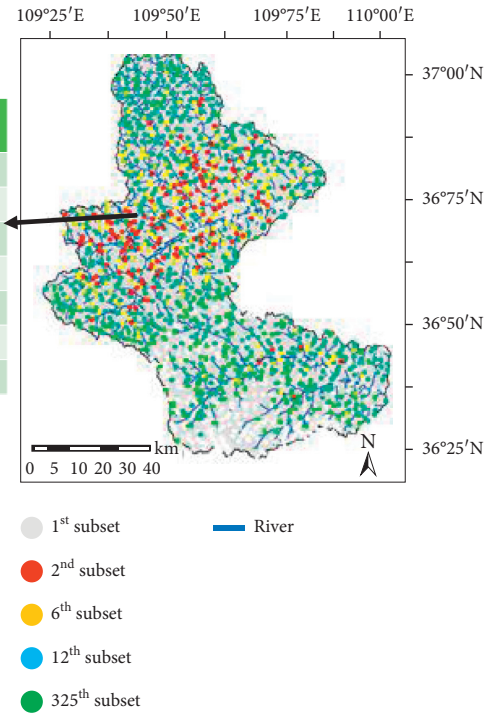
Figure 6: The overall workflow of mapping landslide susceptibility.

Id	Elevation	Slope angle	Slope aspect	Profile curvature	Rock-soil	NDVI	Rainfall
1	37	26	SE	−0.041	1	0.526	20–248
2	33	24	SE	−0.038	2	0.535	26–235
…	…	…	…	…	…	…	…
123	33	25	S	−0.0034	2	0.578	30–194
…	…	…	…	…	…	…	…
223	39	28	S	−0.044	1	0.648	32–182
…	…	…	…	…	…	…	…



(a)

Id	Elevation	Slope angle	Slope aspect	Profile curvature	Rock–soil	NDVI	Rainfall
1	53	43	NW	0.063	0.326	−0.468	15–112
2	56	45	N	0.046	0.331	−0.435	12–108
…	…	…	…	…	…	…	…
123	57	46	N	0.052	0.332	−0.457	14–110
…	…	…	…	…	…	…	…
139	58	48	NE	0.051	0.346	−0.472	18–121
…	…	…	…	…	…	…	…



(b)

Figure 7: The map of the group based on the CA-AQD algorithm in July: (a) the first group and (b) the second group.

sorting tool in ArcGIS. The landslide density of some groups is shown in the right-hand columns of Table 3. Next, the number of clusters is set to five landslide susceptibility classes based on the landslide susceptibility analyses in the study area. Initial cluster centers were chosen randomly among the range of landslide density in all groups. Finally, the above parameters were input to the K-means algorithm, and the landslide susceptibility classes of all groups are shown in Figure 8.

For the cases where a group obtains landslide density of 0, its susceptibility level may be considered as very low; however, when the landslide observation points are not

Table 3: Eigenvalue and landslide density of group based on the CA-AQD algorithm in July.

Number of groups	Eigenvalue (have not been normalized)							Area and landslide density			Remark
	Elevation	Slope angle	Slope aspect	Profile curvature	Rock-soil	NDVI	Rainfall	Area (km^2)	Number (km^2)	Density (km^2)	
1	34.5	25.23	S	−0.04	0.46	0.53	20–248	24.45	1	0.04	Low
2	57.3	46.43	NE	0.05	0.51	−0.44	15–156	8.45	12	1.42	High
. . ..	. . ..	. . ..	. . ..	. . ..	. . ..	. . ..	. . ..	. . ..	. . ..	. . ..	. . ..
121	29.4	20.46	SE	-0.05	0.44	0.62	18–237	8.28	0	0	Determined by experts
. . ..	. . ..	. . ..	. . ..	. . ..	. . ..	. . ..	. . ..	. . ..	. . ..	. . ..	. . ..

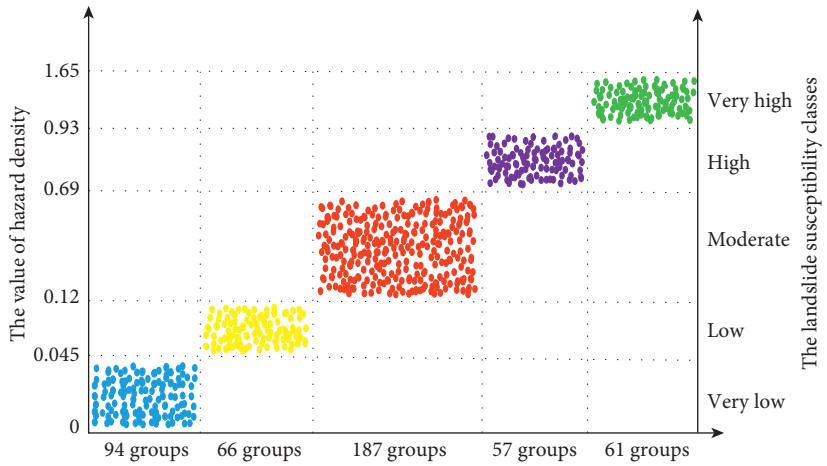


Figure 8: The landslide susceptibility classes of all groups.

covered in the group, there might be a very high susceptibility level. It is a confusing case, thus, to overcome this confusion, the experts ought to use the eigenvalues of the group (the left-hand column of Table 3), to identify the susceptibility level of that group.

To evaluate the influence on the landslide from rainfall, we obtain the rainfall data from the rainfall map in April and July based on their scattered and heavy rain, respectively. According to the susceptibility classes of each group, the susceptibility assessment maps of April and July based on the CA-AQD algorithms are drawn as shown in Figure 9. Figure 9 demonstrates that landslides occurred in July more frequently than in April, which according to the previous study, was almost unanimous in the actual situation [54]. The majority of the landslides occurred as a result of rainwater infiltration because precipitation increased the pore-water pressure and the weight of the loess slope.

4.4. Model Validation. As most landslides occurred from June to October, we validated the prediction model using precipitation data from July. This study compared the CA-AQD model with Chameleon and KPSO models to prove that the CA-AQD model outperformed the Chameleon and KPSO models. To compute the Cohen kappa index as well as performance accuracy, 293 landslides and 213 nonlandslides were randomly applied for model validation. The Cohen kappa indices (k) for the CA-AQD and Chameleon algorithms were greater than 0.8, which demonstrated that the two models were in nearly complete agreement with the field survey. The prediction accuracy values based on CA-AQD, Chameleon, and KPSO algorithms were 0.9249, 0.9110, and 0.6621, respectively, whereas the Cohen kappa indices were 0.8471, 0.8192, and 0.3161, respectively (Table 3). Conversely, the CA-AQD model had the highest accuracy of all three models, which applied the AQD to quantify precipitation for improving the prediction accuracy. As shown in Table 3, comparing the CA-AQD, Chameleon, and KPSO models, the CA-AQD model obtained the highest sensitivity, specificity, and $F1$-measure of 0.9147, 0.9390, and 0.9341, respectively.

Furthermore, from Figure 10, among the three models, the CA-AQD model showed the highest AUC value of 0.884. Also, the AUC values of the CA-AQD and Chameleon models were all closer to 1, which implies that the prediction capability of the CA-AQD and Chameleon models was good from the principle that the landslide prediction model's overall accuracy increased as the AUC value moved closer to 1 [34].

Moreover, to compare the performance of landslides prediction models between classification (supervised) and clustering analysis algorithms (unsupervised), accuracy measure was used to validate the DTU, NBU, SVM, ANN, RF, and CA-AQD algorithms, as shown in Figure 11. Landslide susceptibility map based on DTU and NBU algorithms can be found in the literature [33, 37], in which the 506 points (landslides) were used to develop and evaluate DTU and NBU prediction model. Thus, we divided the 506 points from the literature into training and test data sets. We used 101 points (20% of the data) for the training data set

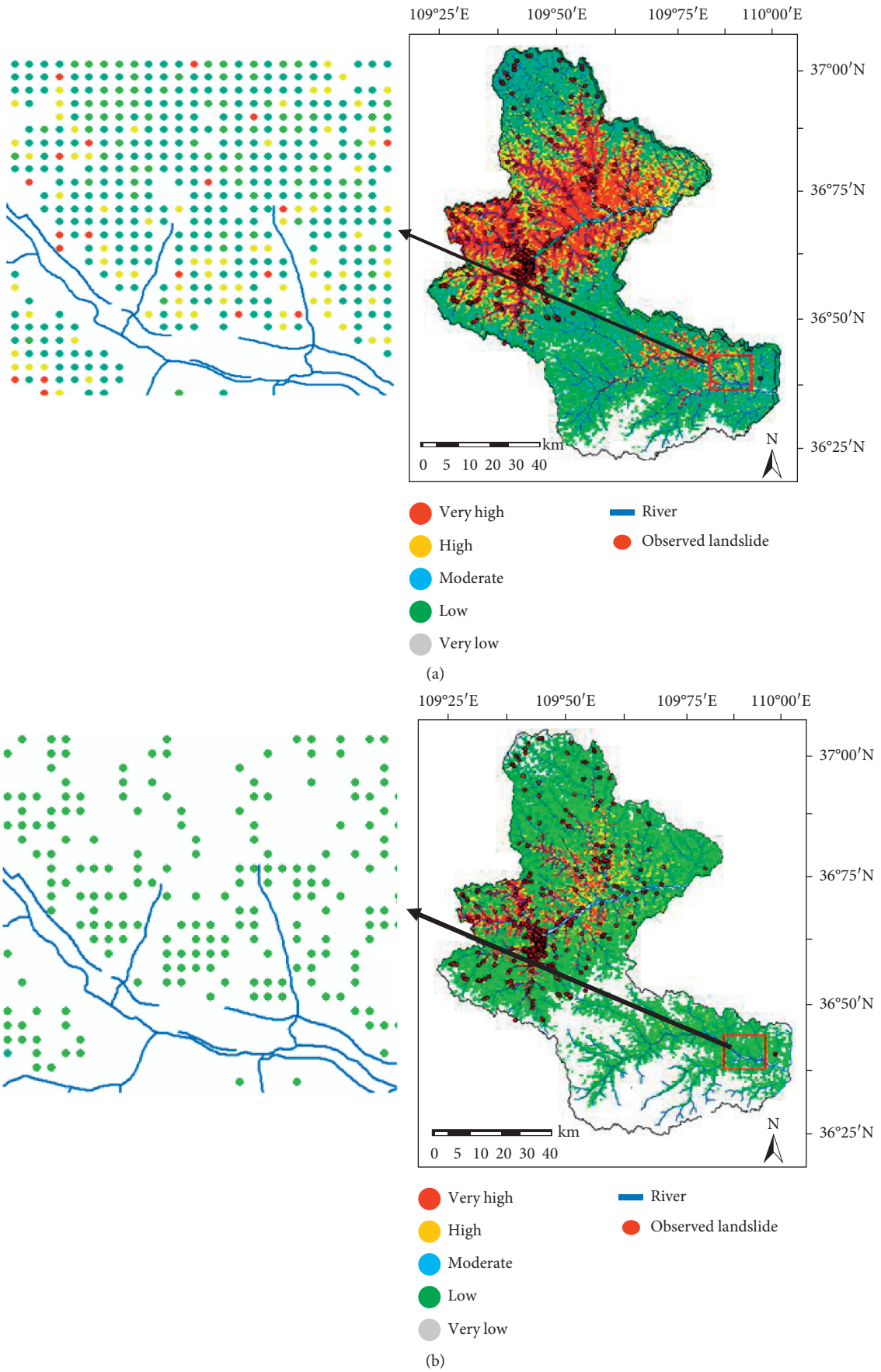


Figure 9: Assessment map of landslide susceptibility based on the CA-AQD algorithm: (a) April and (b) July.

Table 4: The accuracy and Cohen kappa index among CA-AQD, Chameleon, and KPSO.

Models	TP	TN	FP	FN	Sensitivity	Specificity	Accuracy	$F1$-measure	k
CA-AQD	268	200	13	25	0.9147	0.9390	0.9249	0.9341	0.8471
Chameleon	263	198	15	30	0.8976	0.9296	0.9110	0.9212	0.8192
KPSO	197	138	75	96	0.6724	0.6479	0.6621	0.6974	0.3161

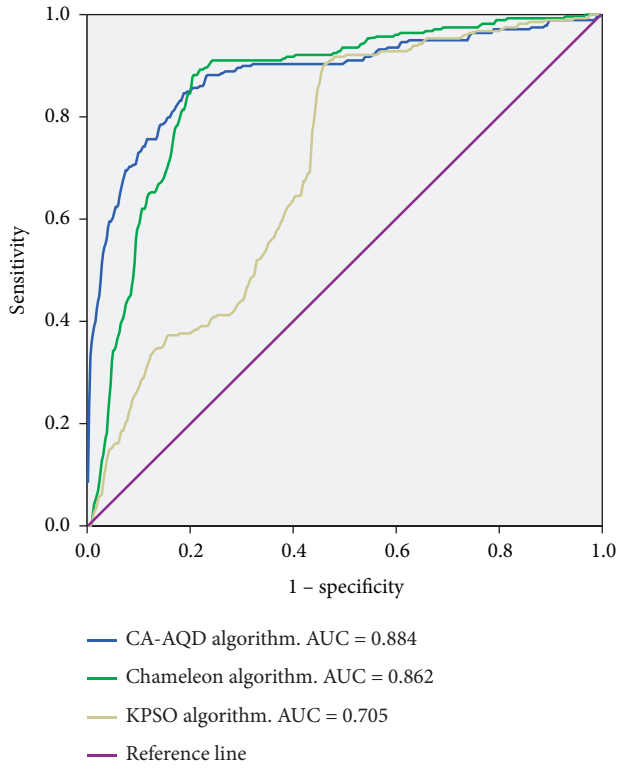


Figure 10: ROC for CA-AQD, Chameleon, and KPSO models.

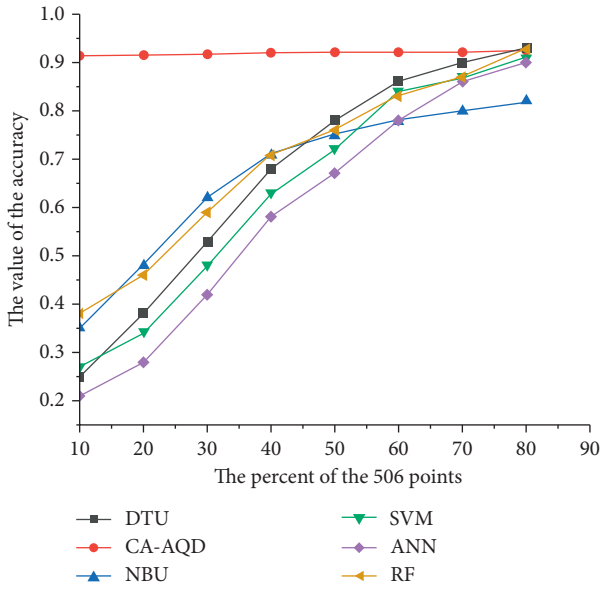


Figure 11: Comparison of the performance accuracy among CA-AQD, DTU, NBU, SVN, ANN, and RF algorithms.

and used 405 points (80% of the data) to evaluate the accuracy of the three prediction models. We then added 10% of the data to the training data set and reduced 10% of the data in the test data set. The training data set had accounted for 80% of the data (405 points). At the same time, SVM, ANN, and RF have used the above training and test data set.

As shown in Figure 10, the larger the training data set, the higher the accuracy and k of the Cohen kappa index in the DTU model, which was followed by the CA-AQD model and the NBU model. As the percentage of total points increased, the k value and accuracy of DTU and NBU algorithm are also increased—in particular, these values were quite low for a few of the points, which indicated the strong dependence on the landslide sample data set to obtain higher prediction accuracy. With increments in the percentage of total points, the k value and accuracy of CA-AQD remained almost the same, which indicated the lack of dependence on the training data set to reach a higher accuracy of prediction.

5. Discussion

Landslides are very complicated processes which are constrained by various topographical as well as environmental factors. In addition to that, the landslide susceptibility model is a very significant visual way for determining the landslide-prone area. Thus, the primary aim of this work is to use the CA-AQD algorithm for assessing landslide susceptibility and compare its performance with that of the chameleon and KPSO clustering algorithms; as well as the DTU and NBU, SVM, ANN, and RF classification algorithms in Baota District, China.

To evaluate the validity of the models, statistical metrics, as well as the AUC-ROC curve, were applied. The results indicated that the CA-AQD and Chameleon models outperformed the KPSO model in assessing landslide susceptibility in the study area, and more reliable landslide susceptibility maps were produced as they showed AUC values are closer to 1. These results suggest that both models are good in classifying well the mapping units to their respective clusters. This is due to their ability to perform well in the large study area as well as detecting well the arbitrary shaped and sized clusters, and efficient handling of noise data, which cannot be carried out well by the KPSO model. Moreover, the CA-AQD model achieved the best performance as compared to the Chameleon and KPSO models in assessing landslide susceptibility. The assessment is significant as the CA-AQD is an improved version of the Chameleon algorithm, which has been based on improving the performance accuracy by taking into consideration the uncertain data processing, which has a significant effect on the clustering results and thus makes it more prominent than others.

On the other side, in comparison with the classification algorithms, the performance accuracy of the DTU, NBU,

SVM, ANN, and RF depended much on the training data sets during the experiments (which are in fact not easy to collect and prepare). This is to say, the performance accuracy of those models increased as there were increments in the training data, thus less training data and less performance accuracy. But the CA-AQD model, showed no such dependence, as it could obtain almost constant performance accuracy throughout the experiment; thus, good performance accuracy can be guaranteed.

6. Conclusion

This study aimed at proposing and designing an improved clustering algorithm for assessing landslide susceptibility using an integration of a Chameleon algorithm and an adaptive quadratic distance (CA-AQD algorithm). It targeted improving the prediction capacity of clustering algorithms in landslide susceptibility modelling by overcoming the limitations found in present clustering models, including strong dependence on the initial partition, noise, and outliers as well as difficulties in quantifying the triggering factors (such as rainfall/precipitation). The model was implemented in Baota District, Shaanxi province, China. The model was validated using statistical metrics as well as AUC-ROC. It was then compared with the Chameleon algorithm and KPSO clustering algorithms, as well as DTU, NBU, SVM, ANN, and RF algorithms. The results suggested that the CA-AQD model achieved the best performance in comparison with the other algorithms in assessing landslide susceptibility in the area. Thus, this work adds to the literature by introducing the first empirical integration and application of the CA-AQD algorithm to the assessment of landslides in the study area, which then is a new insight to the field. Also, the method can be helpful for dealing with landslides for better social and economic development.

Data Availability

The data used in this paper were taken from the Xi'an Center for Geological Survey (CGS)

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the National Key Research and Development Program of China (2018YFC1504700) and National Natural Science Foundation of China (41562019).

References

[1] F. Guzzetti, "Forecasting natural hazards, performance of scientists, ethics, and the need for transparency," *Toxicological & Environmental Chemistry Reviews*, vol. 98, pp. 1043–1059, 2016.

[2] X. Z. Chen, H. Chen, and Y. You, "Weights-of-evidence method based on GIS for assessing susceptibility to debris flows in Kangding County, Sichuan Province, China," *Environmental Earth Sciences*, vol. 75, pp. 70–88, 2016.

[3] H. J. Feng, J. J. Yu, J. L. Zheng, and X. M. Tang, "Evaluation of different models in rainfall-triggered landslide susceptibility mapping: a case study in Chunan, southeast China," *Environmental Earth Science*, vol. 75, pp. 1399–1414, 2016.

[4] Y. L. Chen, D. H. Chen, L. Z. Chun, and J. B. Huang, "Preliminary studies on the dynamic prediction method of rainfall-triggered landslide," *Geological Review*, vol. 13, pp. 1735–1745, 2016.

[5] W. Chen, Z. Sun, X. Zhao, X. Lei, A. Shirzadi, and H. Shahabi, "Performance evaluation and comparison of bivariate statistical-based artificial intelligence algorithms for spatial prediction of landslides," *ISPRS International Journal of Geo-Information*, vol. 9, no. 12, pp. 696–713, 2020.

[6] VH. Nhu, A. Mohammadi, H. Shahabi et al., "Landslide susceptibility mapping using machine learning algorithms and remote sensing data in a tropical environment," *International Journal of Environmental Research and Public Health*, vol. 17, no. 14, pp. 4933–4950, 2020.

[7] B. Pradhan, "Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches," *Journal of the Indian Society of Remote Sensing*, vol. 38, pp. 301–320, 2010.

[8] K. X. Zhang, X. L. Wu, R. Q. Niu, and K. Yang, "The assessment of landslide susceptibility mapping using random forest decision tree methods in the Three Gorges Reservoir area, China," *Environmental Earth Science*, vol. 76, pp. 6731–6751, 2017.

[9] M. Bordoni and Y. Galanti, "The influence of the inventory on the determination of the rainfall-induced shallow landslides susceptibility using generalized additive models," *Catena*, vol. 193, pp. 243–264, 2020.

[10] D. N. Mariano and C. Francesco, "Machine learning ensemble modelling as a tool to improve landslide susceptibility mapping reliability," *Landslides*, vol. 17, pp. 1897–1914, 2020.

[11] F. Djeddaoui, M. Chadli, and R. Gloaguen, "Desertification susceptibility mapping using logistic regression analysis in the Djelfa area, Algeria," *Romote Sensing*, vol. 9, pp. 32–48, 2017.

[12] R. Schlogel, I. Marchesini, and M. Alvioli, "Optimizing landslide susceptibility zonation: effects of DEM spatial resolution and slope unit delineation on logistic regression models," *Geomorphology*, vol. 301, pp. 10–20, 2018.

[13] S. Sunil, S. Man, K. Mukhejee, A. Arabemeri, P. T. T. Ngo, and G. C. Paul, "Predicting the deforestation probability using the binary logistic regression, random forest, ensemble rotational forest, REPTree: a case study at the Gumani River Basin, India," *Science of the Total Environment*, vol. 730, pp. 245–261, 2020.

[14] I. N. Aghdam, B. Pradhan, and M. Panahi, "Landslide susceptibility assessment using a novel hybrid model of statistical bivariate methods (FR and WOE) and adaptive neuro-fuzzy inference system (ANFIS) at southern Zagros Mountains in Iran," *Environmental Earth Science*, vol. 76, pp. 237–256, 2017.

[15] R. Pratap and G. Vikram, "Landslide susceptibility mapping using bivariate statistical method for the hilly township of Mussoorie and its surrounding areas, Uttarakhand Himalaya," *Journal of Earth System Science*, vol. 129, pp. 56–72, 2020.

[16] O. Rahmati, A. Haghizadeh, and H. R. Pourghasemi, "Gully erosion susceptibility mapping: the role of GIS-based bivariate statistical models and their comparison," *Natural Hazards*, vol. 82, pp. 1231–1258, 2016.

[17] M. H. Ghobadi, M. Nouri, and B. Saedi, "The performance evaluation of information value, density, LNRF, and

frequency ratio methods for landslide zonation at Miandarband area, Kermanshah province, Iran," *Arabian Journal of Geosciences*, vol. 10, pp. 430–451, 2017.

[18] L. P. Li, H. X. Lan, and C. B. Guo, "A modified frequency ratio method for landslide susceptibility assessment," *Landslides*, vol. 14, pp. 727–741, 2016.

[19] S. Mondal and S. Mandal, "Application of frequency ratio (FR) model in spatial prediction of landslides in the Balason river basin, Darjeeling Himalaya," *Spatial Information Research*, vol. 25, pp. 1–14, 2017.

[20] K. Singh and V. Kumar, "Landslide hazard mapping along national highway-154A in Himachal Pradesh, India using information value and frequency ratio," *Arabian Journal of Geosciences*, vol. 10, p. 539, 2017.

[21] Y. Wang, D. L. Sun, and H. J. Wen, "Comparison of random forest model and frequency ratio model for landslide susceptibility mapping (LSM) in Yunyang County (Chongqing, China)," *International Journal of Environmental Research and Public*, vol. 17, no. 12, pp. 134–148, 2020.

[22] R. Poonam, R. Naresh, and K. C. Parshant, "Identification of landslide-prone zones in the geomorphically and climatically sensitive Mandakini valley, (central Himalaya), for disaster governance using the weights of evidence method," *Geomorphology*, vol. 284, pp. 41–52, 2017.

[23] S. Teerarungsigul, J. Torizin, and M. Fuchs, "An integrative approach for regional landslide susceptibility assessment using weight of evidence method: a case study of Yom River Basin, Phrae Province, Northern Thailand," *Landslides*, vol. 13, pp. 1151–1165, 2016.

[24] J. Torizin, "Elimination of informational redundancy in the weight of evidence method: an application to landslide susceptibility assessment," *Stochastic Environmental Research and Risk Assessment*, vol. 30, pp. 635–651, 2016.

[25] B. D. Tien, A. Shirzadi, H. Shahabi et al., "New ensemble models for shallow landslide susceptibility modeling in a semi-arid watershed," *Forests*, vol. 10, no. 9, pp. 743–760, 2019.

[26] B. Gordan, D. J. Armaghani, M. Hajihassani, and M. Monjezi, "Prediction of seismic slope stability through combination of particle swarm optimization and neural network," *Engineering with Computers*, vol. 32, pp. 85–97, 2016.

[27] S. Mehebub and R. Sufia, "Exploring effectiveness of frequency ratio and support vector machine models in storm surge flood susceptibility assessment: a study of Sundarban Biosphere Reserve, India," *Catena*, vol. 189, pp. 89–104, 2020.

[28] V. H. Nhu, D. Zandi, H. Shahabi et al., "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," *Applied Sciences*, vol. 10, no. 15, pp. 5047–5061, 2020.

[29] M. S. Roodposhti, T. Safarrad, and H. Shahabi, "Drought sensitivity mapping using two one-class support vector machine algorithms," *Atmospheric Research*, vol. 193, pp. 73–82, 2017.

[30] H. Y. Hong, J. Z. Liu, and T. B. Dieu, "Landslide susceptibility mapping using J48 decision tree with AdaBoost, bagging and rotation forest ensembles in the guangchang area (China)," *Catena*, vol. 163, pp. 399–413, 2018.

[31] H. Y. Hong and J. Z. Liu, "Modeling landslide susceptibility using LogitBoost alternating decision trees and forest by penalizing attributes with the bagging ensemble," *Science of the Total Environment*, vol. 718, pp. 245–261, 2020.

[32] P. R. Kadavi, C. W. Lee, and S. Lee, "Landslide-susceptibility mapping in Gangwon-do, South Korea, using logistic regression and decision tree models," *Environmental Earth Science*, vol. 78, 2019.

[33] Y. M. Mao, M. S. Zhang, P. P. Sun, and G. L. Wang, "Landslide susceptibility assessment using uncertain decision tree model in loess areas," *Environmental Earth Science*, vol. 76, pp. 752–770, 2017.

[34] B. Pradhan, "A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS," *Computers & Geosciences*, vol. 51, pp. 350–365, 2013.

[35] M. S. Tehran, B. Pradhan, and M. N. Jebur, "Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS," *Journal of Hydrology*, vol. 504, pp. 69–79, 2013.

[36] W. Chen, X. S. Yan, Z. Zhao, H. Y. Hong, D. T. Bui, and B. Pradhan, "Spatial prediction of landslide susceptibility using data mining-based kernel logistic regression, naive Bayes and RBFNetwork models for the Long County area (China)," *Bulletin of Engineering Geology and the Environment*, vol. 78, pp. 247–266, 2019.

[37] Y. M. Mao and M. S. Zhang, "Landslide hazards mapping using uncertain Naïve Bayesian classification method," *Journal of Central South University*, vol. 22, pp. 3512–3520, 2015.

[38] T. P. Binh, T. B. Dieu, and I. Prakash, "Hybrid integration of multilayer perceptron neural networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS," *Catena*, vol. 149, pp. 52–63, 2017.

[39] R. X. Tang and H. S. Kulatilake, "Evaluating landslide susceptibility based on cluster analysis, probabilistic methods, and artificial neural networks," *Bulletin of Engineering Geology and the Environment*, vol. 79, pp. 2235–2254, 2020.

[40] Y. Wang, Z. Fang, and H. Hong, "A comparative study of composite kernels for landslide susceptibility mapping: a case study in Yongxin Couty, China," *Catena*, vol. 183, Article ID 104217, 2019.

[41] Y. Wang, Z. Fang, and H. Hong, "Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China," *Science of the Total Environment*, vol. 666, pp. 975–993, 2019.

[42] F. Chen, B. Yu, and B. Li, "A practical trial of landslide detection from single-temporal Landsat8 images using contour-based proposals and random forest: a case study of national Nepal," *Landslides*, vol. 15, no. 3, pp. 453–464, 2018.

[43] T. Chen, L. Zhu, R.-Q. Niu, and J. C. Trinder, "Mapping landslide susceptibility at the Three Gorges Reservoir, China, using gradient boosting decision tree, random forest and information value models," *Journal of Mountain Science*, vol. 17, no. 3, pp. 670–685, 2020.

[44] J. Hu, K. B. Xu, G. L. Wang et al., "A novel landslide susceptibility mapping portrayed by OA-HD and K-medoids clustering algorithms," *Bulletin of Engineering Geology and the Environment*, vol. 80, no. 2, pp. 765–779, 2020.

[45] Q. Q. Ba, Y. M. Chen, and S. S. Deng, "An improved information value model based on gray clustering for landslide susceptibility mapping," *International Journal of Geo-Information*, vol. 6, pp. 267–284, 2017.

[46] Q. Wang, Y. Wang, and R. Q. Niu, "Integration of information theory, k-means cluster analysis and the logistic regression model for landslide susceptibility mapping in the three gorges area, China," *Remote Sensing*, vol. 9, p. 938, 2017.

[47] S. Wan, "Construction of knowledge-based spatial decision support system for landslide mapping using fuzzy clustering and KPSO analysis," *Arabian Journal of Geosciences*, vol. 8, pp. 1041–1055, 2015.
[48] S. Wan, "Entropy-based particle swarm optimization with clustering analysis on landslide susceptibility mapping," *Environmental Earth Science*, vol. 68, pp. 1349–1366, 2013.
[49] M. Ankerst, M. M. Breunig, H. P. Kriegel, and J. S. Sander, "OPTICS: ordering points to identify the clustering structure," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, ACM SIGMOD Record, Philadelphia PA, USA, June 1999.
[50] N. Kinoshita and Y. Endo, "EM-based clustering algorithm for uncertain data," *Knowledge and Systems Engineering*, vol. 245, pp. 69–81, 2014.
[51] G. Karypia, E. H. Han, and V. Kumar, "Chameleon: a hierarchical clustering algorithm using dynamic modeling," *Computer*, vol. 32, pp. 68–75, 1999.
[52] De. Carvalho, A. T. Fd, and C. Tenorio, "Fuzzy k-means clustering algorithms for interval-valued data based on adaptive quadratic distance," *Fuzzy Sets and Systems*, vol. 161, pp. 2978–2999, 2010.
[53] M. Zeraatpisheh, S. Ayoubi, C. W. Brungard, and P. Finke, "Disaggregating and updating a legacy soil map using DSMART, fuzzy c-means and k-means clustering algorithms in Central Iran," *Geoderma*, vol. 340, pp. 249–258, 2019.
[54] M. S. Zhang and J. Liu, "Controlling factors of loess landslide in western China," *Environmental Earth Sciences*, vol. 59, pp. 1671–1680, 2010.
[55] F. K. Hoehler, "Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity," *Journal of Clinical Epidemiology*, vol. 53, pp. 499–503, 2000.
[56] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, pp. 159–174, 1977.