

Research Article

Mapping Landslide Sensitivity Based on Machine Learning: A Case Study in Ankang City, Shaanxi Province, China

Baoxin Zhao ¹, Jingzhong Zhu ², Youbiao Hu ¹, Qimeng Liu ¹ and Yu Liu ³

¹School of Earth and Environment, Anhui University of Science and Technology, Huainan 232001, China

²School of Resources and Geosciences, China University of Mining and Technology, Xuzhou 221116, China

³State Key Lab Mining Response & Disaster Prevention & Control, Anhui University of Science and Technology, Huainan 232001, China

Correspondence should be addressed to Jingzhong Zhu; algzjz0801@126.com and Youbiao Hu; ybhu@aust.edu.cn

Received 4 April 2022; Revised 25 June 2022; Accepted 13 August 2022; Published 8 September 2022

Academic Editor: Constantinos Loupasakis

Copyright © 2022 Baoxin Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The main purpose of this research is to apply the logistic regression (LR) model, the support vector machine (SVM) model based on radial basis function, the random forest (RF) model, and the coupled model of the whale optimization algorithm (WOA) and genetic algorithm (GA) with RF, to make landslide susceptibility mapping for the Ankang City of Shaanxi Province, China. To this end, a landslide inventory map consisting of 4278 identified landslides is randomly divided into training and test landslides in a ratio of 7:3. The 15 landslide influencing factors are selected as follows: slope aspect, slope degree, elevation, terrain curvature, plane curvature, profile curvature, surface roughness, distance to faults, distance to roads, landform, lithology, distance to rivers, rainfall, stream power index (SPI), and normalized difference vegetation index (NDVI), and the potential multicollinearity problem among these factors is detected by Pearson correlation coefficient (PCC), variance inflation factor (VIF), and tolerance (TOL). We evaluate the performance of the model separately by statistical training and test dataset metrics, including sensitivity, specificity, accuracy, kappa, mean absolute error (MSE), root mean square error (RMSE), and area under the receiver operating characteristic curve. The training success rates of LR, SVM, RF, WOA-RF, and GA-RF models are 0.7546, 0.8317, 0.8561, 0.8804, and 0.8957; the testing success rates are 0.7551, 0.8375, 0.8395, 0.8348, and 0.85007. The results show that the GA significantly improves the predictive power of the RF model. This study provides a scientific reference for disaster prevention and control in this area and its surrounding areas.

1. Introduction

Landslide disaster is a common and highly destructive adverse geological phenomenon. Because of the complex and diverse terrain conditions, climate conditions, and frequent engineering geological activities in China, many kinds of adverse geological phenomena occur frequently every year. According to a public report released by the Ministry of Natural Resources in China, there were 7840 geological disasters in 2020. Landslides accounted for about 61.35%, lost \$ 161 million, and killed 139 people. The landslide disasters are mainly concentrated in the central and western regions in our country, which is the most concentrated in southwest China. Ankang City is located in southern Shaanxi and is the most severely affected area by landslides.

According to the statistics of the local government, the landslide disaster in Ankang City has caused more than 500 dead or missing, and the direct economic loss amounted to 1.07 billion US dollars since 1983. The prevention and control of such frequent and destructive landslides are one of the important issues of disaster prevention and mitigation, and the key step is the evaluation of landslide susceptibility. Although there are many kinds of research results, the analysis methods and evaluation criteria are quite different [1–3].

Due to the powerful data processing and mapping capabilities of geographic information system, it has been widely applied to draw landslide susceptibility mapping over the past decades [4]. Currently, the widely used models for predicting landslide susceptibility mainly include qualitative,

quantitative, and artificial intelligence attribution analysis [5, 6]. The use of quantitative attribution analysis model accounted for the main research model. Qualitative attribution analysis is highly correlated with researchers with knowledge reserves and subjectivity, resulting in a large difference in their effects [7], such as analytic hierarchy process (AHP) and Entropy weight method [8]; other statistical models and the coupling models are also widely applied, such as frequency ratio model [9–11], information value model [12, 13], and evidence weight model [14–16]. However, the results of these models are not ideal.

In recent years, with the innovation of artificial intelligence algorithms and the improvement of data processing capabilities, machine learning methods have been widely studied in the field of landslide susceptibility. The machine learning model can handle the geological, hydrological, and other information datasets of high-latitude and large datasets and has higher prediction success rate [17–19]. Common models include random forest (RF), gradient boosting machine (GBM), support vector machine (SVM), artificial neural network (ANN), and logistic regression (LR) [20–24]. Among them, RF, which is widely employed, has achieved remarkable results in model classification and accuracy. For example, Yu and Chen used information model, AHP, and RF method to predict Helong City landslide susceptibility; the effect of RF is obvious [25]. Compared with other common machine learning models, some scholars selected the most suitable algorithm for the prediction results of their respective research areas [26, 27]. The comparison shows that the accuracy of RF training and testing is relatively high. It is proved that the accuracy of the model can be improved by coupling RF with other models. Other traditional machine learning algorithms have been widely applied, such as logistic regression model. Accuracy is widely used in landslide susceptibility evaluation because it can directly generate decision boundaries from original datasets [28]. The advantage of SVM lies in the ability to eliminate overfitting and noise problems in the modeling process, and it has a wide range of practicability [29]. With the introduction of deep learning model, it uses more hidden layers to model the complex relationship between data, and depth replaces breadth, improves fitting degree, and achieves higher precision. Compared with the three baseline models (SVM, NBTree, and REPTree), the GA-optimized deep learning model based on ELM, DBN, and BP exhibits better prediction performance by performing hierarchical analysis on the original dataset to extract the most relevant features [30, 31]. Additionally, the comparison and coupling between models can also further improve the accuracy [32, 33]. However, a large number of research results show that which model is the most suitable for landslide sensitivity evaluation has not been determined.

According to the current research situation of landslide sensitivity analysis, statistical analysis of the temporal and spatial distribution of factors such as rainfall, geomorphology, and earthquake in historical landslide events shows that there is a great correlation between the occurrence of landslides and rainfall [34, 35]. Furthermore, some scholars have also analyzed and compared landslide susceptibility through qualitative and quantitative aspects in Shaanxi Province,

such as FR-AHP [36], certainty factor (CF), and index of entropy (IOE) [37] model. In addition, compared with other models, SVM model shows better prediction results [38]. However, there is no further study on quantitative analysis.

Firstly, this paper screened the influencing factors of landslide by two methods—Pearson correlation coefficient and multicollinearity. The purpose of the study is to investigate the potential application of the optimized RF model by whale optimization algorithm (WOA) and genetic algorithm (GA) in landslide susceptibility analysis in Ankang City and compare and analyze with LR, SVM, and RF algorithm. In the study area, the optimization of RF models with genetic algorithm and whale optimization algorithm is relatively novel. Analyzing various statistical indicators (sensitivity, specificity, accuracy, kappa index, mean absolute error, and root mean squared error), each model is evaluated and analyzed, and GA-RF is the optimal model.

2. Study Area

The study area is located in the southeast of Shaanxi Province, China, with a range of 31°42′N~33°49′N and 108°01′E~110°01′E, belonging to the humid monsoon climate of northern subtropical continent. Generally, Ankang basin is in the middle of two mountains and one river. The north is on Qinling Mountains, and the south is on Bashan Mountains. Hanshui-Chihe-Yuehe-Hanshui are the boundary between Qinling Mountains and Bashan Mountains.

The resident population of the city is more than 24.9 million; the city is composed of nine counties and one district, including Hanbin District, Hanyin County, Shiquan County, Ningshan County, Ziyang County, Langao County, Pingli County, Zhenping County, Xunyang County, and Baihe County. The width of city is about 200 km from east to west, and the length is about 240 km from north to south; the total area is about 23529 km². The elevation of the study area is 97~2897 m, the terrain is undulating, the central terrain is low, and the northern and southern terrain is high.

Ankang City is a subtropical continental monsoon climate with annual average temperature of 15~17°C and annual precipitation of about 1050 mm. As the longest tributary of the Yangtze River, the Hanshui River is about 20 km across Ankang City, and the average annual runoff is 2.01×10^{11} m³. The flow of river from July to October accounts for about 72.2% of total annual flows.

Landslide event refers to the phenomenon that the mechanical balance of rock or soil disappears and falls down the connected shear surface due to factors such as rainfall, slope overflow erosion, and earthquake. The tectonic structure of the study area is composed of the Qinling Indo-China fold belt and the Caledonian fold belt of the Daba Mountain. Sedimentary environment varies greatly; the area of sedimentary rocks is 60.3% of the total area in Ankang City. Besides, a great number of folds and faults are developed in the area, which are the material basis of geological disasters such as landslides and collapses. Therefore, landslides occur frequently in Ankang City, and the disaster points are up to 4278, of which the central region is densely distributed, as shown in Figure 1.

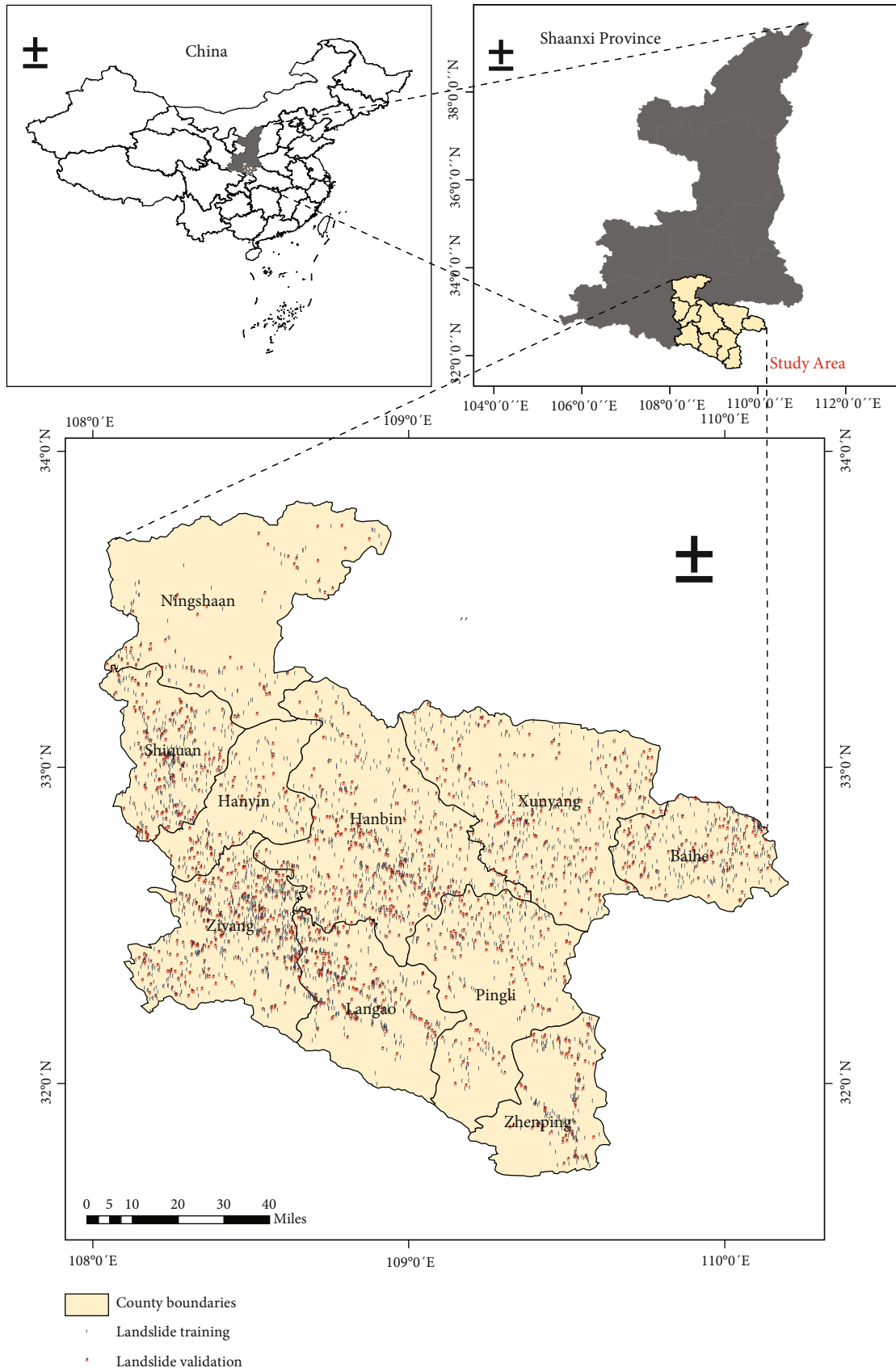


FIGURE 1: The location of the research area and landslide points.

TABLE 1: Data and data sources.

Data name	Data sources	Scale or resolution	Type
DEM	Geospatial data cloud platform	30 m × 30 m	Grid
Landsat8 oli image	Geospatial data cloud platform	30 m × 30 m	Grid
Rivers	DEM	30 m × 30 m	Grid
Roads	Shaanxi Provincial Ministry of Transportation	1 : 100000	Vector
Lithology and faults	China Geological Survey (2002)	1 : 200000	Grid
Annual rainfall	Ankang Meteorological Bureau	30 m × 30 m	Datasheet

3. Datasets

3.1. Landslide Inventory Map. The landslide inventory map is used to determine the number, location, and type of landslides, the accuracy of which affects the results of the landslide susceptibility analysis [39]. The landslide distribution data are from the Chinese Academy of Sciences' Data Center for Resource and Environmental Sciences, and the landslides are individually screened and studied through historical records analysis and satellite images. The 4278 landslide points are distributed in Ankang City. The landslide points are randomly divided into two types of datasets for training (70%) and verification (30%).

3.2. Landslide Conditioning Factors. Different factors and data classification standards in the same region will affect the output results of the model to varying degrees. The selection of landslide factors directly affects the success rate of landslide sensitivity model, because the number of landslide factors and the collinearity between factors will affect the prediction ability of the model. On the premise of collecting 4278 landslide events, according to the human geography and environmental characteristics of Ankang City and the previous research results, 15 landslide factors are collected as follows: slope aspect, slope degree, elevation, terrain curvature, plane curvature, profile curvature, surface roughness, distance to faults, distance to roads, landform, lithology, distance to rivers, rainfall, stream power index (SPI), and normalized difference vegetation index (NDVI).

The digital elevation model (DEM) with 30 × 30 m cell size in the study area is obtained from the 30-m resolution ASTER GDEM (<http://www.gscloud.cn>). Elevation, slope aspect, slope degree, surface roughness, terrain curvature, SPI, plan curvature, profile curvature, and distance to rivers are essentially extracted from the study area DEM data. NDVI is obtained from Landsat 8 OLI image with a resolution of 30 × 30 m. It is processed in ENVI through data reading, radiometric calibration, image clipping, atmospheric correction, and image mosaic. The meteorological data are collected from Shaanxi Meteorological Service. Furthermore, other factors based on the data collected are described and further vectorized by topographic map and geological map.

Combined with the results of previous studies on topography, hydrogeological conditions and human engineering activities. The specific influencing factors include (1) topography factors: elevation, slope aspect, slope degree, terrain curva-

ture, plan curvature, profile curvature, distance to faults, surface roughness, landform, and lithology; (2) geological and hydrogeological factors: distance to rivers, rainfall, and SPI; and (3) human engineering activities factors: NDVI and distance to roads. After preprocessing factors data, the basic unit of data in the study area is set to 30 m × 30 m. The sources of the various factors are shown in Table 1.

3.2.1. Topographic Factors. Elevation is a classical landslide factor extracted directly from DEM data. Ankang City is a mountainous area, which is between the Daba Mountains and Qinling Mountains as a whole, and the central part is low. According to the geographical characteristics of the study area, it is divided into four categories: <600 m, 600~1000 m, 1000~1400 m, 1400~1800 m, and >1800 m (Figure 2(a)).

The different slope aspects show that the azimuth angles of different slopes are different in horizontal direction. The solar radiation, illumination time, temperature, rainfall, and wind speed will also be different in different slope aspects. The difference of slope weathering and erosion is reflected in the change of rock and soil structure and groundwater occurrence conditions and the interaction between sand and mudstone in different degrees [40, 41]. According to the ground slope aspects of the study area, it is divided into flat, north, north-east, east, south-east, south, south-west, west, and north-west (Figure 2(b)).

The slope degree reflects the degree of surface unit, and the stress distribution in rock and soil is different with different slope degree [42]. According to the research slope, the slope degree in the study area is divided into five categories: 0~10°, 10~20°, 20~30°, 30~40°, and >40° (Figure 2(c)).

Terrain curvature is a function of slope point, including profile curvature and plane curvature. It shows the structure and morphology of the terrain, not only reflects the aggregation and dispersion of surface flow, but also affects the acceleration of water flow, thereby affecting surface erosion and deposition [43]. Terrain curvature was extracted from DEM data in the study area using GIS surface tools such as curvature as shown in Figure 2(d). The distribution of topographic curvature in the study area is analyzed according to the classification standard of curvature value which are <0 (concave), =0 (flat), and >0 (convex), and then the distribution map of profile curvature and plane curvature is finally obtained (Figures 2(e) and 2(f)). In addition, three different classes of plane curvature values were defined as follows: <-0.5, -0.5~0.5, and >0.5.

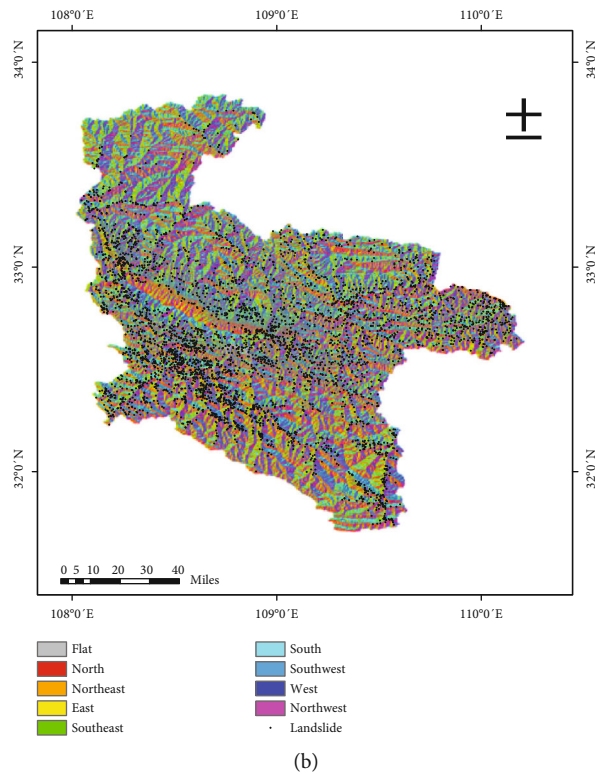
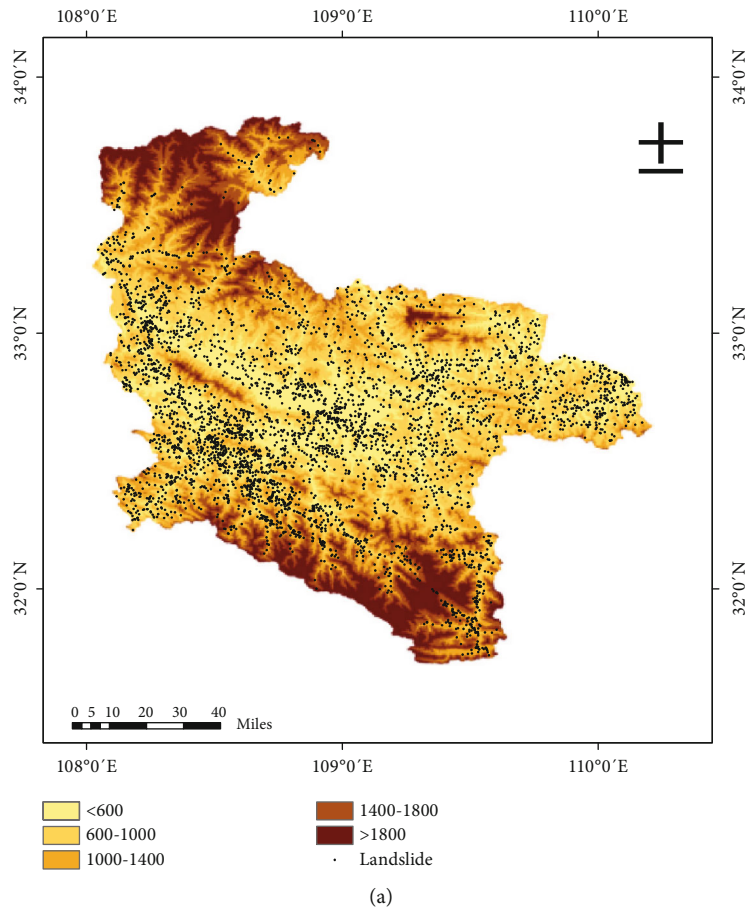


FIGURE 2: Continued.

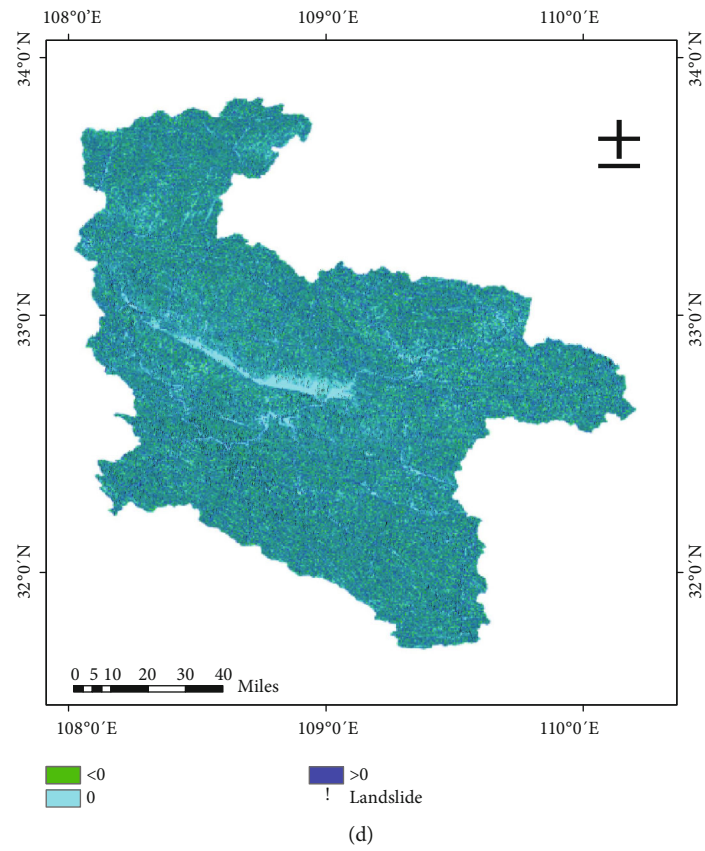
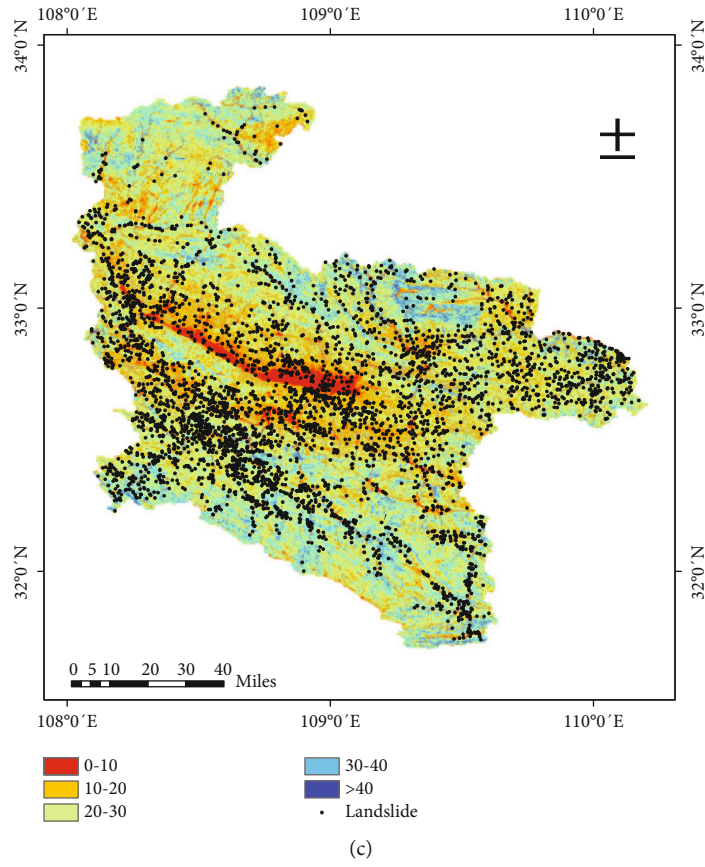
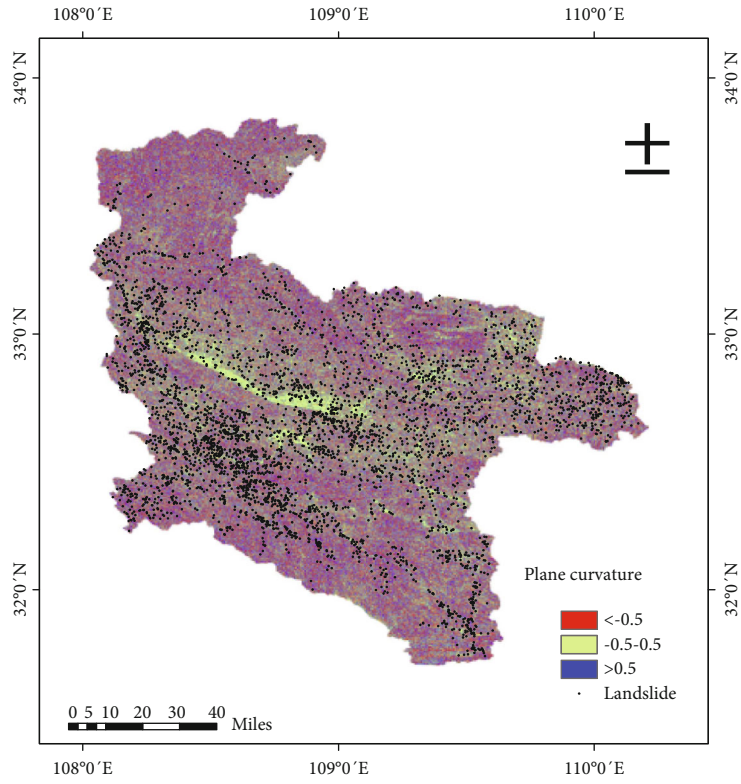
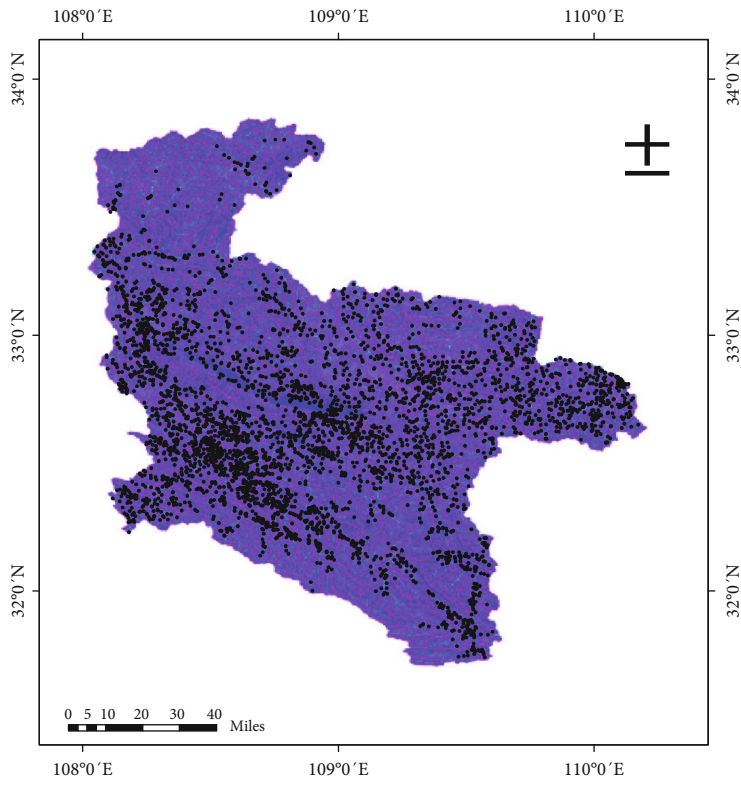


FIGURE 2: Continued.

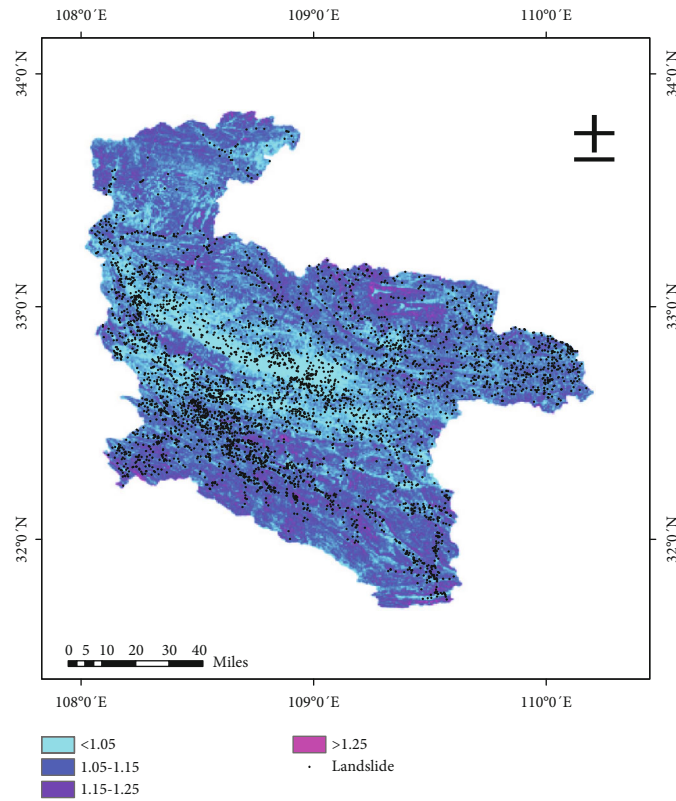


(e)

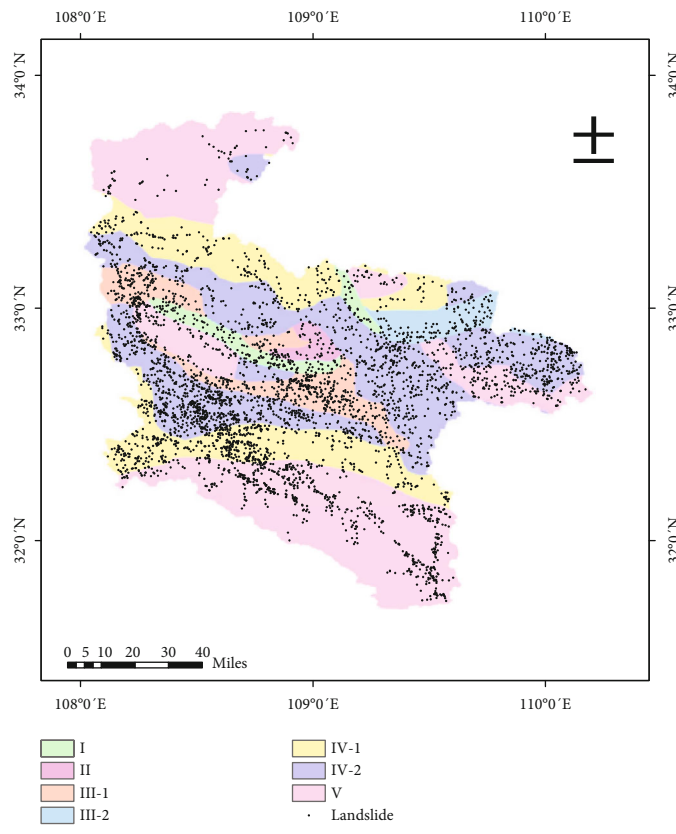


(f)

FIGURE 2: Continued.



(g)



(h)

FIGURE 2: Continued.

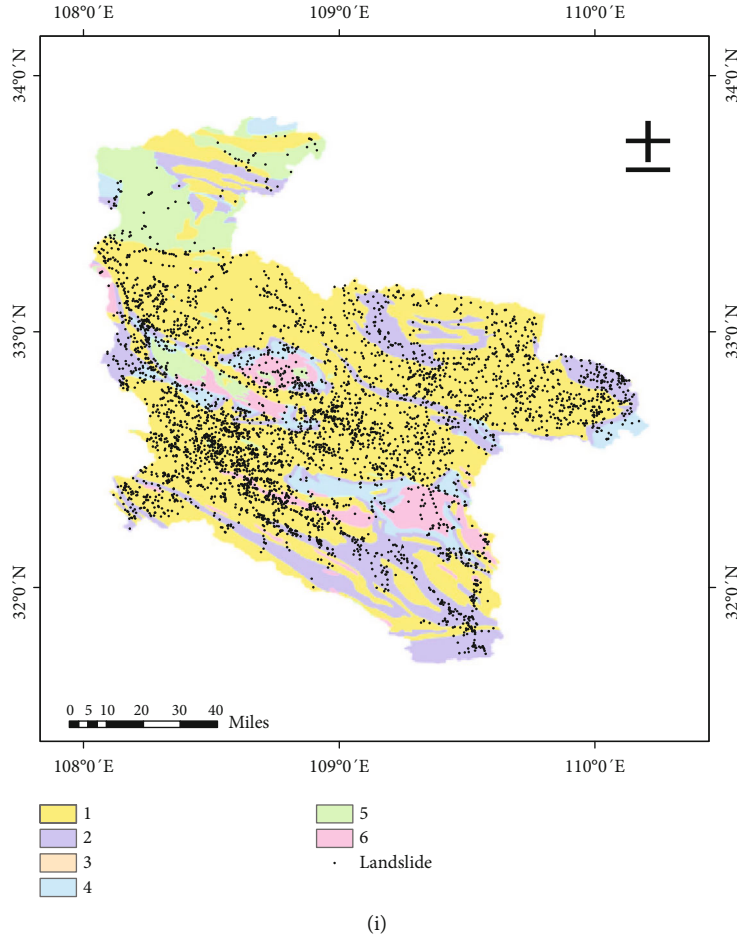


FIGURE 2: Topographic factors mapping of the study area: (a) elevation; (b) slope aspect; (c) slope degree; (d) terrain curvature; (e) plane curvature; (f) profile curvature; (g) surface roughness; (h) landform; (i) lithology.

TABLE 2: Classification of lithologic genesis.

Group	Origin and lithology
1	Sedimentary rock: sandstone, mudstone, and greywacke
2	Carbonate rock: limestone, dolomite, and marl
3	Pyroclastic rock: tuff, volcanic breccias, or ash
4	Metamorphic rock: shales, gneiss, amphibolite, quartzite
5	Intrusive rock: granites, quartz diorites, quartz monzonites, syenite, gabbro, and peridotite
6	Volcanic rock: trachytes or dacite, rhyolites, and basalts

Following surface roughness is the ratio between the slope section area and horizontal projection area in a certain area, which reflects the degree of surface fluctuation and erosion caused by solar radiation [44]. The roughness of the study area is divided into <1.05, 1.05~1.15, 1.15~1.25, and >1.25 (Figure 2(g)). Surface roughness of the study area is calculated using the “Raster Calculator” of

$$S = \frac{1}{\cos \alpha}, \quad (1)$$

where α is the slope degree (unit: radian).

The study area is mostly high mountains in the north and south direction, and the valley basin is in the middle. Different landforms lead to large morphological differences, and steep terrain affects slope stability [45]. The topography in the study area fluctuated greatly, mainly in middle and high altitude mountains, accounting for about 93.86%. Low mountains, plains, and hills accounted for 6.14%. Through investigation and analysis (Figure 2(h)), the landforms of the study area can be divided into I (plain), II (hills), III-1 (small undulating low mountains), III-2 (medium undulating low mountains), IV-1 (small undulating high mountains), IV-2 (medium undulating high mountains), and V (large undulating high mountains).

Lithology is one of the most essential factors for landslide occurrence. The rock components and structures in different parts of the same region are different, and the physical and mechanical properties of rocks are different [46]. Influenced by the same external force, the deformation of rock with different mechanical properties is also different and vice versa. Because the coverage area of sedimentary rock in the study area is about 60%, the rock strength of sedimentary rock is low, it is easy to be eroded by heavy rainfall and rivers, and it is unstable. Therefore, there are more landslides in the sedimentary rock area, accounting for about 69.65% of the total landslide events. According to the genetic classification of rock, it is divided into six categories: sedimentary rocks, carbonate rocks, igneous clastic rocks, metamorphic rock, intrusive rock, and volcanic rocks, as shown in Table 2 and Figure 2(i).

3.2.2. Geological and Hydrogeological Factors. The flow of rivers is mainly erosion and abrasion, which destroys the stress balance of dams and mountains along the rivers. Especially in areas where rivers and precipitation affect soluble rocks such as carbonate rocks, dissolution changes topography [47, 48]. The study area is divided into <1500 m, 1500~3000 m, 3000~4500 m, 4500~6000 m, and >6000 m by distance to rivers (Figure 3(a)).

The research area belongs to the northern subtropical continental monsoon climate, the average annual rainfall is 900~1000 mm, and the largest rainfall occurs in May, July, and September. Due to the large and sudden rainfall in a short time, the shear strength of the rock-soil interface on the slope is significantly reduced. Melillo et al. established an independent algorithm to determine the rainfall threshold and improve the accuracy of landslide prediction through rainfall by analyzing historical rainfall and landslide information and conducting practical verification [49]. According to the spatial distribution of rainfall in Ankang City, the regional rainfall is divided into <800 mm, 800~900 mm, 900~1000 mm, 1000~1100 mm, and >1100 mm (Figure 3(b)).

The stream power index (SPI) is related to rock lithology, particle size, and permeability. Slope surface often produces rill erosion and sediment accumulation, and when slope shear stress exceeds the shear strength of the surface, it may appear instability [50, 51]. SPI in Ankang City can be divided into three categories: <2.5, 2.5~5, and >5 (Figure 3(c)).

The study area is located in the border area between Qinling geosyncline fold system and Bashan orogenic belt, where fold and fault structures are relatively developed. Fold structural belt may produce a large number of interlayer fractures, transverse, and longitudinal tensile fractures. The developed cracks lead to rock and soil breakage, increase the water richness of rock layers, and accelerate rock erosion. In addition, a large number of secondary small faults are often developed near the large fault structural belt [52, 53]. Regional tectonics leads to the complexity of geological conditions in the study area and aggravates the probability of landslide disasters. The distance to faults as a classification reference are divided into four kinds that are <1500 m, 1500~3000 m, 3000~4500 m, 4500~6000 m, and >6000 m (Figure 3(d)).

3.2.3. Human Engineering Activities Factors. NDVI is obtained through ENVI remote sensing image data preprocessing, including radiometric calibration, atmospheric correction, and removal of outliers. Generally, the higher vegetation coverage, the greater shear strength of rock-soil interface. In areas where human activities are frequent or where rock lithology is hard, the smaller NDVI creates conditions for landslides [54]. NDVI in Ankang City is divided into five classes by natural discontinuity method: <0.18, 0.18~0.48, 0.48~0.63, 0.63~0.72, and >0.72 (Figure 4(a)).

Road is one of the indirect factors inducing landslide [55, 56]. The total highway mileage in the study area is 24500 km, the railway mileage is 900 km, and the traffic is very developed. Mountain roads may change the original diversion channel, and the continuity of rock and soil is destroyed, and some remote areas may set up less protective measures, or not timely repair, which will affect the stability of the landslide. GIS is used to make multiple buffers for roads, and the study area is classified into <1500 m, 1500~3000 m, 3000~4500 m, 4500~6000 m, and >6000 m by the distance to roads (Figure 4(b)).

4. Methodology

4.1. Modeling Process. (1) The random area safety points are created by GIS, the number of which is roughly the same as the number of landslide points, and the safety points and random points are combined together. The collected factor data are normalized, and the variables are classified using the Multipoint Extraction tool. (2) We analyzed and compared 15 landslide factors using Pearson's correlation coefficient method and multicollinearity method and finally eliminated curvature, roughness, and SPI. The remaining 12 factors (slope aspect, slope degree, elevation, plane curvature, profile curvature, distance to faults, distance to roads, landform, lithology, distance to rivers, rainfall, and NDVI) are used. The Scikit-learn library in Python is used to randomly select 2994 points (70%) and 1284 points (30%) for training and model validation of landslide points and safe points [57]. The RF, LR, and SVM models are determined after evaluating the AUC values. (3) After the coupling optimization of the GA and WOA and the RF model, each model is evaluated by comparing the values of RMSE, MSE, and AUC, and the optimal parameter combination is selected. (4) Convert the grid data of the study area into more than 30million vector points through grid-to-point and predict the information value of each factor to obtain the output probability value of the landslide in the study area. Using the machine learning algorithm in python, the probability values are rearranged into the grid size, and the probability grid data of landslide occurrence in this area is obtained. (5) Draw landslide susceptibility mapping, and analyze and discuss the performance of each model and the main control disaster factors.

4.2. Selection of Landslide Conditioning Factors. In the area of landslide sensitivity evaluation, the selection of landslide condition factors is not based on the quantity. Unnecessary factors will cause excessive spatial data and increase the

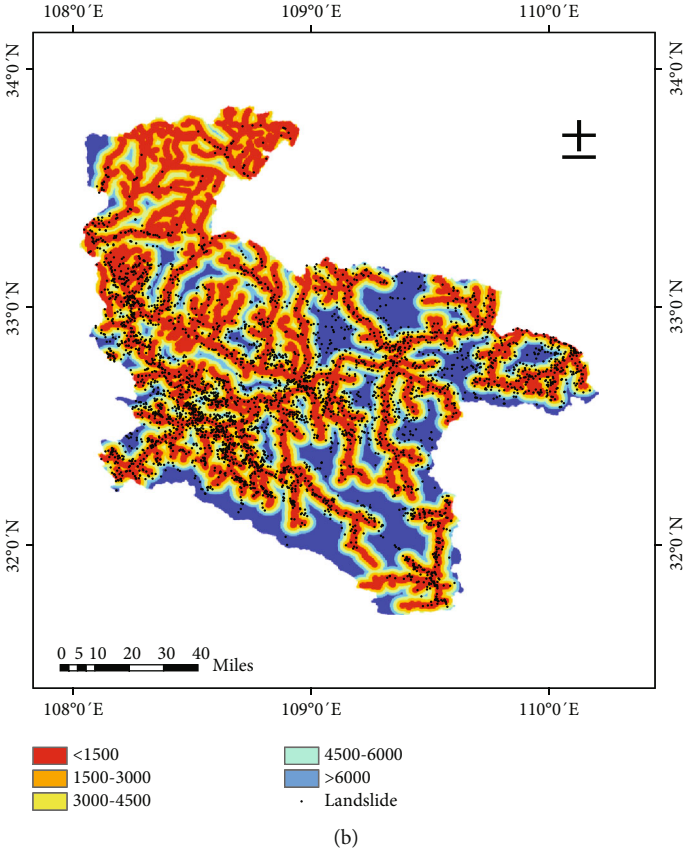
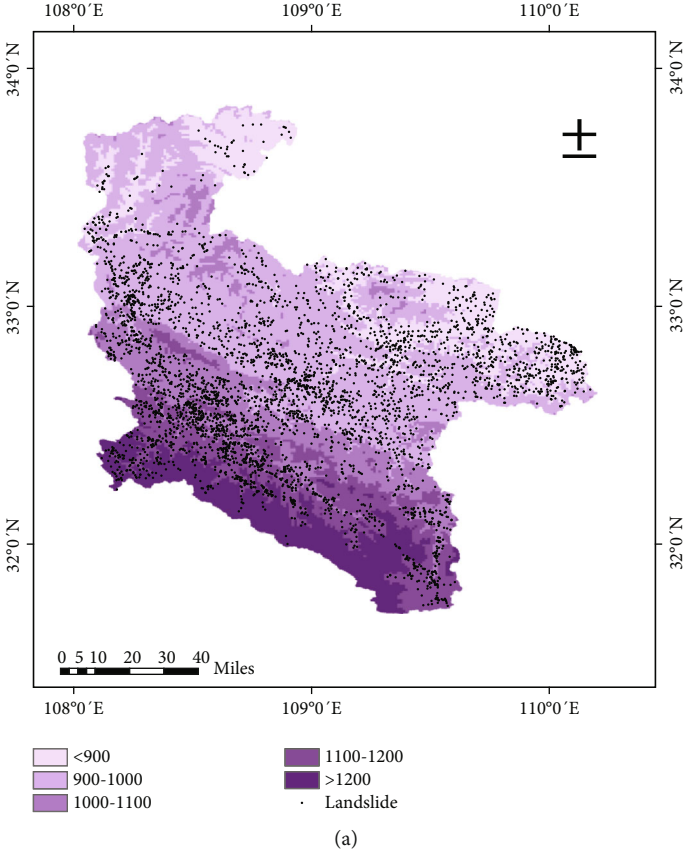


FIGURE 3: Continued.

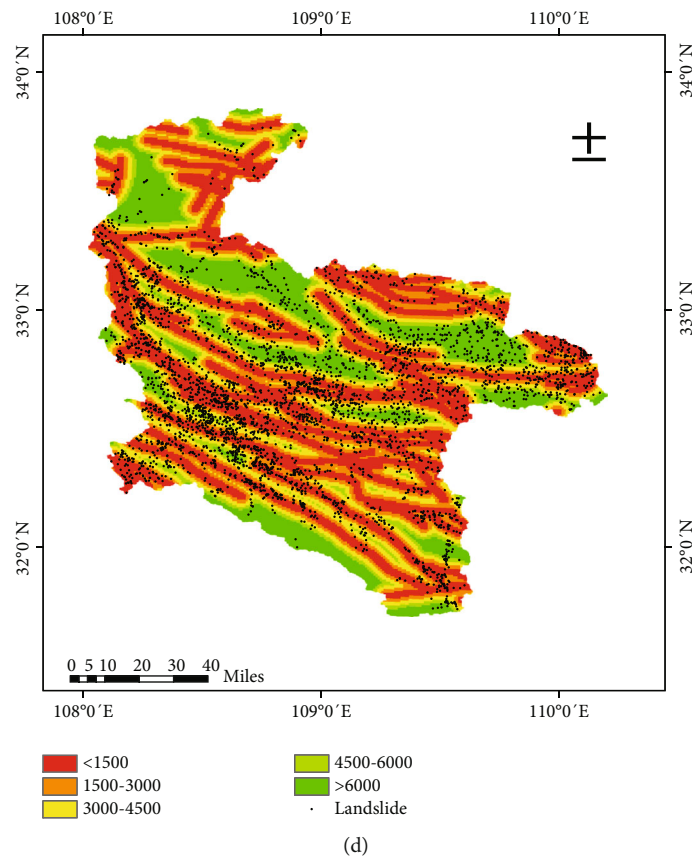
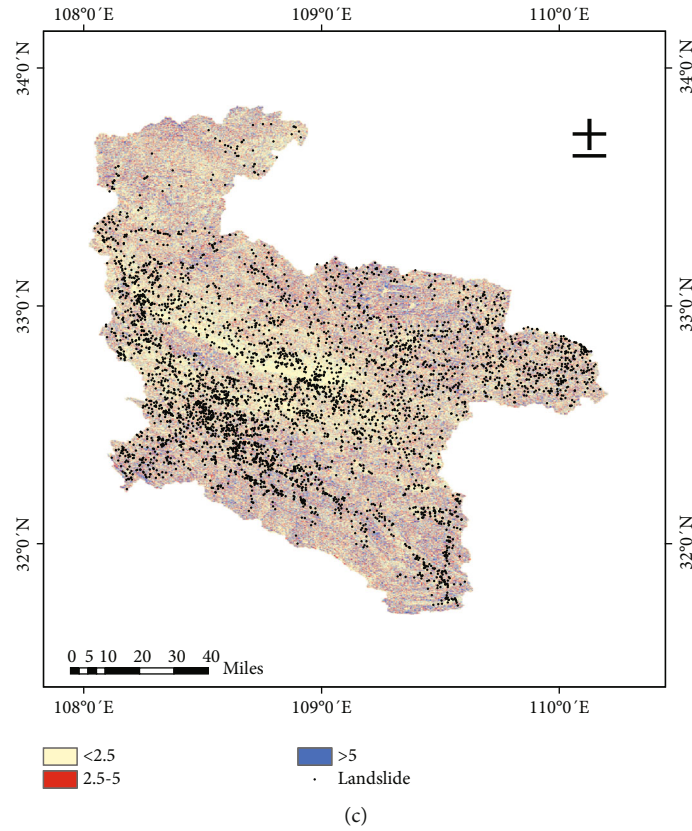
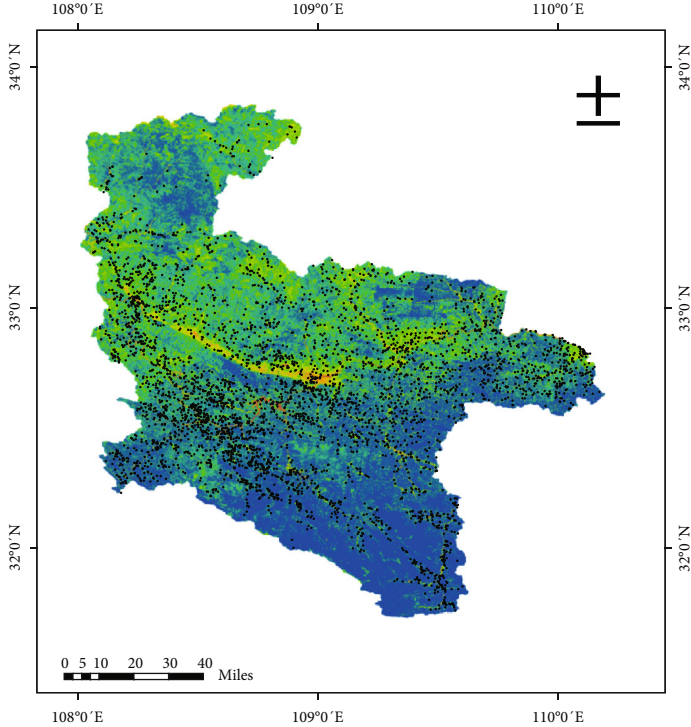
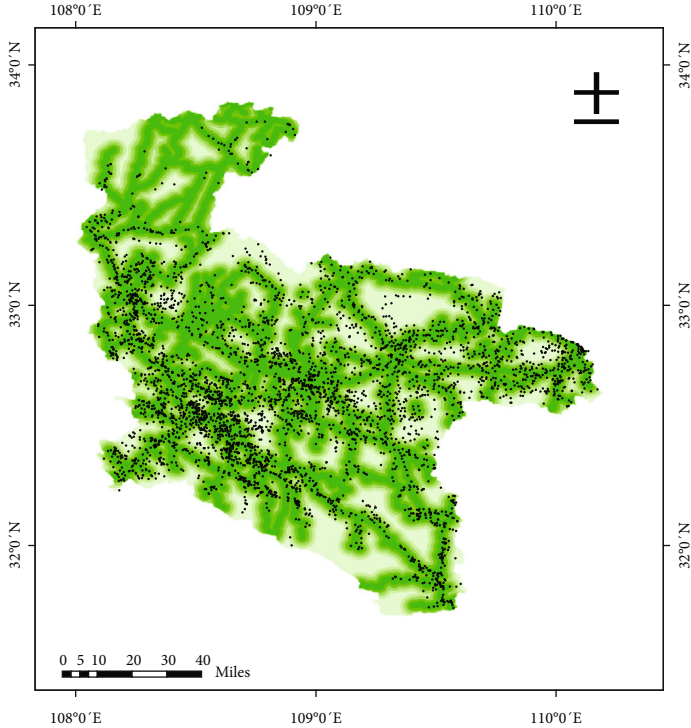


FIGURE 3: Geological and hydrogeological factors mapping of the study area: (a) rainfall; (b) distance to rivers; (c) stream power index; (d) distance to faults.



(a)



(b)

FIGURE 4: Human engineering activities factors mapping of the study area: (a) normalized difference vegetation index; (b) distance to roads.

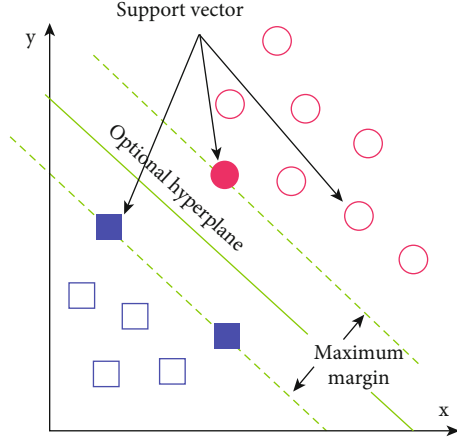


FIGURE 5: Support vector machine.

meaningless workload, which will often have an adverse impact on the evaluation accuracy [58]. There is no fixed reference for the selection of the number of influencing factors. According to the relevant literature and the geographical characteristics of the study area, 15 factors are selected in this study. It is also necessary to ensure that each index factor in the evaluation system is independent of each other, so as to avoid mutual interference between factors with strong correlation, which will affect the accuracy of prediction results. In order to remove the factors with large correlation and unimportant factors, two most commonly used tests, namely, multicollinearity analysis and Pearson correlation coefficient method, are used [59].

4.3. Logistic Regression. On the basis of linear regression, logistic regression adds a sigmoid function (logic equation) to realize the division of decision boundary of classification problems. The principle is to map the range $(-\infty, +\infty)$ of linear regression to the range $(0, 1)$ of Sigmoid function by logical function [60]. The expression of Sigmoid function is as follows:

$$f(y) = \frac{1}{1 + e^{-y}}, \quad (2)$$

where $y = \beta_0 + \alpha_0 * x_0 + \alpha_1 * x_1 + \dots + \alpha_m * x_m$; x_m represents 14 landslide factors; α_m is the regression coefficient through variables; y is the intermediate variable. $f(y)$ represents the probability of landslide; $1 - f(y)$ indicates the probability of no landslide.

4.4. Support Vector Machine. Support vector machine (SVM) classifies 12 landslide factor data in a linear way relative to the optimal boundary [61]. The sample of the study area is the support vector that determines the boundary line. The best boundary (hyperplane) is two lines of maximum distance between different vectors. It ensures the correctness and difference of classification (Figure 5). The formula is as

follows:

$$y = a + \sum_{i=1}^m \beta_i x_i = a + \beta^T X, \quad (3)$$

where m is the number of landslide impact factors, x_i is the landslide influencing factor in the study area, and y is the objective function.

When the accurate optimal boundary cannot be found by using soft interval, the support vector is generally mapped to high-dimensional space to solve this problem. However, the calculation of feature space is generally difficult. In order to optimize the problem, the Kernel function $K(x)$ is introduced to avoid the specific form of feature space $\gamma(x)$ [62]. It is only necessary to determine that the Kernel function is the inner product of two infinite dimensional vectors. The function formula is expressed as

$$K(x_i, x) = \gamma(x_i)\gamma(x),$$

$$y = w^T \gamma(x_i) + b = \sum_{i=1}^m \alpha_i y_i K(x_i, x), \quad (4)$$

where y_i is the training output of 12 factors and y is the training target.

By comparing the performance of the five kernel classifiers through various statistical indicators, it is concluded that RBF-SVM is relatively excellent [63], and then RBF-SVM is selected as one of the prediction models in this paper.

Radial basis function is $k(x_i, x) = (-\gamma(x_i - x_j))$, $\gamma > 0$, where b and γ are parameters of the kernel functions.

4.5. Random Forest. The random forest based on decision tree principle is mainly modeled by bagging algorithm. Bootstrap method is used for K times of random sampling of n training set samples, and n times are repeated to obtain K decision trees. Then d ($d < D$) features are randomly selected from the D features of the training subset, and the best features are selected from d features as the classification basis, and n times are repeated. Count the mode of the decision tree results as the final classification results [64]. The classification result of RF is more accurate, simple, and easy to understand, which is suitable for the big dataset processing and analysis (Figure 6). Based on the law of probability and statistics, the probability of each sample being drawn to a bootstrap set is P , and the formula of P is as follows:

$$P = 1 - \left(1 - \frac{1}{n}\right)^n, \quad (5)$$

where n is large enough in this study, p is about 63%, and the remaining 37% data are not involved in the modeling, called out-of-bag.

Compared with the bagging algorithm, the random forest has more feature restrictions. The parameter d is too small or too large, and the final fitting effect is relatively poor, so the optimal feature number is the biggest problem

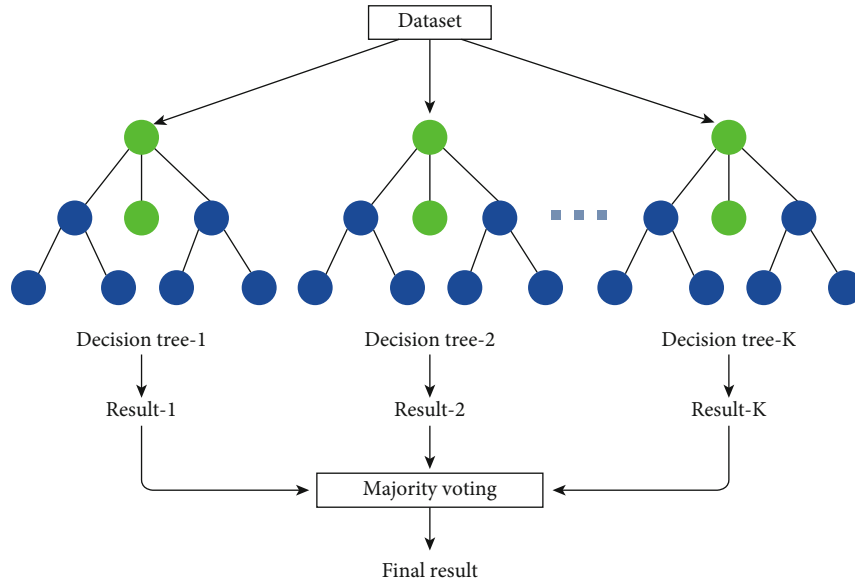


FIGURE 6: Random forest.

TABLE 3: Statistical metrics and their descriptions.

No	Metric	Formula	Description
1	TP	True positive	TP refers to the number of landslides that are accurately described as landslides [71]
2	TN	True negative	TN refers to the number of nonlandslides that are accurately classified as nonlandslide [72]
3	FP	False positive	FP refers to the number of nonlandslides that are wrong predicted as landslides [73]
4	FN	False negative	FN refers to the number of landslides that are accurately described as nonlandslide [74]
5	SST	$SST = \frac{TP}{TP + FN}$	The proportion of landslides that are correctly classified as landslide. This shows how good the model is [75]
6	SPF	$SPF = \frac{TN}{TN + FP}$	The proportion of nonlandslides that are correctly classified as nonlandslide. This shows how good the model is. It describes how well the landslide model used to classify nonlandslides can predict [76]
7	ACC	$ACC = \frac{TP + TN}{TP + TN + FP + FN}$	The proportion of landslides and nonlandslides correctly predicted in the total. The higher the accuracy value, the better the performance of the landslide model [77]
8	K	$K = \frac{P_p - P_{exp}}{1 - P_{exp}}$	The kappa value is between -1 and 1. The closer the value is to 1, the better the prediction effect of the model is [78]
9	MSE	$MSE = \frac{1}{N} \sum_{i=1}^N (x_m - x_p)^2$	MSE and RMSE values are calculated from measurements (x_m) and predictions (x_p), the lower the performance of the landslide model, the better [79]
	RMSE	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_m - x_p)^2}$	

TABLE 4: Analysis result table based on multicollinearity diagnosis method.

Factors	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
VIF	1.28	1.08	1.72	1.02	1.52	1.27	1.83	8.57	2.54	1.01	1.27	8.29	1.84	1.44	24.91
TOL	0.78	0.93	0.58	0.98	0.66	0.79	0.68	0.12	0.77	0.99	0.79	0.12	0.55	0.69	0.00

Note: 1: NDVI; 2: distance to faults; 3: distance to rivers; 4: lithology; 5: rainfall; 6: distance to roads; 7: stream power index; 8: plane curvature; 9: slope degree; 10: slope aspect; 11: profile curvature; 12: surface roughness; 13: elevation; 14: landform; 15: terrain curvature.

TABLE 5: Correlation matrix of evaluation factors based on Pearson coefficient method.

Factors	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.00														
2	0.16	1.00													
3	0.13	-0.05	1.00												
4	0.41	0.14	0.11	1.00											
5	-0.05	0.05	0.01	-0.02	1.00										
6	0.03	0.02	-0.03	0.4	0.02	1.00									
7	0.01	0.29	-0.20	0.05	0.02	0.21	1.00								
8	-0.05	0.04	0.01	-0.04	0.08	0.13	-0.01	1.00							
9	0.03	0.32	0.01	-0.01	-0.04	-0.23	0.16	0.04	1.00						
10	0.02	0.01	0.02	0.001	-0.02	0.01	-0.03	-0.04	0.01	1.00					
11	0.01	-0.07	-0.01	0.01	-0.03	0.06	0.02	-0.46	-0.01	0.05	1.00				
12	-0.03	0.07	0.01	-0.03	0.02	0.02	-0.02	0.80	0.03	-0.05	-0.89	1.00			
13	0.007	0.27	-0.01	-0.03	-0.03	0.03	0.16	0.03	0.94	0.02	-0.01	0.02	1.00		
14	0.225	0.32	-0.01	0.30	0.09	0.11	0.49	-0.01	0.19	-0.01	-0.01	0.01	0.17	1.00	
15	-0.07	0.14	-0.13	-0.06	0.09	0.04	0.40	0.001	0.13	-0.02	0.01	-0.01	0.12	0.47	1.00

Note: 1: distance to roads; 2: NDVI; 3: distance to faults; 4: distance to rivers; 5: lithology; 6: stream power index; 7: rainfall; 8: plane curvature; 9: slope degree; 10: slope aspect; 11: profile curvature; 12: terrain curvature; 13: surface roughness; 14: elevation; 15: landform.

TABLE 6: Model performance on the test dataset.

Metric	LR	SVM	RF	WOA-RF	GA-RF
TP	1003	1076	1039	1051	1068
TN	936	896	924	914	917
FP	331	371	343	353	350
FN	297	261	309	249	232
Sensitivity	0.772	0.805	0.799	0.808	0.822
Specificity	0.739	0.707	0.729	0.721	0.724
Accuracy	0.755	0.768	0.765	0.765	0.773
AUC	0.755	0.837	0.839	0.835	0.851
Kappa	0.51	0.505	0.529	0.53	0.546
RMSE	0.495	0.481	0.485	0.484	0.476
MAE	0.245	0.232	0.235	0.235	0.227

TABLE 7: Model performance on the training dataset.

Metric	LR	SVM	RF	WOA-RF	GA-RF
TP	2353	2641	2661	2700	2768
TN	2165	2338	2465	2572	2595
FP	846	673	546	439	416
FN	625	337	317	278	210
Sensitivity	0.79	0.887	0.894	0.907	0.929
Specificity	0.719	0.776	0.819	0.854	0.862
Accuracy	0.754	0.831	0.856	0.88	0.895
AUC	0.755	0.832	0.856	0.88	0.896
Kappa	0.509	0.663	0.712	0.761	0.791
RMSE	0.496	0.411	0.38	0.346	0.323
MAE	0.246	0.169	0.144	0.12	0.105

[65]. In the model, Gini coefficient is used to continuously test all the segmentation points of the feature subset of the same tree, and the branch corresponding to the smallest feature is selected. Then the model is verified by comparing different out-of-bag errors [66].

4.6. Genetic Algorithm. Genetic algorithm is a stochastic global search optimization method that simulates the phenomena of duplication, crossover, and mutation that occur in natural selection and inheritance [67]. The solution of the solution space is represented as the genotype string (that is, chromosome) structural data of the genetic algorithm, and different combinations of these string structural data constitute different points. Starting from any initial population (Population), through random selection of chromosomes with higher fitness, crossover, and mutation operations, a group of individuals more suitable for the environment is generated, so that the group evolves into a better and better area in the search space [68]. Such generations continue to multiply and evolve and finally converge to a group of individuals that are most adapted to the environment, so as to obtain a high-quality solution to the problem.

The main steps of the genetic algorithm are as follows:

Step 1. Generation of the initial population, random generation of hyperparameter combinations within the approximate hyperparameter range after grid search.

Step 2. Individual fitness value evaluation and detection, based on the score of each individual's random forest training results, evaluate the fitness of each individual.

Step 3. Breeding selection, establishing criteria to select individuals from the parent to participate in breeding. While selecting elite individuals as much as possible, the diversity

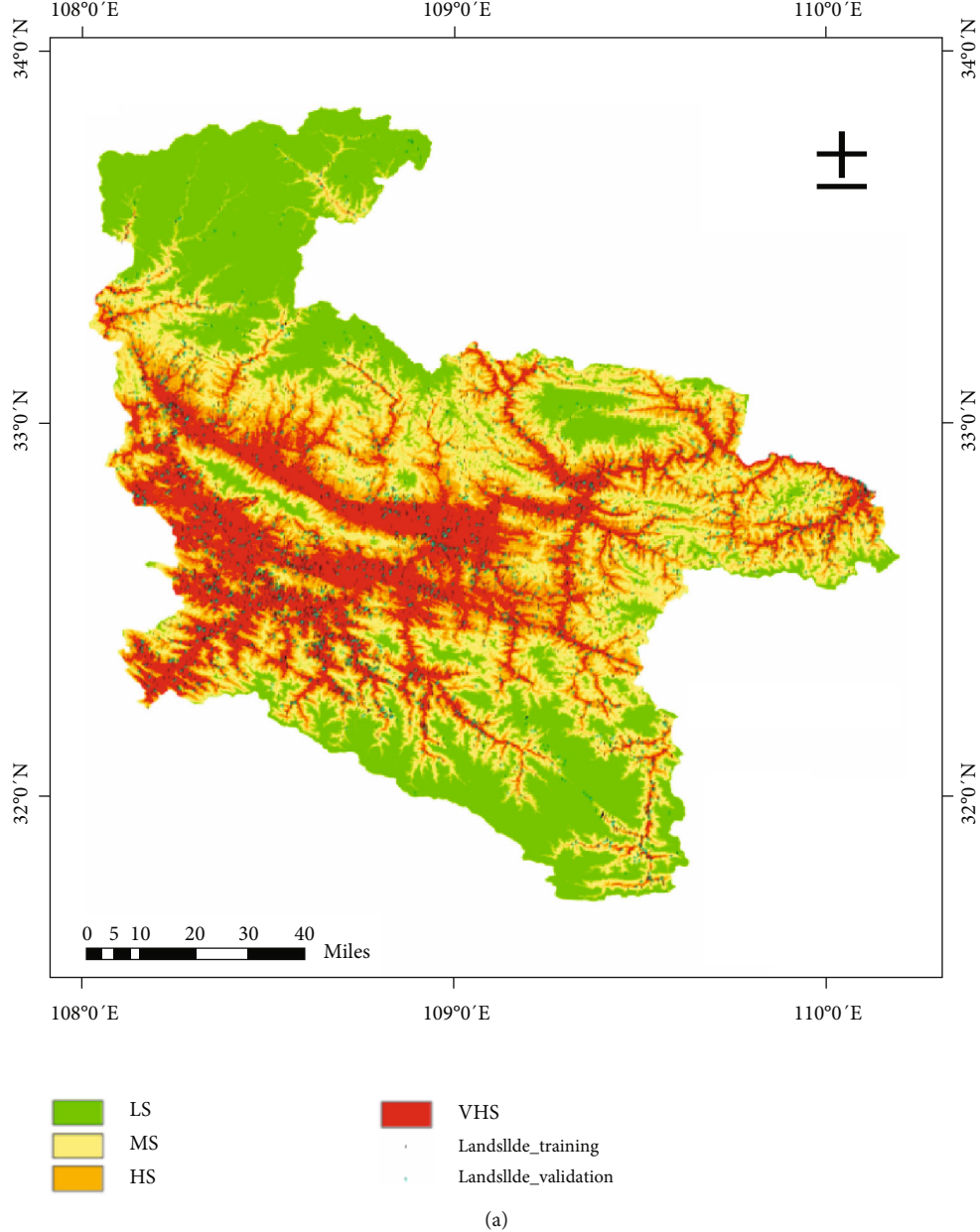


FIGURE 7: Continued.

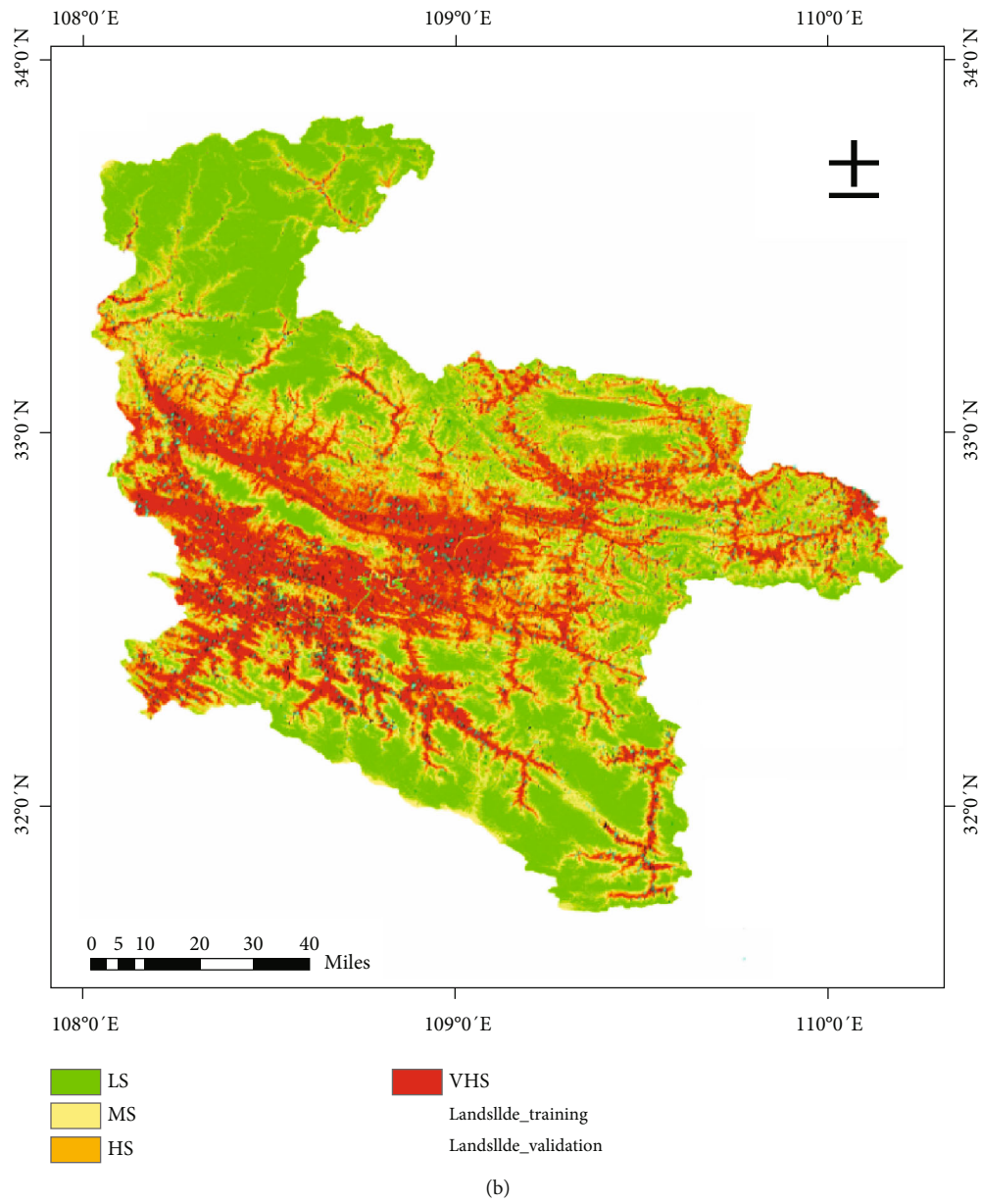


FIGURE 7: Continued.

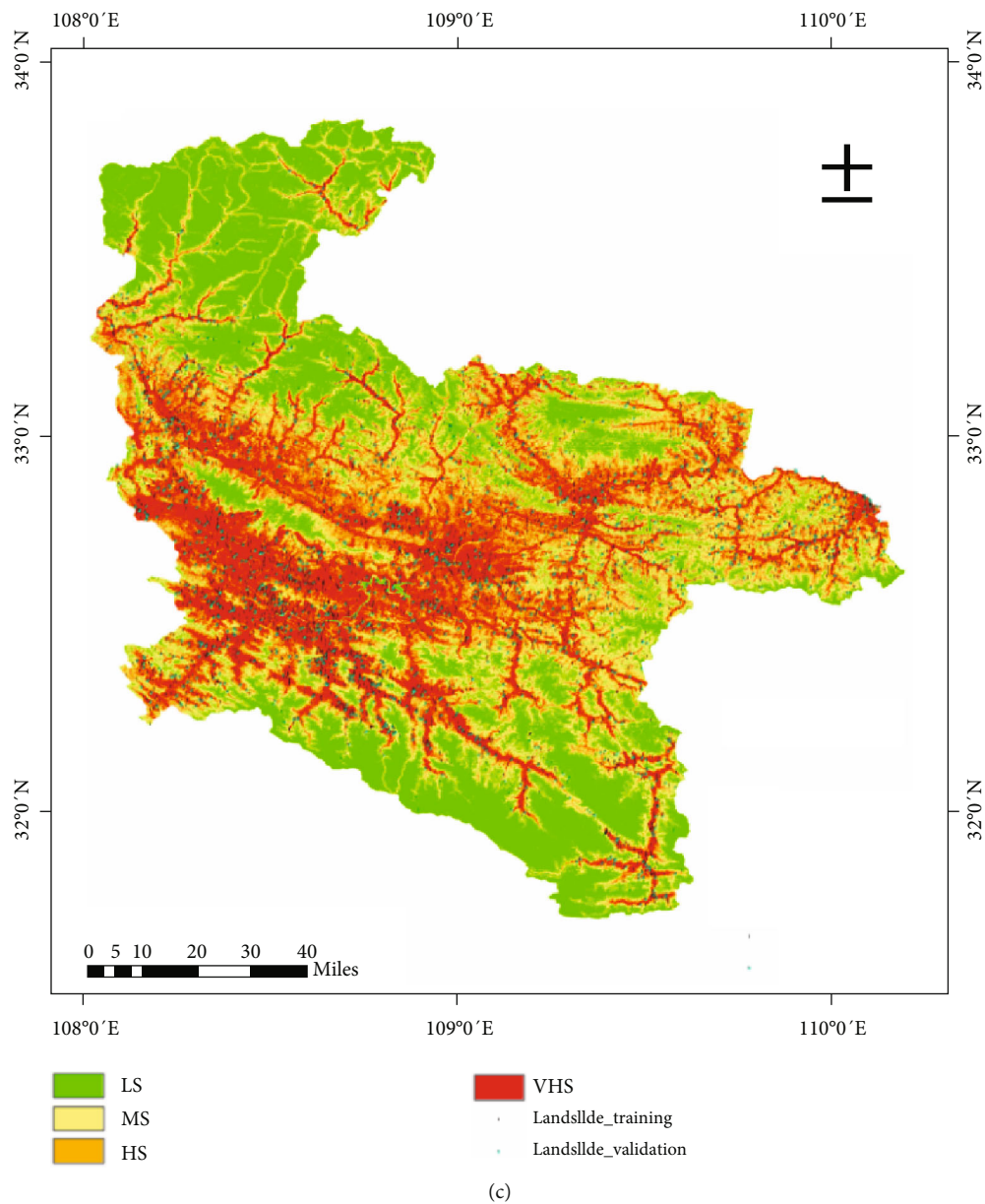
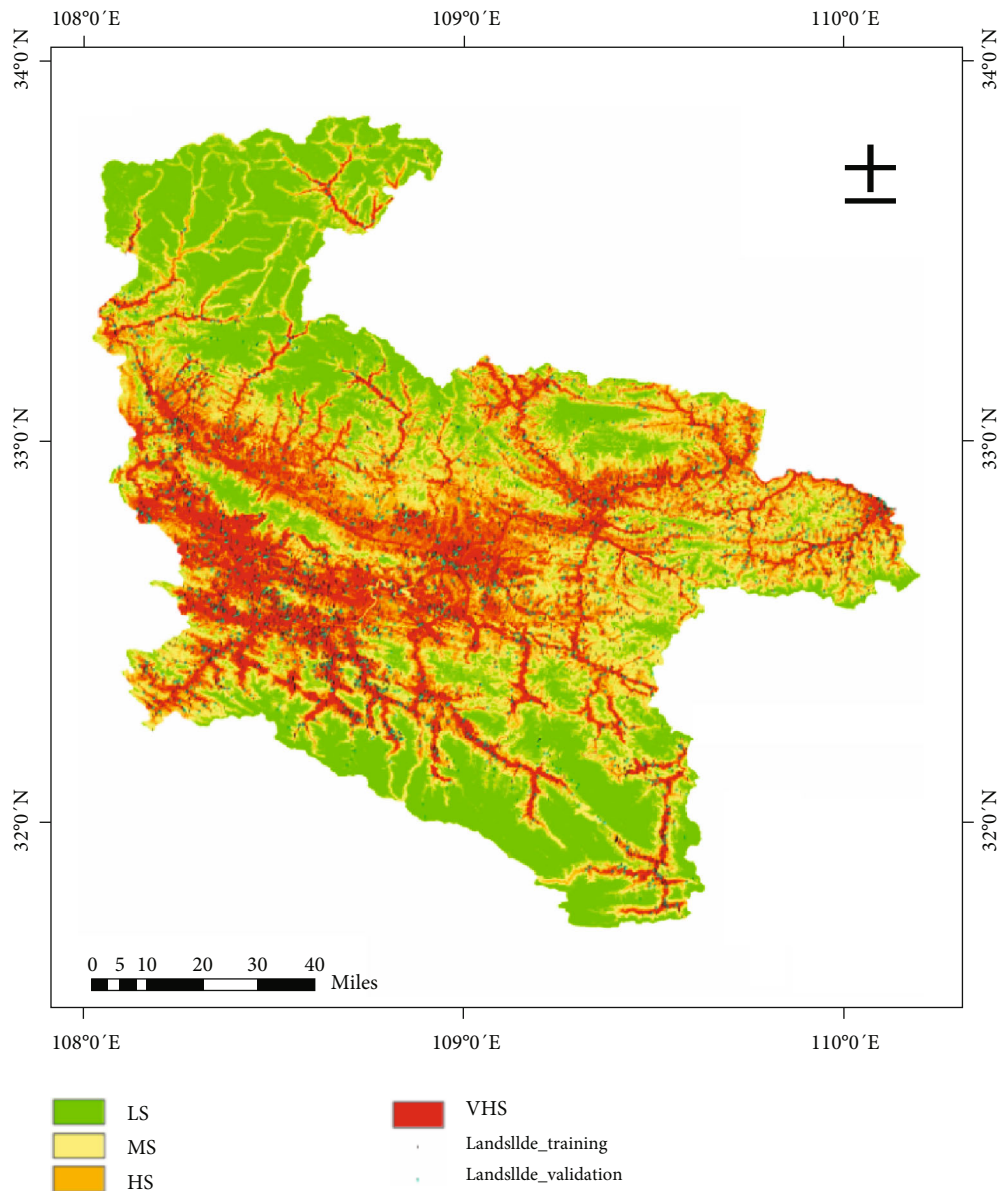


FIGURE 7: Continued.



(d)

FIGURE 7: Continued.

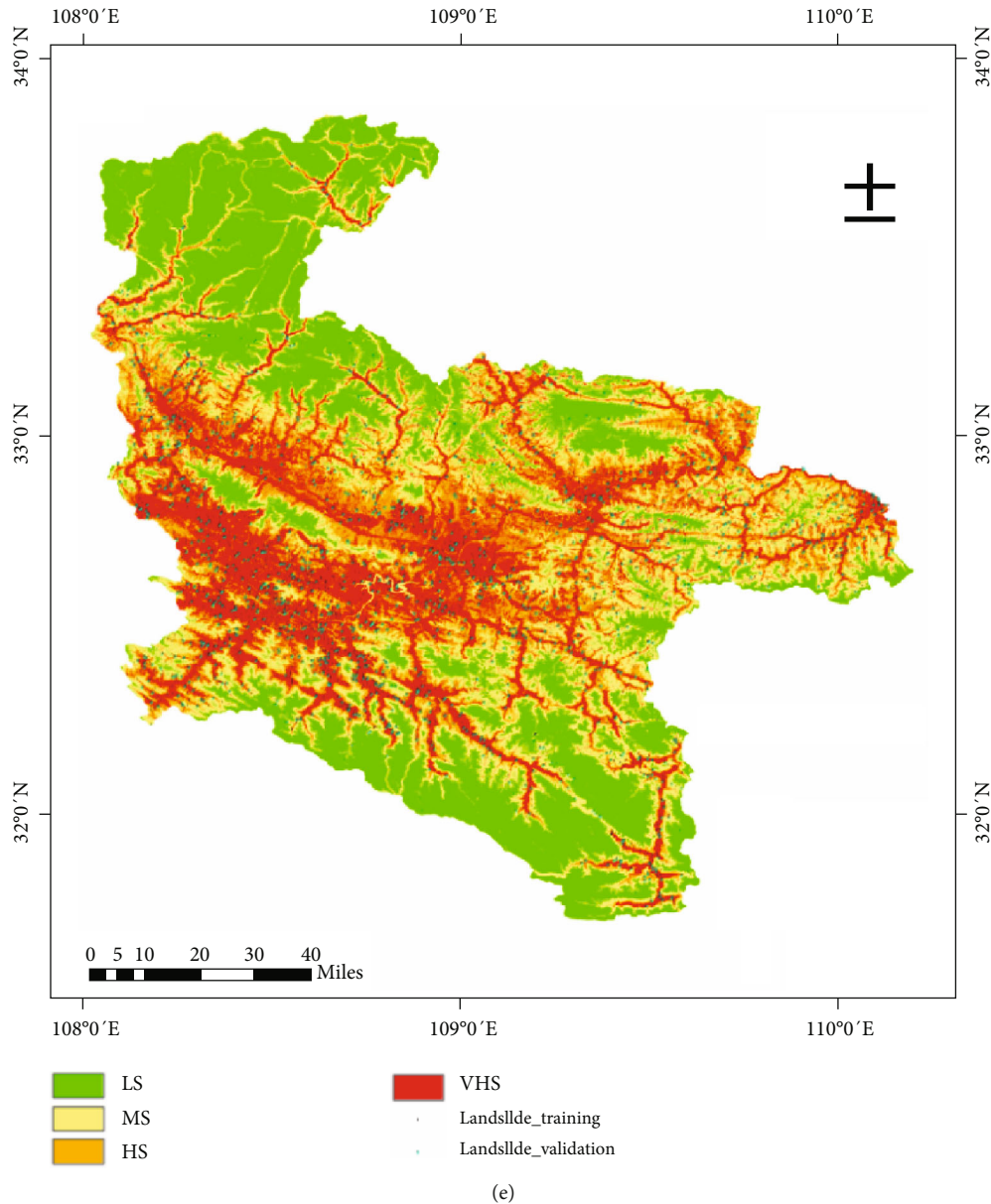


FIGURE 7: Landslide susceptibility maps based on different models: (a) LSM by LR; (b) LSM by SVM; (c) LSM by RF; (d) LSM by WOA-RF; (e) LSM by GA-RF.

of the population should also be maintained to prevent the algorithm from falling into local optimum prematurely.

Step 4. Mutation, the mutation process includes a series of biologically inspired operations, such as recombination and mutation. Through the mutation operation, the individual codes of the parent are inherited and recombined in a certain way to form a descendant group.

Step 5. Environmental selection to regroup parents and offspring into new groups. The offspring bred in this process are reinserted into the parent population, replacing part or all of the parent population, forming a new population with a similar size to the previous generation.

Step 6. The stopping criterion determines when the algorithm stops. There are usually two situations: the algorithm has found the optimal solution, or the algorithm has been selected into the local optimum and cannot continue to search in the solution space.

4.7. Whale Optimization Algorithm. The whale optimization algorithm (WOA) is a swarm intelligence optimization algorithm that imitates the whale’s predation behavior in nature, and the whale’s predation behavior is mainly divided into three categories: surround prey, foam net attack, and search and prey [69]. Therefore, before using WOA algorithm to solve the problem, the above three types of predation behaviors are expressed mathematically [70]. The purpose of whales’ predation is to capture prey. When a group of whales

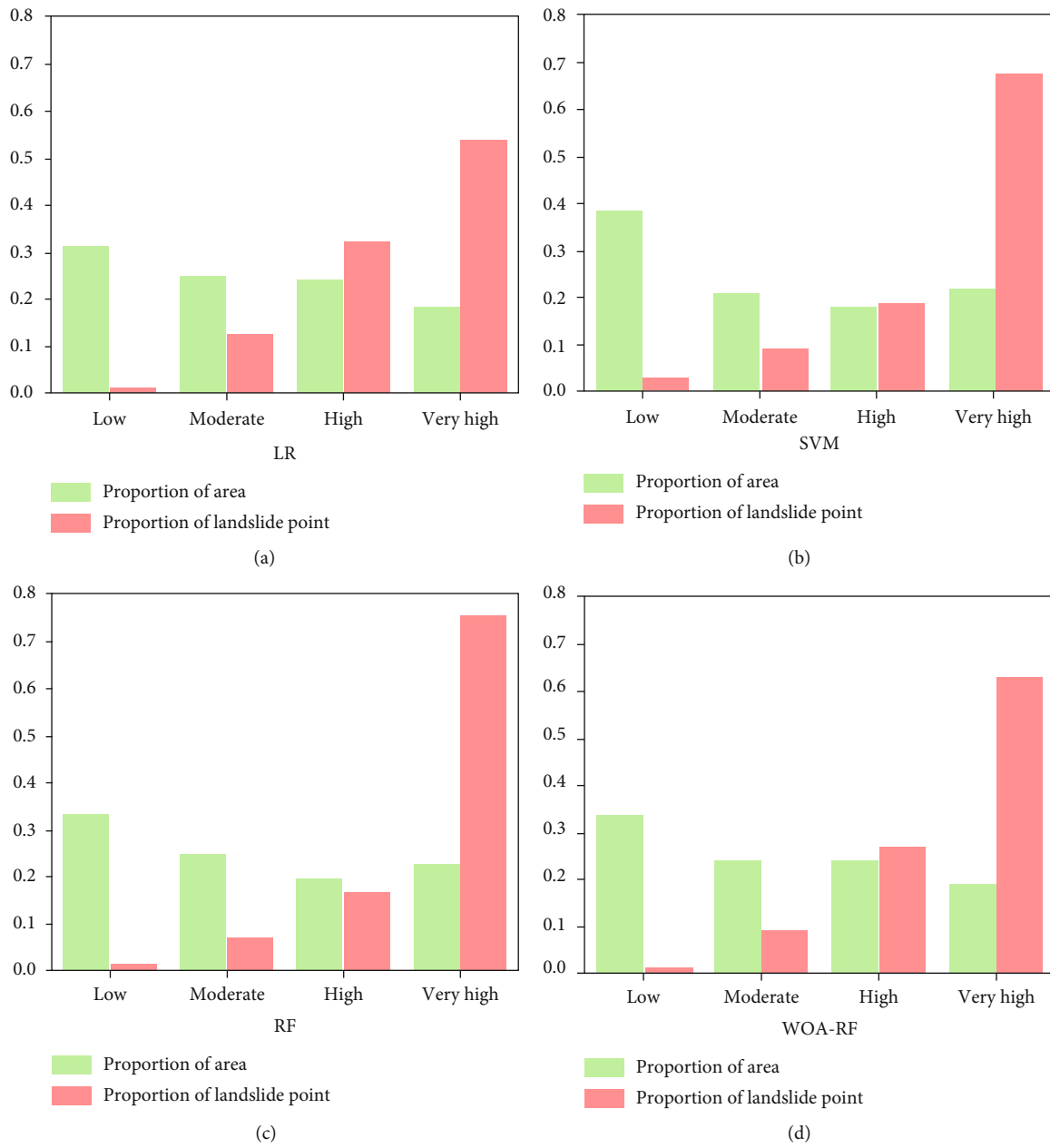


FIGURE 8: Continued.

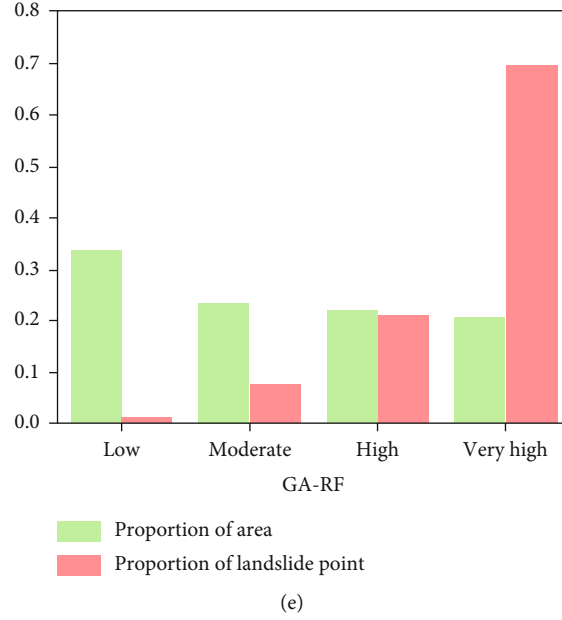


FIGURE 8: Raster distributions of landslides of different classification: (a) LR; (b) SVM; (c) RF; (d) WOA-RF; (e) GA-RF.

TABLE 8: Landslide densities in each zone of landslide susceptibility.

Classes	LR	SVM	RF	WOA-RF	GA-RF
Low	0.04	0.08	0.04	0.04	0.03
Moderate	0.49	0.45	0.29	0.38	0.32
High	1.33	1.04	0.85	1.12	0.96
Very high	2.82	3.08	3.32	3.33	3.39

are looking for prey together, there must be a whale that finds the prey first. At this time, other whales will definitely swim to the whale that found the prey to compete for the prey. In this paper, the number of whale populations is 10, and in the termination condition, the maximum number of iterations is 100. The main steps of the WOA algorithm are as follows:

4.7.1. Encircling Prey. The next position of individual whale (X^j) under the influence of the best individual whale (X^{j+1}) is calculated as

$$\begin{aligned}
 X_k^{j+1} &= X_k^* - A \cdot D_k, \\
 D_k &= \left| C \cdot X_k^* - X_k^j \right|, \\
 C &= 2r, \\
 A &= 2a \cdot r - a, \\
 a &= 2 - \frac{2i}{i \max},
 \end{aligned} \tag{6}$$

where a is linearly decreasing from 2 to 0 as the number of iterations i increases, A is the distance adjustment factor, C is a random number, and D_k is the distance difference between the best whale and the current individual whale.

4.7.2. Search for Prey. In the mathematical model of hunting behavior surrounded by search, the value of A is restricted to $(-1,1)$, individual whale is best. However, when the value of A is not at $(-1,1)$, the current individual whale may not have reached the current maximum. The best whale individuals approach but randomly select a whale individual from the current whale individuals to approach, which is the idea of search and prey. Search and predation may make the current whale individual deviate from the target prey, but it will increase the global search ability of the individual whale.

$$\text{target whale} = \begin{cases} \text{best whale, } |A| < 1, \\ \text{random whale, } |A| \geq 1. \end{cases} \tag{7}$$

4.7.3. Spiral Position Update. The current whale approaches the current best individual whale in a spiral fashion. The whale location update is formulated as follows:

$$\begin{aligned}
 X_k^{j+1} &= X_k^* + D_k \cdot e^{bl} \cdot \cos(2\pi l), \\
 D_k &= \left| X_k^* - X_k^j \right|.
 \end{aligned} \tag{8}$$

When whales hunt their prey, they will not only shrink the encircling circle, but also swim to the prey in a spiral form, so they choose to shrink the encircling circle or swim toward the prey in a spiral fashion with a 50% probability.

$$X_k^{j+1} = \begin{cases} X_k^* - A \cdot D_k, & p < 0.5, \\ X_k^* + D_k \cdot e^{bl} \cdot \cos(2\pi l), & p \geq 0.5, \end{cases} \tag{9}$$

where b is a constant defining the shape of the logarithmic spiral and l is a random number in $[-1,1]$.

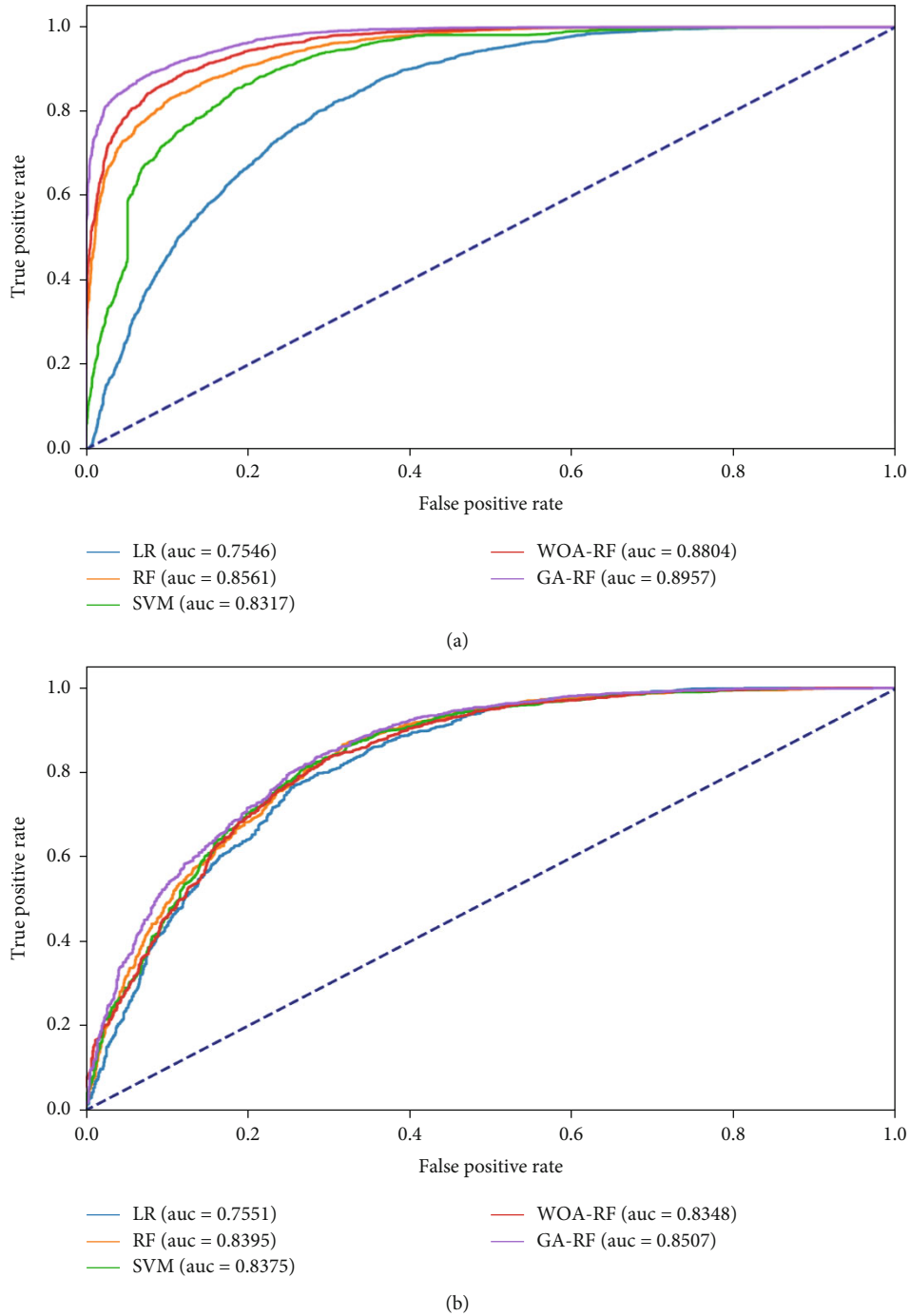


FIGURE 9: ROC curves of five landslide susceptibility models using (a) training and (b) testing.

4.8. Evaluation and Comparison Methods

4.8.1. Statistical Metrics. Through the quantitative analysis of training samples and test samples, the effect of the model on data fitting and prediction ability is evaluated. In this paper, several commonly used statistical indicators are selected for quantitative analysis and comparison of different model training sample sets and test sample sets (Table 3), such as the true positive value (TP), true negative value (TN), false positive value (FP), false negative value (FN), sensitivity

(SST), specificity (SPF), accuracy (ACC), kappa index (K), mean absolute error (MSE), and root mean squared error (RMSE).

4.8.2. Receiver Operating Characteristic Curve. Scholars at home and abroad often use ROC test model to predict the accuracy of landslide sensitivity and measure the matching effect of the model on the data, and landslide prediction results are divided into positive and negative categories [80]. x -axis is the false positive rate, indicating

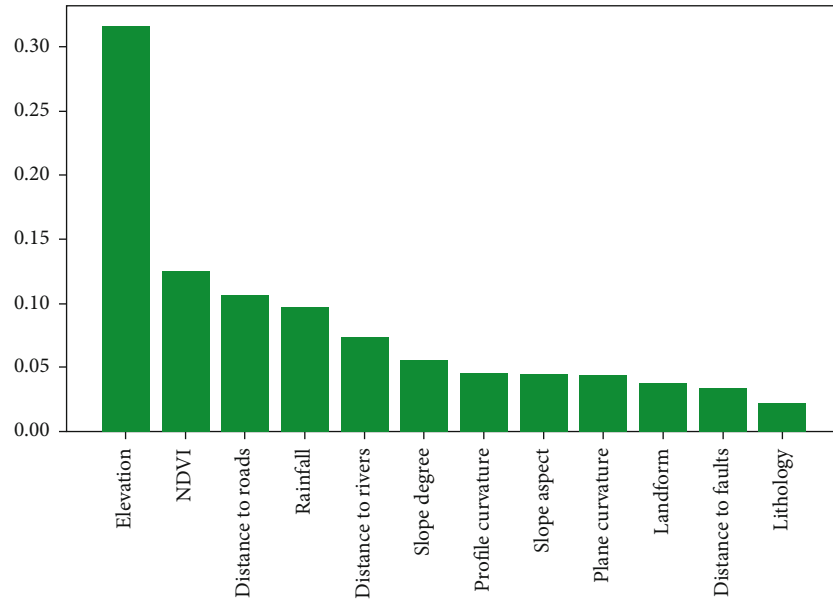


FIGURE 10: Distribution of factors with GA-RF model.

the proportion of nonlandslide points divided into positive categories in all nonlandslide points; y -axis representative is divided into the positive class and is divided into the number of landslide points accounted for the proportion of all landslide points. The farther the ROC is away from the 45° diagonal line, the more accurate the model is [81]. The area under ROC, namely AUC value, is referred to deep study on the quantitative comparative analysis of prediction success rate accuracy [82]. The closer the AUC value is to 1.0, the more accurate the discrimination model is.

5. Results and Discussion

5.1. Elimination and Selection of Landslide Affecting Factors. The results of Pearson correlation coefficient and multicollinearity analysis [83] show that the tolerance of surface roughness (TOL), terrain curvature, and SPI is less than 0.5, and their variance inflation coefficient (VIF) is more than 5 in Table 4. In addition, in terms of relevance, curvature is strongly correlated with plan curvature, curvature with profile, surface roughness, and slope. The comprehensive results show that it is best to eliminate these three factors (terrain curvature, surface roughness, and SPI) before importing the model in Table 5. Therefore, in this paper, 12 factors including slope aspect, slope degree, elevation, plane curvature, profile curvature, distance to faults, distance to roads, landform, lithology, distance to rivers, rainfall, and normalized difference vegetation index (NDVI) are finally selected as the evaluation indicators to participate in this landslide sensitivity study.

5.2. Modeling Process and Evaluations. After removing the factors with high correlation and strong collinearity, we combined the landslide pixels and nonlandslide pixels to extract the values of 12 factors. In this paper, the modeling

and comparison analysis of the test and training datasets are carried out, respectively. GA and WOA heuristic algorithms are proposed to optimize the RF model and find the optimal parameter combination. Then, the values of statistical indicators such as RMSE, MSE, and accuracy are used to verify the success rate and performance superiority of the model. The optimal parameter combination has the maximum evolutionary generation number (NGEN) = 40, the population size (pop size) = 300, the number of individuals selected for the next generation (MU) = 5, the number of children to be produced in each generation (LAMBDA) = 10, the crossover probability (CXPB) = 0.7, and the mutation probability (MUTPB) = 0.2.

After selecting the optimal parameter values for each model, several metrics after training and test sets are run separately, including sensitivity, specificity, accuracy, kappa index, mean absolute error, and root mean square error. The following table compares and analyzes the index values of each model statistics and shows that the prediction effect of the GA-RF model is significantly higher than that of the LR, SVM, RF, and WOA-RF models.

For each model, the statistical measures of the test dataset are shown in Table 6. In summary, The GA-RF ensemble model had the highest sensitivity (0.822), followed by the WOA-RF (0.808), SVM (0.805), RF (0.799), and LR (0.772) models. The highest specificity (0.739) was achieved by the LR model, followed by the RF (0.729) and GA-RF (0.724), WOA-RF (0.721), and SVM (0.707) models. The GA-RF model also achieved the highest accuracy (0.773) and the LR model (0.755) the lowest. Accuracies of the SVM, WOA-RF, and RF models are, respectively, 0.768, 0.765, and 0.765. The highest kappa value (0.546) was obtained by the GA-RF model, followed by the WOA-RF (0.53), RF (0.529), LR (0.510), and SVM (0.505) models. The GA-RF model has the lowest RMSE and MSE values, 0.476 and 0.227, respectively.

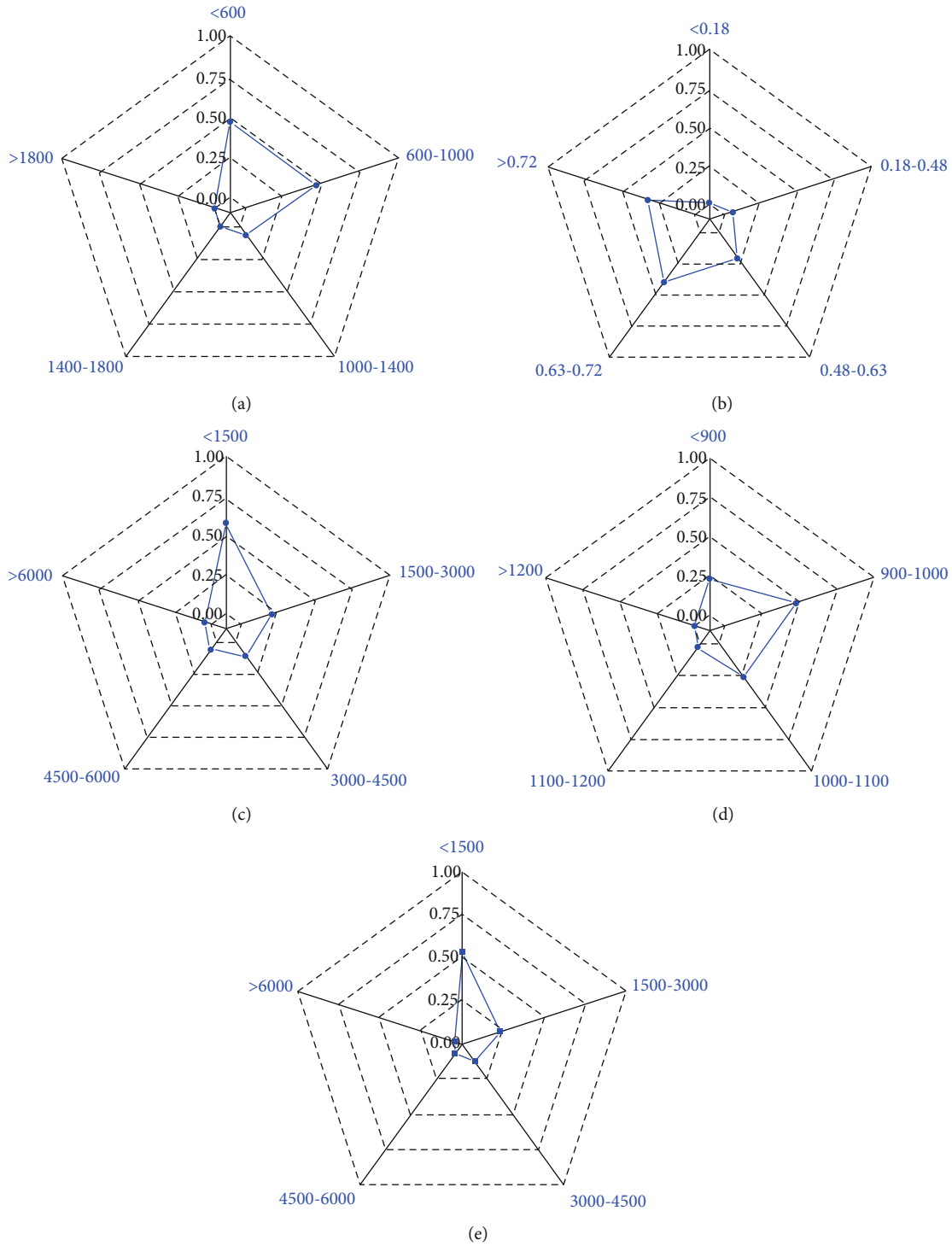


FIGURE 11: Distribution with high and very high landslide sensitivity based on GA-RF model on (a) elevation thematic map, (b) NDVI thematic map, (c) distance to roads thematic map, (d) rainfall thematic map, and (e) distance to rivers thematic map.

On the other hand, the indicators of the training set statistics clearly show that the GA-RF model has the best performance. The statistical indicators values obtained by the GA-RF model are divided into sensitivity (0.929), specificity (0.862), accuracy (0.895), kappa index (0.791), mean abso-

lute error (0.323), and root mean squared error (0.105). Comparing the optimized GA-RF model with the RF model, the improvement degrees of each index of the GA-RF model are sensitivity (0.073), specificity (0.043), accuracy (0.039), RMSE (-0.009), MSE (-0.008), and kappa index (0.017).

The value of each index intuitively reflects the optimization improvement of the GA-RF model compared with the performance before optimization (Table 7).

5.3. Landslide Susceptibility Map of Multiple Models. This paper uses GA and WOA algorithm to optimize the RF model after successful testing and obtain the best parameter combination and then together with the LR and SVM models to predict the probability of each pixel in the study area. In order to realize the visualization of landslide susceptibility mapping, this paper employs Jenks natural break-point method to divide the grid data output from the model into four categories: low sensitivity (LS), moderate sensitivity (MS), high sensitivity (HS), and very high sensitivity (VHS), as shown in Figure 7. The GA-RF model are classified as follows: low sensitivity (LS), probability range of ≤ 0.0001 – 0.1929 ; moderate sensitivity (MS), probability range of 0.1929 – 0.4240 ; high sensitivity (HS), probability range of 0.4240 – 0.6436 ; and very high sensitivity (VHS), probability range of 0.6436 – 0.9826 ; these classes in WOA-RF model map for the LS, MS, HS, and VHS are ≤ 0.0001 – 0.1922 , 0.1922 – 0.4275 , 0.4275 – 0.6627 , and 0.6627 – 1.0 . For RF model map, these classes are LS (≤ 0.0001 – 0.2000), MS (0.2000 – 0.4471), HS (0.4471 – 0.6471), and VHS (0.6471 – 1). Consequently, in SVM the classes were LS (0.0009 – 0.1891), MS (0.1891 – 0.4157), HS (0.4157 – 0.6615), and VHS (0.6615 – 0.9802). For LR model map, these classes are LS (0.0004 – 0.1921), MS (0.1921 – 0.4269), HS (0.4269 – 0.26695), and VHS (0.6695 – 0.9981).

The results show that, according to the SVM model, the low sensitivity area is 38.40%, medium class is 21.21%, high class is 18.25%, and extremely high class is 22.14%. In the GA-RF model, the low sensitivity area is the largest (33.64%), followed by the medium class (23.40%), the high class (22.34%), and the very high class (20.62%). The extremely high sensitivity area in the WOA-RF model is the smallest among all models (18.85%), and the medium class is 23.81%, and high class is 23.83%. Low class is 31.53%, medium class is 25.07%, high class is 24.18%, and very high class is 19.22% in the LR model. Low class is 33.34%, middle class is 24.67%, high class is 19.40%, and very high class is 22.62% in the RF model (Figure 8). Generally, the five models have the same trend of change.

The models in the study area intuitively reflect that the landslide sensitivity between Daba Mountain and Qinling Mountains are higher than that in the southern and northern regions. High and very high sensitivity areas have large terrain fluctuations. Overall, the sensitivity map visualizes the landslide susceptibility distribution.

5.4. Model Prediction Evaluation. In order to quantify each category, this study draws the proportion of the sensitive area of four levels in the total area and the proportion of historical landslide events in the total number of landslides. Green represents the proportion of each level of landslide sensitivity mapping, and red represents the proportion of historical landslide events. The sensitivity area of high and very high risks is less than 50% in the models, but the pro-

portion of historical landslides is more than 90%. Landslide densities (LD) is the ratio of the percentage of the number of landslide events to the percentage of all pixel areas for each category on the sensitivity mapping. Comparing the values of LD of several models, the higher the value corresponding to the high and the extremely high sensitivity, the better the prediction effect of the model. The higher the red column is, the better the prediction effect of the model is. Meanwhile, the landslide density values corresponding to the very high sensitive areas in the five models LR, SVM, RF, WOA-RF, and GA-RF are calculated, and the results are 2.82, 3.08, 3.32, 3.33, and 3.39, respectively. That is to say, GA-RF > WOA-RF > RF > SVM > LR as shown in Table 8 and Figure 8.

This research analyzes the prediction accuracy of RF, SVM, RF, WOA-RF, and GA-RF models using under the ROC curve based on the training dataset and validation dataset. The training success rates of LR, SVM, RF, WOA-RF, and GA-RF models are 0.7546, 0.8317, 0.8561, 0.8804, and 0.8957; the testing success rates are 0.7551, 0.8375, 0.8395, 0.8348, and 0.85007 (Figure 9). Combined with the analysis of the above raster histogram results, LR is the worst of the five models, and the success rate of GA-RF prediction is the highest, followed by RF, SVM, and WOA-RF are relatively low. And combined with the value of accuracy of the training set and the test set in Section 5.2, the comprehensive results show that the GA-RF model has the best effect.

5.5. Contribution Value of Disaster-Causing Factors. The main controlling factors of landslide susceptibility still have no deterministic specification, and the contribution degree of evaluation factors is of great significance to the predictors of landslide susceptibility. Some landslide factors are eliminated through Pearson correlation coefficient and multicollinearity analysis. The results show that the various models of landslides significantly improve the index values of AUC and accuracy. Due to the different contribution rate of different features to the output results, the factor contribution rate sorting has a certain effect on the interpretability and visualization of the model. After landslide susceptibility prediction and evaluation, Python is used to rank feature importance of GA-RF model. The greater the decrease in the accuracy outside the bag is, the greater the impact of this feature on the prediction classification results is. After the GA-RF model is classified, the feature importance is calculated based on the Gini coefficient, in which the overall importance of feature j is measured by the average importance of feature j in a tree.

The contribution rates of each feature of GA-RF model with high success rates to landslide prediction results are plotted. We consider that these factors all contribute positively to the simulation of debris flow sensitivity (Figure 10). The five factors of elevation, NDVI, distance to roads, rainfall, and distance to rivers have the highest contribution rate. Previous studies in the area and surrounding areas have also shown that there is also a strong correlation between elevation, rainfall, and roads through the occurrence of landslide events.

The distribution of landslides with high and very high sensitivity classes on five most important factor thematic maps (elevation, NDVI, distance to roads, rainfall, and distance to rivers) is shown in Figure 11. Most have high and very high landslide susceptibility closely related to the following conditions: the average elevation is below 1000 m, the NDVI is between 0.63 and 0.72, the distance to roads is below 1500 m, the rainfall is between 900 mm and 1000 mm, and the distance to rivers is also below 1500 m. Therefore, in these types of highly coupled areas, strengthen the early warning of heavy rainfall and increase prevention and control measures to protect people's property and personal safety.

5.6. Discussion. Compared with traditional qualitative analysis, some classical machine learning methods are more effective in predicting landslide susceptibility, such as LR and SVM models [84]. The LR model is directly based on the original dataset; multicollinearity is not a problem, and it has good prediction performances. SVM outperforms logistic regression in landslide susceptibility evaluation. SVM and RF models are widely used in landslide prediction with good results [85–87]. Because RF and SVM models will be overfitted, landslide factors need to be controlled within a certain range to ensure the performance of the landslide prediction model; 8–12 landslide factors are the best [88]. This study analyzes and compares 15 landslide influencing factors and excludes curvature, roughness, and SPI by Pearson correlation coefficient, variance expansion coefficient, and tolerance. We find that each model has a certain improvement in AUC value.

Then we counted the specific parameters of the training and testing sets, and the results show that all models have achieved good results. With the coupling of heuristic algorithms and machine learning, the accuracy of landslide prediction models can be effectively improved [89]. The GA and WOA can quickly and efficiently acquire hyperparameter combinations [90]. The WOA and GA in this study are actually looking for the best parameter combination of the RF model to improve the prediction ability of the RF model. We evaluate the performance of the model separately by statistical training and test dataset metrics, including sensitivity, specificity, accuracy, kappa, mean absolute error root mean square error, and area under the receiver operating characteristic curve in the case of training dataset and test set, and the results show that the GA-RF model has the best performance. The weight of each factor based on the Gini index of the GA-RF model has been obtained. The distribution of landslides with high and very high sensitivity classes is on five most important factor thematic maps (elevation, NDVI, distance to roads, rainfall, and distance to rivers). According to the results obtained, relevant preventive measures are formulated to achieve early prevention and control of landslides.

6. Conclusion

The selection of landslide factors is a key step in landslide mapping. Since the main control factors of landslides have

not been determined, the choice of the number of landslides has to be solved. Choosing too many numbers may lead to overfitting of the model, while too small numbers may lead to insufficient accuracy. A large number of scholars have studied the coupled optimization model. Among them, the GA and WOA in this study are applied. The research mainly studies the optimization of the RF model by GA and WOA, that is, to find the best parameter combination of the RF model.

In this study, we use the GA and WOA not to be separate models but to optimize the RF model to find its optimal parameter combination. First, a landslide inventory map consisting of 4278 identified landslides is randomly divided into training and test landslides in a ratio of 7:3. The 15 landslide impact factors detected potential multicollinearity among the factors using PCC, VIF, and TOL. The performance of the model separately by statistical training and test dataset metrics are evaluated, including sensitivity, specificity, accuracy, kappa, MSE, RMSE, and AUC.

The training success rates of LR, SVM, RF, WOA-RF, and GA-RF models are 0.7546, 0.8317, 0.8561, 0.8804, and 0.8957; the testing success rates are 0.7551, 0.8375, 0.8395, 0.8348, and 0.85007. The results also show that the GA algorithm significantly improves the predictive power of the RF model. The training success rates of LR, SVM, RF, WOA-RF, and GA-RF models are 0.7546, 0.8317, 0.8561, 0.8804, and 0.8957; the testing success rates are 0.7551, 0.8375, 0.8395, 0.8348, and 0.85007. The landslide density values corresponding to the very high sensitive areas in the five models of LR, SVM, RF, WOA-RF, and GA-RF are calculated, and the results are 2.82, 3.08, 3.32, 3.33, and 3.39, respectively. Comparing the optimized GA-RF model with the RF model, the improvement degrees of each index of the GA-RF model are sensitivity (0.073), specificity (0.043), accuracy (0.039), RMSE (-0.009), MSE (-0.008), and Kappa index (0.017). It shows that genetic algorithm is better than other algorithms in optimizing RF performance. The landslide susceptibility map produced can be used for correct decision-making and risk management of landslides. After comprehensive evaluation, GA-RF model is the most suitable prediction model in the five models, which has great reference significance in disaster prevention and control in Ankang City and its surrounding areas.

In terms of preventing overfitting and improving performance, it may be possible to further optimize and compare more deep learning and whether factor features need to be reduced in dimension to improve accuracy, which requires further research and analysis.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflict of interest

Authors' Contributions

B.X. Zhao designed and wrote the paper, B.X. Zhao and J.Z. Zhu collected the data for the paper, B.X. Zhao and Y. Liu are responsible for visualization of landslide sensitivity, and Y.B. Hu and Q.M. Liu reviewed and supervised the paper writing. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

The authors acknowledge the financial support of the Natural Science Foundation of Anhui Province (grant no. 1908085ME145) and Anhui University of Science and Technology Graduate Innovation Fund Project (grant no. 2020CX2004).

References

- [1] Y. Liu, C. Xu, B. Huang et al., "Landslide displacement prediction based on multi-source data fusion and sensitivity states," *Engineering Geology*, vol. 271, article 105608, 2020.
- [2] M. F. Fadhillah, W. L. Hakim, M. Panahi, F. Rezaie, C. W. Lee, and S. Lee, "Mapping of landslide potential in Pyeongchang-gun, South Korea, using machine learning meta-based optimization algorithms," *The Egyptian Journal of Remote Sensing and Space Science*, vol. 25, no. 2, pp. 463–472, 2022.
- [3] J. Wang, L. Xiao, and S. N. Ward, "Tsunami Squares modeling of landslide tsunami generation considering the 'Push Ahead' effects in slide/water interactions: theory, experimental validation, and sensitivity analyses," *Engineering Geology*, vol. 288, article 106141, 2021.
- [4] L. Bragagnolo, R. V. da Silva, and J. M. V. Grzybowski, "Landslide susceptibility mapping with *r.landslide*: a free open-source GIS-integrated tool based on Artificial Neural Networks," *Environmental Modelling & Software*, vol. 123, article 104565, 2020.
- [5] F. Huang, Z. S. Cao, J. F. Guo, S. H. Jiang, S. Li, and Z. Guo, "Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping," *Catena*, vol. 191, article 104580, 2020.
- [6] A. Małka, "Landslide susceptibility mapping of Gdynia using geographic information system-based statistical models," *Natural Hazards*, vol. 107, no. 1, pp. 639–674, 2021.
- [7] S. Das, S. Sarkar, and D. P. Kanungo, "GIS-based landslide susceptibility zonation mapping using the analytic hierarchy process (AHP) method in parts of Kalimpong Region of Darjeeling Himalaya," *Environmental Monitoring and Assessment*, vol. 194, no. 4, article 234, 2022.
- [8] Q. Ba, Y. Chen, S. Deng, Q. Wu, J. Yang, and J. Zhang, "An improved information value model based on gray clustering for landslide susceptibility mapping," *ISPRS International Journal of Geo-Information*, vol. 6, no. 1, p. 18, 2017.
- [9] M. L. Chanu and O. Bakimchandra, "Landslide susceptibility assessment using AHP model and multi resolution DEMs along a highway in Manipur, India," *Environmental Earth Sciences*, vol. 81, no. 5, p. 156, 2022.
- [10] B. Li, N. Wang, and J. Chen, "GIS-based landslide susceptibility mapping using information, frequency ratio, and artificial neural network methods in Qinghai Province, Northwestern China," *Advances in Civil Engineering*, vol. 2021, Article ID 4758062, 14 pages, 2021.
- [11] A. Rehman, J. Song, F. Haq et al., "Multi-hazard susceptibility assessment using the analytical hierarchy process and frequency ratio techniques in the Northwest Himalayas, Pakistan," *Remote Sensing*, vol. 14, no. 3, p. 554, 2022.
- [12] H. Akinci and A. Yavuz Ozalp, "Landslide susceptibility mapping and hazard assessment in Artvin (Turkey) using frequency ratio and modified information value model," *Acta Geophysica*, vol. 69, no. 3, pp. 725–745, 2021.
- [13] R. K. Dash, P. O. Falae, and D. P. Kanungo, "Debris flow susceptibility zonation using statistical models in parts of Northwest Indian Himalayas-implementation, validation, and comparative evaluation," *Natural Hazards*, vol. 111, no. 2, pp. 2011–2058, 2022.
- [14] I. Ilija and P. Tsangaratos, "Applying weight of evidence method and sensitivity analysis to produce a landslide susceptibility map," *Landslides*, vol. 13, no. 2, pp. 379–397, 2016.
- [15] S. Lambie, S. Awatere, A. Daigneault et al., "Trade-offs between environmental and economic factors in conversion from exotic pine production to natural regeneration on erosion prone land," *New Zealand Journal of Forestry Science*, vol. 51, article 14, 2021.
- [16] G. Wang, X. Chen, and W. Chen, "Spatial prediction of landslide susceptibility based on GIS and discriminant functions," *ISPRS International Journal of Geo-Information*, vol. 9, no. 3, p. 144, 2020.
- [17] Q. He, Z. Jiang, M. Wang, and K. Liu, "Landslide and wildfire susceptibility assessment in Southeast Asia using ensemble machine learning methods," *Remote Sensing*, vol. 13, no. 8, p. 1572, 2021.
- [18] M. A. Hussain, Z. Chen, R. Wang, and M. Shoaib, "PS-InSAR-based validated landslide susceptibility mapping along Karakorum Highway, Pakistan," *Remote Sensing*, vol. 13, no. 20, p. 4129, 2021.
- [19] E. R. Sardooi, A. Azareh, T. Mesbahzadeh, F. S. Sardoo, E. J. R. Parteli, and B. Pradhan, "A hybrid model using data mining and multi-criteria decision-making methods for landslide risk mapping at Golestan Province, Iran," *Environmental Earth Sciences*, vol. 80, no. 15, article 487, 2021.
- [20] C. Wang, Q. Lin, L. Wang et al., "The influences of the spatial extent selection for non-landslide samples on statistical-based landslide susceptibility modelling: a case study of Anhui Province in China," *Natural Hazards*, vol. 112, no. 3, pp. 1967–1988, 2022.
- [21] Z. Liang, C. Wang, Z. Duan, H. Liu, X. Liu, and K. Ullah Jan Khan, "A hybrid model consisting of supervised and unsupervised learning for landslide susceptibility mapping," *Remote Sensing*, vol. 13, no. 8, p. 1464, 2021.
- [22] Y. Wang, X. Wu, Z. Chen, F. Ren, L. Feng, and Q. du, "Optimizing the predictive ability of machine learning methods for landslide susceptibility mapping using SMOTE for Lishui City in Zhejiang Province, China," *International Journal of Environmental Research and Public Health*, vol. 16, no. 3, p. 368, 2019.
- [23] W. Xie, X. Li, W. Jian et al., "A novel hybrid method for landslide susceptibility mapping-based geodetector and machine learning cluster: a case of Xiaojin County, China," *ISPRS International Journal of Geo-Information*, vol. 10, no. 2, p. 93, 2021.
- [24] K. Y. Almansi, A. R. M. Shariff, A. F. Abdullah, and S. N. Syed Ismail, "Hospital site suitability assessment using three machine learning approaches: evidence from the Gaza Strip in Palestine," *Applied Sciences*, vol. 11, no. 22, article 11054, 2021.

- [25] C. Yu and J. Chen, "Application of a GIS-based slope unit method for landslide susceptibility mapping in Helong City: comparative assessment of ICM, AHP, and RF model," *Symmetry*, vol. 12, no. 11, p. 1848, 2020.
- [26] Z. Bai, Q. Liu, and Y. Liu, "Landslide susceptibility mapping using GIS-based machine learning algorithms for the Northeast Chongqing Area, China," *Arabian Journal of Geosciences*, vol. 14, no. 24, p. 2831, 2021.
- [27] Z. Wang, Q. Liu, and Y. Liu, "Mapping landslide susceptibility using machine learning algorithms and GIS: a case study in Shexian County, Anhui Province, China," *Symmetry*, vol. 12, no. 12, p. 1954, 2020.
- [28] M. Zhang, X. Cao, L. Peng, and R. Niu, "Landslide susceptibility mapping based on global and local logistic regression models in Three Gorges Reservoir area, China," *Environmental Earth Sciences*, vol. 75, no. 11, p. 958, 2016.
- [29] D. T. Bui, H. Shahabi, A. Shirzadi et al., "Landslide detection and susceptibility mapping by AIRSAR data using support vector machine and index of entropy models in Cameron Highlands, Malaysia," *Remote Sensing*, vol. 10, no. 10, p. 1527, 2018.
- [30] B. Ghasemian, H. Shahabi, A. Shirzadi et al., "A robust deep-learning model for landslide susceptibility mapping: a case study of Kurdistan Province, Iran," *Sensors*, vol. 22, no. 4, p. 1573, 2022.
- [31] S. Bera, V. K. Upadhyay, B. Guru, and T. Oommen, "Landslide inventory and susceptibility models considering the landslide typology using deep learning: Himalayas, India," *Natural Hazards*, vol. 108, no. 1, pp. 1257–1289, 2021.
- [32] A. Shirzadi, H. Shahabi, K. Chapi et al., "A comparative study between popular statistical and machine learning methods for simulating volume of landslides," *Catena*, vol. 157, pp. 213–226, 2017.
- [33] V. H. Nhu, A. Shirzadi, H. Shahabi et al., "Shallow landslide susceptibility mapping by random forest base classifier and its ensembles in a semi-arid region of Iran," *Forests*, vol. 11, no. 4, p. 421, 2020.
- [34] S. Z. X. Liang, D. Chen, D. H. Li, Y. Qi, and Z. Zhao, "Spatial and temporal distribution of geologic hazards in Shaanxi Province," *Remote Sensing*, vol. 13, no. 21, p. 4259, 2021.
- [35] H. J. Qiu, Y. F. Cui, S. Hu, D. Yang, Y. Pei, and W. Yang, "Temporal and spatial distributions of landslides in the Qinba Mountains, Shaanxi Province, China," *Geomatics, Natural Hazards and Risk*, vol. 10, no. 1, pp. 599–621, 2019.
- [36] Q. Q. Wang, W. P. Li, W. Chen, and H. Y. Bai, "GIS-based assessment of landslide susceptibility using certainty factor and index of entropy models for the Qianyang County of Baoji city, China," *Journal of Earth System Science*, vol. 124, no. 7, article 19399, 2015.
- [37] R. Y. Gao, C. M. Wang, Z. Liang, S. Han, and B. Li, "A research on susceptibility mapping of multiple geological hazards in Yanzi River Basin, China," *ISPRS International Journal of Geo-Information*, vol. 10, no. 4, p. 218, 2021.
- [38] Y. T. Wang, A. C. Seijmonsbergen, W. Bouten, and Q. T. Chen, "Using statistical learning algorithms in regional landslide susceptibility zonation with limited landslide field data," *Journal of Mountain Science*, vol. 12, no. 2, pp. 268–288, 2015.
- [39] S. H. Xu, M. Zhang, Y. Ma et al., "Multiclassification method of landslide risk assessment in consideration of disaster levels: a case study of Xianyang City, Shaanxi Province," *ISPRS International Journal of Geo-Information*, vol. 10, no. 10, article 646, 2021.
- [40] C.-W. Chen, H. Chen, L.-W. Wei, G. W. Lin, T. Iida, and R. Yamada, "Evaluating the susceptibility of landslide landforms in Japan using slope stability analysis: a case study of the 2016 Kumamoto earthquake," *Landslides*, vol. 14, no. 5, pp. 1793–1801, 2017.
- [41] L. Li, H. Lan, A. Strom, and R. Macciotta, "Landslide longitudinal shape: a new concept for complementing landslide aspect ratio," *Landslides*, vol. 19, no. 5, pp. 1143–1163, 2022.
- [42] Z. Zhao, Z. Y. Liu, and C. Xu, "Slope unit-based landslide susceptibility mapping using certainty factor, support vector machine, random Forest, CF-SVM and CF-RF models," *Frontiers in Earth Science*, vol. 9, 2021.
- [43] Z. Gao and M. Ding, "Application of convolutional neural network fused with machine learning modeling framework for geospatial comparative analysis of landslide susceptibility," *Natural Hazards*, 2022.
- [44] N. R. Regmi and J. I. Walter, "Detailed mapping of shallow landslides in eastern Oklahoma and western Arkansas and potential triggering by Oklahoma earthquakes," *Geomorphology*, vol. 366, Article ID 106806, 2020.
- [45] C. Martinello, C. Cappadonia, C. Conoscenti, and E. Rotigliano, "Landform classification: a high-performing mapping unit partitioning tool for landslide susceptibility assessment—a test in the Imera River basin (northern Sicily, Italy)," *Landslides*, vol. 19, no. 3, pp. 539–553, 2022.
- [46] I. C. Nicu, "Frequency ratio and GIS-based evaluation of landslide susceptibility applied to cultural heritage assessment," *Journal of Cultural Heritage*, vol. 28, pp. 172–176, 2017.
- [47] B. T. Pham, T. Nguyen-Thoi, C. Qi et al., "Coupling RBF neural network with ensemble learning techniques for landslide susceptibility mapping," *Catena*, vol. 195, article 104805, 2020.
- [48] T. Xiao, S. Segoni, L. Chen, K. Yin, and N. Casagli, "A step beyond landslide susceptibility maps: a simple method to investigate and explain the different outcomes obtained by different approaches," *Landslides*, vol. 17, no. 3, pp. 627–640, 2020.
- [49] M. Melillo, M. T. Brunetti, S. Peruccacci, S. L. Gariano, and F. Guzzetti, "An algorithm for the objective reconstruction of rainfall events responsible for landslides," *Landslides*, vol. 12, no. 2, pp. 311–320, 2015.
- [50] M. Avand, S. Janizadeh, D. T. Bui, V. H. Pham, P. T. T. Ngo, and V.-H. Nhu, "A tree-based intelligence ensemble approach for spatial prediction of potential groundwater," *International Journal of Digital Earth*, vol. 13, no. 12, pp. 1408–1429, 2020.
- [51] S. Dağ, A. Akgün, A. Kaya, S. Alemdağ, and H. T. Bostancı, "Medium scale earthflow susceptibility modelling by remote sensing and geographical information systems based multivariate statistics approach: an example from Northeastern Turkey," *Environmental Earth Sciences*, vol. 79, no. 19, p. 468, 2020.
- [52] S. Li, Z. Ni, Y. Zhao et al., "Susceptibility analysis of geohazards in the Longmen Mountain Region after the Wenchuan Earthquake," *International Journal of Environmental Research and Public Health*, vol. 19, no. 6, article 3229, 2022.
- [53] J.-W. Lin, M.-H. Hsieh, and Y.-J. Li, "Factor analysis for the statistical modeling of earthquake-induced landslides," *Frontiers of Structural and Civil Engineering*, vol. 14, no. 1, pp. 123–126, 2020.

- [54] X. Sun, J. Chen, X. Han, Y. Bao, X. Zhou, and W. Peng, "Landslide susceptibility mapping along the upper Jinsha River, South-Western China: a comparison of hydrological and curvature watershed methods for slope unit classification," *Bulletin of Engineering Geology and the Environment*, vol. 79, no. 9, pp. 4657–4670, 2020.
- [55] G. J. Hearn and R. P. Martin, "The performance of the Dharan-Dhankuta Road, east Nepal, in the context of the Fookes et al. (1985) mountain model: a tribute to the late Professor PG Fookes," *Bulletin of Engineering Geology and the Environment*, vol. 81, no. 4, article 132, 2022.
- [56] J. Sekajugo, G. Kagoro-Rugunda, R. Mutyebera et al., "Can citizen scientists provide a reliable geo-hydrological hazard inventory? An analysis of biases, sensitivity and precision for the Rwenzori Mountains, Uganda," *Environmental Research Letters*, vol. 17, no. 4, article 045011, 2022.
- [57] C. Zhou, K. Yin, Y. Cao et al., "Landslide susceptibility modeling applying machine learning methods: a case study from Longju in the Three Gorges Reservoir area, China," *Computers & Geosciences*, vol. 112, pp. 23–37, 2018.
- [58] Y. Li, F. Lin, L. Lou et al., "Performance analysis of landslide susceptibility assessment under different factor-filtering models," *Arabian Journal of Geosciences*, vol. 14, no. 12, p. 1160, 2021.
- [59] T. Y. Zhang, L. Han, W. Chen, and H. Shahabi, "Hybrid integration approach of entropy with logistic regression and support vector machine for landslide susceptibility modeling," *Entropy*, vol. 20, no. 11, p. 884, 2018.
- [60] X. Hu, H. Zhang, H. Mei, D. Xiao, Y. Li, and M. Li, "Landslide susceptibility mapping using the stacking ensemble machine learning method in Lushui, Southwest China," *Applied Sciences*, vol. 10, no. 11, p. 4016, 2020.
- [61] E. Moradi, M. Abdolshahnejad, M. Borji Hassangavyar et al., "Machine learning approach to predict susceptible growth regions of *Moringa peregrina* (Forssk)," *Ecological Informatics*, vol. 62, article 101267, 2021.
- [62] M. Mokhtari and S. Abedian, "Spatial prediction of landslide susceptibility in Taleghan basin, Iran," *Stochastic Environmental Research and Risk Assessment*, vol. 33, no. 7, pp. 1297–1325, 2019.
- [63] W. Chen, H. C. Chai, Z. Zhao, Q. Wang, and H. Hong, "Landslide susceptibility mapping based on GIS and support vector machine models for the Qianyang County, China," *Environmental Earth Sciences*, vol. 75, no. 6, article 474, 2016.
- [64] M. Ada and B. T. San, "Comparison of machine-learning techniques for landslide susceptibility mapping using two-level random sampling (2LRS) in Alakir catchment area, Antalya, Turkey," *Natural Hazards*, vol. 90, no. 1, pp. 237–263, 2018.
- [65] A. Tella and A. L. Balogun, "GIS-based air quality modelling: spatial prediction of PM10 for Selangor State, Malaysia using machine learning algorithms," *Environmental Science and Pollution Research*, 2021.
- [66] A. Arabameri, M. Santosh, S. Saha et al., "Spatial prediction of shallow landslide: application of novel rotational forest-based reduced error pruning tree," *Geomatics, Natural Hazards and Risk*, vol. 12, no. 1, pp. 1343–1370, 2021.
- [67] H. L. Nguyen, B. T. Pham, L. Son et al., "Adaptive network based fuzzy inference system with meta-heuristic optimizations for international roughness index prediction," *Applied Sciences*, vol. 9, no. 21, p. 4715, 2019.
- [68] M. Mehrabi, B. Pradhan, H. Moayedi, and A. Alamri, "Optimizing an adaptive neuro-fuzzy inference system for spatial prediction of landslide susceptibility using four state-of-the-art metaheuristic techniques," *Sensors*, vol. 20, no. 6, p. 1723, 2020.
- [69] A. Jaafari, M. Panahi, D. Mafi-Gholami et al., "Swarm intelligence optimization of the group method of data handling using the cuckoo search and whale optimization algorithms to model and predict landslides," *Applied Soft Computing*, vol. 116, article 108254, 2022.
- [70] W. Chen, H. Y. Hong, M. Panahi et al., "Spatial prediction of landslide susceptibility using GIS-based data mining techniques of ANFIS with whale optimization algorithm (WOA) and grey wolf optimizer (GWO)," *Applied Sciences*, vol. 9, no. 18, p. 3755, 2019.
- [71] T. Kavzoglu and A. Teke, "Predictive performances of ensemble machine learning algorithms in landslide susceptibility mapping using random forest, extreme gradient boosting (XGBoost) and natural gradient boosting (NGBoost)," *Arabian Journal for Science and Engineering*, vol. 47, no. 6, pp. 7367–7385, 2022.
- [72] H. J. Oh and B. Pradhan, "Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area," *COMPUTERS & GEOSCIENCES*, vol. 37, no. 9, pp. 1264–1276, 2011.
- [73] W. Chen, X. Zhao, H. Shahabi et al., "Spatial prediction of landslide susceptibility by combining evidential belief function, logistic regression and logistic model tree," *Geocarto International*, vol. 34, no. 11, pp. 1177–1201, 2019.
- [74] K. Xiong, B. R. Adhikari, C. A. Stamatopoulos et al., "Comparison of different machine learning methods for debris flow susceptibility mapping: a case study in the Sichuan Province, China," *Remote Sensing*, vol. 12, no. 2, p. 295, 2020.
- [75] V. G. Biju and C. Prashanth, "Friedman and Wilcoxon evaluations comparing SVM, bagging, boosting, K-NN and decision tree classifiers," *Journal Applied Computer Science Methods*, vol. 9, no. 1, pp. 23–47, 2017.
- [76] T. Chen, L. Zhu, R. Niu, C. J. Trinder, L. Peng, and T. Lei, "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.
- [77] J. Chen, S. T. Yang, H. W. Li, B. Zhang, and J. R. Lv, "Research on geographical environment unit division based on the method of natural breaks (Jenks)," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-4/W3, pp. 47–50, 2013.
- [78] A. Jaafari, A. Najafi, H. R. Pourghasemi, J. Rezaeian, and A. Sattarian, "GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran," *International Journal of Environmental Science and Technology*, vol. 11, no. 4, pp. 909–926, 2014.
- [79] M. Marjanović, M. Kovačević, B. Bajat, and V. Voženilek, "Landslide susceptibility assessment using SVM machine learning algorithm," *Engineering Geology*, vol. 123, no. 3, pp. 225–234, 2011.
- [80] F. E. S. Silalahi, Y. A. Pamela, and F. Hidayat, "Landslide susceptibility assessment using frequency ratio model in Bogor, West Java, Indonesia," *Geoscience Letters*, vol. 6, no. 1, 2019.
- [81] A. G. Bruzón, P. Arrogante-Funes, F. Arrogante-Funes et al., "Landslide susceptibility assessment using an AutoML framework," *International Journal of Environmental Research and Public Health*, vol. 18, no. 20, article 10971, 2021.

- [82] A. Grozavu and C. V. Patriche, "Mapping landslide susceptibility at national scale by spatial multi-criteria evaluation," *Geomatics, Natural Hazards and Risk*, vol. 12, no. 1, pp. 1127–1152, 2021.
- [83] Q. Yasin, M. Majdański, G. M. Sohail, and H. Vo Thanh, "Fault and fracture network characterization using seismic data: a study based on neural network models assessment," *Geomechanics Geophysics Geo-Energy and Geo-Resources*, vol. 8, no. 2, 2022.
- [84] B. T. Pham, B. Pradhan, D. T. Bui, I. Prakash, and M. B. Dholakia, "A comparative study of different machine learning methods for landslide susceptibility assessment: a case study of Uttarakhand area (India)," *Environmental Modelling & Software*, vol. 84, pp. 240–250, 2016.
- [85] Y. Song, R. Niu, S. Xu et al., "Landslide susceptibility mapping based on weighted gradient boosting decision tree in Wanzhou section of the Three Gorges Reservoir area (China)," *ISPRS International Journal of Geo-Information*, vol. 8, no. 1, p. 4, 2019.
- [86] R. Genuer, J. M. Poggi, and C. Tuleau-Malot, "Variable selection using random forests," *Pattern Recognition Letters*, vol. 31, no. 14, pp. 2225–2236, 2010.
- [87] U. Knauer, C. S. von Rekowski, M. Stecklina et al., "Tree species classification based on hybrid ensembles of a convolutional neural network (CNN) and random forest classifiers," *Remote Sensing*, vol. 11, no. 23, p. 2788, 2019.
- [88] Z. X. Zou, C. R. Xiong, H. M. Tang, R. E. Criss, A. Su, and X. Liu, "Prediction of landslide runout based on influencing factor analysis," *Environmental Earth Sciences*, vol. 76, no. 21, 2017.
- [89] H. Y. Hong, P. Tsangaratos, I. Ilia, C. Loupasakis, and Y. Wang, "Introducing a novel multi-layer perceptron network based on stochastic gradient descent optimized by a meta-heuristic algorithm for landslide susceptibility mapping," *Science of the Total Environment*, vol. 742, article 140549, 2020.
- [90] A. A. Shahri and F. M. Moud, "Landslide susceptibility mapping using hybridized block modular intelligence model," *Bulletin of Engineering Geology and the Environment*, vol. 80, no. 1, pp. 267–284, 2021.