scientific reports

OPEN

Evaluation of landslide susceptibility of mountain highway based on RF and SVM models

Qingfeng He[1]✉, Shoulong Wu[1], Xia Zhao[1]✉, Zhankun Hui[1], Zhengang Wang[2], Paraskevas Tsangaratos[3], Ioanna Ilia[3], Wei Chen[1], Yixin Chen[1] & Yiheng Hao[1]

Geological complexities along mountain highways frequently trigger landslides, posing significant threats to transportation safety and infrastructure. This study evaluates landslide susceptibility along the Lizha-Jiezi section of China's G345 national highway using Random Forest (RF) and Support Vector Machine (SVM) models. Eleven conditioning factors including altitude, slope, aspect, plan curvature, profile curvature, lithology, distance to fault, rainfall, distance to river, normalized difference vegetation index (NDVI), and distance to road were analyzed using remote sensing and field surveys. A landslide inventory of 67 events was divided into training (70%) and validation (30%) datasets, with non-landslide samples selected at least 100 m away from landslide locations to minimize spatial overlap. Factor contribution analysis identified distance to road as the most significant predictor, highlighting anthropogenic impacts on slope destabilization. Model validation via receiver operating characteristic (ROC) curves demonstrated RF's superior performance (AUC = 0.887) over SVM (AUC = 0.735). The RF-derived susceptibility map classified five risk levels, revealing high-risk zones concentrated within 200 m of roads, consistent with field observations. Results emphasize the necessity of integrating anthropogenic factors into landslide risk management for mountainous infrastructure. This study provides actionable insights for mitigation strategies and land-use planning, offering a scalable framework adaptable to similar regions.

Keywords Random forest, Support vector machine, Landslide susceptibility mapping, Mountain highway

Landslides, as prevalent natural geological processes in mountainous and hilly regions, pose significant threats to human safety, transportation infrastructure, economic development, and ecological environments. Currently, landslides are increasing globally, and all countries in the world are suffering from the threat of landslide disasters. The La Clapiere landslide is the largest landslide in France[1], and its rate of decline peaked between 1987 and 1988, raising fears of a catastrophic rupture. The Apennine region of Italy has many landslides formed due to tectonic activity and extreme weather, causing great economic losses[2,3]. Korea, where mountains and hills occupy most of its land area, has considerable landslide hazards[4–7]. China is also a typical disaster-stricken country, with mountains occupying two-thirds of its land area, especially in the Qinling region of China, with majestic mountains, gullies and ravines, and most of the roads built over the mountains and with large ups and downs, the geological environment conditions along the roads are complex and changeable, and the ecological environment is fragile, leading to frequent landslides. Landslides along the road can cause road burial, or even damage the road, blocking traffic and causing serious economic losses[8]. Therefore, landslide susceptibility evaluation[9] along mountainous advanced highways plays a vital role in ensuring the reliability of road construction and carrying out landslide prevention and control work.

In general, the landslide susceptibility evaluation can provide some basic basis for the prevention and control of landslides, and achieve comprehensive management with emphasis. Since the 1960s, domestic and foreign experts and scholars have studied many methods for landslide susceptibility evaluation. The most commonly used landslide susceptibility analysis models include the following two types: (1) Knowledge-driven models, which include expert scoring method, analytic hierarchy process (AHP)[10,11], fuzzy logic method[12–14] and fuzzy comprehensive evaluation method et al.[10,15]; (2) Data-driven models, which include the traditional data-driven

[1]College of Geology and Environment, Xi'an University of Science and Technology, Xi'an 710054, Shaanxi, China. [2]China Coal Science and Technology Ecological Environment Technology Co., LTD, Beijing 100013, China. [3]Laboratory of Engineering Geology and Hydrogeology, Department of Geological Sciences, School of Mining and Metallurgical Engineering, National Technical University of Athens, 15780 Athens, Greece. ✉email: heqf@xust.edu.cn; 19209071022@stu.xust.edu.cn

1

model: information model[16–18], frequency ratio model (FR)[19–22], logistic regression model[23–26], and machine learning models: artificial neural network model (ANN)[22,27], support vector machine model (SVM)[28–33], random forest model (RF)[19,34–37], alternative decision tree model et al.[38–47].

The modeling and quantification process of knowledge-driven model is simple, and its modeling principle is closely related to the characteristics of landslide conditioning factors, while it is subjective. Zhu et al.[48] took Kaixian and the Three Gorges study area as the research object and proposed landslide susceptibility mapping based on expert knowledge. Moragues et al.[49] used the AHP and weighted linear combination method to evaluate the susceptibility to slope instability process of the Argentine Hubei branch of the South Patagonia ice sheet. Pourghasemi et al.[50] used fuzzy logic method and AHP model to make LSMs of the Kharaz basin in Iran. Zhao et al.[10] used Shannon entropy theory, fuzzy comprehensive evaluation method and AHP to establish the landslide susceptibility model in the study area and evaluate the susceptibility. The above research shows that with the development of technology and the deepening of research, people's understanding of landslide is further improved, and more and more knowledge fields are involved in landslide data. The knowledge-driven model based on the knowledge and experience in experts is greatly affected by the level of experts, subjective and unable to effectively deal with the huge landslide data and the respective characteristics of landslide conditioning factors.

The data-driven model excavates the internal relationship between conditioning factors and landslide through objective mathematical statistics analysis, so as to realize the prediction of landslide susceptibility. Akbar and Ha[51] obtained the landslide data of Kahan Valley in Western Himalayas of Pakistan through 3 S technology, and evaluated the landslide susceptibility of the region based on the information content model; Berhane et al.[52] constructed landslide inventory maps through field investigation, stereo aerial photography analysis and image interpretation, and then used frequency ratio model to generate LSM to complete susceptibility evaluation; Schlogel et al.[53] combined three different resolution DEMs with the selection of landslide or source area to obtain sample data sets, and then established a logistic regression model to evaluate the landslide susceptibility in the study area. The above research is based on the traditional data-driven model for landslide susceptibility evaluation. Those evaluation results show that the traditional data-driven model lacks the ability to excavate and analyze the interaction between landslide conditioning factors, and fails to consider the linear and uncertainty characteristics of landslide caused by conditioning factor. As a new type of data-driven model, machine learning model makes up for the shortcomings of strong subjectivity and low prediction accuracy of knowledge-driven model, as well as the deficiency of the traditional data-driven model in analyzing the interaction between conditioning factors. Machine learning model guarantees the accuracy and reliability of calculation and analysis results through objective mathematical statistics analysis. Park et al.[54] compared the ability of FR, AHP, logistic regression and ANN models to generate LSM. The results showed that the accuracy of these four models was roughly similar, but the ANN was slightly higher than the other three models. Pourghasemi and Kerle[55] used GIS-based RF for LSM in the western part of Mazandalan province, northern Iran; Chen et al.[56] used three data mining techniques, namely, adaptive neuro-fuzzy inference system combined with frequency ratio (ANFIS-FR), generalized additive model (GAM) and SVM, to evaluate the landslide susceptibility of Hanyuan, and the SVM model had the highest prediction accuracy[57]; Zhao et al.[19] examine the effectiveness of various machine learning models for landslide susceptibility prediction at different spatial resolutions, finding that higher resolutions and integrated models significantly improve accuracy, with the RS-ADT model at 12.5 m resolution performing best; Pandey et al.[58] used boosted regression tree (BRT), generalized linear model (GLM), RF and SVM models to analyze the susceptibility of landslides along the highway corridor from Nahan to Rajgarh. The results showed that RF model had the highest prediction accuracy, followed by SVM model. Many scholars have shown that RF and SVM models have good applicability and high prediction accuracy[59].

Highway plays an important role in the economic development of a region[60]. However, the frequent occurrence of landslides along the highway[61] has brought certain difficulties to the construction, operation and maintenance of highway projects[62]. Therefore, landslide susceptibility assessment along mountain highways plays a vital role in reducing landslide disasters along mountain highways. The G345 highway Zhen'an section is one of the key highway construction projects in Shaanxi Province, located in Zhen'an County, Shaanxi Province. Zhen'an County is located in the middle of the southern of Qinling Mountains. The geological structure is complex, the geological environment is poor and the mountains are crisscrossed in Zhen'an. At the same time, the precipitation is abundant, and there are many types of landslides. It is one of the high-prone areas of landslides in Shaanxi Province. In this paper, the Lizha-Jiezi section of G345 national highway and its surrounding area within a 2 km range is taken as the study area. Comprehensively considered the topography, geological structure, meteorological hydrology, vegetation cover and human activities, 11 conditioning factors were selected: altitude, slope, aspect, plan curvature, profile curvature, lithology, distance to fault, rainfall, distance to river, normalized difference vegetation index (NDVI), and distance to road. The RF and SVM models are applied to obtain the LSM of the study area, which can provide important reference for landslides prevention and mitigation, risk assessment, engineering construction, land use planning and economic development along highways in the study area.

Study area

The total length of the national highway in the study area is 85.771 km, which is located in Zhen'an County, Shangluo City, Shaanxi Province, China (Fig. 1). The geographical coordinates are 108°33′58″–109°06′32″E, 33°17′04″–33°30′45″ N, east from Jiezi, west to Lizha, which is an important part of G345 national highway.

The study area belongs to the subtropical climate zone and is warm and humid with abundant rainfall. However, the mountainous terrain is complex, and the higher the altitude is, the lower the temperature is, showing great changes in the vertical direction. The inter-annual precipitation changes greatly, and the annual
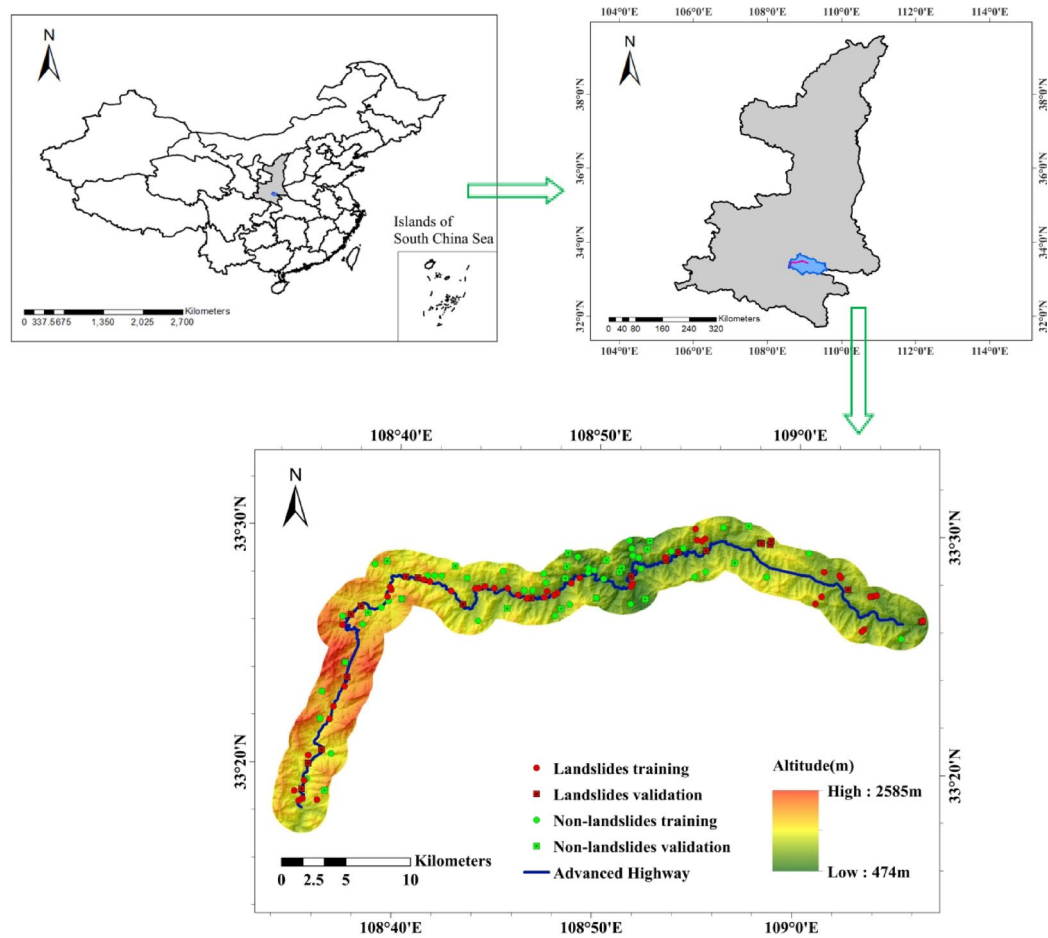
Fig. 1. Study area. (Note: this figure is made using ArcGIS desktop 10.8 (ArcMap component) https://www.esri.com/zh-cn/arcgis/products/arcgis-desktop/overview).

precipitation distribution is also very uneven. The seasonality is obvious, showing a single peak. The precipitation is large from June to October, and the highest is in July.

The terrain of the study area is generally high in the northwest and low in the southeast, and the terrain of the approach area is undulating. The highest altitude is 2585 m, the lowest is 474 m, and the height difference is 2111 m. Zhen'an-Banyan inverse fault is developed in the study area, Its occurrence is about 5°~25°∠50°~85°, and it basically runs through the longitudinal highway. Moreover, the fault zone is generally northward.

Data preparation

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn.

Landslide inventory mapping

There are many methods to construct the landslide inventory map. This study obtained the data of the study area and constructed the landslide inventory map by consulting historical data, remote sensing image map, geological map and detailed field investigation. A total of 67 landslides were found during the field investigation in the study area, including 42 slides and 25 falls. Among them, 58 small landslides accounted for 86.57% of the total number of landslides, and 9 medium-sized landslides accounted for 13.43% of the total number of landslides. Typical slides are slide3 (Fig. 2a), with a volume of $1.2 \times 10^4 m^3$, typical falls as fall7 (Fig. 2b), volume of $1.62 \times 10^4 m^3$. Through the analysis and comparison of landslide data, 67 landslides in the study area were randomly divided into 70% of the training set samples (47)[63–66] and 30% of the validation set samples (20)[67–69]. In order to construct the landslide susceptibility evaluation model, an equal number of non-landslide samples (67) were randomly selected from the study area 100 m away from the landslides, and randomly divided into 70% of the training set samples (47) and 30% of the validation set samples (20). The training set samples were used to construct the RF and SVM models, the validation set samples were used to verify the generalization ability of the model[70], and the reliability of the model was investigated.

Fig. 2. Example of landslides photos in the study area.

Landslide conditioning factors

Considering the topography, geological structure, meteorological hydrology, vegetation cover and human activities in the study area, a total of 11 conditioning factors were selected[71–74]: altitude, slope, aspect, plan curvature, profile curvature, lithology, distance to fault, rainfall, distance to river, NDVI and distance to road.

The altitude reflects the potential energy of the slope and largely controls the stability of the slope. Altitude affects the water content and stress in the slope, as well as the intensity of human engineering activities and the vegetation distribution on the slope surface. The altitude map (Fig. 3a) is realized by ASTER GDEM data with a resolution of 30 m collected by Geospatial Data Cloud (http://www.gscloud.cn/search). In addition, DEM data are used to generate slope (Fig. 3b), aspect (Fig. 3c), plan curvature (Fig. 3d) and profile curvature (Fig. 3e) by GIS.

As the material basis of landslide, lithology largely affects the composition of rock and soil, particle size, uniformity, loose degree, weathering degree and other aspects. The strength of different lithology is very different, and the impact on landslide is also different. Therefore, the lithology affects the stability of the slope, which is one of the important factors in the formation of landslide. Lithology obtained from field surveys and detailed survey data is divided into four categories (Fig. 3f). Under the combined influence of internal and external dynamic geological effects, the rock strata will deform. Once the stress exceeds its own strength limit, the rock cannot maintain the original complete shape and ruptures. When the dislocation of rock position reaches a critical value, a fault will be formed. Faults affect the development of landslides, and the closer to the fault, the greater the possibility of landslides, showing strong control. According to the topographic map, the fault distribution is obtained, and the distance to the fault is obtained by the buffer analysis function of GIS (Fig. 3g).

Rainfall is the main trigger factor of landslides[75], which is manifested in the infiltration of rainwater into rock and soil, resulting in rock softening. With the addition of a large amount of rainwater, the pore pressure will increase while the moisture content of rock and soil increases, which will reduce the effective stress of rock and soil and provide conditions for the development of landslides. Rainfall data come from Shaanxi Hydrology and Water Resources Information Network (http://www.shxsw.com.cn/) (Fig. 3h). Long-term river cutting and side erosion of bank slope, especially in flood season, increased the slope front surface to a large extent, created natural conditions for the occurrence of landslides, mainly in the spatial distribution of landslides. According to the previous research experience and field investigation in this study area, the distance to the river is obtained by using GIS buffer analysis function (Fig. 3i).

The main role of vegetation is to protect the stability of slope and reduce soil erosion. In general, the lusher the vegetation is, the lower the development degree of landslides will be, but it does not play a decisive role in the development of landslides. NDVI (Fig. 3j) extracted from Landsat 8 OLI_TIRS (http://www.gscloud.cn/search) remote sensing image can accurately reflect the coverage of surface vegetation.

The reconstruction and expansion project of mountainous highways inevitably excavates the slope, thus forming artificial high and steep slope. Under the influence of external factors, especially rainfall, the loose rock and soil on the slope collapses, forming landslides such as slides and falls, which may cause traffic congestion and hurt pedestrians. According to the topographic map combined with the field investigation in this study area, the distance between the study area and the road is obtained by using GIS buffer analysis function (Fig. 3k).

Methods

This study is divided into five steps: (1) Using GIS for data preparation, get grid point attribute data. (2) Correlation analysis of indicators. (3) Establishing RF and SVM respectively to get landslide susceptibility index. (4) Validation of the models by ROC. (5) The landslide susceptibility data is imported into GIS for generating the LSM.

Correlation analysis of conditioning factor

Landslide is affected by a variety of factors, and the 11 conditioning factors may have a certain correlation. In the later construction of the models, multiple collinearity problems may occur, which affect the accuracy of the models. Therefore, it is very important to analyze the correlation between indicators and eliminate redundant conditioning factors. Therefore, this paper uses Pearson correlation coefficient[76–78] to analyze the correlation of each factor in the study area. In GIS, the attribute data of each factor layer in the study area are extracted and
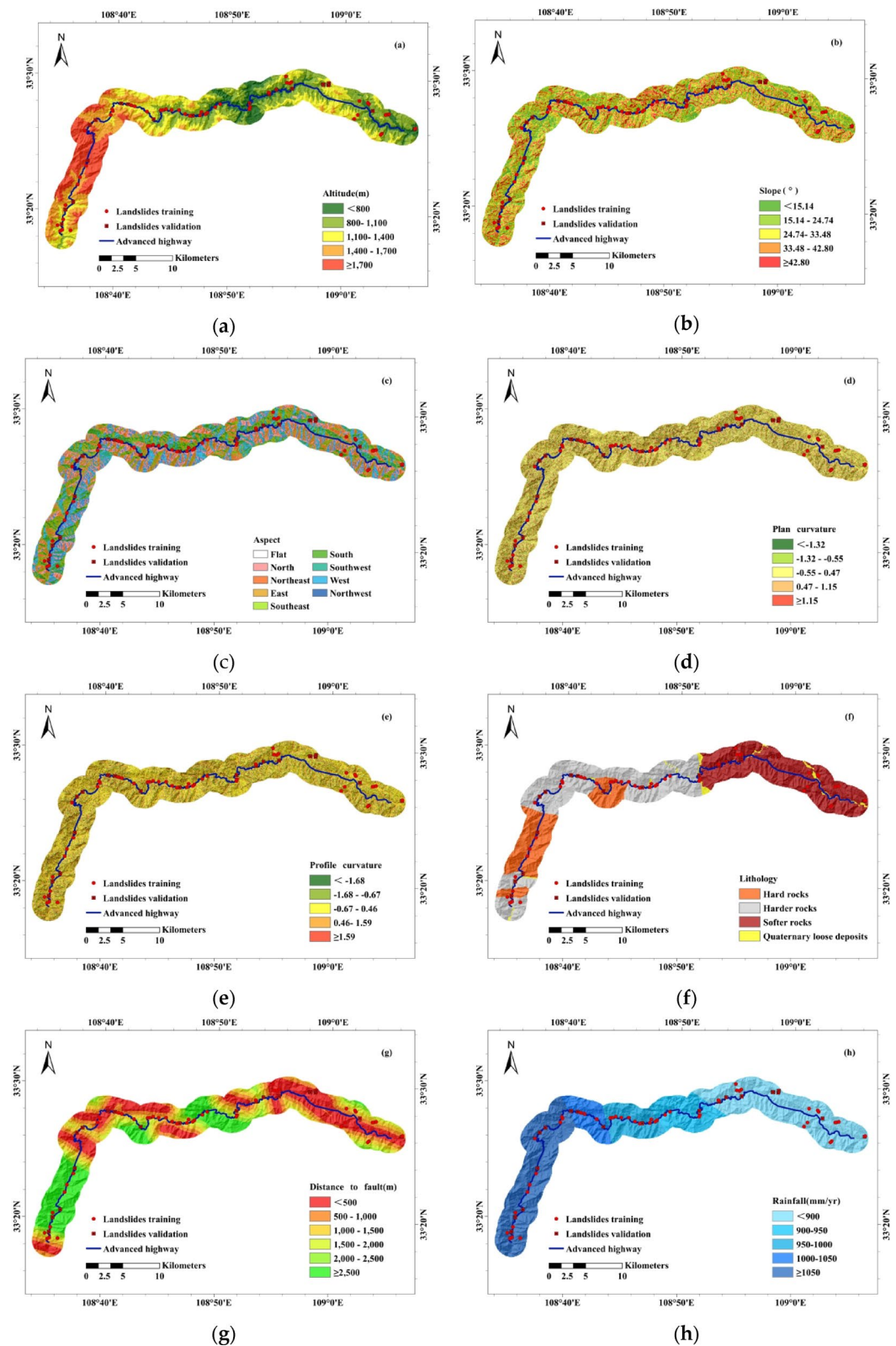
Fig. 3. Spatial distribution of landslides conditioning factors (Note: this figure is made using ArcGIS desktop 10.8 (ArcMap component) https://www.esri.com/zh-cn/arcgis/products/arcgis-desktop/overview). (**a**) Altitude, (**b**) slope, (**c**) aspect, (**d**) plan curvature, (**e**) profile curvature, (**f**) lithology, (**g**) distance to fault, (**h**) rainfall, (**i**) distance to river, (**j**) NDVI, (**k**) distance to road.

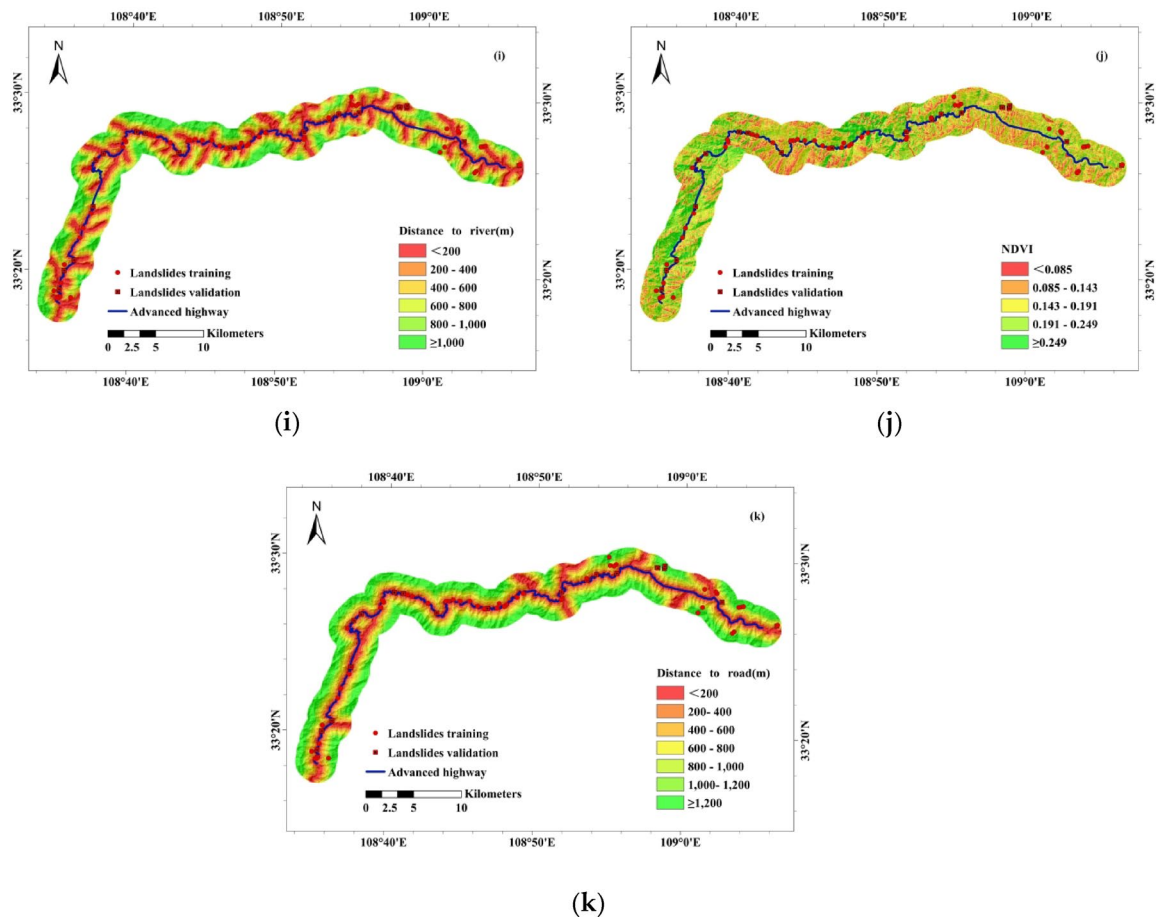(**i**)

(**j**)

(**k**)

Fig. 3. (continued)

imported into Matlab. The corr function is used to analyze the correlation between each factor. The correlation coefficient R is used to measure the correlation between each factor, and the value range of R is [-1,1][79]. Here |R| is used to measure the correlation between factors. As shown in Table 1, the greater the |R| is, the closer it is to 1, which indicates that the correlation between the two factors is higher. A set of conditioning factor system suitable for this study area is established by excluding highly relevant indicators.

Random forest

RF was first proposed by Breiman[80] as an ensemble algorithm. RF is a set of decision trees constructed by random way to form a forest. The number of decision trees in the forest can be customized, and there is no direct relationship between the constructed decision trees. The RF uses the bootstrap method to model the decision tree for each bootstrap[81] sample from the original sample, and then combines the decision trees[82] together to obtain the final classification or prediction results by voting. RF model reduces the correlation between any two decision trees and avoids model overfitting by randomizing sample data and feature selection. Figure 4 illustrates the process of building a RF model. A large number of theoretical and empirical studies have proved that RF algorithm has high prediction accuracy, strong generalization ability and fast training speed.

Support vector machine

SVM is a new learning machine based on statistical learning theory[83,84]. In general, the traditional statistical analysis method uses the principle of minimum empirical risk to construct a mathematical model. This method is applicable under the condition of large enough samples. However, when the sample size is small, both the size

	Value ranges	Degree of relatedness
	<0.4	Low correlation
|R|	0.4~0.7	Explicit relevance
	0.7~1	High correlation

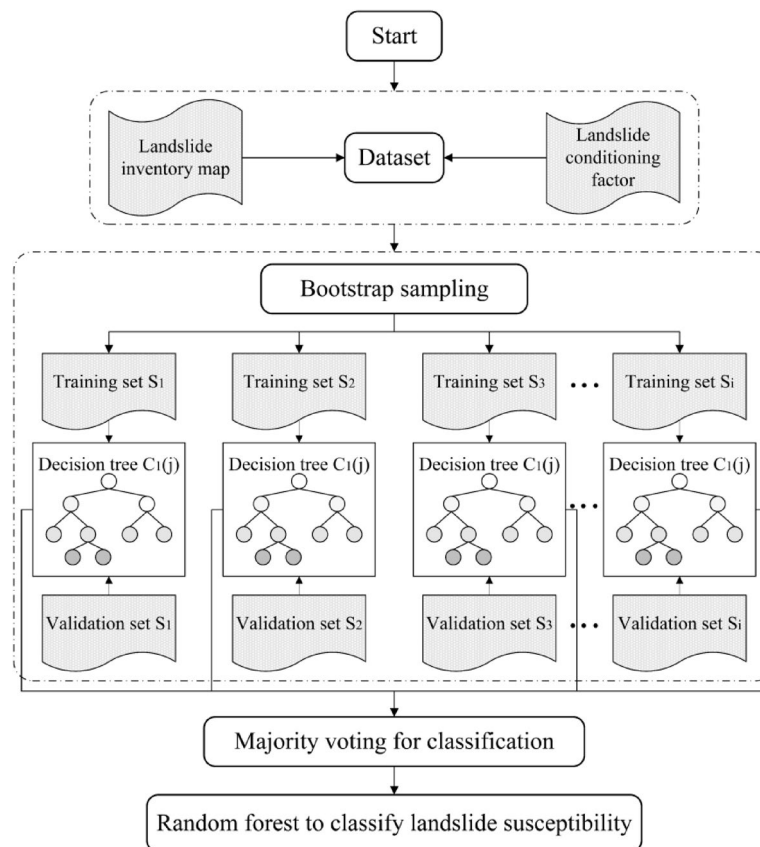**Table 1.** Conditioning factor correlation table.

Fig. 4. Flowchart of the random forest method.

of empirical risk and the complexity of the model should be considered, and efforts should be made to minimize the complexity of the model. Based on the above situation, the Vapnik comprehensively considered the empirical risk and the VC dimension of the model (representing the complexity of the model), and compromised the size of the two to minimize the approximate actual risk. In practical application, the main goal of this method is to minimize the sum of generalization error rate of the model and the amount of VC dimension. In general, compared with the traditional statistical methods, SVM has higher prediction accuracy and good generalization ability.

The main idea of SVM is to find an optimal surface so that the interval between all samples and the surface is maximized and the hyperplane is expressed as follow:

$$f\left(x\right) = \left[\omega \times \phi\left(x\right)\right] + b \tag{1}$$

In the formula: ω is feature vector of hyperplane in high dimensional space, ϕ is mapping function from low dimensional space to high dimensional space, b is threshold.

Landslide susceptibility evaluation is a typical nonlinear problem, which is affected by many factors. Therefore, this paper uses nonlinear mapping function to transform low-dimensional space to high-level space. After the original dimension data are converted into high-dimensional spatial data, the hyperplane is searched by using the kernel function to maximize the hyperplane interval of the sample. Radial basis function (RBF) kernel can provide more accurate prediction results in most classification models, especially in nonlinear environments[85–88]. Therefore, RBF kernel is selected for SVM modeling in this landslide susceptibility evaluation.

$$K\left(x_i, x\right) = exp\left(-\gamma \left\|x_i - x\right\|^2\right) \tag{2}$$

In the formula: $K\left(x_i, x\right)$ is kernel function, γ is parameter of a kernel function.

Results
Conditioning factor analysis
With the increase of altitude, the density of landslides decreased significantly (Fig. 5a). In the range of altitude < 800 m, the density of landslides is the largest, which is 0.4998/km². Landslides are mainly distributed in the range of altitude < 1400 m, a total of 60, accounting for 89.55% of the total. In this range, there are many rivers, and human engineering activities are frequent. It is easy to form steep slopes, if induced by rainfall and river scour, it is easy to cause landslides. Within the slope range of < 33.48° (Fig. 5b), 53 landslides were developed
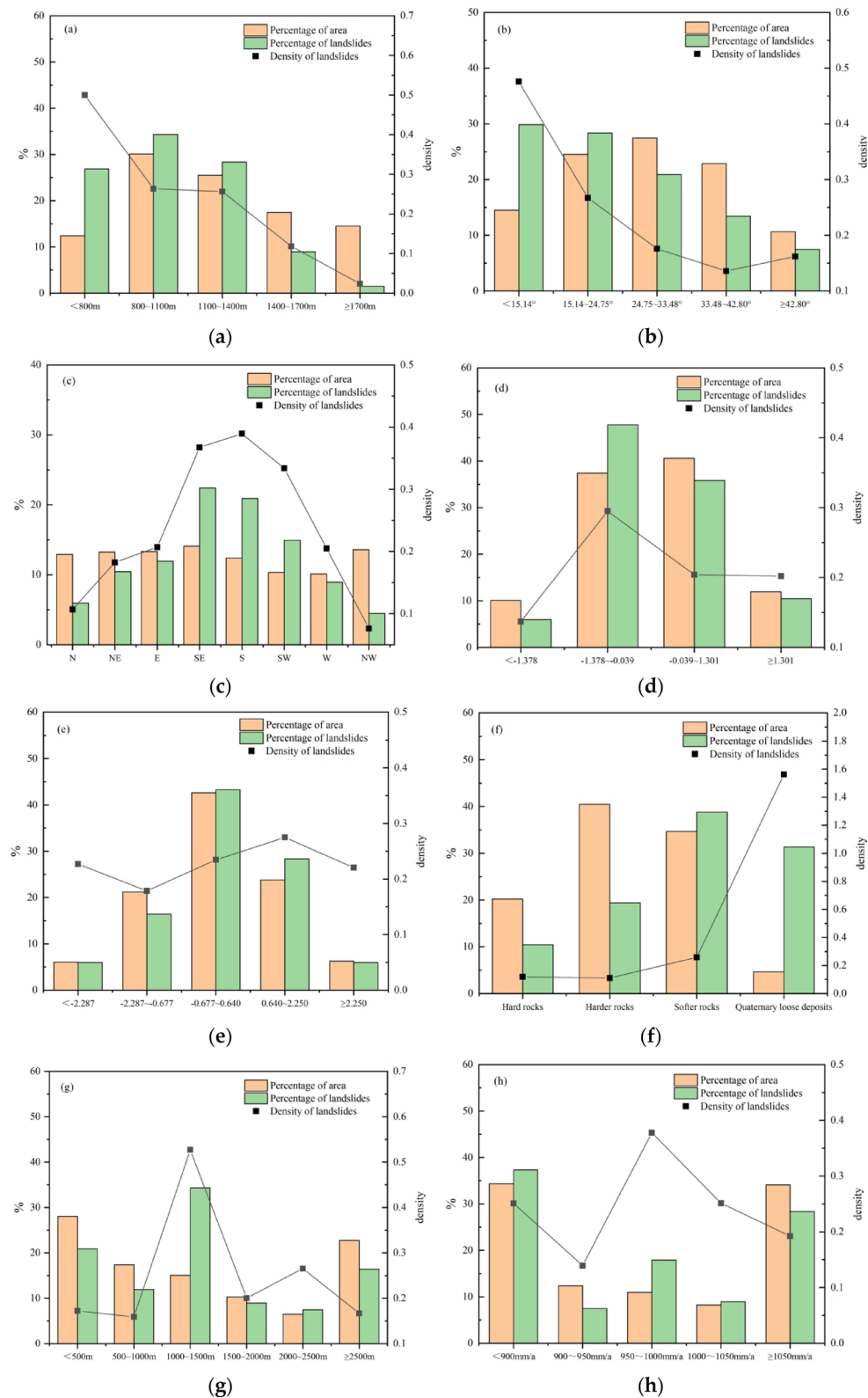
Fig. 5. Relationship between conditioning factors and landslide (**a**) altitude, (**b**) slope, (**c**) aspect, (**d**) plan curvature, (**e**) profile curvature, (**f**) lithology, (**g**) distance to fault, (**h**) rainfall, (**i**) distance to river, (**j**) NDVI, (**k**) distance to road.
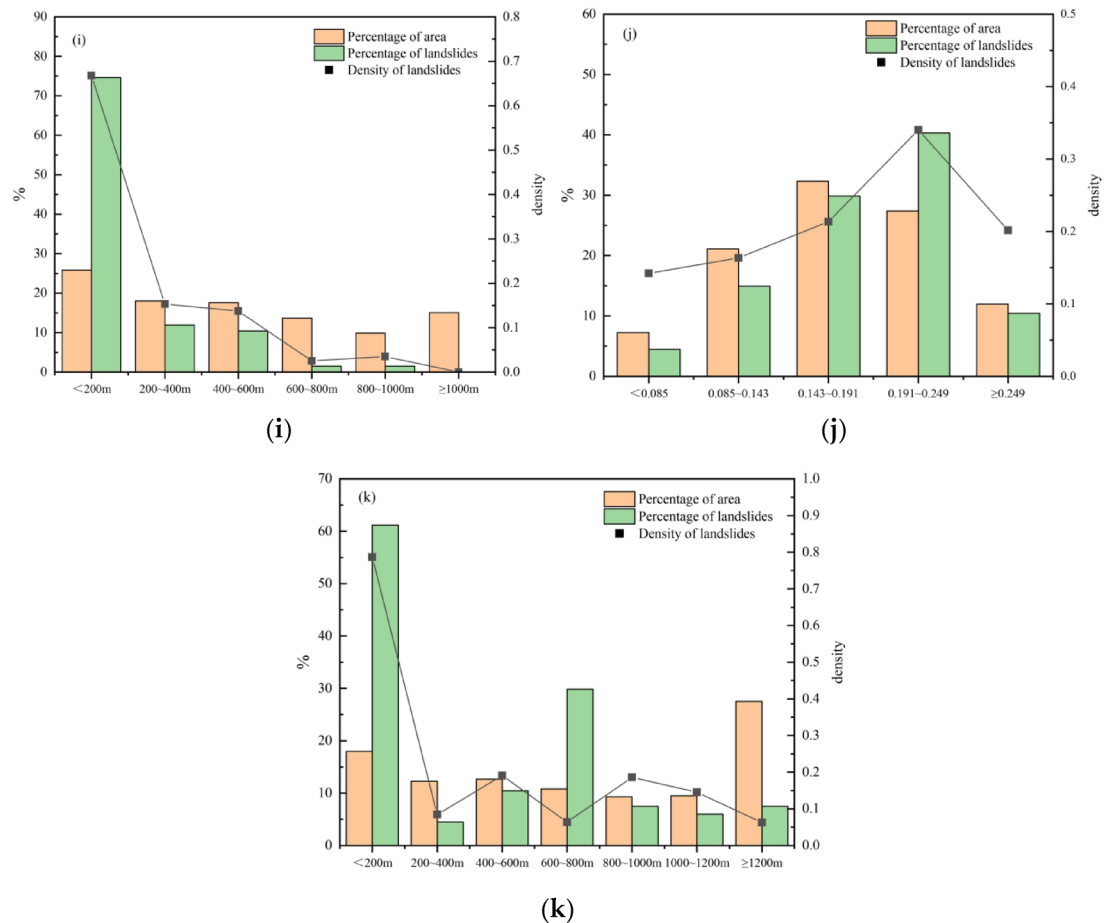
(i)

(j)

(k)

Fig. 5. (continued)

in the study area, accounting for 79.1% of the total. When the slope is within the range of < 15.14°, the density of landslide is the largest, which is 0.4757/km^2. In this range, frequent human engineering activities combined with the slope affect the effective surface of the slope. If induced by rainfall, it is easy to cause landslides. The change of landslide density in slope direction shows obvious single peak, and the maximum landslide density is 0.3894/km^2 in S (157.5°~202.5°) (Fig. 5c). The landslide is mainly distributed in SE (112.5°~157.5°), S (157.5°~202.5°), SW (202.5°~247.5°), a total of 39, accounting for 58.21% of the total. In the case of similar geological environment conditions and human engineering activities, the hydrothermal condition of sunny slope is better, the internal water of rock and soil is more likely to reach saturation, and it is easier to form landslides. In the range of plane curvature − 1.378 ~ 1.301, landslide distribution is more (Fig. 5d), a total of 56, accounting for 83.58% of the total. In the range of plan curvature − 1.378~ -0.039, the density of landslide is the largest, which is 0.2951/km^2. In the range of section curvature − 2.287 ~ 2.250, landslide distribution is more (Fig. 5e), a total of 59, accounting for 88.06% of the total. In the range of section curvature 0.640 ~ 2.250, the density of landslide is the largest, which is 0.2749/km^2.

The lithology in the study area is mainly hard rock group, and 13 landslides are developed (Fig. 5f), accounting for 19.4% of the total. However, the density of the landslide is only 0.1107/km^2. The lithology of the stratum is mainly limestone, slate and dolomite. The local weathering is strong, the joints and fissures are developed, and the rock mass is relatively broken. If the induced factors such as rainfall are encountered, it is easy to form landslides. From the perspective of landslide density, the landslide density of Quaternary loose deposits is 1.5625/km^2, and the density is the largest. It is mainly Quaternary alluvial-flood deposits and residual slope deposits, which are distributed in floodplains, valleys, foothill slopes and other parts. Here, human engineering activities are frequent. Building houses, road construction and excavation of slope toe have a great impact on the geological environment, and are easy to form landslides. Overall, with the weakening of rock hardness, the overall trend of disaster density is increasing. The distance to the fault in the study area is mainly within the range of < 1500 m, and 45 landslides (Fig. 5g) are developed, accounting for 67.16% of the total. Overall, the closer to the fault, the greater the landslide density is. Especially, it is controlled by several major faults such as Zhen'an-Banyan fault, and it is easier to form landslides due to the joint action of several faults.

The landslides in the study area are distributed in 25 places (Fig. 5h) where the annual precipitation is less than 900 mm/a, accounting for 37.31% of the total. Because the lithology is soft rock and the faults are densely distributed there, the landslides are most distributed. In the range of 950 ~ 1000 mm/a, the distribution area is small and affected by two faults, so the density of landslide is the largest. Rainfall, as the main external inducing

condition, is easy to form landslides. However, because the lithology of the western mountainous area is intrusive rock group, metamorphic rock group and a small amount of Quaternary loose strata, the density of landslides decreases. The main distribution distance of landslide in the study area is less than 200 m from the river, a total of 50 (Fig. 5i), accounting for 74.63% of the total, and the density of landslide is the largest. With the increase of distance to the river, the number of landslides and the density of landslides decreased significantly. Overall, affected by river erosion, it is easy to form landslides.

Landslides in the study area are mainly distributed in NDVI of 0.085 ~ 0.249 (Fig. 5j), a total of 57, accounting for 85.07% of the total. In the range of 0.191 ~ 0.249, the density of landslide is the largest, and the degree of vegetation development can protect slope stability and reduce soil erosion to a certain extent, but it does not play a decisive role.

The landslides in the study area are mainly distributed within the distance of < 200 m to the road (Fig. 5k), There are a total of 41, accounting for 61.19% of the total. And the density of landslides is the largest, which is 0.8806/km^2. With the increase of distance to the road, the number of landslides and the density of landslide decreased significantly. Overall, human engineering activities on slope foot damage, more likely to form landslides.

The attribute data of each factor in the study area were extracted by GIS, and the data were imported into Matlab. The correlation analysis of each index was carried out by corr function, and the Pearson correlation coefficient matrix Table 2 between the two indexes was obtained.

It can be concluded from the Table 2 that the correlation coefficient between most indicators is |R| < 0.4, indicating that most indicators are lowly correlated, only the correlation coefficient between individual indicators is 0.4≤|R|<0.7, which indicates that individual indicators are significantly correlated. And there is no high correlation between indicators. In summary, the 11 conditioning factors selected in this paper can form an evaluation system.

Random forest

In the validation set of 40 samples, the prediction results of 9 samples are inconsistent with the actual situation (Fig. 6a), Moreover, the prediction accuracy is 77.5% (Fig. 6b), indicating the prediction results are relatively good in landslide susceptibility evaluation. The grid attribute data of landslide conditioning factors in the study area extracted from GIS were imported into the RF classifier after training based on Matlab, and the landslide susceptibility index of each grid point data in the whole study area was obtained by prediction and calculation. The landslide susceptibility index was imported into GIS to generate the LSM of the study area. Then the natural discontinuity point method is used to classify the LSM into five grades[89,90], including: very low, low, moderate, high, very high. Finally, the LSM based on RF model is completed (Fig. 6c).

Support vector machine

In the SVM model, the selection of the kernel function and the corresponding parameters of the kernel function have a decisive influence on the later construction of the model[91]. In this study, the RBF is used as the kernel function. Parameters c and g can be selected by using the cross-validation method in the model construction stage. The range of parameters c and g[92] in this study is [-3.5, 3]. After inputting all kinds of boundary range parameters into the Matlab cross-validation model, when the parameters c and g are 0.93303 and 1, respectively, the cross-validation accuracy reaches the maximum, and the optimal accuracy is 76.5957%. Therefore, best c = 0.93303, best g = 1 (Fig. 7a).

The fitting degree of the classifier model based on SVM to the validation set samples is 72.5% (Fig. 7b), indicating that the model has high fitting degree and high prediction accuracy.

The grid attribute data of landslide influence factors in the study area extracted from GIS are imported into the SVM classifier based on Matlab[93–99]. The landslide susceptibility index of each grid point data in the whole study area is predicted and calculated. The LSM is obtained by GIS, and the natural discontinuity method is used to classify the LSM into five grades, including: very low, low, moderate, high, very high. Finally, the LSM based on SVM (Fig. 7c) is completed.

	a	b	c	d	e	f	g	h	i	j	k
a	1										
b	0.029	1									
c	0.019	0.049	1								
d	0.028	0.000	0.016	1							
e	− 0.062	0.001	− 0.010	− 0.615	1						
f	0.664	0.027	− 0.044	− 0.001	− 0.001	1					
g	− 0.626	− 0.018	− 0.076	− 0.004	0.021	− 0.619	1				
h	0.087	− 0.288	− 0.052	− 0.021	0.016	0.187	− 0.166	1			
i	0.388	− 0.009	0.078	− 0.006	− 0.047	0.090	− 0.167	− 0.080	1		
j	0.302	0.022	− 0.003	− 0.001	− 0.001	0.402	− 0.483	0.210	0.030	1	
k	0.350	0.004	0.051	− 0.011	− 0.034	0.075	− 0.189	− 0.064	0.696	0.089	1

Table 2. Correlation coefficient matrix of evaluation indexes.

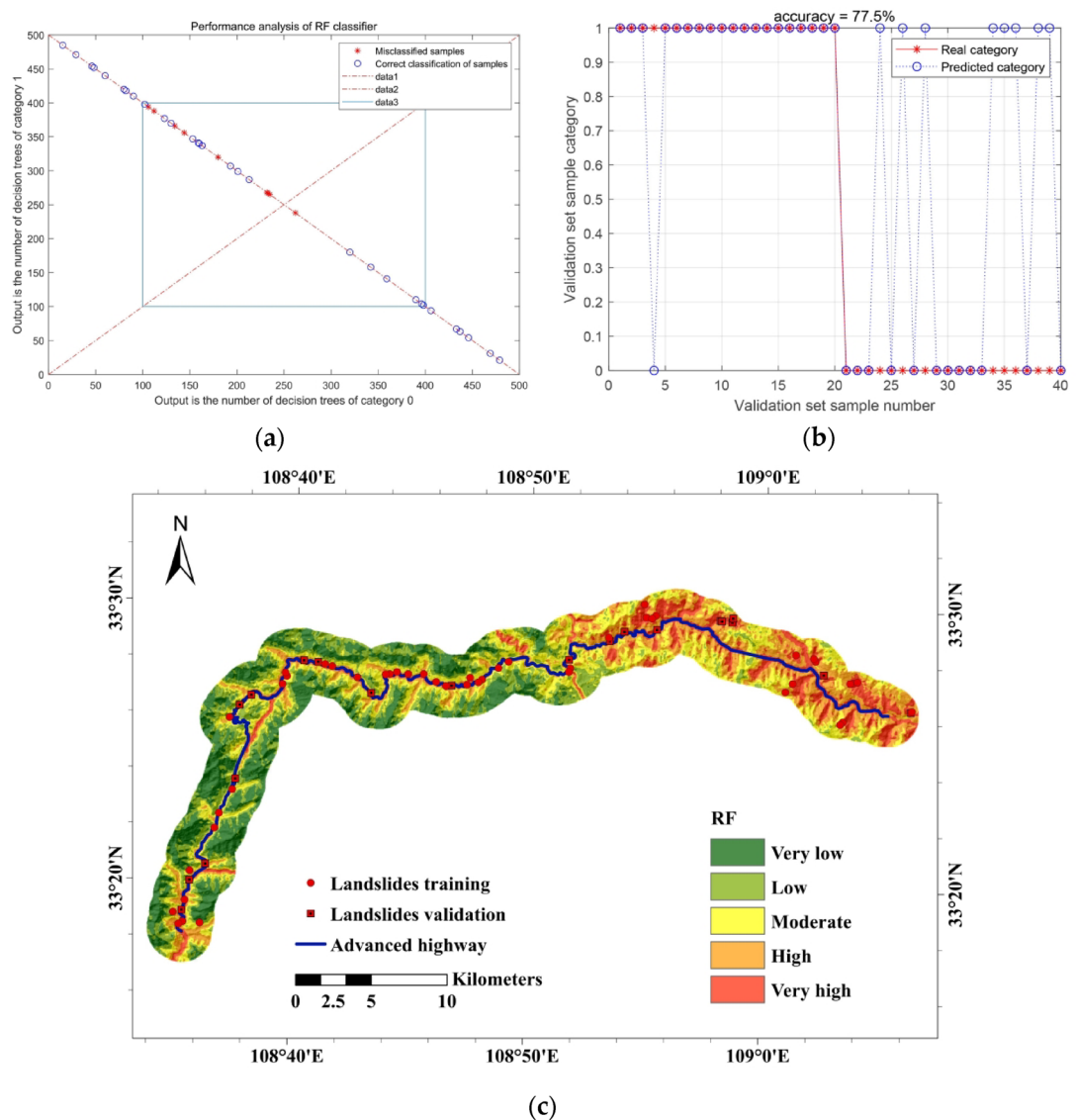**Fig. 6**. (**a**) Performance analysis of RF classifier; (**b**) validation set prediction results of RF model; (**c**) landslides susceptibility map of RF model. (Note: this figure is made using ArcGIS desktop 10.8 (ArcMap component) https://www.esri.com/zh-cn/arcgis/products/arcgis-desktop/overview).

Model validation

When evaluating the performance of predictive models, the confusion matrix serves as an effective analytical tool. This matrix is presented as a two-dimensional square array, with its core structure comprising four key elements: TN (True Negative, indicating correct identification of negative samples), FN (False Negative, denoting erroneous classification of positive samples as negative), TP (True Positive, representing accurate identification of positive samples), and FP (False Positive, referring to incorrect classification of negative samples as positive).

Based on these four fundamental parameters, this study calculates the Positive Predictive Value (PPV), Negative Predictive Value (NPV), and Accuracy (ACC) as shown in Table 3, which serve as statistical error evaluation criteria to enable objective assessment and comparison of different model performances. Consequently, in the study area, the RF model demonstrates higher ACC values than the SVM model. Specifically, the RF model correctly classified 78.1% (PPV) of the dataset as landslide points and 82.8% (NPV) as non-landslide points.

$$PPV = \frac{TP}{FP + TP} \tag{3}$$

$$NPV = \frac{TN}{FN + TN} \tag{4}$$

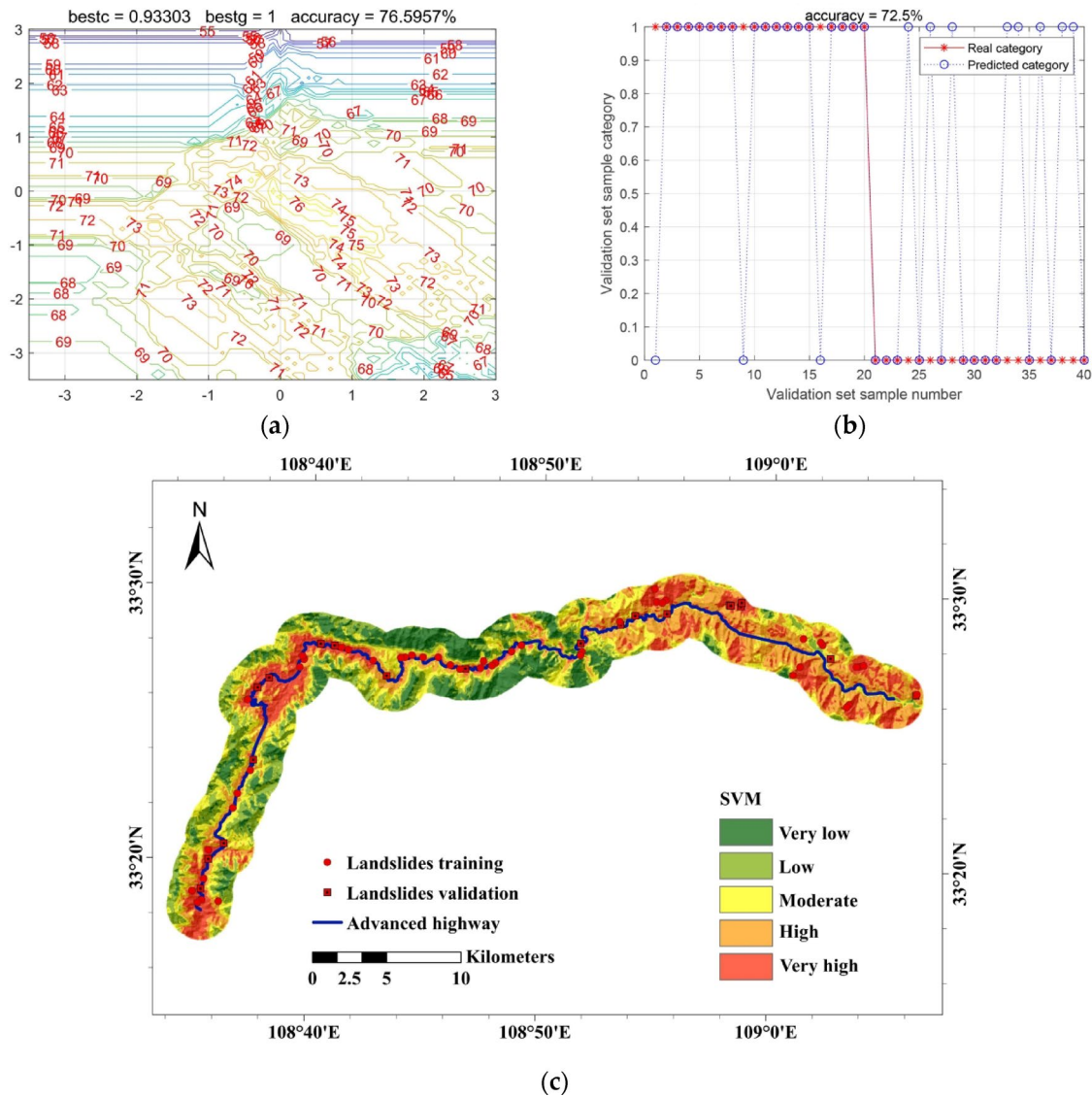$$ACC = \frac{TP + TN}{P + N} \tag{5}$$

(a)

(b)

(c)

Fig. 7. (**a**) Result of SVM parameter selection; (**b**) validation set prediction results of SVM model; (**c**) landslides susceptibility map of SVM model. (Note: this figure is made using ArcGIS desktop 10.8 (ArcMap component) https://www.esri.com/zh-cn/arcgis/products/arcgis-desktop/overview).

Statistical indicators	RF	SVM
PPV	0.781	0.699
NPV	0.828	0.738
ACC	0.803	0.716

Table 3. Statistical indicators of each model.

This study used ROC analysis models to predict performance[100–104]. The X-axis specificity represents the probability value of the prediction error for the non-landslide data. The Y-axis susceptibility represents the probability of predicting landslide results. The larger the area on the lower right side of the curve is, the higher the prediction accuracy of the model is. Therefore, the area between the prediction accuracy curve of the model and the abscissa is called the AUC value and the value range is 0–1[105,106]. The closer the AUC value is to 1, the higher the prediction accuracy of the representative model[107,108]. The AUC of the RF was 0.887, and that of the SVM model was 0.735 (Fig. 8). Therefore, both RF and SVM were suitable for landslide susceptibility evaluation in this study area, but RF had better prediction ability.
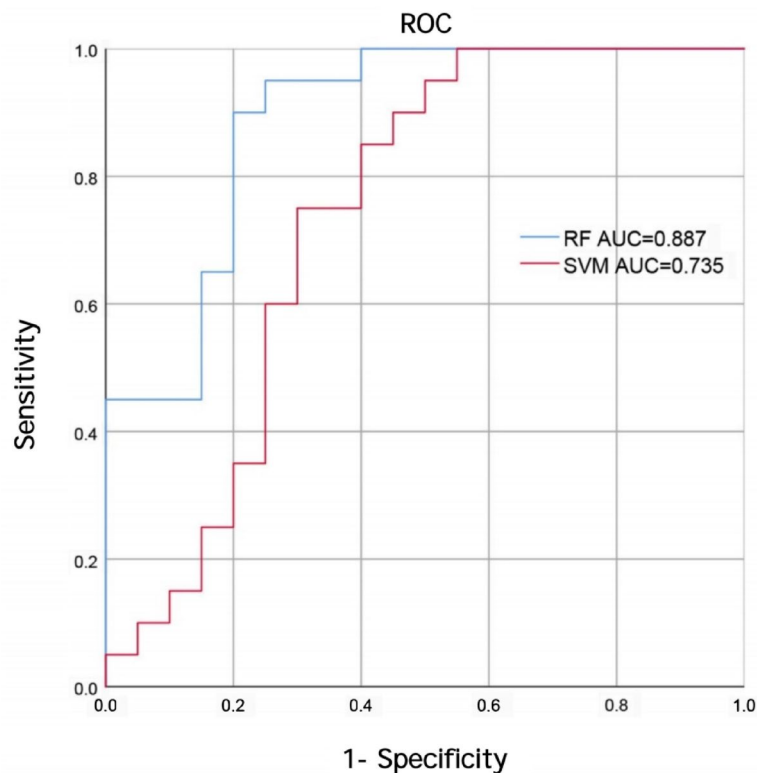
Fig. 8. ROC curves for the model validation set.

Discussion

In this paper, RF and SVM models are established, and 11 conditioning factors are selected. The main purpose is to establish the landslide susceptibility model of Lizha-Jiezi section of G345 national highway and evaluate the landslide susceptibility of the study area.

The evaluation unit contains the landslides data of each conditioning factor. The division of the evaluation unit is the premise of using GIS to evaluate the susceptibility of the study area, including the determination of the type, size and number of the evaluation unit. At present, the methods of evaluation unit division include grid unit and slope unit. Grid unit is the most commonly used evaluation unit division method at present. This method can quickly and effectively carry out the later data extraction and overlay analysis, but the shape of grid unit cells is regular, which cannot well reflect the characteristics of terrain change. Due to the discontinuity of slope unit, the evaluation results of landslide susceptibility evaluation are often poor in accuracy and thus difficult to meet the actual requirements. In this study, the grid unit is used to evaluate the susceptibility of landslide. Although it cannot respond well to the surface fluctuation, the evaluation results transit smoothly and have higher prediction accuracy than the slope unit.

The prediction ability of landslide conditioning factors (Fig. 9) shows that each conditioning factor contributes to landslide susceptibility modeling, but the contribution is different. The dominance of distance to road as the most influential factor in both models underscores the significant role of human engineering activities in slope destabilization. Road construction often involves slope cutting and vegetation removal, exacerbating geological vulnerability, a phenomenon widely observed in mountainous highways[58,61]. Notably, this finding contrasts with studies emphasizing rainfall or lithology as primary drivers[24], suggesting that region-specific human interventions may override natural triggers in certain contexts. In the SVM model, altitude, slope, and lithology also show high contribution. Altitude reflects the potential energy of the slope, slope reflects the steepness of terrain, and lithology is the material basis of landslide, which controls the stability of the slope to a certain extent. However, the high contribution of distance to river and rainfall in the RF model aligns with global patterns where hydrological factors critically influence slope stability[52,53]. Rainfall increases the moisture content of rock and soil, thereby elevating pore water pressure and decreasing their effective stress. This process increases the sliding force on unstable slopes and provides conditions for landslide initiation. Long-term river cutting and lateral erosion directly damage riverbank slopes, weakening their stability and thereby promoting landslide occurrence. Although the contribution of aspect, plan curvature, profile curvature, distance to fault and NDVI to the model is relatively small, the role of these factors cannot be ignored.

Compared with knowledge-driven model and traditional data-driven model, machine learning model can obtain higher prediction accuracy more quickly and effectively. In this study, RF and SVM models in machine learning model are used to generate LSM in the study area. These two methods have been widely used in landslide susceptibility assessment[109]. In this study, the susceptibility levels of each model are divided
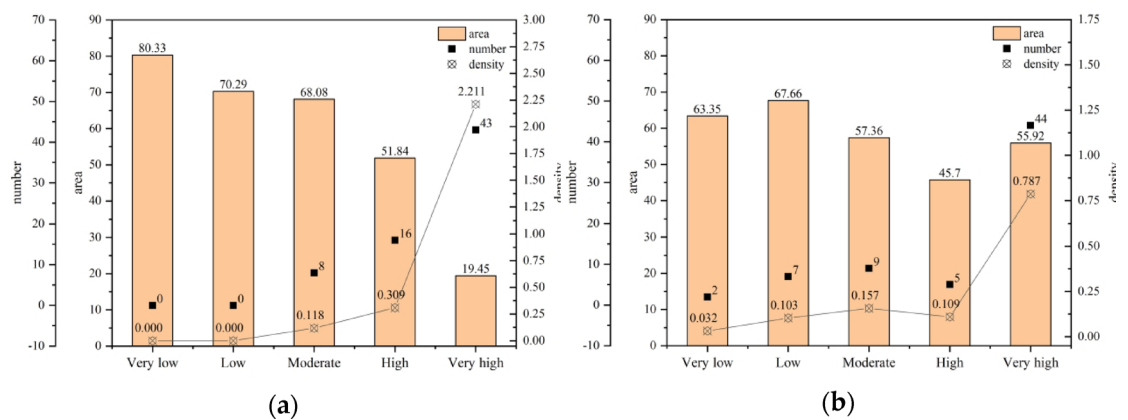
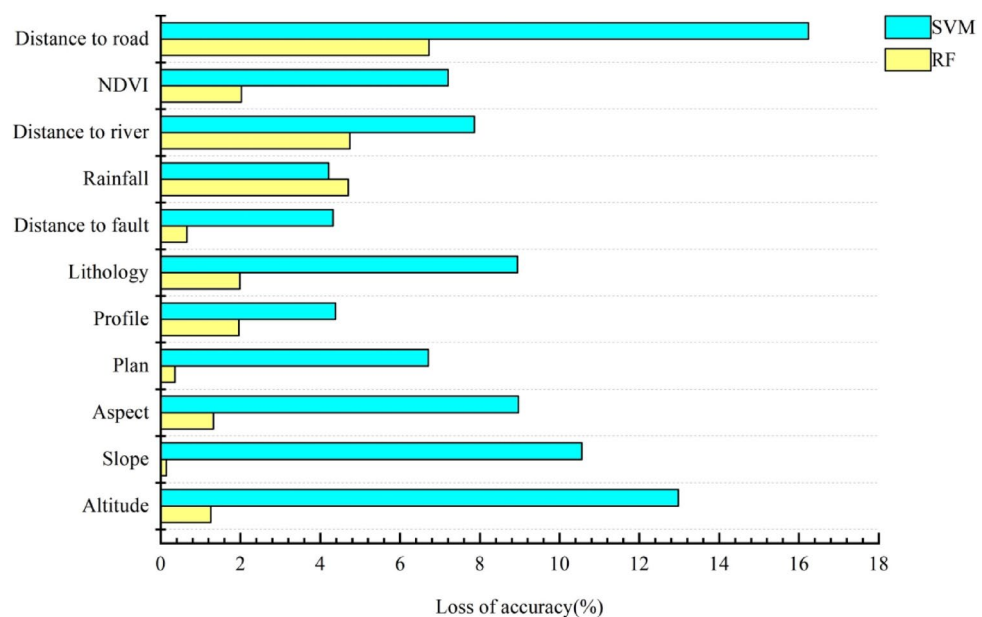**Fig. 9**. The prediction ability of the landslide conditioning factors.

Fig. 10. (**a**) Description of what is contained in the first panel; (**b**) histograms of evaluation results of SVM model.

into five categories: very low, low, moderate, high and very high. The area occupied by different categories, the number of landslides and the density of landslides of each model are counted and calculated (Fig. 10). Combined with the model accuracy verification (Fig. 8), it is found that the RF model has higher prediction accuracy of landslide susceptibility in the study area. RF model is not very susceptible to parameters. It is easy to determine which parameters to use[110]. It has strong generalization ability and is not easy to produce overfitting. It has high prediction accuracy in landslide susceptibility evaluation in other areas[111–114]. Although RF and SVM have achieved good prediction accuracy, it is very important to select a more suitable prediction model for the study area according to the unique geological environmental conditioning factors and landslide inducing factors in the landslide susceptibility assessment.

Due to the limited information available, this study only considers the distance analysis when analyzing the conditioning factors of faults, rivers and roads, and ignores the influence of scale and influence range on landslide susceptibility. Furthermore, while this study employed randomized data splitting (70% training, 30% validation) and ensured non-landslide samples were selected from areas > 100 m away from landslide locations to minimize spatial overlap, potential spatial autocorrelation within the inventory dataset was not explicitly quantified. Spatial autocorrelation could lead to inflated model performance metrics if spatially clustered landslides share similar environmental characteristics, thereby reducing the model's generalizability to regions with distinct spatial dependencies. More data can be collected in future studies to study the relationship between the scale, influence range and landslide susceptibility of faults, rivers and roads. Additionally, future studies

could also enhance robustness by incorporating spatial cross-validation or integrating spatial covariates to explicitly address autocorrelation, particularly when extrapolating predictions to broader geographic contexts.

Conclusion

This study advances landslide susceptibility assessment in mountainous highway environments by integrating multi-source geospatial data with robust machine learning models, providing novel insights and methodological contributions to the field. The application of Random Forest (RF) and Support Vector Machine (SVM) models to the Lizha-Jiezi section of the G345 national highway revealed that RF (AUC = 0.887) outperformed SVM (AUC = 0.735), demonstrating its superior adaptability in capturing complex interactions among conditioning factors. Notably, the dominance of distance to road as the most influential factor underscores the critical role of human engineering activities in destabilizing slopes, which challenges the conventional emphasis on natural triggers like rainfall or lithology in similar terrains. This highlights the necessity of prioritizing anthropogenic impacts in landslide risk management for mountainous infrastructure projects, a perspective less explored in prior studies.

Furthermore, our systematic evaluation of 11 conditioning factors, including topography, hydrology, and human activity variables, establishes a scalable framework for integrating multi-dimensional data into susceptibility mapping. The methodology's success in identifying high-risk zones within 2 km of the highway offers actionable insights for targeted landslide mitigation, enabling authorities to optimize resource allocation and engineering interventions.

Future research could improve spatial resolution and integrate dynamic variables (e.g., real-time rainfall, seismic activity, vegetation changes) to enhance prediction accuracy. Additionally, future studies should employ spatial cross-validation, spatial econometric models, or advanced sampling strategies to reduce biases from clustered landslide occurrences and improve the models' extrapolation capability to broader regions. Subsequent work must also validate the framework's adaptability across diverse geological and climatic environments, particularly in areas experiencing rapid urbanization or frequent extreme weather events. By integrating advanced machine learning techniques with actionable risk management, this approach not only advances academic discourse but also provides policymakers and engineers with tools to proactively protect critical infrastructure and address geological threats exacerbated by human activities.

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

References

1. Palis, E., Lebourg, T., Tric, E., Malet, J. P. & Vidal, M. Long-term monitoring of a large deep-seated landslide (La clapiere, South-East French Alps): Initial study. *Landslides* **14**, 155–170. https://doi.org/10.1007/s10346-016-0705-7 (2017).
2. Bentivenga, M. et al. Geomorphological and geophysical surveys with InSAR analysis applied to the Picerno Earth flow (southern Apennines, Italy). *Landslides* **18**, 471–483. https://doi.org/10.1007/s10346-020-01499-z (2021).
3. Guerriero, L. et al. PS-driven inventory of town-damaging landslides in the benevento, Avellino and Salerno provinces, Southern Italy. *J. Maps.* **15**, 619–625. https://doi.org/10.1080/17445647.2019.1651770 (2019).
4. Cha, D. S., Hwang, J. S. & Choi, B. K. Landslides detection and volume estimation in Jinbu area of Korea. *For. Sci. Technol.* **14**, 61–65. https://doi.org/10.1080/21580103.2018.1446367 (2018).
5. Lee, J. S., Kang, H., Suk, J. W. & Yun-Tae, K. Development of hazard level-based rainfall threshold for prediction of rainfall-induced landslide occurrence in Korea. *J. Korean Soc. Hazard. Mitigation.* **19**, 225–236. https://doi.org/10.9798/kosham.2019.19.5.225 (2019).
6. Kim, J. C., Lee, S., Jung, H. S. & Lee, S. Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. *Geocarto Int.* **33**, 1000–1015. https://doi.org/10.1080/10106049.2017.1323964 (2018).
7. Lee, J. U. et al. The effects of different geological conditions on landslide-triggering rainfall conditions in South Korea. *Water* **14**, 14. https://doi.org/10.3390/w14132051 (2022).
8. Jiang, W. G., Rao, P. Z., Cao, R., Tang, Z. H. & Chen, K. Comparative evaluation of geological disaster susceptibility using multi-regression methods and spatial accuracy validation. *J. Geogr. Sci.* **27**, 439–462. https://doi.org/10.1007/s11442-017-1386-4 (2017).
9. Ding, Q. F., Chen, W. & Hong, H. Y. Application of frequency ratio, weights of evidence and evidential belief function models in landslide susceptibility mapping. *Geocarto Int.* **32**, 619–639. https://doi.org/10.1080/10106049.2016.1165294 (2017).
10. Zhao, H. L., Yao, L. H., Mei, G., Liu, T. Y. & Ning, Y. S. A fuzzy comprehensive evaluation method based on AHP and entropy for a landslide susceptibility map. *Entropy* **19**(16). https://doi.org/10.3390/e19080396 (2017).
11. Yu, C. L., Chen, J. P. & Model, R. F. Application of a GIS-based slope unit method for landslide susceptibility mapping in helong city: Comparative assessment of icM, AHP, and RF model. *Symmetry-Basel* **12**, 21. https://doi.org/10.3390/sym12111848 (2020).
12. Aksoy, B. & Ercanoglu, M. Landslide identification and classification by object-based image analysis and fuzzy logic: An example from the Azdavay region (Kastamonu, Turkey). *Comput. Geosci.* **38**, 87–98. https://doi.org/10.1016/j.cageo.2011.05.010 (2012).
13. Firuzi, E., Ansari, A., Amini Hosseini, K. & Kheirkhah, N. Developing an earthquake damaged-based multi-severity casualty method by using Monte Carlo simulation and fuzzy logic; case study: Mosha fault seismic scenario, tehran, Iran. *Stoch. Env. Res. Risk Assess.* **38**, 2019–2039. https://doi.org/10.1007/s00477-024-02667-6 (2024).
14. Aggarwal, S., Rallapalli, S. & Adinarayana, J. Uncertainty-based fuzzified environmental-socio-economic risk assessment of precision agricultural practices. *Stoch. Env. Res. Risk Assess.* https://doi.org/10.1007/s00477-024-02864-3 (2024).
15. Sun, R., Fei, K., Reheman, Y., Zhou, J. & Jiao, D. Comprehensive uncertainty evaluation of dam break consequences considering multi-source information fusion. *Environ. Earth Sci.* **83**, 323. https://doi.org/10.1007/s12665-024-11610-5 (2024).
16. Dou, Q. et al. A method for improving controlling factors based on information fusion for debris flow susceptibility mapping: A case study in Jilin province, China. *Entropy* **21** (22). https://doi.org/10.3390/e21070695 (2019).
17. Zhang, Q. et al. A landslide susceptibility assessment method considering the similarity of geographic environments based on graph neural network. *Gondwana Res.* **132**, 323–342. https://doi.org/10.1016/j.gr.2024.04.013 (2024).

18. Yang, Y. et al. An information quantity and machine learning integrated model for landslide susceptibility mapping in Jiuzhaigou, China. *Nat. Hazards* (2024).

19. Zhao, X., Chen, W., Tsangaratos, P. & Ilia, I. Evaluating landslide susceptibility: the impact of resolution and hybrid integration approaches. *Geomatics Nat. Hazards Risk* **15**, 2409198. https://doi.org/10.1080/19475705.2024.2409198 (2024).

20. Zhang, Y. X. et al. Optimizing the frequency ratio method for landslide susceptibility assessment: A case study of the Caiyuan basin in the Southeast mountainous area of China. *J. Mt. Sci.* **17**, 340–357. https://doi.org/10.1007/s11629-019-5702-6 (2020).

21. Barman, J. & Das, J. Assessing classification system for landslide susceptibility using frequency ratio, analytical hierarchical process and geospatial technology mapping in Aizawl district, NE India. *Adv. Space Res.* **74**, 1197–1224. https://doi.org/10.1016/j.asr.2024.05.007 (2024).

22. Nwazelibe, V. E., Egbueri, J. C., Chinanu, O. GIS-based landslide susceptibility mapping of Western Rwanda: An integrated artificial neural network, frequency ratio, and Shannon entropy approach. *Environ. Earth Sci.* **82**, 439.431–439.423. https://doi.org/10.1007/s12665-023-11134-4 (2023).

23. Colkesen, I., Sahin, E. K. & Kavzoglu, T. Susceptibility mapping of shallow landslides using kernel-based Gaussian process, support vector machines and logistic regression. *J. Afr. Earth Sci.* **118**, 53–64. https://doi.org/10.1016/j.jafrearsci.2016.02.019 (2016).

24. Devkota, K. C., Regmi, A. D. & Althuwaynee, O. F. Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal himalaya. *Nat. Hazards.* **65**, 135–165. https://doi.org/10.1007/S11069-012-0347-6 (2013).

25. Dutta, A., Sarkar, K. & Tarun, K. Machine learning regression algorithms for predicting the susceptibility of jointed rock slopes to planar failure. *Earth Sci. Inf.* **17**, 2477–2493. https://doi.org/10.1007/s12145-024-01296-5 (2024).

26. Hoa, P. V. et al. One-dimensional deep learning driven Geospatial analysis for flash flood susceptibility mapping: A case study in North central Vietnam. *Earth Sci. Inf.* **17**, 4419–4440. https://doi.org/10.1007/s12145-024-01285-8 (2024).

27. Nanehkaran, Y. A. et al. Riverside landslide susceptibility overview: Leveraging artificial neural networks and machine learning in accordance with the United Nations (UN) sustainable development goals. *Water* **15**, 2707 (2023).

28. Bui, D. et al. Landslide detection and susceptibility mapping by AIRSAR data using support vector machine and index of entropy models in Cameron highlands, Malaysia. *Remote Sens.* **10**, 32. https://doi.org/10.3390/rs10101527 (2018).

29. Chen, W., Pourghasemi, H. R. & Naghibi, S. A. A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy data mining models in China. *Bull. Eng. Geol. Environ.* **77**, 647–664. https://doi.org/10.1007/s10064-017-1010-y (2018).

30. Feizizadeh, B., Roodposhti, M. S., Blaschke, T. & Aryal, J. Comparing GIS-based support vector machine kernel functions for landslide susceptibility mapping. *Arab. J. Geosci.* **10**, 13. https://doi.org/10.1007/s12517-017-2918-z (2017).

31. Hong, H. Y. et al. Spatial prediction of landslide hazard at the Luxi area (China) using support vector machines. *Environ. Earth Sci.* **75**, 14. https://doi.org/10.1007/s12665-015-4866-9 (2016).

32. Mabdeh, A. N., Al-Fugara, A., Abualigah, L., Saleem, K. & Snasel, V. Enhanced forest fire susceptibility mapping by integrating feature selection genetic algorithm and bagging-based support vector machine with artificial neural networks. *Stoch. Env. Res. Risk Assess.* **38**, 5039–5058. https://doi.org/10.1007/s00477-024-02851-8 (2024).

33. Wani, F. M., Vemuri, J., Reddy, K. S. K. K. & Rajaram, C. Forecasting duration characteristics of near fault pulse-like ground motions using machine learning algorithms. *Stoch. Env. Res. Risk Assess.* https://doi.org/10.1007/s00477-024-02729-9 (2024).

34. Dou, J. et al. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima volcanic island, Japan. *Sci. Total Environ.* **662**, 332–346. https://doi.org/10.1016/j.scitotenv.2019.01.221 (2019).

35. Garcia-Carretero, R., Holgado-Cuadrado, R. & Barquero-Perez, O. Assessment of classification models and relevant features on nonalcoholic steatohepatitis using random forest. *Entropy* **23**(23). https://doi.org/10.3390/e23060763 (2021).

36. Li, J. Y., Wang, W. D., Li, Y. E., Han, Z. & Chen, G. Q. Spatiotemporal landslide susceptibility mapping incorporating the effects of heavy rainfall: A case study of the heavy rainfall in August 2021 in Kitakyushu, Fukuoka, Japan. *Water* **13** (16). https://doi.org/10.3390/w13223312 (2021).

37. Doke, R. et al. Monitoring of landslide displacements in owakudani, Hakone volcano, Japan, using SAR interferometry. *Landslides* **21**, 1207–1219. https://doi.org/10.1007/s10346-024-02224-w (2024).

38. Hong, H. Y., Liu, J. Z. & Zhu, A. X. Modeling landslide susceptibility using logitboost alternating decision trees and forest by penalizing attributes with the bagging ensemble. *Sci. Total Environ.* **718**, 15. https://doi.org/10.1016/j.scitotenv.2020.137231 (2020).

39. Hong, H. Y., Pradhan, B. & Xu, C. Tien bui, D. Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. *Catena* **133**, 266–281. https://doi.org/10.1016/j.catena.2015.05.019 (2015).

40. Pham, B. T., Bui, D. T. & Prakash, I. Landslide susceptibility assessment using bagging ensemble based alternating decision trees, logistic regression and J48 decision trees methods: A comparative study. *Geotech. Geol. Eng.* https://doi.org/10.1007/s10706-017-0264-2 (2017).

41. Dyer, A. S. et al. Offshore application of landslide susceptibility mapping using gradient-boosted decision trees: a Gulf of Mexico case study. *Nat. Hazards* **120**. https://doi.org/10.1007/s11069-024-06492-6 (2024).

42. Mao, Y. et al. Predicting the elasticity modulus of sedimentary rocks using deep random forest optimization (DRFO) algorithm. *Environ. Earth Sci.* **83** https://doi.org/10.1007/s12665-024-11768-y (2024).

43. Dey, S., Das, S. & Roy, S. K. Landslide susceptibility assessment in Eastern himalayas, india: a comprehensive exploration of four novel hybrid ensemble data driven techniques integrating explainable artificial intelligence approach. *Environ. Earth Sci.* **83** https://doi.org/10.1007/s12665-024-11945-z (2024).

44. Yang, S., Li, D., Sun, Y. & She, X. Effect of landslide spatial representation and raster resolution on the landslide susceptibility assessment. *Environ. Earth Sci.* **83** https://doi.org/10.1007/s12665-024-11442-3 (2024).

45. Svitelman, V., Saveleva, E. & Neuvazhaev, G. Comparison of feature importance measures and variance-based indices for sensitivity analysis: Case study of radioactive waste disposal flow and transport model. *Stoch. Env. Res. Risk Assess.* https://doi.org/10.1007/s00477-024-02869-y (2024).

46. Tran, T. D. & Kim, J. Guidance on the construction and selection of relatively simple to complex data-driven models for multi-task streamflow forecasting. *Stoch. Env. Res. Risk Assess.* **38**, 3657–3675. https://doi.org/10.1007/s00477-024-02776-2 (2024).

47. Wu, S., Wang, H., Zhang, J. & Qin, H. Hybrid method for rainfall-induced regional landslide susceptibility mapping. *Stoch. Env. Res. Risk Assess.* **38**, 4193–4208. https://doi.org/10.1007/s00477-024-02753-9 (2024).

48. Zhu, A. X. et al. An expert knowledge-based approach to landslide susceptibility mapping using GIS and fuzzy logic. *Geomorphology* **214**, 128–138. https://doi.org/10.1016/j.geomorph.2014.02.003 (2014).

49. Moragues, S. et al. Analytic hierarchy process applied to landslide susceptibility mapping of the North branch of Argentino lake, Argentina. *Nat. Hazards* **105**, 915–941. https://doi.org/10.1007/s11069-020-04343-8 (2021).

50. Pourghasemi, H. R., Pradhan, B. & Gokceoglu, C. Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. *Nat. Hazards.* **63**, 965–996. https://doi.org/10.1007/s11069-012-0217-2 (2012).

51. Akbar, T. A. & Ha, S. R. Landslide hazard zoning along Himalayan Kaghan Valley of Pakistan—by integration of GPS, GIS, and remote sensing technology. *Landslides* **8**, 527–540. https://doi.org/10.1007/s10346-011-0260-1 (2011).

52. Berhane, G. et al. Landslide susceptibility zonation mapping using GIS-based frequency ratio model with multi-class Spatial data-sets in the Adwa-Adigrat mountain chains, Northern Ethiopia. *J. Afr. Earth Sci.* **164**, 15. https://doi.org/10.1016/j.jafrearsci.2020.103795 (2020).

53. Schlogel, R. et al. Optimizing landslide susceptibility zonation: Effects of DEM spatial resolution and slope unit delineation on logistic regression models. *Geomorphology* **301**, 10–20. https://doi.org/10.1016/j.geomorph.2017.10.018 (2018).

54. Park, S., Choi, C., Kim, B. & Kim, J. Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. *Environ. Earth Ences.* **68**, 1443–1464. https://doi.org/10.1007/s12665-012-1842-5 (2013).

55. Pourghasemi, H. R. & Kerle, N. Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran province, Iran. *Environ. Earth Sci.* **75**, 17. https://doi.org/10.1007/s12665-015-4950-1 (2016).

56. Chen, W. et al. Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. *Geomorphology* **297**, 69–85. https://doi.org/10.1016/j.geomorph.2017.09.007 (2017).

57. Mao, Y. et al. Utilizing hybrid machine learning and soft computing techniques for landslide susceptibility mapping in a drainage basin. *Water* **16**, 380 (2024).

58. Pandey, V. K., Sharma, K. K., Pourghasemi, H. R. & Bandooni, S. K. Sedimentological characteristics and application of machine learning techniques for landslide susceptibility modelling along the highway corridor Nahan to Rajgarh (Himachal Pradesh), India. *Catena* **182**, 18. https://doi.org/10.1016/j.catena.2019.104150 (2019).

59. Ahangari Nanehkaran, Y. et al. Application of machine learning techniques for the estimation of the safety factor in slope stability analysis. *Water* **14**, 3743 (2022).

60. Banerjee, P., Ghose, M. K. & Pradhan, R. Analytic hierarchy process and information value method-based landslide susceptibility mapping and vehicle vulnerability assessment along a highway in Sikkim himalaya. *Arab. J. Geosci.* **11**, 18. https://doi.org/10.1007/s12517-018-3488-4 (2018).

61. Liu, Y., Zhao, L. J., Bao, A. M., Li, J. L. & Yan, X. B. Chinese high resolution satellite data and GIS-based assessment of landslide susceptibility along highway G30 in Guozi-gou valley using logistic regression and MaxEnt model. *Remote Sens.* **14**, 23. https://doi.org/10.3390/rs14153620 (2022).

62. Chen, L. F. et al. Landslide susceptibility assessment using weights-of-evidence model and cluster analysis along the highways in the Hubei section of the three Gorges reservoir area. *Comput. Geosci.* **156**, 13. https://doi.org/10.1016/j.cageo.2021.104899 (2021).

63. Ada, M. & San, B. T. Comparison of machine-learning techniques for landslide susceptibility mapping using two-level random sampling (2LRS) in Alakir catchment area, antalya, Turkey. *Nat. Hazards*. **90**, 237–263. https://doi.org/10.1007/s11069-017-3043-8 (2018).

64. Aghdam, I. N., Varzandeh, M. H. M. & Pradhan, B. Landslide susceptibility mapping using an ensemble statistical index (Wi) and adaptive neuro-fuzzy inference system (ANFIS) model at Alborz mountains (Iran). *Environ. Earth Sci.* **75**, 20. https://doi.org/10.1007/s12665-015-5233-6 (2016).

65. Chen, W., Chen, X., Peng, J. B., Panahi, M. & Lee, S. Landslide susceptibility modeling based on ANFIS with teaching-learning-based optimization and satin Bowerbird optimizer. *Geosci. Front.* **12**, 93–107. https://doi.org/10.1016/j.gsf.2020.07.012 (2021).

66. Chen, W. & Li, Y. GIS-based evaluation of landslide susceptibility using hybrid computational intelligence models. *Catena* **195**, 16. https://doi.org/10.1016/j.catena.2020.104777 (2020).

67. Fang, Z. C., Wang, Y., Peng, L. & Hong, H. Y. Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping. *Comput. Geosci.* **139**, 15. https://doi.org/10.1016/j.cageo.2020.104470 (2020).

68. Li, Y. & Chen, W. Landslide susceptibility evaluation using hybrid integration of evidential belief function and machine learning techniques. *Water* **12**, 29. https://doi.org/10.3390/w12010113 (2020).

69. Zhang, G. L., Wang, M. & Liu, K. Forest fire susceptibility modeling using a convolutional neural network for Yunnan Province of China. *Int. J. Disaster Risk Sci.* **10**, 386–403. https://doi.org/10.1007/s13753-019-00233-1 (2019).

70. Chung, C. J. F. & Fabbri, A. G. Validation of spatial prediction models for landslide hazard mapping. *Nat. Hazards*. **30**, 451–472. https://doi.org/10.1023/B:NHAZ.0000007172.62651.2b (2003).

71. Ahmed, M. F., Hussain, M., Rogers, J. D. & Khan, M. S. Initial screening of regional landslide hazards in the Hunza river watershed, Gilgit baltistan, Pakistan. *Nat. Hazards Rev.* **22**, 17. https://doi.org/10.1061/(asce)nh.1527-6996.0000497 (2021).

72. Althuwaynee, O. F., Pradhan, B. & Lee, S. Application of an evidential belief function model in landslide susceptibility mapping. *Comput. Geosci.* **44**, 120–135. https://doi.org/10.1016/j.cageo.2012.03.003 (2012).

73. Hong, H. Y. et al. Landslide susceptibility assessment at the Wuning area, China: A comparison between multi-criteria decision making, bivariate statistical and machine learning methods. *Nat. Hazards* **96**, 173–212. https://doi.org/10.1007/s11069-018-3536-0 (2019).

74. Yi, Y. N., Zhang, Z. J., Zhang, W. C., Jia, H. H. & Zhang, J. Q. Landslide susceptibility mapping using multiscale sampling strategy and convolutional neural network: A case study in Jiuzhaigou region. *Catena* **195**, 13. https://doi.org/10.1016/j.catena.2020.104851 (2020).

75. Pourkhosravani, M., Ali, M., Saied, P., Derakhshani, R. & and Monitoring of Maskun landslide and determining its quantitative relationship to different Climatic conditions using D-InSAR and PSI techniques. *Geomatics Nat. Hazards Risk* **13**, 1134–1153. https://doi.org/10.1080/19475705.2022.2065939 (2022).

76. Feng, W. L., Zhu, Q. Y., Zhuang, J. & Yu, S. M. An expert recommendation algorithm based on pearson correlation coefficient and FP-growth. *Cluster Comput.* **22**, S7401–S7412. https://doi.org/10.1007/s10586-017-1576-y (2019).

77. Xu, H. H. & Deng, Y. Dependent evidence combination based on Shearman coefficient and pearson coefficient. *IEEE Access.* **6**, 11634–11640. https://doi.org/10.1109/access.2017.2783320 (2018).

78. Yan, Y. Y. et al. Using the google earth engine to rapidly monitor impacts of geohazards on ecological quality in highly susceptible areas. *Ecol. Indic.* **132**, 12. https://doi.org/10.1016/j.ecolind.2021.108258 (2021).

79. Edelmann, D., Mori, T. F. & Szekely, G. J. On relationships between the pearson and the distance correlation coefficients. *Stat. Probab. Lett.* **169**, 6. https://doi.org/10.1016/j.spl.2020.108960 (2021).

80. Breiman, L. Random forests. *Mach. Learn.* https://doi.org/10.1023/A:1010933404324 (2001).

81. Goetz, J. N., Brenning, A., Petschko, H. & Leopold, P. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput. Geosci.* **81**, 1–11. https://doi.org/10.1016/j.cageo.2015.04.007 (2015).

82. Li, T. Y. & Zhou, M. ECG classification using wavelet packet entropy and random forests. *Entropy* **18**(16). https://doi.org/10.3390/e18080285 (2016).

83. Chen, W. et al. Performance evaluation and comparison of bivariate statistical-based artificial intelligence algorithms for spatial prediction of landslides. *ISPRS Int. Geo-Inf.* **9**, 21. https://doi.org/10.3390/ijgi9120696 (2020).

84. Tehrany, M. S., Pradhan, B., Mansor, S. & Ahmad, N. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *Catena* **125**, 91–101. https://doi.org/10.1016/j.catena.2014.10.017 (2015).

85. Han, J. et al. Performance of logistic regression and support vector machines for seismic vulnerability assessment and mapping: A case study of the 12 September 2016 ML5.8 Gyeongju earthquake, South Korea. *Sustainability* **11**(19). https://doi.org/10.3390/su11247038 (2019).

86. He, Q. F. et al. Landslide spatial modelling using novel bivariate statistical based Naive bayes, RBF classifier, and RBF network machine learning algorithms. *Sci. Total Environ.* **663**, 1–15. https://doi.org/10.1016/j.scitotenv.2019.01.329 (2019).

87. Jebur, M. N., Pradhan, B. & Tehrany, M. S. Manifestation of LiDAR-derived parameters in the spatial prediction of landslides using novel ensemble evidential belief functions and support vector machine models in GIS. *IEEE J. Sel. Top. Appl. Earth Observ Remote Sens.* **8**, 674–690. https://doi.org/10.1109/jstars.2014.2341276 (2015).

88. Mirzaei, G., Soltani, A., Soltani, M. & Darabi, M. An integrated data-mining and multi-criteria decision-making approach for hazard-based object ranking with a focus on landslides and floods. *Environ. Earth Sci.* **77**, 23. https://doi.org/10.1007/s12665-018-7762-2 (2018).

89. Arabameri, A. et al. Landslide susceptibility evaluation and management using different machine learning methods in the gallicash river watershed, Iran. *Remote Sens.* **12**, 29. https://doi.org/10.3390/rs12030475 (2020).

90. Lin, J. H., Chen, W. H., Qi, X. H. & Hou, H. R. Risk assessment and its influencing factors analysis of geological hazards in typical mountain environment. *J. Clean. Prod.* **309**, 10. https://doi.org/10.1016/j.jclepro.2021.127077 (2021).

91. Kalantar, B., Pradhan, B., Naghibi, S. A., Motevalli, A. & Mansor, S. Assessment of the effects of training data selection on the landslide susceptibility mapping: A comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN). *Geomatics Nat. Hazards Risk.* **9**, 49–69. https://doi.org/10.1080/19475705.2017.1407368 (2018).

92. Shi, H. B., Fu, W. L., Li, B. L., Shao, K. X. & Yang, D. H. Intelligent fault identification for rolling bearings fusing average refined composite multiscale dispersion entropy-assisted feature extraction and SVM with Multi-Strategy enhanced swarm optimization. *Entropy* **23**(27). https://doi.org/10.3390/e23050527 (2021).

93. Ballabio, C. & Sterlacchini, S. Support vector machines for landslide susceptibility mapping: the staffora river basin case study, Italy. *Math. Geosci.* **44** https://doi.org/10.1007/s11004-011-9379-9 (2012).

94. Libert, A. & Van Hulle, M. M. Predicting premature video skipping and viewer interest from EEG recordings. *Entropy* **21**, 11. https://doi.org/10.3390/e21101014 (2019).

95. Lin, H. F., Shin, W. Y. & Joung, J. Support vector Machine-Based transmit antenna allocation for multiuser communication systems. *Entropy* **21**(17). https://doi.org/10.3390/e21050471 (2019).

96. Wei, Y. J., Fang, S. L. & Wang, X. Y. Automatic modulation classification of digital communication signals using SVM based on hybrid features, cyclostationary, and information entropy. *Entropy* **21**(17). https://doi.org/10.3390/e21080745 (2019).

97. Zhang, T. Y. et al. GIS-based landslide susceptibility mapping using hybrid integration approaches of fractal dimension with index of entropy and support vector machine. *J. Mt. Sci.* **16**, 1275–1288. https://doi.org/10.1007/s11629-018-5337-z (2019).

98. Zhao, B. B., Ge, Y. F. & Chen, H. Z. Landslide susceptibility assessment for a transmission line in Gansu province, China by using a hybrid approach of fractal theory, information value, and random forest models. *Environ. Earth Sci.* **80**, 23. https://doi.org/10.1007/s12665-021-09737-w (2021).

99. Zhu, K. H., Chen, L. & Hu, X. Rolling element bearing fault diagnosis by combining adaptive local iterative filtering, modified fuzzy entropy and support vector machine. *Entropy* **20**, 12. https://doi.org/10.3390/e20120926 (2018).

100. Pham, B. T., Prakash, I. & Bui, D. T. Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees. *Geomorphology* **303**, 256–270. https://doi.org/10.1016/j.geomorph.2017.12.008 (2018).

101. Saha, A. & Saha, S. Comparing the efficiency of weight of evidence, support vector machine and their ensemble approaches in landslide susceptibility modelling: A study on Kurseong region of Darjeeling himalaya, India. *Remote Sens. Applications: Soc. Environ.* **19** https://doi.org/10.1016/j.rsase.2020.100323 (2020).

102. Xiao, T., Yin, K. & Liu, S. Spatial prediction of landslide susceptibility using GIS-based statistical and machine learning models in Wanzhou county,three Gorges reservoir. *China Acta Geochim. V* **38**, 46–61. https://doi.org/10.1007/s11631-019-00341-1 (2019).

103. Yu, L. B., Cao, Y., Zhou, C., Wang, Y. & Huo, Z. T. Landslide susceptibility mapping combining information gain ratio and support vector machines: A case study from Wushan segment in the three Gorges reservoir area, China. *Appl. Sci. -Basel.* **9**, 19. https://doi.org/10.3390/app9224756 (2019).

104. Thanh, L. N. et al. Using landslide statistical index technique for landslide susceptibility mapping: Case Study: Ban Khoang Commune, Lao Cai Province, Vietnam. *Water* **14**, 22. https://doi.org/10.3390/w14182814 (2022).

105. Cao, J. et al. Multi-geohazards susceptibility mapping based on machine learning-a case study in jiuzhaigou, China. *Nat. Hazards.* **102**, 851–871. https://doi.org/10.1007/s11069-020-03927-8 (2020).

106. He, Q. F. et al. Novel entropy and rotation Forest-Based credal decision tree classifier for landslide susceptibility modeling. *Entropy* **21**(24). https://doi.org/10.3390/e21020106 (2019).

107. Hakim, W. L. et al. Convolutional neural network (CNN) with metaheuristic optimization algorithms for landslide susceptibility mapping in icheon, South Korea. *J. Environ. Manage.* **305**, 14. https://doi.org/10.1016/j.jenvman.2021.114367 (2022).

108. Persichillo, M. G. et al. Shallow landslides susceptibility assessment in different environments. *Geomatics Nat. Hazards Risk.* **8**, 748–771. https://doi.org/10.1080/19475705.2016.1265011 (2017).

109. Pham, B. T., Pradhan, B., Bui, D. T., Prakash, I. & Dholakia, M. B. A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). *Environ. Modell Softw.* **84**, 240–250. https://doi.org/10.1016/j.envsoft.2016.07.005 (2016).

110. Hong, H. Y., Pourghasemi, H. R. & Pourtaghi, Z. S. Landslide susceptibility assessment in Lianhua County (China): A comparison between a random forest data mining technique and bivariate and multivariate statistical models. *Geomorphology* **259**, 105–118. https://doi.org/10.1016/j.geomorph.2016.02.012 (2016).

111. Adnan, M. S. G. et al. Improving spatial agreement in machine learning-based landslide susceptibility mapping. *Remote Sens.* **12**, 23. https://doi.org/10.3390/rs12203347 (2020).

112. Chen, W., Zhang, S., Li, R. W. & Shahabi, H. Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and Naive Bayes tree for landslide susceptibility modeling. *Sci. Total Environ.* **644**, 1006–1018. https://doi.org/10.1016/j.scitotenv.2018.06.389 (2018).

113. Dang, V. H., Dieu, T. B., Tran, X. L. & Hoang, N. D. Enhancing the accuracy of rainfall-induced landslide prediction along mountain roads with a GIS-based random forest classifier. *Bull. Eng. Geol. Environ.* **78**, 2835–2849. https://doi.org/10.1007/s10064-018-1273-y (2019).

114. Zhao, F. M. et al. Landslide susceptibility mapping of Karakorum highway combined with the application of SBAS-InSAR technology. *Sensors* **19**, 18. https://doi.org/10.3390/s19122685 (2019).

Acknowledgements

Author contributions

Conceptualization, Q.H. S.W., P.T., I.I. and C.W.; methodology, Q.H., S.W., Y.C., and X.Z.; software, Q.H., Z.H., and X.Z.; validation, Q.H., S.W., and X.Z.; formal analysis, Q.H., Z.H., X.Z., S.W., and Y.C.; investigation, Q.H.,and Z.H., X.Z.; writing—original draft preparation, Q.H., Z.H., and X.Z.; writing—review and editing, Q.H., X.Z., S.W., Z.W.,C.W., and Y.H.; All authors have read and agreed to the published version of the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to Q.H. or X.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.