Machine learning solution for regional landslide susceptibility based on fault zone division strategy

Yunhao WANG, Luqi WANG, Songlin LIU, Weixin SUN, Pengfei LIU, Lin ZHU, Wenyu YANG, Tong GUO

Articles you may be interested in

Integrating predictive modeling techniques with geospatial data for landslide susceptibility assessment in northern Pakistan

Journal of Mountain Science. 2023, 20(9): 2603   https://doi.org/10.1007/s11629−023−8029−2

Comparative performance assessment of landslide susceptibility models with presence−only, presenceabsence, and pseudo−absence data

Journal of Mountain Science. 2020, 17(12): 2961   https://doi.org/10.1007/s11629−020−6277−y

Debris flow susceptibility and propagation assessment in West Koyulhisar, Turkey

Journal of Mountain Science. 2020, 17(11): 2611   https://doi.org/10.1007/s11629−020−6261−6

GIS−based landslide susceptibility mapping using hybrid integration approaches of fractal dimension with index of entropy and support vector machine

Journal of Mountain Science. 2019, 16(6): 1275   https://doi.org/10.1007/s11629−018−5337−z

Landslide susceptibility mapping using Genetic Algorithm for the Rule Set Production (GARP) model

Journal of Mountain Science. 2018, 15(9): 2013   https://doi.org/10.1007/s11629−018−4833−5

Original Article

Machine learning solution for regional landslide susceptibility based on fault zone division strategy

WANG Yunhao[1] https://orcid.org/0009-0005-7433-0533; e-mail: wangyh@stu.cqu.edu.cn

WANG Luqi[1,2,3]* https://orcid.org/0000-0001-5108-250X; ✉ e-mail: wlq93@cqu.edu.cn

LIU Songlin[1] https://orcid.org/0009-0001-6147-7166; e-mail: songlinl@cqu.edu.cn

SUN Weixin[1] https://orcid.org/0000-0002-1711-6718; e-mail: 20201601052@cqu.edu.cn

LIU Pengfei[4] https://orcid.org/0009-0006-3224-7608; e-mail: 273888264@qq.com

ZHU Lin[5] https://orcid.org/0009-0008-2351-0181; e-mail: zhuoling@163.com

YANG Wenyu[1] https://orcid.org/0009-0004-8372-5337; e-mail: yangwy@cqu.edu.cn

GUO Tong[5] https://orcid.org/0009-0002-4108-0935; e-mail: 445004120@qq.com

* Corresponding author

1 School of Civil Engineering, Chongqing University, Chongqing 400045, China

2 Key Laboratory of New Technology for Construction of Cities in Mountain Area, Chongqing University, Ministry of Education, Chongqing 400045, China

3 National Joint Engineering Research Center of Geohazards Prevention in the Reservoir Areas, Chongqing University, Chongqing 400045, China

4 Institute of Geological Environment Monitoring, Chongqing 401147, China

5 Chongqing 205 Geological Team, Chongqing 401121, China

Abstract: Landslide susceptibility assessment is an essential tool for disaster prevention and management. In areas with multiple fault zones, the impact of fault zone on slope stability cannot be disregarded. This study performed qualitative analysis of fault zones and proposed a zoning method to assess the landslide susceptibility in Chengkou County, Chongqing Municipality, China. The region within a distance of 1 km from the faults was designated as sub-zone A, while the remaining area was labeled as sub-zone B. To accomplish the assessment, a dataset comprising 388 historical landslides and 388 non-landslide points was used to train the random forest model. 10-fold cross-validation was utilized to select the training and testing datasets for the model. The results of the models were analyzed and discussed, with a focus on model performance and prediction uncertainty. By implementing the proposed division strategy based on fault zone, the accuracy, precision, recall, F-score, and AUC of both two sub-zones surpassed those of the whole region. In comparison to the results obtained for the whole region, sub-zone B exhibited an increase in AUC by 6.15%, while sub-zone A demonstrated a corresponding increase of 1.66%. Moreover, the results of 100 random realizations indicated that the division strategy has little effect on the prediction uncertainty. This study introduces a novel approach to enhance the prediction accuracy of the landslide susceptibility

mapping model in areas with multiple fault zones.

1 Introduction

Landslides, being the most frequently occurring geological disasters, pose a significant threat to the safety of both human lives and property (Lima et al. 2022; Wang et al. 2023b). The Three Gorges Reservoir area in China emerges as one of the regions significantly impacted by landslide disasters, thereby posing considerable challenges for the management of geological disaster risk (Tang et al. 2019; Liu et al. 2023b; Wang et al. 2023a; Zhou et al. 2023). With the aim of mitigating these risks and selecting suitable sites for future development, landslide susceptibility mapping (LSM) has emerged as a crucial tool. The LSM study operates under the common assumption that the factors conducive to future landslides closely coincide with those that precipitated historical landslide occurrences. Based on the distribution of historical landslides and associated conditioning factors, such assessments aid in determining the spatial probability of disaster occurrence, thereby holding immense significance for disaster prevention and land use planning (Zhang et al. 2022b).

Currently, regional LSM studies can be categorized into three distinct types: knowledge-based, data-driven, and physics-informed methods. Each of these methods possesses specific merits and limitations (Achu et al. 2023). The knowledge-based method features a straightforward process, though it is susceptible to significant subjectivity. The physics-informed method can directly simulate the force interactions, yet it is constrained by oversimplified mechanics laws (Wei et al. 2023). Over the past two decades, the integration and application of Geographic Information System (GIS) have garnered considerable attention, capitalizing on the rapid advancements in geospatial software, remote sensing, and global positioning system technology (Demir 2019). Due to the superiority of machine learning (ML) models in addressing intricate and nonlinear problems, the data-driven method combining GIS has gained popularity in LSM studies (Huang et al. 2020; Kainthura and Sharma 2022; Li et al. 2022).

Among the data-driven methods commonly employed in LSM studies, notable examples include random forest (RF), extreme gradient boosting (XGBoost), logistic regression (LR), support vector machine (SVM), and artificial neural network (ANN) (Aditian et al. 2018; Wang et al. 2023c; Yang et al. 2023). In order to improve the predictive accuracy of the ML model, scholars have employed various methods. Liu et al. (2023a) proposed a novel strategy for selecting negative samples, which enhances the performance of model while also increasing the interpretability. Hu et al. (2020) compared the impact of three non-landslide sample extracting methods on model performance and pointed out the advantage of fractal theory model. Kavzoglu and Teke (2022) employed optimization algorithms to adjust the hyper-parameter settings of XGBoost, resulting in improved accuracy. Lin et al. (2023) applied the convolutional neural networks based on Bayesian optimization to conduct LSM and obtained excellent performance. Zhang et al. (2022a) combined qualitative and quantitative analysis to assess the landslide susceptibility of Yunyang County, Chongqing City, China, thereby improving the mapping accuracy and demonstrating the feasibility of the division strategy based on qualitative analysis. Additionally, numerous studies have focus on the prediction uncertainty of LSM (Achu et al. 2023; Wei et al. 2023). Nevertheless, in regions with scarce historical landslide data, the performance of data-driven models often proves unsatisfactory. To address this issue, Wang et al. (2022) employed a transfer learning method to transfer the knowledge from regions characterized sufficient landslide records to those with limited data. Fu et al. (2023) applied TrAdaBoost transfer learning algorithm to effectively leverage landslide data outside the target area. Al-Najjar and Pradhan (2021) used Generative Adversarial Networks to enrich inventory data, resulting in an enhancement of prediction performance.

Against the backdrop of intricate geological conditions, the formation mechanism of geological disasters becomes highly complex. In addition to variables such as rainfall and engineering activities, the presence of active fractures emerges as a pivotal influencing factor. The rock mass within the fault zone experiences significant fragmentation, leading to a higher concentration of landslides in the vicinity of these zones. Previous studies have extensively investigated landslides along fault zones. For instance, Chen et al. (2022) used a hybrid ML model to conduct LSM along the Anninghe fault zone in China. Demir

(2019) applied a GIS-based approach to evaluate the landslide susceptibility in a part of the North Anatolian fault zone, indicating the significant influence of distance from faults on landslide occurrence. Zhang et al. (2023) employed the XGBoost model to assess the landslide susceptibility in Fengjie County, Chongqing City, China, emphasizing the distance from faults as the second most influential conditioning factor. Zhou et al. (2021b) utilized a hybrid RF model to assess landslide susceptibility in Wuxi County, Chongqing City, China, identifying the distance from faults as the primary contributing factor to landslides.

Based on the aforementioned analysis, it becomes evident that the influence of fault zones on slope cannot be disregarded. However, few studies have been undertaken to enhance the predictive accuracy of ML models based on qualitative analysis of fault zones. In this study, a division strategy based on fault zones is employed to partition the Chengkou County, Chongqing City, China into two sub-zones. The region within a 1 km distance from faults is designated as sub-zone A, while the remaining area is defined as sub-zone B. RF model serves as the classifier for assessing the landslide susceptibility of the study area. Five performance indicators are selected to evaluate the

performance of ML model. The results of 100 random realizations of RF model are used to measure the prediction uncertainty. The impact of the division strategy based on fault zones on both model performance and prediction uncertainty is investigated.

2 Materials

2.1 Description of the study area

Chongqing is located in the upper reaches of the Yangtze River, in the transitional zone between the Qinghai-Tibet Plateau and the middle-to-lower reaches of the Yangtze River Plain. As shown in Fig. 1, the study area, Chengkou County (31°37'N - 32°12'N, 108°15'E - 109°16'E), is located in the northeast of Chongqing. This area spans 96 km in an east-to-west direction and 66 km in a north-to-south direction, encompassing a total area of 3289.08 km². Chengkou County falls under a subtropical humid monsoon climate. The study area is characterized by its typical mountainous terrain, comprising prominent landforms such as ridges, valleys, and slopes, with elevation ranging from 483 m to 2687 m.
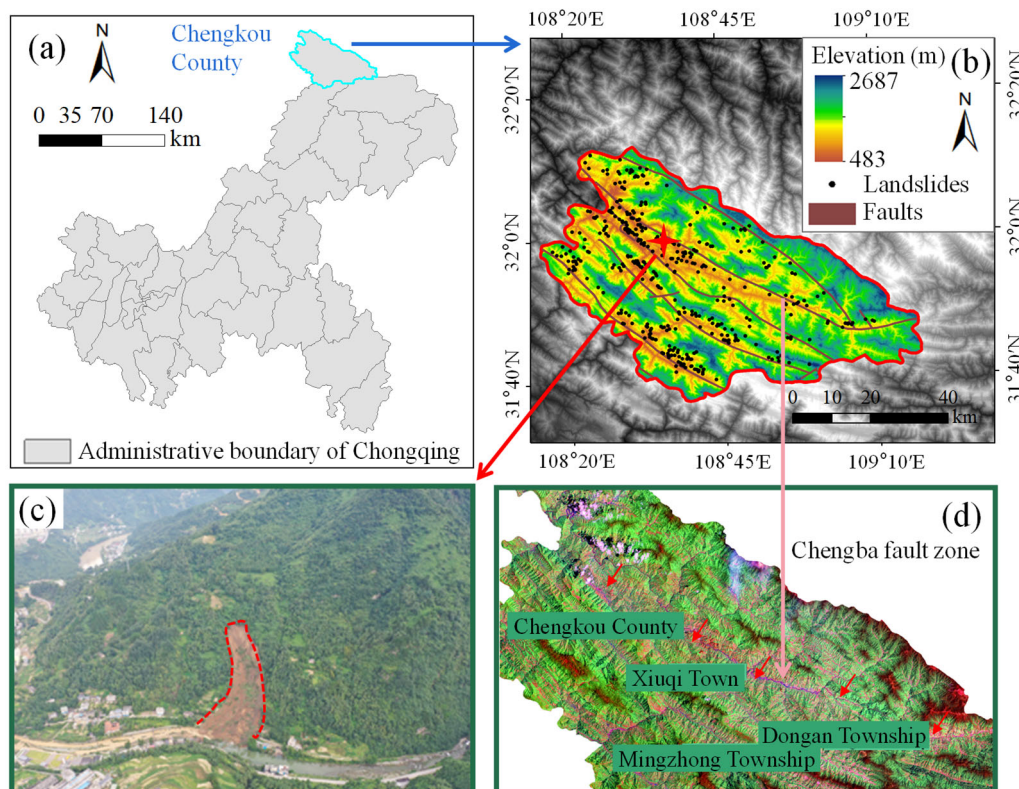
Fig. 1 Study area. (a) Location of the study area; (b) Landslide inventory map; (c) A landslide event (2019.08.14); (d) Remote sensing image of Chengba fault zone.

The rocks in the study area are formed in the Mesozoic Triassic, Paleozoic Permian, Silurian, Ordovician, and Cambrian, and the lithological categories mainly include gabbro and amphibolite. The faults in Chengkou County are highly developed, coupled with complex lithology, rendering the area particularly susceptibility to geological disasters. Fig. 1d depicts the remote sensing image of Chengba fault zone. The section of the Chengba fault in Chongqing is wave-shaped, north-dipping, with a dip angle of 40° - 80°. This fault exhibits significant scale and has a long history of activity, exerting control over the sedimentation processes on both sides. The terrain types on either side of the fault exhibit apparent differences. The southwestern side features rugged topography, with mountain ridges perpendicular to the Chengba fault, while the northeastern side boasts relatively intact terrain, with ridges predominantly parallel to the fault. Owing to the combined effects of multiple factors, landslide disasters along the Chengba fault zone occur frequently, resulting in substantial impacts on local residents.

2.2 Landslide inventory

Cataloguing landslides is a crucial step in LSM, and the completeness and quality of landslide data directly impact the ultimate predictive accuracy (Zhang et al. 2023). In this study, the landslide inventory is sourced from the Chongqing Geological Monitoring Station, incorporating information obtained from field surveys and historical hazard reports conducted by front-line surveyors. The landslide inventory datasheet encompasses various essential details, including the occurrence time, location, geometric characteristics of slope body, main triggering factor, landslide type, elevation information, area, volume, and hazard grade. A total of 401 landslides were recorded in Chengkou County between 1970 and 2020. The distribution pattern of these landslides is depicted in Fig. 1b.

Based on the landslide inventory, a statistical analysis is presented in Fig. 2. Fig. 2 indicates that the year with the highest number of landslide disasters is 2010, witnessing a total of 187 landslides. In terms of months, July stands out as having the highest incidence of landslides (75.56%), followed by June (9.98%), with the remaining months collectively contributing 14.46%. This pattern can be attributed to the significant correlation with precipitation. From the

perspective of area, 37.4% of historical landslides have an area between 10^4 and 10^5 m², 1.5% greater than 10^5 m² and 61.1% smaller than 10^4 m². Furthermore, 76 landslides exhibit a volume exceeding 10^5 m³, according for 18.95% of total inventory, 60.85% of total inventory between 10^4 and 10^5 m³ and 20.2% smaller than 10^4 m³. Notably, a severe landslide incident occurred in Longtian Township, Chengkou County, on August 14, 2019, as depicted in Fig. 1c. This disaster resulted in the destruction of three buildings and tragically claimed the lives of seven individuals.
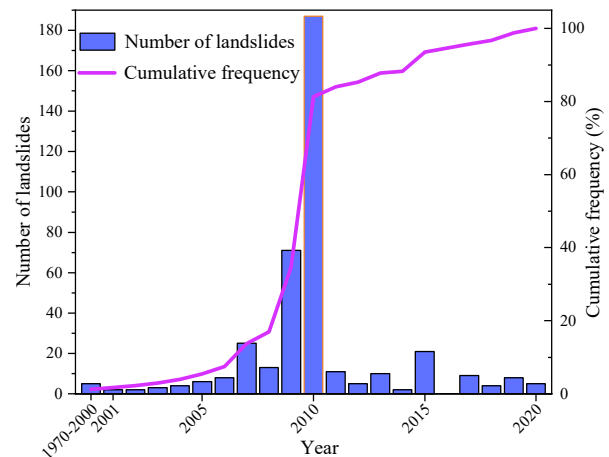
Fig. 2 Cumulative frequency and number of historical landslides.

2.3 Data collection and preparation

Landslide susceptibility assessment is a typical binary classification problem. In this study, the ratio of the positive/negative samples is 1:1. The positive samples are 401 landslide cases. The negative samples are extracted based on 500m buffer zone strategy, which represents selecting non-landslide points 500m away from landslide points (Wang et al. 2023c; Zhang et al. 2023). The occurrence of landslide is attributed to a combination of factors, including rainfall, human activities, and various other elements. There exists no standardized guideline for selecting conditioning factors. Taking all influencing factors into account proves to be a time-consuming and inefficient process. Therefore, a total of 18 conditioning factors from four different aspects are selected as inputs for the ML model, as outlined in Table 1 and depicted in Fig. 3 (Zhou et al. 2021b). The sources and types of GIS data used in this study are presented in Table 2.

A 30m grid resolution is employed for LSM in this study. The grid covering Chengkou County comprises a total of 3659408 cells. Apart from elevation, the

Table 1 Classification of conditioning factors

Factor category	Conditioning factor	Type	Classification standard
Topographic factors	Elevation (m)	9	1. < 500; 2. 500-750; 3. 750-1000; 4. 1000-1250; 5. 1250-1500; 6. 1500-1750; 7. 1750-2000; 8. 2000-2250; 9. > 2250
	Relief degree of land surface (RDLS) (m)	8	1. < 20; 2. 20-30; 3. 30-40; 4. 40-50; 5. 50-60; 6. 60-70; 7. 70-80; 8. > 80
	Slope (°)	7	1. < 5; 2. 5-15; 3. 15-25; 4. 25-35; 5. 35-45; 6. 45-55; 7> 55
	Aspect	9	1. Flat; 2. North; 3. Northeast; 4. East; 5. Southeast; 6. South; 7. Southwest; 8. West; 9. Northwest
	Curvature	6	1. < -1; 2. -1 to -0.5; 3. -0.5 to 0; 4. 0-0.5; 5. 0.5-1; 6. > 1
	Plan Curvature	6	1. < -1; 2. -1 to -0.5; 3. -0.5 to 0; 4. 0-0.5; 5. 0.5-1; 6. > 1
	Profile Curvature	6	1. < -1; 2. -1 to -0.5; 3. -0.5 to 0; 4. 0-0.5; 5. 0.5-1; 6. > 1
	Topographic wetness index (TWI)	5	1. < 4; 2. 4-6; 3. 6-8; 4. 8-10; 5. > 10
	Stream power index (SPI)	7	1. < 15; 2. 15-30; 3. 30-45; 4. 45-60; 5. 60-100; 6. 100-1000; 7. > 1000
Geological factors	Soil thickness (m)	4	1. < 75; 2. 75-95; 3. 95-115; 4. 115-125
	Distance from faults (m)	7	1. < 500; 2. 500-1000; 3. 1000-1500; 4. 1500-2000; 5. 2000-2500; 6. 2500-3000; 7. > 3000
Environmental factors	Normalized vegetation index (NDVI)	6	1. 0-0.5; 2. 0.5-0.6; 3. 0.6-0.7; 4. 0.7-0.8; 5. 0.8-0.9; 6. 0.9-1.0
	Distance from rivers (m)	7	1. < 200; 2. 200-400; 3. 400-600; 4. 600-800; 5. 800-1000; 6. 1000-1200; 7. > 1200
	Aridity	6	1. < 700; 2. 700-800; 3. 800-900; 4. 900-1000; 5. 1000-1100; 6. > 1100
	Index of moisture (IM)	4	1. < 4500; 2. 4500-5500; 3. 5500-6500; 4. > 6500
Human engineering activities	Distance from roads (m)	7	1. < 200; 2. 200-400; 3. 400-600; 4. 600-800; 5. 800-1000; 6. 1000-1200; 7. > 1200
	GDP	/	/
	Human activity index (HAI)	3	1. 0-0.25; 2. 0.25-0.5; 3. 0.5-1

Table 2 Data and data sources

Data name	Data sources	Type
Historical landslides	Chongqing Geological Monitoring Station	Datasheet
Elevation	Resource and Environment Science and Data Center	Grid
Soil thickness	Resource and Environment Science and Data Center	Grid
Faults	Resource and Environment Science and Data Center	Vector
Normalized vegetation index	Resource and Environment Science and Data Center	Grid
River network	Resource and Environment Science and Data Center	Vector
Aridity	Resource and Environment Science and Data Center	Grid
Index of moisture	Resource and Environment Science and Data Center	Grid
Road network	Resource and Environment Science and Data Center	Vector
GDP	Resource and Environment Science and Data Center	Grid
Human activity index	Nation Earth System Science Data Center	Grid
Other topographic factor	GIS processing of Elevation	Grid

remaining eight topographic factors are derived from DEM data through appropriate processing techniques. The values of relief degree of land surface (RDLS), slope, aspect, curvature, plane curvature, and profile curvature can be calculated by the built-in function of GIS software. The distances from faults, rivers, and roads are determined through nearest analysis module within the software. Topographic wetness index (TWI) and stream power index (SPI) can be obtained using Eq. (1) and Eq. (2), respectively.

$$TWI = \ln(A_s / \tan \beta) \tag{1}$$

$$SPI = A_s \times \tan \beta \tag{2}$$

where A_s represents the slope contributing area, and β is the slope inclination.

3 Methodology

The purpose of this study is to investigate how the
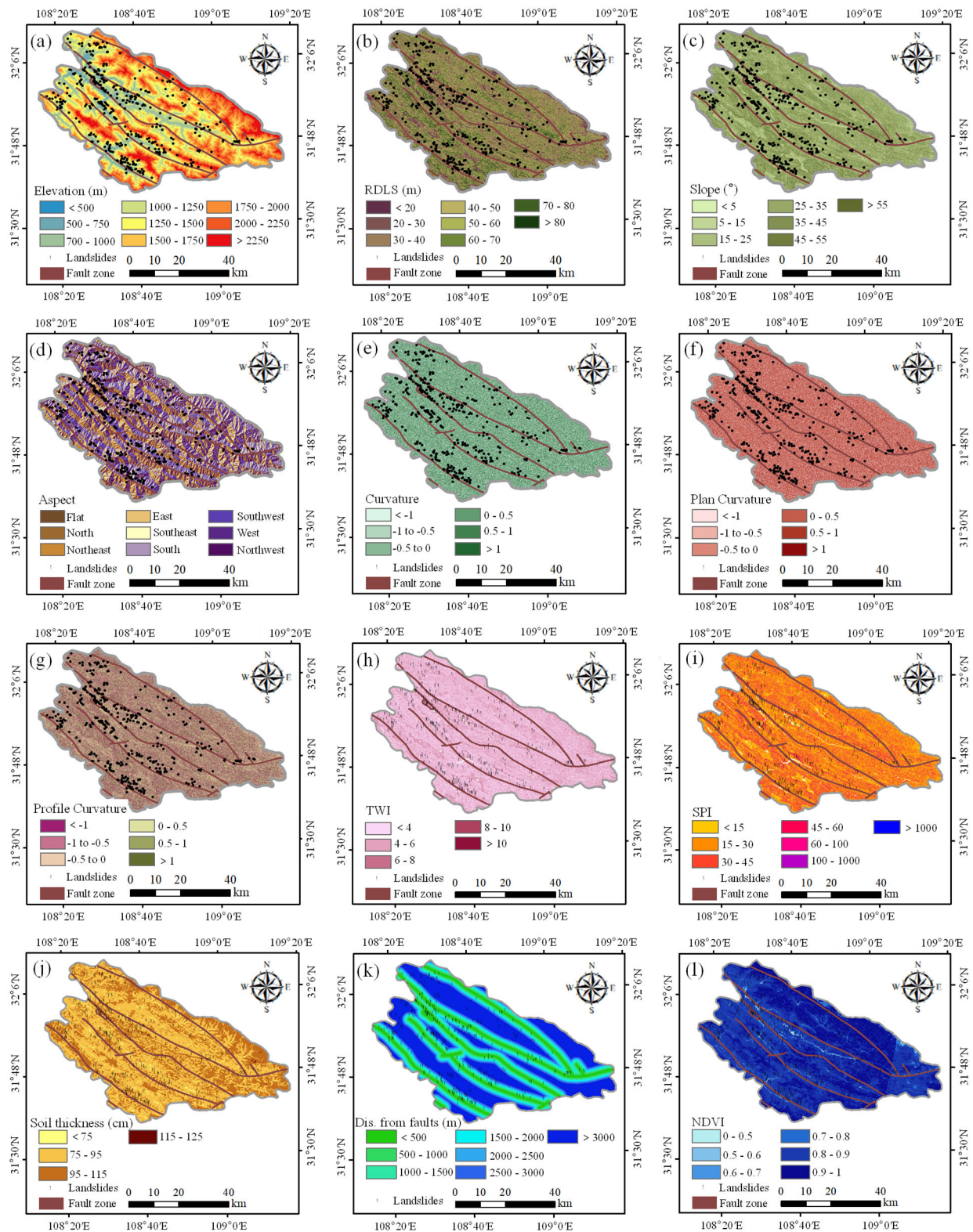
Fig. 3 Conditioning factors selected as inputs for the ML model: (a) Elevation; (b) Relief degree of land surface; (c) Slope; (d) Aspect; (e) Curvature; (f) Plan curvature; (g) Profile curvature; (h) Topographic wetness index; (i) Stream power index; (j) Soil thickness; (k) Distance from faults; (l) Normalized vegetation index; (m) Distance from rivers; (n) Aridity; (o) Index of moisture; (p) Distance from roads; (q) GDP; (r) Human activity index.

(-To be continued-)

(-Continued-)
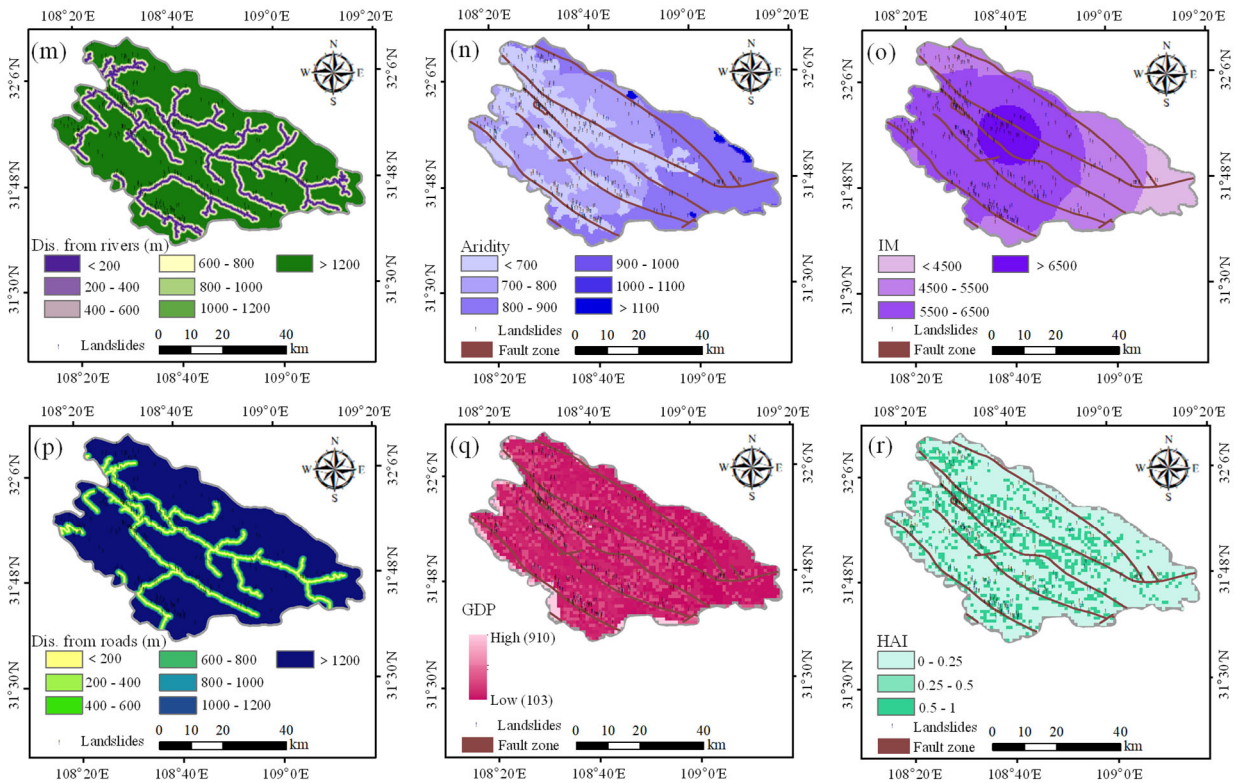
Fig. 3 Conditioning factors selected as inputs for the ML model: (a) Elevation; (b) Relief degree of land surface; (c) Slope; (d) Aspect; (e) Curvature; (f) Plan curvature; (g) Profile curvature; (h) Topographic wetness index; (i) Stream power index; (j) Soil thickness; (k) Distance from faults; (l) Normalized vegetation index; (m) Distance from rivers; (n) Aridity; (o) Index of moisture; (p) Distance from roads; (q) GDP; (r) Human activity index.

division strategy based on fault zone affects both model performance and prediction uncertainty. Fig. 4 depicts the flowchart of the proposed LSM method based on division strategy. It consists of five sequential steps, namely data collection and preparation, GIS processing, data-driven module considering division strategy, mapping, and model evaluation. In the data-driven module, we use Sklearn-ensemble-Random Forest Classifier in Python to establish an RF model. The model is trained using 388 landslide records from 1970 to 2018, along with an equal number of non-landslide samples. The remaining 13 landslide cases from 2019 to 2020 are employed to validate the generated landslide susceptibility maps.

3.1 Random forest

RF is a bagging ensemble algorithm proposed by Breiman (2001) based on classification and regression tree. The flowchart of the RF algorithm is illustrated in Fig. 5. RF model operates by constructing a multitude of decision trees during training. Each tree is built on a random subset of the training data, and at each split, only a random subset of features is considered. In RF model, the ultimate prediction for classification tasks is established using a voting mechanism. Each tree casts a vote for a particular class, and the predicted class is the one that receives the most votes. The voting mechanism can be expressed through the following formula:

$$\hat{Y} = \text{argmax}_c \sum_{i=1}^{N} I(\hat{Y}_i = c) \tag{3}$$

where $\hat{Y}$ represents final predicted class, N is the total number of decision trees in the RF model, c represents a particular class, $I(\hat{Y}_i = c)$ is the indicator function, yielding 1 if $\hat{Y}_i = c$ and 0 otherwise.

For the base evaluator, the important role of the decision tree is to classify specific features from the input data with labels and features classification and present the decision rules in a tree diagram (Zhang et al. 2023). During the growth of the decision tree, the Gini coefficient plays a crucial role as a measure of impurity for the classification trees within the ensemble. The Gini impurity is a metric that quantifies
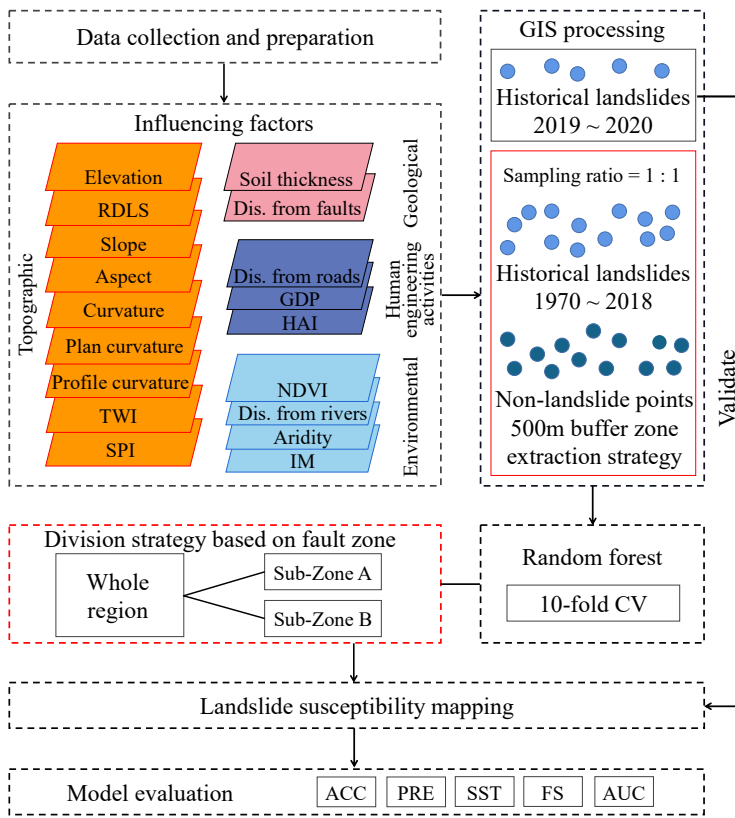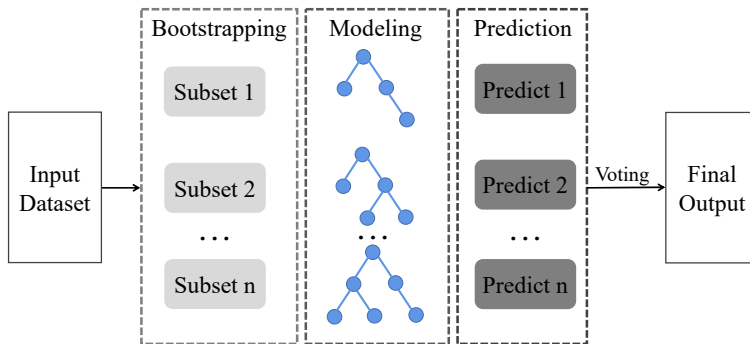
Fig. 4 Flowchart of the methodology.

Fig. 5 Schematic diagram of random forest model.

the likelihood of misclassifying a randomly chosen element in the dataset. It is minimized by selecting feature and threshold combinations that result in more homogeneous child nodes. Specifically, the Gini coefficient is defined as follows:

$$Gini(t) = 1 - \sum_{i=1}^{C} p(i|t)^2 \qquad (4)$$

where C represents the number of classes, $p(i|t)$ is the proportion of samples belonging to class i in node t.

The *max_depth* and *n_estimators*, closely associated with the Gini coefficient, are the hyper-parameters that need to be focused on in the RF model (Liu et al. 2023a). The *max_depth* is mainly used to

control the number of layers in the tree. A smaller *max_depth* may result in shallow trees that might not capture complex relationships in the data, while a larger *max_depth* can lead to overfitting. The number of decision trees is represented by *n_estimators*, which needs to be balanced between model effectiveness and training difficulty. Properly tuning *n_estimators* can improve predictive performance, especially when dealing with complex datasets. In addition, *random_state* mainly ensures that the same forest is grown for each training, thereby ensuring the randomness of training.

3.2 K-fold cross-validation

Ideally, when sufficient data is available, it is recommended to randomly split the dataset into three components: a training set, a validation set, and a testing set. The training set enables the model to learn from labeled samples, the validation set is crucial for fine-tuning a ML model, and the testing set is used to assess the model's overall performance on new, unseen data. Obviously, the dataset for this study is not adequate. Only 388 positive samples and 388 negative samples are used to construct the model. K-fold cross-validation (CV) is a technique used in ML to assess the performance of the model, especially when the available dataset is limited. Instead of using a single split between training and testing sets, k-fold CV randomly divides the dataset D into k equally sized and mutually exclusive subsets:

$$D = D_1 \cup D_2 \cup \dots \cup D_k \,, D_i \cap D_j \\ = \emptyset \ (i \neq j) \qquad (5)$$

In each iteration, k-1 subsets are randomly selected as the training set, and the remaining one subset is used as the testing set. This process results in k distinct training/testing sets, enabling k rounds of training and testing. While there is no strict rule for determining the value of k, common choices in applied ML include k=5 or 10. Referring to the previous studies (Sun et al. 2020; Sun et al. 2023), the value of k is set to 10 in this study. The schematic diagram of 10-fold CV is depicted in Fig. 6.
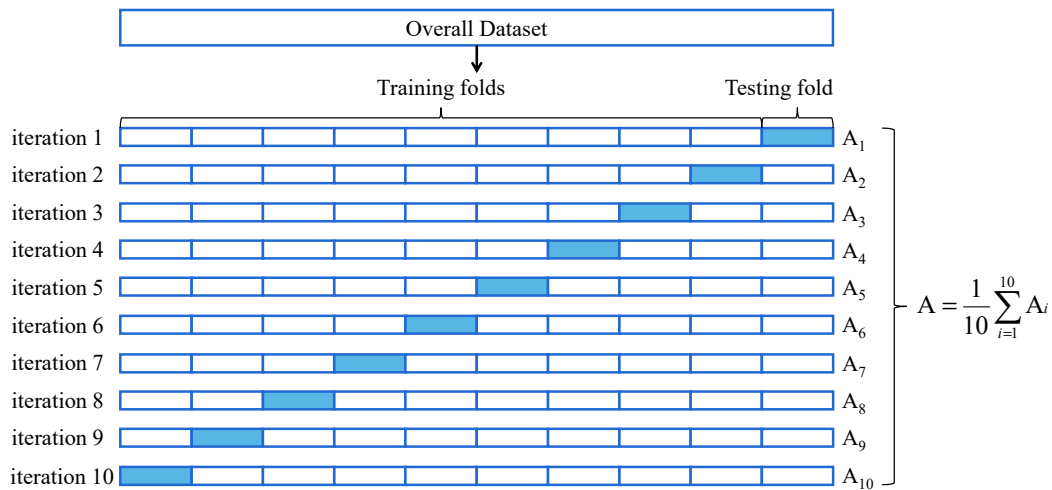
Fig. 6 Schematic diagram of 10-fold cross-validation.

Table 3 Distribution of sample points in different regions

Region	Landslide points	Non-landslide points	Fishnet cells
Whole region	388	388	3659408
Sub-Zone A	129	129	864998
Sub-Zone B	259	259	2794410

3.3 Division strategy based on fault zone

In the case of Chengkou County, which features multiple fault zones, this study utilizes a division strategy based on fault zone, coupled with the RF model, to establish a comprehensive framework for regional LSM. As illustrated in Fig. 7, the study area is partitioned into two sub-zones based on the situation of fault zones (Huang et al. 2023). The region within a distance of 1 km from the faults is designated as sub-zone A, while the remaining area is labeled as sub-zone B. For both sub-zone A and sub-zone B, buffer zones of 500 m are established around historical landslide points. Negative sample points are randomly selected outside these buffer zones. After partitioning, the distribution of samples in different regions is detailed in Table 3.

3.4 Frequency ratio

Frequency ratio (FR) is utilized in this study to illustrate the contribution of conditioning factors to landslide occurrence. In the assessment of landslide susceptibility, it is assumed that future landslides will occur under similar conditions as historical landslides (Lee and Pradhan 2007; Chen and Chen 2021). When the FR exceeds 1, it signifies that the particular class of the conditioning factor is conducive to landslide occurrence. The FR can be determined using Eq. (6) (Zhang et al. 2019).

Fig. 7 Division strategy based on fault zone.

$$\mathrm{FR} = \frac{L'/L}{A'/A} \tag{6}$$

where L' represents the number of landslide cases in a specific domain for each class, while L denotes the total number of landslides in the study area. Similarly, A' represents the number of pixels in the same domain for each class, while A stands for the total number of pixels.

3.5 Model performance

The confusion matrix (Fig. 8) is a visual tool commonly utilized in supervised learning to analyze the prediction accuracy of binary classification
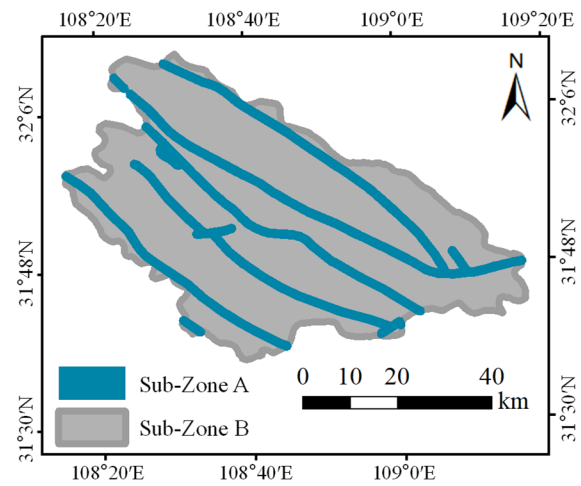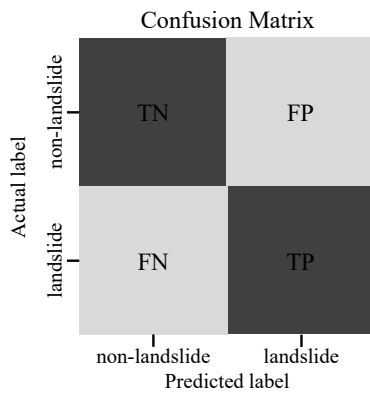
Fig. 8 Confusion matrix used to analyze the prediction accuracy of binary classification problems.

problems. The receiver operating characteristic curve (ROC) is a composite indicator that reflects a continuous variable of sensitivity and specificity. AUC is defined as the area under the ROC curve. The ROC curve may not always provide a straightforward distinction between classifie to determine superiority. In such cases, the AUC serves as a valuable quantitative metric for comparing classifier performance. Accuracy (ACC), Precision (PRE), Recall (SST), F-score (FS), and AUC are selected as performance indicators to evaluate the ML model (Hong et al. 2019; Chen et al. 2021). The value of ACC, PRE, SST, and FS can be calculated using Eqs. (7-10).

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$PRE = \frac{TP}{TP + FP} \quad (8)$$

$$SST = \frac{TP}{TP + FN} \quad (9)$$

$$FS = \frac{2PRE \times SST}{PRE + SST} \quad (10)$$

where TP represents that the actual label and the predicted label are both positive, TN indicates that two labels are both negative, FP means that the actual label is negative while the predicted label is positive, and FN denotes that the actual label is positive while the predicted label is negative (Wang et al. 2021).

4   Results and Analyses

4.1 Landslide susceptibility mapping

Prior to conduct LSM, the susceptibility values are classified into five levels using Jenks natural fracture method: very high, high, moderate, low, and very low (Wang et al. 2020). The mapping results and landslide kernel density analysis are presented in Fig. 9. A comparison of Figs. 9a and Figs. 9b shows that Fig. 9a overestimates the landslide susceptibility of the study area. Fig. 9a illustrates that the very high susceptibility areas in Chengkou County are mainly concentrated in the northwestern region. Fig. 9b shows that the very high and high susceptibility areas are mainly concentrated along the fault zone. The map displayed in Fig. 9b generally aligns with the patterns observed in Fig. 9c.

The landslide susceptibility thresholds and quantitative analysis results of different regions are summarized in Table 4. It is evident that the value of landslide ratio/area ratio exhibit a positive correlation with the landslide susceptibility level, consistent with the findings of Liu et al. (2023a). As the susceptibility level improves, the number of landslides occurring under a unit area increases. Compared with the statistical results of the whole region, after applying division strategy based on fault zone, the area ratio of very low and low susceptibility areas increase, while the area ratio of very high and high susceptibility areas
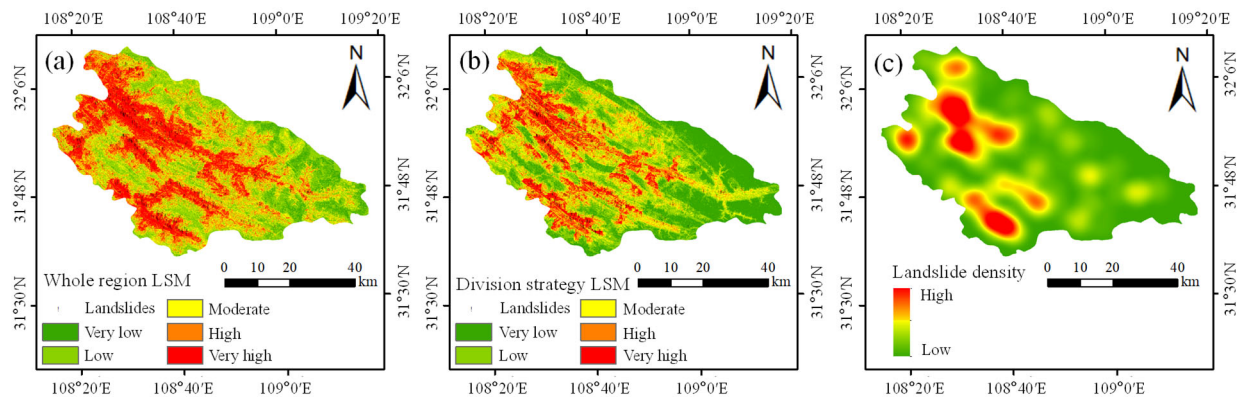
Fig. 9 Landslide susceptibility maps: (a) Whole region; (b) Division strategy; (c) Historical landslides density map.

Table 4 Characteristics of five susceptibility levels in different maps

| Region	Landslide susceptibility level	Landslide susceptibility threshold	Landslide ratio (%)	Area ratio (%)	Landslide ratio /area ratio
Whole region	Very low	0 - 0.098	0	6.76	0
	Low	0.098 - 0.298	1.29	25.46	0.051
	Moderate	0.298 - 0.498	5.41	25.41	0.213
	High	0.498 - 0.698	15.46	20.50	0.754
	Very high	0.698 - 1	77.84	21.87	3.559
Division strategy	Very low	0 - 0.188	0.52	24.86	0.021
	Low	0.188 - 0.380	2.84	25.59	0.111
	Moderate	0.380 - 0.561	10.05	22.49	0.447
	High	0.561 - 0.745	22.16	15.87	1.396
	Very high	0.745 - 1	64.43	11.19	5.758

decrease. Table 4 demonstrates that 86.59% historical landslides are distributed in very high or high susceptibility area after employing division strategy. Additionally, the validation of the generated susceptibility map with historical landslides (2019-2020) reveals that 4 out of 13 landslides are distributed in the very high susceptibility area, and 2 in high susceptibility area. These highlight the reliability of the division strategy. Moreover, after applying division strategy, the value of landslide ratio/ area ratio of high and very high susceptibility areas improve significantly. This means that using smaller high-risk areas can cover more historical landslides.

**4.2 Model performance and prediction uncertainty**

In order to compare the model performance and prediction uncertainty, we analyze the result of the RF model through 100 random realizations (Achu et al. 2023; Wei et al. 2023). Each realization obtains an individual trained model. The AUC values of 100 random realizations for different regions are presented in Fig. 10. The trained model with highest AUC value is executed to generate landslide susceptibility map. The standard deviation and range of AUC values represent the prediction uncertainty. As depicted in Fig. 10, after applying division strategy, there is a tendency for AUC values to increase, especially for sub-zone B. The result shows that AUC values of whole region range from 0.728 to 0.781, the corresponding standard deviation is 0.011. The statistics of sub-zone A show a slightly larger standard deviation and a wider AUC bound, spanning from 0.738

to 0.794. The model in sub-zone B exhibits superior result, indicated by a narrower bound ranging from 0.787 to 0.829 and a smaller standard deviation of 0.008. In terms of the value of the standard deviation, the division strategy has little effect on the prediction uncertainty.

The ROC curves from 10-fold CV across different regions are depicted in Fig. 11. Comparing with the AUC results for the whole region (0.781), the mean AUC of sub-zone B experiences an increase of 6.15%, while sub-zone A demonstrates a corresponding increase of 1.66%. According to the confusion matrices illustrated in Fig. 12, the values of the four evaluation indicators (ACC, PRE, SST, and FS) can be determined using Eqs. (7-10). The results of the indicators for the three models are summarized in Table 5. The analysis reveals that all four evaluation criterion values are significantly enhanced by applying division strategy based on fault zone for LSM. Comparing to the results of the whole region, sub-zone A exhibits a 6.44% increase in ACC, and the F-score experiences a notable
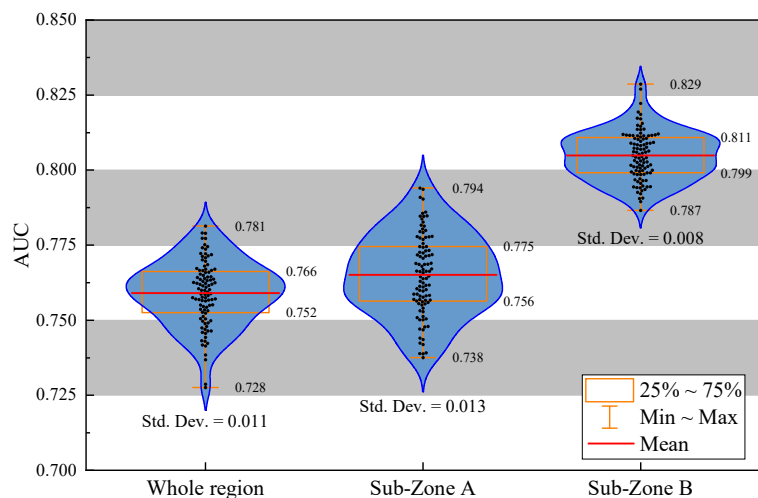
Fig. 10 Receiver operating characteristic curve-area under the ROC curve of 100 random realizations.
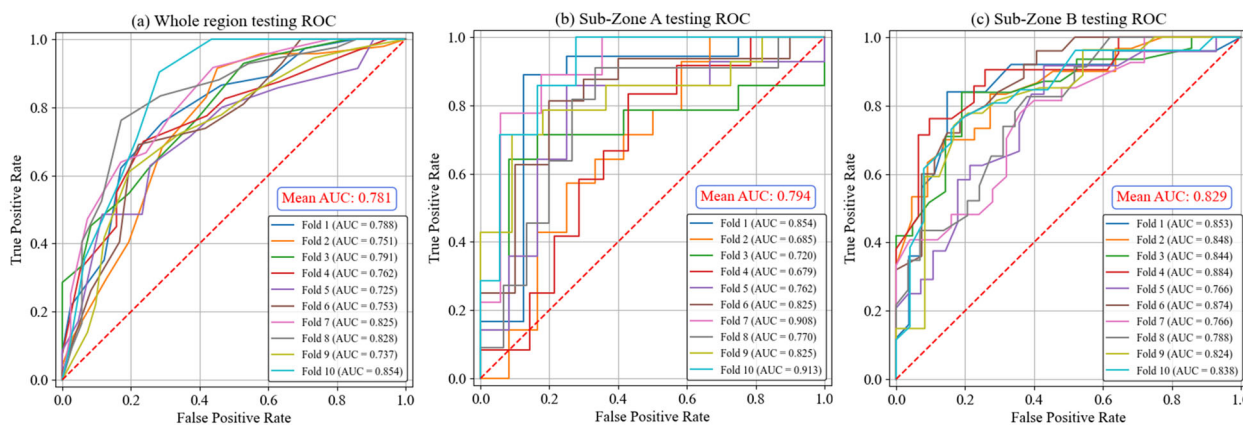
Fig. 11 Receiver operating characteristic curve-area under the ROC curve of different regions.
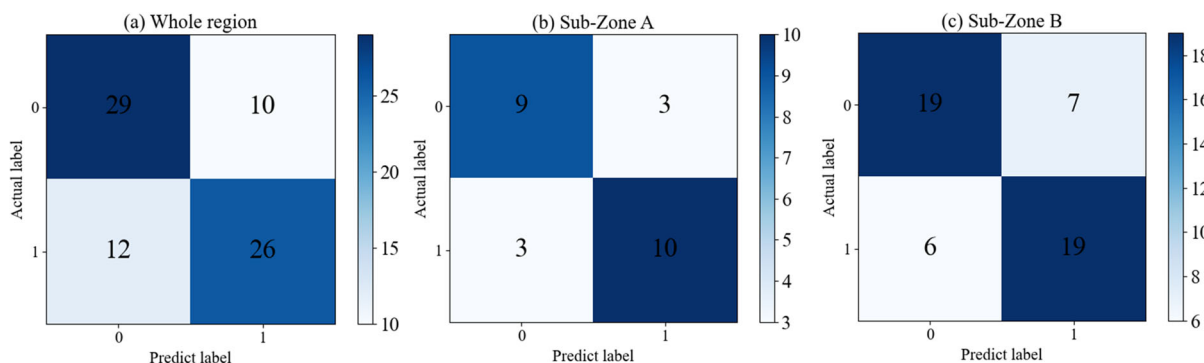
Fig. 12 Confusion matrices of different regions.

Table 5 Results of evaluation criterion

Model	Evaluation criterion				Mean AUC
	ACC	PRE	SST	F-score	
Whole region	0.714	0.722	0.684	0.703	0.781
Sub-Zone A	0.760	0.769	0.769	0.769	0.794
Sub-Zone B	0.745	0.731	0.760	0.745	0.829

improvement of 9.39%. The ACC of sub-zone B increases by 4.34%, and the F-score increases by 5.97% compared to the whole region.

4.3 Importance analysis of conditioning factors

Analyzing the importance of conditioning factors can offer valuable insights for disaster prevention and hazard management. The analysis is conducted through the feature-importance module of the RF model. The importance ranking of 18 factors in different regions are illustrated in Fig. 13. For the whole region, the top five most significant factors are aridity, elevation, distance from roads, IM, and distance from rivers. In terms of sub-zone A, distance from roads remains a dominant factor, followed by aridity, elevation, distance from rivers, and slope.

Within sub-zone B, the five most crucial factors are aridity, elevation, IM, aspect, and distance from roads. It is noteworthy that the importance of distance from faults increases in two sub-zones. This may be related to the division strategy adopted. In addition, all three regions exhibit the same characteristics, with elevation ranking highest among topographic factors, distance from faults ranking as the foremost geological factor, aridity prevailing as the paramount environmental factor, and distance from roads ranking highest among human engineering activity factors.

5 Discussion

### 5.1 Correlation analysis between the distribution of landslides and conditioning factors

The analysis of FR values aids in identifying the spatial relationship between conditioning factors and landslides (Chen and Chen 2021). The conditioning factors that exhibit high significance are thoroughly examined, including aridity, elevation, distance from roads, aspect, IM, slope, distance from rivers, distance

from faults. The results are summarized in Table 6. Among the six classes of aridity, it is observed that 72.32% of the landslides occurred in the first category. One reason for the influence of elevation on the distribution of landslides is the presence of different vegetation types and vegetation coverage within varying elevation ranges. Elevation also exhibits a strong correlation with precipitation. Changes in groundwater levels, resulting from rainfall, can impact slope stability. In the study area, landslides are predominantly concentrated within the elevation range of 500-2000 m. It is important to note that the impact of human engineering activities on landslides

Fig. 13 Importance of the conditioning factors: (a) Whole region; (b) Sub-Zone A; (c) Sub-Zone B.

Table 6 Analysis of the relationship between conditioning factors and landslides

Factors	Class	No. of landslide	Landslide ratio (%)	No. of pixels in domain	Area ratio (%)	FR
Aridity	< 700	290	72.32	1237612	33.82	2.14
	700-800	69	17.21	1153445	31.52	0.55
	800-900	42	10.47	1225536	33.49	0.31
	900-1000	0	0	8051	0.22	0
	1000-1100	0	0	5855	0.16	0
	> 1100	0	0	28909	0.79	0
Ele. (m)	< 500	0	0	82	≈ 0	0
	500-750	17	4.24	71823	1.96	2.16
	750-1000	76	18.95	342693	9.36	2.02
	1000-1250	120	29.93	584531	15.97	1.87
	1250-1500	100	24.94	716316	19.57	1.27
	1500-1750	50	12.47	724001	19.78	0.63
	1750-2000	26	6.48	612746	16.74	0.39
	2000-2250	11	2.74	404025	11.04	0.25
	> 2250	1	0.25	203191	5.55	0.04
Dis. from roads (m)	< 200	27	6.73	177782	4.86	1.39
	200-400	20	4.99	165512	4.52	1.10
	400-600	32	7.98	159548	4.36	1.83
	600-800	32	7.98	154870	4.23	1.89
	800-1000	39	9.73	150732	4.12	2.36
	1000-1200	35	8.73	146315	4.00	2.18
	> 1200	216	53.86	2704649	73.91	0.73
Aspect	Flat	26	6.48	423670	11.58	0.56
	North	38	9.48	447077	12.22	0.77
	Northeast	44	10.97	356180	9.73	1.12
	East	54	13.47	353285	9.65	1.40
	Southeast	63	15.71	423299	11.57	1.35
	South	77	19.20	498667	13.63	1.41
	Southwest	42	10.47	413793	11.31	0.92
	West	35	8.73	346938	9.48	0.93
	Northwest	22	5.49	396499	10.83	0.50
IM	< 4500	3	0.75	263172	7.19	0.10
	4500-5500	68	16.96	1102158	30.12	0.56
	5500-6500	287	71.57	1963097	53.65	1.33
	> 6500	43	10.72	330981	9.04	1.19
Slope (°)	< 5	4	1.04	41554	1.14	0.91
	5-15	40	9.87	343578	9.39	1.05
	15-25	105	26.23	830124	22.68	1.16
	25-35	114	28.31	1202846	32.87	0.86
	35-45	94	23.38	916493	25.04	0.93
	45-55	39	9.87	288379	7.88	1.25
	> 55	5	1.30	36434	1.00	1.30
Dis. from rivers (m)	< 200	36	8.98	258562	7.07	1.27
	200-400	27	6.73	239298	6.54	1.03
	400-600	36	8.98	228244	6.24	1.44
	600-800	39	9.73	220748	6.03	1.61
	800-1000	39	9.73	212394	5.80	1.68
	1000-1200	33	8.23	204404	5.59	1.47
	> 1200	191	47.63	2295758	62.73	0.76
Dis. from faults (m)	< 500	73	18.21	438527	11.98	1.52
	500-1000	56	13.97	426480	11.65	1.20
	1000-1500	45	11.22	411166	11.24	1.00
	1500-2000	44	10.97	387036	10.58	1.04
	2000-2500	44	10.97	362832	9.92	1.11
	2500-3000	43	10.72	337798	9.23	1.16
	> 3000	96	23.94	1295569	35.40	0.68
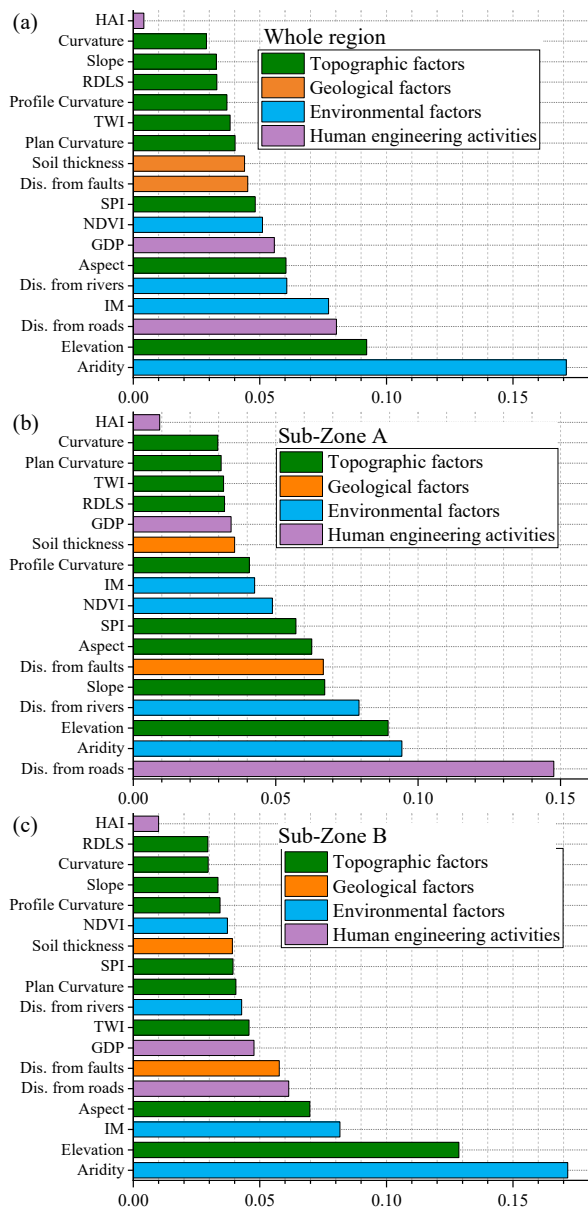
extends beyond road construction. Therefore, the relationship between distance from roads and landslides may not be evident in this study. Although solar radiation intensity varies across different aspects, the distribution of landslides among the nine aspect classes shows little variation. The most favorable aspect condition for landslides is found to be south-facing. Within the study area, landslides occur frequently when the value of IM ranges between 5500 and 6500. The slope value directly determines the stress distribution, which affects the slope stability. It is observed that the majority of landslides in the study area occur on slopes ranging from 15° to 45°, indicating that this particular range is highly conducive to landslide development. Regarding the distance from rivers, the class 800-1000 m has the highest FR value. For the distance from faults, a total of 129 landslide cases occurred within the range of 0-1 km from faults, according for 32.2% of the total landslides recorded. As the distance from faults increases, the number of landslides gradually decreases. Notably, the FR value of the first class (< 500 m) is 1.52, representing the highest value observed.

5.2 Role of division strategy based on fault zone

As previously demonstrated, in terms of the generated maps, the conventional method overestimates the landslide susceptibility of the study area. The area classified as very low and low susceptibility accounts for only 32.22% of the whole region. The application of a division strategy based on fault zone proves effective in mitigating this overestimation. Table 4 demonstrates that, following the implementation of the division strategy, 50.45% of the area falls in the very low or low susceptibility regions. The value of landslide ratio/area ratio for high and very high susceptibility areas indicates that the proposed division strategy can encompass a greater number of landslides within a smaller high-risk area. In addition, the division strategy optimizes the performance of the ML model. In comparison to the AUC results of the whole region, both sub-zones show an increase in mean AUC value. According to Fig. 12, when employing the division strategy, 23.1% of the positive samples are incorrectly predicted as negative in sub-zone A, and 24% in sub-zone B. However, for the whole region, 31.6% of the positive samples are incorrectly predicted as negative. After implementing the division strategy, sub-zone A demonstrates a 6.44% increase in ACC, accompanied by a notable

improvement of 9.39% in the F-score. Sub-zone B experiences a 4.34% increase in ACC, and its F-score shows a noteworthy improvement of 5.97%. Fig. 10 shows that sub-zone B exhibits superior prediction uncertainty, as evidenced by a narrower range and a smaller standard deviation. The standard deviation of sub-zone A is not significantly different from that of the whole region. Fig. 13 indicates that the division strategy does not significantly change the importance ranking of conditioning factors. Aridity, elevation, and distance from roads remain dominant factors in all three regions.

Chengkou County is one of the areas in Chongqing with highly intricate geological structures, where fault zones play a predominant and controlling role in the development features and spatial distribution of landslide disasters. Fault activities lead to a reduction in the strength of rock and soil, creating favorable conditions for the development of landslides through the provision of structural planes. Furthermore, landslides in the study area are influenced by factors such as the river systems, precipitation, and engineering activities. The development of landslides in this area exhibits a close correlation with the distance from faults. Table 6 indicates a gradual decrease in the number of landslides as the distance from the faults increases. More than 76% of historical landslides are distributed within 3 km of the fault zones, a pattern similar to the statistical characteristic observed by Zhou et al. (2021a) along the Anninghe fault zone.

However, adopting the proposed division strategy inevitably leads to a limited number of landslide samples in sub-zone A, thereby affecting the construction of the model's dataset. By incorporating more detailed landslide data from the vicinity of the fault zones, the precision of the proposed strategy can be further enhanced. Future studies could employ data augmentation techniques, such as collecting landslide data along the Xianshuihe fault zone (Guo et al. 2015) and the fault zone in Qinghai-Tibet Plateau (Qi et al. 2021), to train the model in sub-zone A. Transfer learning (Fu et al. 2023) may be a potential tool for addressing this problem.

6 Conclusions

In this study, a division strategy based on fault zone is proposed to assess the landslide susceptibility in Chengkou County, Chongqing City, China. A

comparative analysis is conducted, utilizing 388 historical landslides from 1970 to 2018 in the study area, to demonstrate the superiority of this strategy. The 13 landslide cases from 2019 to 2020 are employed to validate the generated susceptibility map. This study comprehensively examines the model performance and the importance of 18 conditioning factors. Moreover, 100 random realizations of the random forest model are conducted to investigate the prediction uncertainty. The main findings can be summarized as follows:

(1) The proposed division strategy based on fault zone introduces a novel approach to enhance the prediction accuracy of the landslide susceptibility mapping model. By implementing the division strategy, the accuracy, precision, recall, F-score, and AUC of both two sub-zones surpass those of whole region, indicating the effectiveness of this method. Comparing with the results of the whole region, the AUC of sub-zone B experiences an increase of 6.15%, while sub-zone A demonstrates a corresponding increase of 1.66%.

(2) The AUC distribution results of 100 random realizations show that the division strategy based on fault zone has little effect on the prediction uncertainty.

(3) The conditioning factors that have a significant impact on landslide are aridity, elevation, and distance from roads. The whole region and two sub-zones all exhibit the same characteristics, with elevation, distance from faults, aridity, and distance from roads ranking highest among topographic factor, geological factor, environmental factor, and human engineering activity factors, respectively.

(4) The proposed method can reduce the overestimation of landslide susceptibility in the study area. The value of landslide ratio/area ratio for high and very high susceptibility areas suggests that the proposed division strategy can encompass a greater number of landslides within a smaller high-risk area.

Acknowledgments

Author Contribution

WANG Yunhao: Methodology; WANG Luqi: Supervision; LIU Songlin: Software; SUN Weixin: Validation; LIU Pengfei, Zhu Lin, Guo Tong: Field investigation. YANG Wenyu: Conceptualization. All authors have read and agreed to the published version of the manuscript.

Ethics Declaration

Availability of Data/Materials: Data are available from the corresponding author on request.

Conflict of Interest: The authors declare no conflict of interest.

References

Achu AL, Aju CD, Di Napoli M, et al. (2023) Machine-learning based landslide susceptibility modelling with emphasis on uncertainty analysis. Geosci Front 14(6): 101657. https://doi.org/10.1016/j.gsf.2023.101657

Aditian A, Kubota T, Shinohara Y (2018) Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. Geomorphology 318: 101-111. https://doi.org/10.1016/j.geomorph.2018.06.006

Al-Najjar HAH, Pradhan B (2021) Spatial landslide susceptibility assessment using machine learning techniques assisted by additional data created with generative adversarial networks. Geosci Front 12: 625-637. https://doi.org/10.1016/j.gsf.2020.09.002

Breiman L (2001) Random forests. Mach Learn 45: 5-32. https://doi.org/10.1023/A:1010933404324

Chen W, Chen X, Peng JB, et al. (2021) 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

Chen X, Chen W (2021) GIS-based landslide susceptibility assessment using optimized hybrid machine learning methods. Catena 196: 104833. https://doi.org/10.1016/j.catena.2020.104833

Chen Z, Zhou H, Ye F, et al. (2022) Landslide susceptibility mapping along the Anninghe fault zone in China using SVM and ACO-PSO-SVM models. Lithosphere 2022(1): 5216125. https://doi.org/10.2113/2022/5216125

Demir G (2019) GIS-based landslide susceptibility mapping for a part of the north Anatolian fault zone between Resadiye and Koyulhisar (Turkey). Catena 183: 104211. https://doi.org/10.1016/j.catena.2019.104211

Fu Z, Li C, Yao W (2023) Landslide susceptibility assessment through tradaboost transfer learning models using two landslide inventories. Catena 222: 106799. https://doi.org/10.1016/j.catena.2022.106799

Guo C, Montgomery DR, Zhang Y, et al. (2015) Quantitative assessment of landslide susceptibility along the Xianshuihe

fault zone, Tibetan Plateau, China. Geomorphology 248: 93-110. https://doi.org/10.1016/j.geomorph.2015.07.012

Hong HY, Miao YM, Liu JZ, et al. (2019) Exploring the effects of the design and quantity of absence data on the performance of random forest-based landslide susceptibility mapping. Catena 176: 45-64. https://doi.org/10.1016/j.catena.2018.12.035

Hu Q, Zhou Y, Wang S, et al. (2020) Machine learning and fractal theory models for landslide susceptibility mapping: Case study from the Jinsha river basin. Geomorphology 351: 106975. https://doi.org/10.1016/j.geomorph.2019.106975

Huang FM, Cao ZS, Jiang SH, et al. (2020) Landslide susceptibility prediction based on a semi-supervised multiple-layer perceptron model. Landslides 17: 2919-2930. https://doi.org/10.1007/s10346-020-01473-9

Huang W, Ding M, Li Z, et al. (2023) Landslide susceptibility mapping and dynamic response along the Sichuan-Tibet transportation corridor using deep learning algorithms. Catena 222: 106866. https://doi.org/10.1016/j.catena.2022.106866

Kainthura P, Sharma N (2022) Machine learning driven landslide susceptibility prediction for the Uttarkashi region of Uttarakhand in India. Georisk 16: 570-583. https://doi.org/10.1080/17499518.2021.1957484

Kavzoglu T, Teke A (2022) Advanced hyperparameter optimization for improved spatial prediction of shallow landslides using extreme gradient boosting (XGBoost). B Eng Geol Environ 81(5): 201. https://doi.org/10.1007/s10064-022-02708-w

Lee S, Pradhan B (2007) Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. Landslides 4: 33-41. https://doi.org/10.1007/s10346-006-0047-y

Li BH, Liu K, Wang M, et al. (2022) Global dynamic rainfall-induced landslide susceptibility mapping using machine learning. Remote Sensing 14(22): 5795. https://doi.org/10.3390/rs14225795

Lima P, Steger S, Glade T, et al. (2022) Literature review and bibliometric analysis on data-driven assessment of landslide susceptibility. J Mt Sci 19: 1670-1698. https://doi.org/10.1007/s11629-021-7254-9

Lin M, Teng S, Chen G, et al. (2023) Application of convolutional neural networks based on bayesian optimization to landslide susceptibility mapping of transmission tower foundation. B Eng Geol Environ 82(2): 51. https://doi.org/10.1007/s10064-023-03069-8

Liu S, Wang L, Zhang W, et al. (2023a) A physics-informed data-driven model for landslide susceptibility assessment in the Three Gorges Reservoir area. Geosci Front 14(5): 101621. https://doi.org/10.1016/j.gsf.2023.101621

Liu SL, Wang LQ, Zhang WA, et al. (2023b) A comprehensive review of machine learning-based methods in landslide susceptibility mapping. Geol J 58(6): 2283-2301. https://doi.org/10.1002/gj.4666

Qi T, Meng X, Qing F, et al. (2021) Distribution and characteristics of large landslides in a fault zone: A case study of the Qinghai-Tibet Plateau. Geomorphology 379: 107592. https://doi.org/10.1016/j.geomorph.2021.107592

Sun D, Gu Q, Wen H, et al. (2023) Assessment of landslide susceptibility along mountain highways based on different machine learning algorithms and mapping units by hybrid factors screening and sample optimization. Gondwana Res 123: 89-106. https://doi.org/10.1016/j.gr.2022.07.013

Sun D, Xu J, Wen H, et al. (2020) An optimized random forest model and its generalization ability in landslide susceptibility mapping: Application in two areas of Three Gorges Reservoir, China. J Earth Sci-China 31: 1068-1086. https://doi.org/10.1007/s12583-020-1072-9

Tang HM, Wasowski J, Juang CH (2019) Geohazards in the Three Gorges Reservoir area, China lessons learned from decades of research. Eng Geol 261: 105267. https://doi.org/10.1016/j.enggeo.2019.105267

Wang H, Wang L, Zhang L (2022) Transfer learning improves landslide susceptibility assessment. Gondwana Res 123: 238-254. https://doi.org/10.1016/j.gr.2022.07.008

Wang H, Zhang L, Yin K, et al. (2021) Landslide identification using machine learning. Geosci Front 12: 351-364. https://doi.org/10.1016/j.gsf.2020.02.012

Wang L, Wu C, Yang Z, et al. (2023a) Deep learning methods for time-dependent reliability analysis of reservoir slopes in spatially variable soils. Comput Geotech 159: 105413. https://doi.org/10.1016/j.compgeo.2023.105413

Wang L, Xiao T, Liu S, et al. (2023b) Quantification of model uncertainty and variability for landslide displacement prediction based on Monte Carlo simulation. Gondwana Res 123: 27-40. https://doi.org/10.1016/j.gr.2023.03.006

Wang Y, Wang L, Liu S, et al. (2023c) A comparative study of regional landslide susceptibility mapping with multiple machine learning models. Geol J. https://doi.org/10.1002/gj.4902

Wang YM, Feng LW, Li SJ, et al. (2020) A hybrid model considering spatial heterogeneity for landslide susceptibility mapping in Zhejiang Province, China. Catena 188: 104425. https://doi.org/10.1016/j.catena.2019.104425

Wei X, Zhang L, Gardoni P, et al. (2023) Comparison of hybrid data-driven and physical models for landslide susceptibility mapping at regional scales. Acta Geotech 18(8): 4453-4476. https://doi.org/10.1007/s11440-023-01841-4

Yang C, Liu L-L, Huang F, et al. (2023) Machine learning-based landslide susceptibility assessment with optimized ratio of landslide to non-landslide samples. Gondwana Res 123: 198-216. https://doi.org/10.1016/j.gr.2022.05.012

Zhang T, Han L, Han J, et al. (2019) Assessment of landslide susceptibility using integrated ensemble fractal dimension with kernel logistic regression model. Entropy 21(2): 218. https://doi.org/10.3390/e21020218

Zhang W, Liu S, Wang L, et al. (2022a) Landslide susceptibility research combining qualitative analysis and quantitative evaluation: A case study of Yunyang County in Chongqing, China. Forests 13(7): 1055. https://doi.org/10.3390/f13071055

Zhang WA, He YW, Wang LQ, et al. (2023) Landslide susceptibility mapping using random forest and extreme gradient boosting: A case study of Fengjie, Chongqing. Geol J 58(6): 2372-2387. https://doi.org/10.1002/gj.4683

Zhang WG, Li HR, Han L, et al. (2022b) Slope stability prediction using ensemble learning techniques: A case study in Yunyang County, Chongqing, China. J Rock Mech Geotech Eng 14(4): 1089-1099. https://doi.org/10.1016/j.jrmge.2021.12.011

Zhou C, Hu Y, Xiao T, et al. (2023) Analytical model for reinforcement effect and load transfer of pre-stressed anchor cable with bore deviation. Constr Build Mater 379: 131219. https://doi.org/10.1016/j.conbuildmat.2023.131219

Zhou HF, Liu B, Ye F, et al. (2021a) Landslide distribution and sliding mode control along the Anninghe fault zone at the eastern edge of the Tibetan Plateau. J Mt Sci 18: 2094-2107. https://doi.org/10.1007/s11629-020-6573-6

Zhou X, Wen H, Zhang Y, et al. (2021b) Landslide susceptibility mapping using hybrid random forest with GeoDetector and RFE for factor optimization. Geosci Front 12(5): 101211. https://doi.org/10.1016/j.gsf.2021.101211