



OPEN Study on landslide hazard risk in Wenzhou based on slope units and machine learning approaches

Dengjie Kang¹, Sheng Dan², Zhang Hua^{1✉}, Lu Jingyi³, Wang Chenlu³, Wang Zhenguo⁴ & Wang Shaohua⁴

Landslides are a prevalent and devastating form of geological disaster. These events occur when gravity causes rock and soil masses to slide along specific surfaces or zones, often triggered by intense rainfall, seismic activity, or human engineering activities. Assessing landslide hazard risk is crucial for effective disaster management, yet traditional approaches often rely on administrative or grid units, which lack the granularity needed for site-specific hazard management. This results in uniformly high-risk classifications for hilly areas, complicating practical engagement and increasing management costs. The study further combines historical landslide data and applies machine learning models such as Random Forest, XGBoost, and LightGBM to analyze landslide susceptibility in Wenzhou City, proposing a slope unit-based landslide hazard assessment method. The results are as follows: (1) Landslide Susceptibility across different slope units was categorized as low, low-moderate, moderate, moderate-high, high, and very high, with the very high-risk slope units accounting for 5.35% of the total area and the low-risk slope units covering the largest area (975.41 km²). (2) Among the machine learning models used for landslide susceptibility analysis at the slope unit level, the Random Forest model performed the best, demonstrating higher prediction reliability, with an accuracy of 77.94% for Random Forest, 76.95% for XGBoost, and 78.30% for LightGBM. (3) Extreme rainfall events significantly increased the proportion of high-risk slope units, particularly in mountainous and hilly areas. According to different rainfall return periods, the proportion of very high-risk slope units increased from 5.35 to 40.39% under the 100-year return period. (4) A case study of Xuekou Village validated the practical application of the slope unit risk assessment results and proposed preventive measures for medium-to-high-risk units, such as regular monitoring and enhanced vegetation coverage, to mitigate landslide risks.

Keywords Slope units, Machine learning, Landslide susceptibility, Landslide risk

Geological risks constitute the most prevalent natural disasters in China, primarily due to the country's complex and diverse geographical setting, which is characterized by highly variable topography and frequent geological activity^{1,2}. Geological disasters pose significant threats to public safety and the stability of the national economy. Recent statistics indicate that an average of 6751 geological hazard incidents have occurred annually over the past decade, with annual economic losses exceeding 3 billion yuan (Table 1). A considerable proportion of these incidents are landslides, accounting for approximately 57.95% of all reported geological disasters each year.

Landslides are geological disasters that occur when masses of rock and soil on slopes mobilize along specific sliding surfaces or zones due to gravity, following inward and downward trajectories. They are typically observed in steep, geologically complex areas experiencing intense rainfall^{3,4}. Landslides are prevalent worldwide, especially in developing nations, and often lead to significant casualties and extensive property damage⁵. For instance, on June 24, 2017, around 6:00 AM, an abrupt slope failure in XinMo village, Sichuan province, obstructed a river channel for 2 km, burying over 100 people⁶. In general, landslide-inducing processes are complex and may involve a range of hydrological, geological, meteorological, and anthropogenic factors⁷. Rainfall, seismic activity, and human engineering endeavors are direct triggering mechanisms that can lead to landslides^{8–11}. Prolonged or intense precipitation within a short duration can induce soil saturation, increasing the mass of the slope while reducing the soil's shear strength, thereby initiating landslides. Additionally, earthquakes, as abrupt and forceful

¹School of National Safety and Emergency Management, Beijing Normal University, Beijing 100875, China.

²Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China. ³School of Systems Science, Beijing Normal University, Beijing 100875, China. ⁴State Grid Zhejiang Electric Power Research Institute, Hangzhou 310007, China. ✉email: zhanghua2011@bnu.edu.cn

Year	Expenditure on geological exploration (billion yuan)	Number of geological disasters	Direct economic loss (billion yuan)	Number of successful forecasts	Avoided loss amount (billion yuan)	Number of landslides	Proportion of landslides (%)
2014	1145.00	10,970	54.10	417	18.1	8128	74.09
2015	360.00	8224	24.90	452	5	5452	66.29
2016	247.29	9710	31.70	676	7.1	3653	37.62
2017	198.36	7521	35.90	1642	14.5	5524	73.45
2018	173.72	2966	14.70	496	9.6	1631	54.99
2019	172.11	6181	27.70	948	8.3	4220	68.27
2020	161.61	7840	50.20	534	10.2	4810	61.35
2021	1008.40	4772	32.00	905	13.5	2335	48.93
2022	1105.00	5659	15.00	321	6	3919	69.25
2023	1105.00	3668	10.00	427	5	925	25.22
Mean	567.65	6751.1	34.17	681.8	9.73	4059.7	57.95

Table 1. National geological disaster situation table. [Data sourced from the disaster reports published on the China Geological Environmental Information Network].

tectonic events, can destabilize the prevailing geological equilibrium, which in turn may initiate or accelerate landslide movements^{12,13}.

The slope unit, a fundamental natural unit representing mountain formations, is one of the geographic units used to assess landslide susceptibility. Characteristics such as sunny or shaded slopes, steepness, and soil or rock composition all directly influence landslides¹⁴. However, because traditional medium to large-scale or regional geological hazard risk assessments often rely on administrative or grid units, the results are difficult to apply for site-specific geological hazard management. Hilly areas tend to appear uniformly high-risk, which hinders practical engagement and increases management costs. This issue arises primarily because some management units are either too large for effective handling or are not suited to directly implement these outcomes.

Selecting appropriate mapping units is a primary challenge for researchers in geological hazard risk assessment, as it influences the accuracy and feasibility of hazard zoning¹⁵. The following criteria should be observed during the selection process: adjacent units should exhibit significant heterogeneity or variations, while each mapping unit should exhibit maximum homogeneity or similarity within itself. Raster units, homogeneous geological units, and slope units are currently the most frequently used mapping units in landslide hazard assessment. Various methods have been developed for delineating slope units, each with its own strengths and limitations, depending on the characteristics of the terrain and the specific requirements of the analysis. Traditional hydrological methods, which utilize Digital Elevation Models (DEMs) to analyze surface water flow directions, identify watersheds, and define drainage lines, are commonly employed. These methods are intuitive and based on established hydrological principles, making them easy to understand and apply. However, they are sensitive to the resolution of the DEM and tend to perform poorly in flat areas, where the delineation of slope units is less accurate^{10,16}. On the other hand, multiscale segmentation methods allow for the recognition of slope units by progressively refining the analysis across different scales, adapting to the diverse nature of the terrain. These methods are highly flexible and can effectively handle complex topographies, but the subjective nature of parameter selection and scale determination, along with the high computational complexity, can pose challenges^{17,18}. Another technique, curvature-based segmentation, uses terrain curvature to differentiate landforms such as ridges and valleys, providing a detailed analysis of topographic variations. While this approach excels in capturing fine terrain details, it requires high-quality DEMs and is time-consuming, with a risk of generating overly detailed slope units that may not be useful for large-scale assessments^{19,20}. Texture-based watershed methods, which apply image processing techniques to simulate water flow and delineate slope units based on terrain texture and hydrological properties, are particularly suited for identifying subtle variations in complex terrains. However, their technical complexity requires advanced data processing skills, limiting their accessibility^{21–23}. The GRASS GIS r.slopeunits plugin, which automates the extraction of slope units by analyzing slope and direction in DEM data, supports iterative optimization to improve results in flat areas. This method is highly automated and efficient for large datasets but requires GIS expertise and relies heavily on high-quality DEM data^{24,25}.

Additionally, case-specific dynamic analysis, which is widely applied for small-scale research units, can facilitate an in-depth quantitative analysis of the causes and mechanisms of landslides⁴. However, this approach is limited in applicability for developing macro-scale general prevention or defense models due to its complexity. Given the pressing need for practicable solutions, slope unit-based landslide hazard assessment is highly beneficial. The slope unit serves as an analytical unit that enhances the applicability of landslide susceptibility assessments to actual geological settings. It is delineated by terrain boundaries. For example, studies have utilized slope units in combination with various models (such as Random Forest, Support Vector Machine models, and deterministic factor models) for landslide susceptibility mapping, yielding favorable outcomes in comprehensive, high-precision evaluations. As it enhances predictive accuracy and better represents the geographical characteristics of landslides, this combination modeling approach is applicable to land-use planning and high-risk region management²⁶. Slope units are increasingly incorporated into earthquake-induced landslide risk models to

simulate the full landslide movement process, including initiation and deposition zones. Implemented on GIS and R platforms, the enhanced Newmark model replicates actual landslide zones and movement directions, thus providing more precise impact predictions for landslides. This provides important information for assessing the risk of geological hazards and developing mitigation plans²⁷.

Based on the extensive research mentioned above, the study of landslide disasters exhibits the following limitations. Most notably, traditional methods for delineating slope units, such as hydrological analysis using Digital Elevation Models (DEMs), struggle in flat areas, where accuracy is compromised. Additionally, multiscale segmentation and curvature-based methods, while effective in handling complex topographies, are computationally intensive and highly subjective in parameter selection, limiting their applicability in large-scale assessments. Moreover, the dynamic case-specific analyses that are widely applied for small-scale studies are often too intricate to be adopted in broader risk management frameworks. These limitations highlight the need for improved methods that can better adapt to diverse terrains and provide more accurate and scalable results. Consequently, this article aims to address these gaps by enhancing large-scale slope unit extraction techniques, creating a comprehensive evaluation index that incorporates terrain, hydrology, and human activity, and leveraging machine learning models for more accurate landslide susceptibility mapping. The research's environmental significance lies in its potential to inform landslide prevention and mitigation strategies, ultimately reducing the economic and human toll of such disasters.

This study examines Wenzhou as the research area and seeks to enhance large-scale slope unit extraction techniques to accurately define assessment units for landslide geological hazards. By incorporating terrain, hydrology, and human activity, the study establishes a comprehensive evaluation index framework to identify key factors contributing to landslide susceptibility. This approach is combined with historical landslide data to construct a training dataset for the research area, which is then used to train machine learning models such as Random Forest, XGBoost, and LightGBM. The optimal model, determined by comparing model accuracy and other significant metrics, generates a landslide susceptibility map and a feature importance ranking (Fig. 1). The study also provides scientific management advice for landslide prevention and mitigation by examining assessment results from assessing landslide danger in relative terms approaches.

Data and methodology

Data source and processing

The data used in this study were obtained from multiple sources: the European Space Agency's World Cover project (<https://esa-worldcover.org/en>) supplied land use data with a spatial resolution of 10 m; the Moderate Resolution Imaging Spectroradiometer (MODIS) platform (<https://modis.gsfc.nasa.gov/>) supplied Normalized Difference Vegetation Index (NDVI) data with a 250 m resolution; the United States Geological Survey (USGS, <https://www.usgs.gov/>) supplied geological lithology maps, soil data, and geomorphological data; and the National Aeronautics and Space Administration (NASA) (<https://search.asf.alaska.edu/>) supplied Digital Elevation Model (DEM) data with a 30 m resolution, which were subsequently utilized to derive slope, terrain ruggedness (relative elevation), plan curvature, and profile curvature, all with a 30 m resolution. We used OpenStreetMap to obtain road network and hydrological data, the Wenzhou Meteorological Bureau for 30 m resolution meteorological data, and the Zhejiang Geological Environment Monitoring Station for the acquisition and integration of landslide data.

The study employs Min–Max scaling, which transforms feature scales to a predefined range, typically between 0 and 1. Additionally, the study employs the Variance Inflation Factor (VIF) as a statistical measure to assess the extent of multicollinearity among the explanatory variables in the regression model. Data normalization is an essential preprocessing step in machine learning and statistical modeling, ensuring that different features are quantified on a consistent scale and thereby mitigating biases introduced by scale discrepancies.

Slope unit extraction method and optimization

The concept of slope units was initially introduced by Carrara²⁸, who classified slope units as a specific category of terrain units. They defined a slope unit as a region delineated by the watershed divide and the drainage line. In practical applications, the delineation of slope units is based on geomorphological features, such as valleys and ridgelines. The division of slope units is carried out through surface hydrological analysis using Digital Elevation Models (DEMs). The underlying principle involves extracting valleys and ridgelines from both the positive and negative terrain (representing the drainage line and watershed divide, respectively). The resulting catchment basins and reverse catchment basins are then merged, followed by manual adjustments to rectify any irregularities in the units, ultimately defining the slope unit as the region bounded by the watershed divide and drainage line. However, the area enclosed by the watershed divide and drainage line can consist of a single slope, multiple slopes, or even a small watershed. In practical research, some scholars have refined this definition, suggesting that a slope unit should exhibit a high degree of internal homogeneity while ensuring a notable heterogeneity between different units. Xie²⁹ proposed that a slope unit represents a region with distinguishable topographical features relative to adjacent areas. Each slope unit demonstrates clear differences in terrain characteristics (such as boundary lines, streams, slope, and slope angle) when compared to neighboring units. Furthermore, slope units are regarded as the fundamental units for assessing geological hazards (e.g., landslides, rockfalls, debris flows), with individual slope units being the smallest spatial unit where such hazards can occur. In the context of geological hazard assessment, adopting slope units as the primary unit for evaluating hazard risks is considered to improve the consistency with actual terrain and landform characteristics³⁰. As illustrated in Fig. 2a,b, the developmental stages of valleys and rivers are considered to play a critical role in the formation of landslides and rockfalls³¹.

The *r.slopeunits* plugin in GRASS GIS software was selected for the initial extraction of slope units after a range of methodologies was evaluated to identify the optimal technique for the study area. The *r.slopeunits*

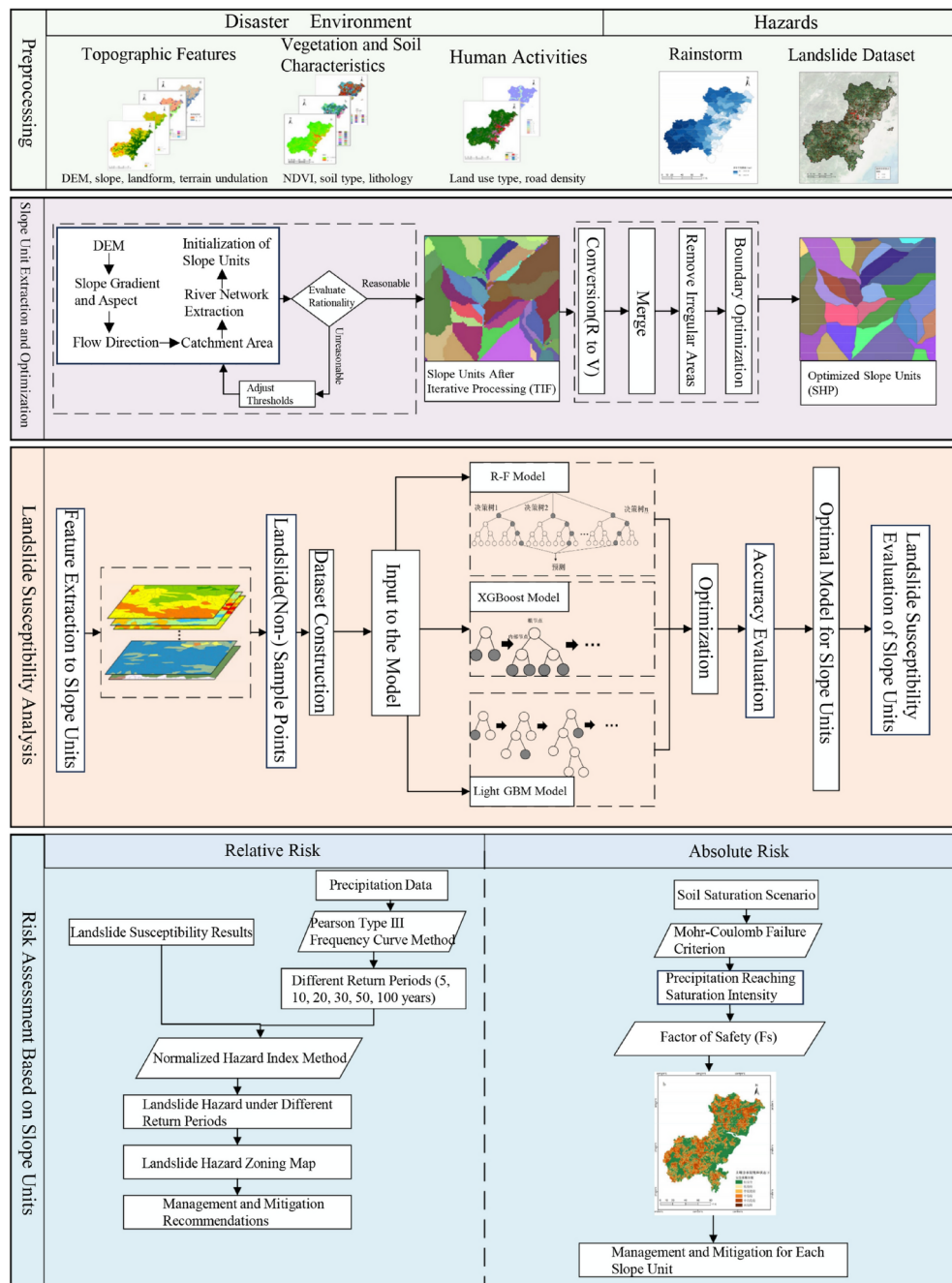


Fig. 1. Technical flowchart.

plugin utilizes data from Digital Elevation Models (DEM) and applies advanced algorithms within the GRASS GIS environment to accurately delineate and extract slope units. Additionally, due to rasterization, slope unit boundaries may exhibit irregular boundaries, and minor and spurious slope units can be generated in flat terrain areas. To address these issues, further refinement of the results extracted by *r.slopeunits* was conducted in this study.

To refine the Wenzhou results generated by *r.slopeunits*, a Python script was employed to merge the slope units of multiple districts and transform the raster data (TIFF) into vector data (Shapefile). This procedure reduced frequent angular boundaries and implausible slope unit delineations (Fig. 2c,d). By enhancing the spatial continuity of slope unit and aligning it more closely with actual terrain features, the refinement improved the precision and reliability of terrain analysis, while also increasing the efficacy of data processing workflows.

This adjustments significantly enhanced the accuracy of slope unit extraction and expanded the applicability of the method to encompass a broader range of terrain types. The improved method facilitates high-precision slope unit extraction across larger areas, ensuring the accuracy and applicability of the results while enhancing extraction efficiency. This holds substantial potential for geological hazard studies. In summary, through the

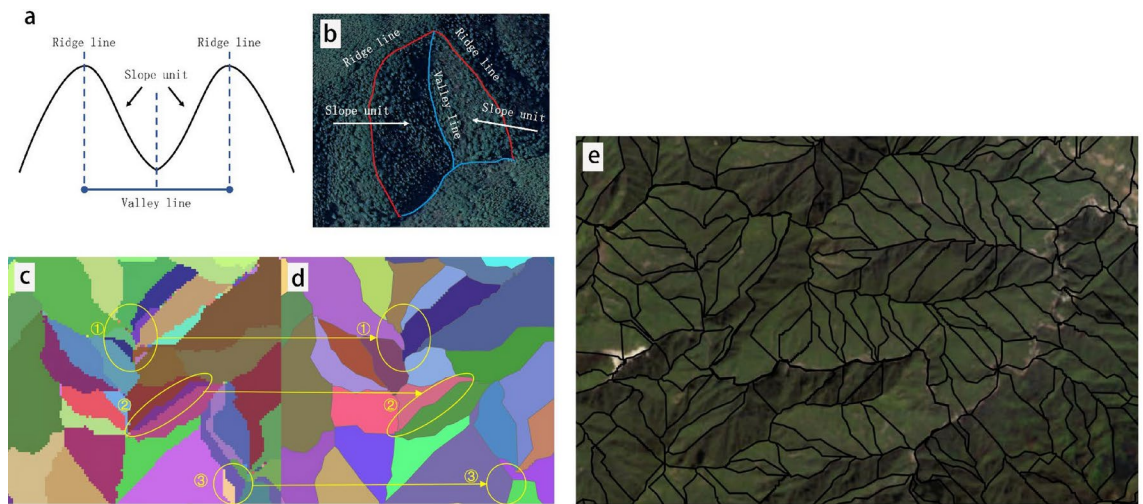


Fig. 2. Schematic diagram of slope unit. (a) slope unit profile diagram; (b) Slope unit plan view; (c) and (d) comparison before and after optimization of slope elements; (e) representation of slope unit extraction in parts of Wenzhou. (Fig. 2 was created using ArcGIS 10.8.2 (<https://www.esri.com>), with satellite imagery provided by the software's default base map).

Terrain classification	Slope	Number of slope units	Area (km ²)	Average elevation (m)
Plain	0°–0.5°	577	29.38	16.47
Slight slope	0.5°–2°	5837	353.89	7.48
Gentle slope	2°–5°	16,418	1015.32	18.32
Slope	5°–15°	34,291	2245.77	324.45
Steep slope	15°–35°	95,413	7219.77	440.58
Precipice	35°–55°	1100	68.25	478.07

Table 2. Table of statistical characteristics of slope units.

optimization of existing slope unit extraction methods, this study achieved dual improvements in detail handling and scale control (Fig. 2e). Over 150,000 slope units were successfully extracted for Wenzhou, providing a more precise and practical slope unit extraction method for fields such as terrain analysis and land use planning. The relevant characteristic values of the extracted slope units are shown in Table 2.

Machine learning models

Machine learning techniques have been widely and effectively applied across diverse domains, particularly in landslide catastrophe research, where their utility continues to grow³². Random Forest, XGBoost, and LightGBM constitute the three primary machine learning models utilized in this study to evaluate landslide susceptibility. These models collectively establish a methodology for evaluating landslide vulnerability.

The Random Forest algorithm's output is derived from the aggregate outcome of an ensemble of decision trees. For classification problems, the mode class among the trees is selected as the final output through a majority voting scheme; for regression problems, the mean of the predictions is computed. The mathematical expressions are as follows:

$$Y_{class} = mode(\{y_1, y_2, \dots, y_N\})$$

$$Y_{reg} = \frac{1}{N} \sum_{i=1}^N y_i$$

where Y_{class} represents the prediction result for classification problems, and Y_{reg} represents the prediction result for regression problems. y_i denotes the prediction result of the i tree, and N is the number of trees in the ensemble. By tuning the number of trees and the tree depth parameters, the model's performance can be optimized.

The objective function of XGBoost consists of two parts: one part is a traditional loss function that measures the difference between the predicted value and the true value of the model, while the other part is a regularization

term designed to penalize model complexity, aiming to prevent overfitting. Specifically, the objective function is defined as:

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

where l is the loss function that measures the difference between the y_i of the i sample and the $\hat{y}_i$, and Ω represents the regularization term that penalizes model complexity (such as the depth of the tree, the number of leaf nodes, etc.). f_k represents the k tree in the ensemble, and K denotes the number of trees.

Traditional GBDT uses a greedy approach to find the best split, which can be computationally expensive on large datasets. LightGBM addresses this by spreading continuous feature values into discrete bins (using histogram-based algorithms), and then it performs the best split search based on these discrete bins. This approach not only reduces memory usage but also improves computational efficiency. The split gain is calculated as follows:

$$G = \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} + \frac{(G_L + G_R)^2}{H_L + H_R + \lambda}$$

where G_L and G_R represent the gradients of the left and right child nodes, respectively, and H_L and H_R represent the second-order derivatives of the left and right child nodes, respectively, and λ is the regularization parameter.

Landslide hazard assessment method

Calculation method for relative value of landslide hazard

The calculation method for the relative value of landslide hazard is based on the susceptibility output from the random forest model. Additionally, it utilizes the Pearson Type III frequency curve method to estimate rainfall intensity for different return periods. By overlaying susceptibility with rainfall intensity for various return periods, the relative value of landslide hazard in the study area is obtained. The specific calculation formula for the relative hazard index H_{iT} is as follows:

$$H_{iT} = \frac{Y_i}{Y_{max}} \times p_i$$

$$p_i = \frac{L_i}{L_{max}}$$

where H_{iT} represents the hazard index for the i slope unit under a specific condition, Y_i is the susceptibility value of the i slope unit, and Y_{max} is the maximum susceptibility value. p_i denotes the probability of slope instability due to precipitation for the i slope unit under different return periods, where L_i is the rainfall amount for the i slope unit at various return periods, and L_{max} is the maximum rainfall intensity observed during the analysis period. Thus, the hazard index H_{iT} combines two key factors: rainfall intensity and landslide susceptibility, providing a quantitative indicator for landslide risk assessment under different return periods. The hazard data obtained through this method allows for a comprehensive evaluation of landslide risk.

Case study

General situation and background of the study area

Wenzhou City, situated along China's southeastern coast, features a diverse and balanced spatial distribution of geomorphological characteristics, including mountains, hills, basins, plains, and islands (Fig. 3). With its abundant natural resources, complex geological structure, and robust economy, Wenzhou represents a vital component of China's economic framework. However, Wenzhou is among the cities in China most vulnerable to recurrent geological hazards³³. In recent years, the frequency of geological disasters has escalated. According to the 2022 Geological Environment Overview of Wenzhou, published by the Wenzhou Natural Resources and Planning Bureau, by the end of 2022, the city had established 2222 registered geological disaster risk prevention zones, impacting 11,800 households, 38,200 residents, and threatening assets valued at 1.2 billion yuan. Additionally, 44 geological hazard points were identified, posing risks to 3091 individuals, 856 households, and assets valued at 1.96 billion yuan. In 2022, Wenzhou experienced 14 minor, unanticipated geological disasters, all of natural origin, which resulted in economic losses of 333,000 yuan and two fatalities. Wenzhou's fragile geological environment, coupled with its complex topography and geomorphology, leads to a high concentration of geological hazard sites in mountainous areas, posing significant risks²⁷. Rainfall and anthropogenic engineering activities are primary drivers of these hazards, which are closely associated with geological factors such as topography, geomorphology, and stratigraphy³⁴.

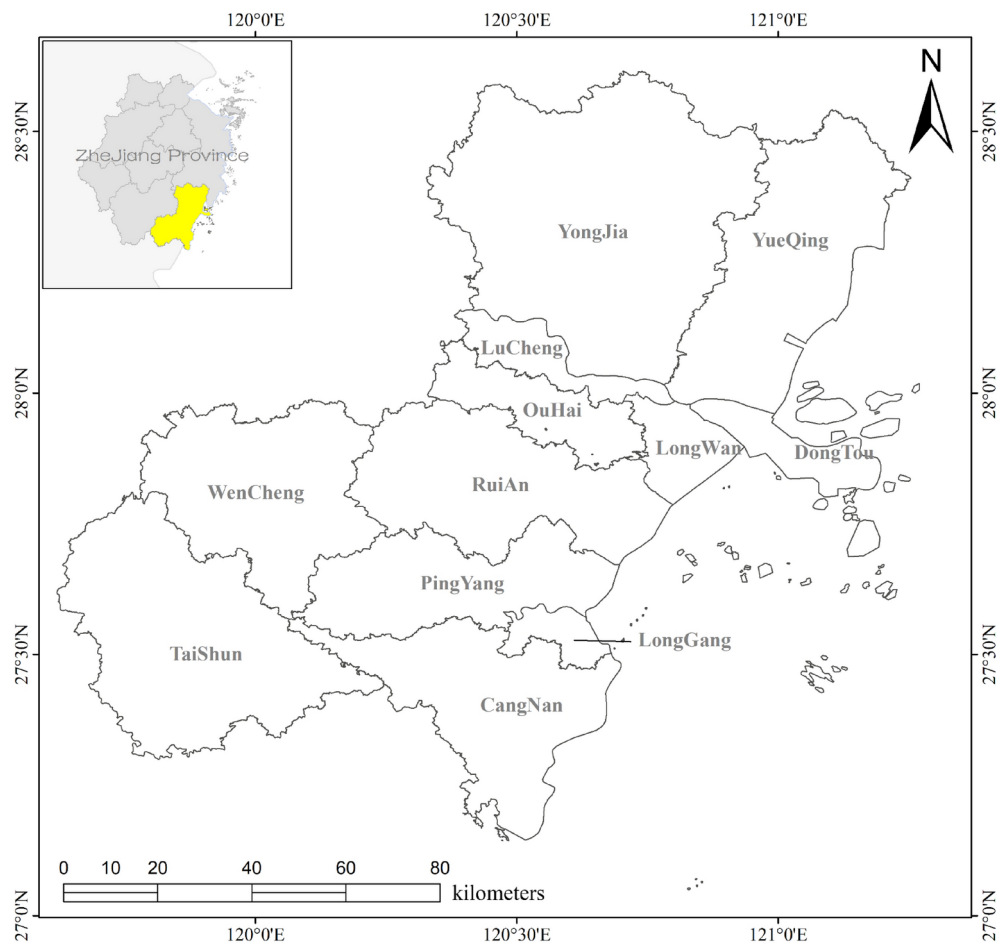


Fig. 3. Location map of the research area. (Fig. 3 was created using ArcGIS 10.8.2 (<https://www.esri.com>)).

Landslide susceptibility analysis based on machine learning

Building a feature dataset

This study comprehensively considers factors such as topography, geomorphology, vegetation cover, lithology, soil type, land use, infrastructure distribution, and hydrogeological conditions to construct a landslide susceptibility assessment index system. Selected indicators include elevation, geomorphology, slope, relief degree, plan curvature, profile curvature, Normalized Difference Vegetation Index (NDVI), lithology, soil type, land use, road proximity, and distance to rivers. Collectively, these indicators provide a detailed assessment of landslide susceptibility in a specific region by encapsulating both anthropogenic activities and environmental conditions influencing landslide occurrence.

The heatmap illustrates the Pearson correlation coefficients between various indicators (Fig. 4), where values close to 0 imply minimal linear association, while values near +1 or −1 indicate a strong positive or negative correlation, respectively. A strong positive correlation (0.96) between slope and relief degree highlights a significant similarity between these indicators. Conversely, a moderate negative correlation (−0.60) between land use data and NDVI suggests an inverse relationship, where higher land use values are associated with lower NDVI, consistent with anticipated trends. Certain indicators, such as soil type and geomorphology (0.01) and NDVI and soil type (0.01), display very weak correlations, indicating minimal direct linear association. To address multicollinearity and reduce redundancy, slope is selected over relief degree, thereby excluding relief degree from further analysis (Fig. 5). This selection results in a feature dataset comprising eleven key indicators: elevation, slope, geomorphology, NDVI, land use data, plan and profile curvatures, lithology, soil type, and distance to rivers.

Construction of landslide sample dataset

Wenzhou City contains 951 geological hazard locations. To minimize sampling bias, this study initially conducted a 500 m buffer analysis around these hazard sites to objectively select non-hazard sample points, thereby avoiding locations too close to the hazard zones. To precisely delineate waterbody boundaries, the study also incorporated comprehensive river and lake data within the study area. Following this preprocessing, non-hazard sample points located outside all river and lake areas and more than 500 m from landslide sites were selected using a random sampling technique. This method aims to approximate natural randomness, thereby enhancing the generalizability of subsequent analysis. Through multiple trials assessing various landslide-to-

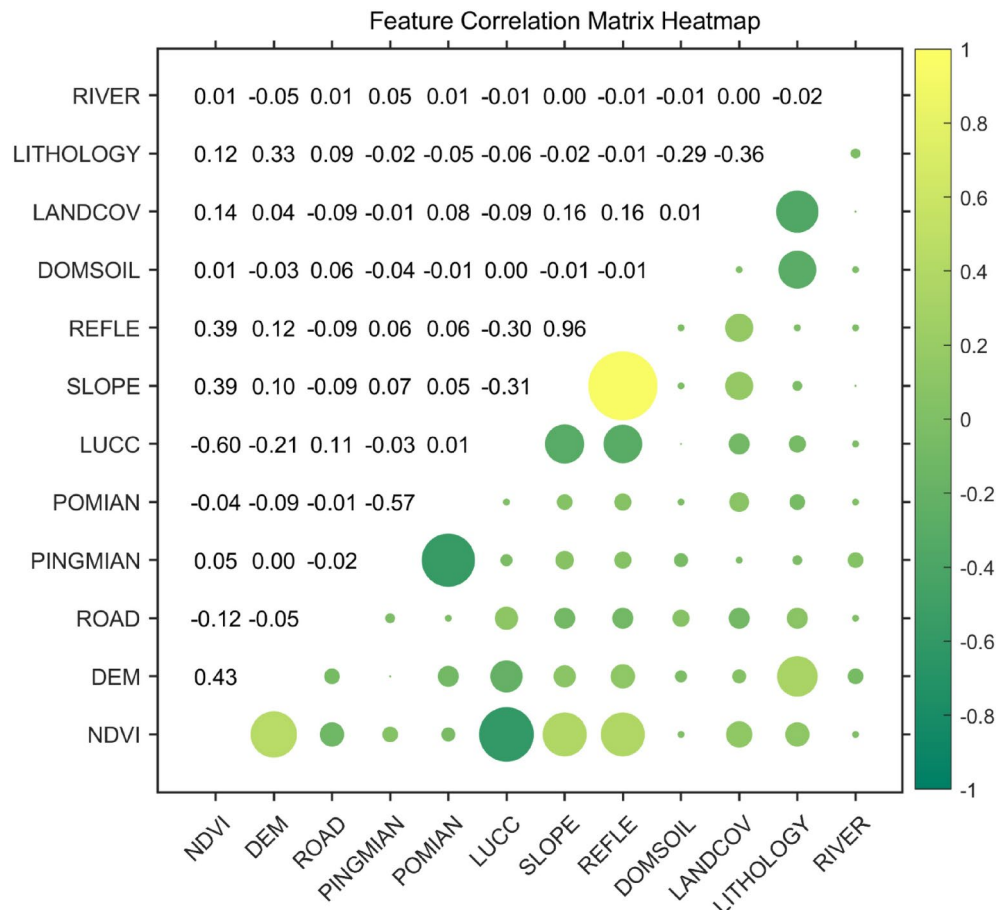


Fig. 4. Heatmap of correlation coefficients for evaluation metrics.

non-landslide ratios, an optimal 1:1 ratio was identified. As shown in Fig. 6, the final landslide sample dataset includes 1902 points. This balanced ratio ensures an equitable distribution of positive and negative samples, establishing a robust foundation for constructing an effective prediction model.

Accuracy evaluation of slope unit model

This study utilized various evaluation metrics, including accuracy, ROC curve, precision-recall curve, F1 score, and average precision, to assess the performance of three machine learning models on slope units. These metrics offer a comprehensive assessment of each model's efficacy in predicting landslide susceptibility. Figure 7 presents the evaluation metrics specifically for the Random Forest model.

The Random Forest model demonstrates strong predictive capability, achieving an accuracy of 77.94%, indicating effective classification of both positive and negative samples. Due to its relatively high accuracy, the Random Forest model shows potential for application across diverse fields. The model effectively minimizes false negatives, as indicated by a recall rate of 78.93%, meaning approximately 78.93% of actual positive instances were correctly identified. The F1 score, a harmonic mean of precision and recall, is 76.15%, reflecting the model's balanced performance between precision and recall. The relatively high F1 score indicates that the Random Forest model effectively detects positive samples in classification tasks while managing misclassifications. The model demonstrates robust predictive reliability, with an average precision of 85.74%, successfully distinguishing actual positive samples from predicted positives.

Figure 8 presents the evaluation metrics for the XGBoost model. The XGBoost model achieved an accuracy of 76.95%, slightly lower than Random Forest and LightGBM, which suggests a marginally higher tendency for misclassification. Nevertheless, XGBoost demonstrated the highest recall rate at 78.16% among the three models, showcasing its strong capability to correctly identify positive samples—a valuable trait in scenarios where minimizing missed positives is crucial. Despite the slightly reduced overall accuracy, XGBoost's F1 score of 75.84% reflects a well-balanced performance between precision and recall, highlighting the model's robustness. Furthermore, an average precision of 84.33% confirms XGBoost's reliable predictive power for positive samples, placing it second among the models and reinforcing its efficacy across various applications.

Figure 9 displays the evaluation metrics for the LightGBM model. The LightGBM model achieved an accuracy of 78.30%, indicating strong overall classification reliability. However, with a recall rate of 74.81%—the lowest among the three models—LightGBM is less effective at capturing all positive cases, which may limit its utility in applications where minimizing missed positives is essential. Its F1 score of 74.08% reflects a reasonable balance

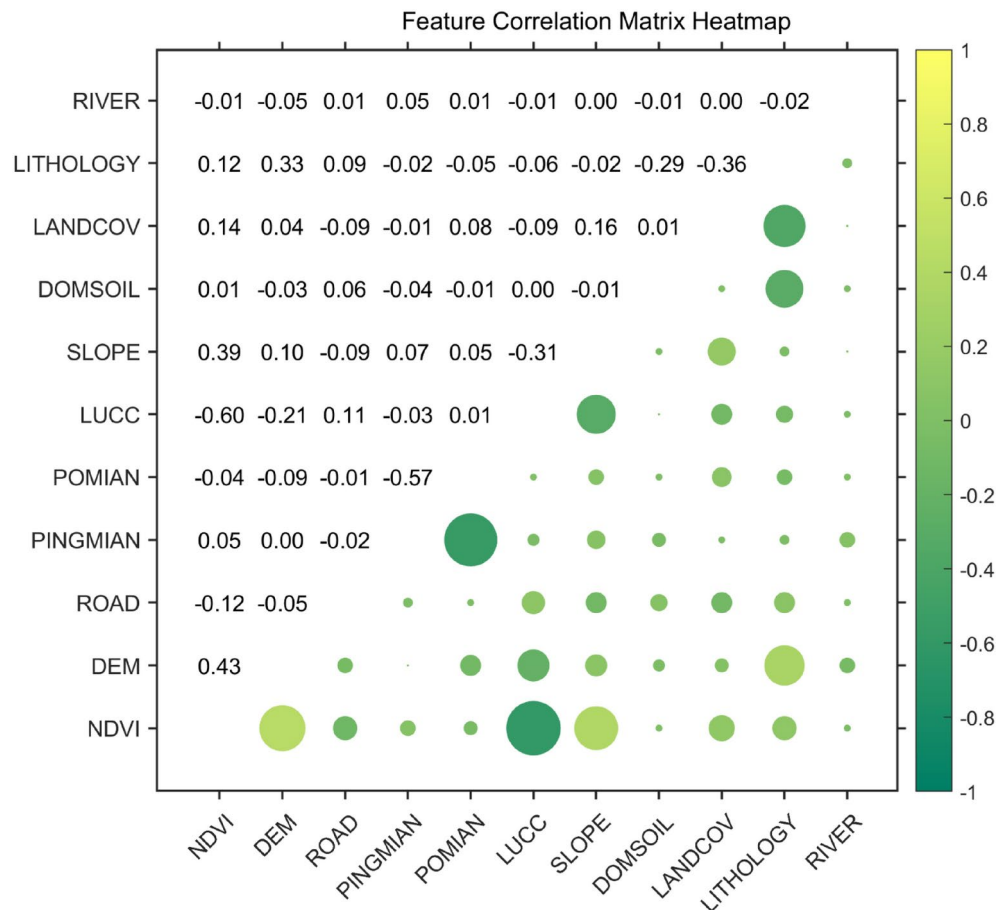


Fig. 5. Heatmap of correlation coefficients for revised evaluation metrics.

between precision and recall, though slightly below Random Forest's performance in this regard. LightGBM's average precision of 82.91% is also the lowest among the models, suggesting somewhat reduced reliability in predicting positive samples.

In summary, a comparison of the three models' performance metrics shows that the Random Forest model demonstrates superior results across accuracy, recall, F1 score, and average precision, establishing it as the most effective model overall for predicting landslide susceptibility.

Analysis of the importance of features in the optimal model

The slope is the most influential factor (Table 3), followed by Digital Elevation Model (DEM) and the Normalized Difference Vegetation Index (NDVI). Slope curvature (Pomian) also shows a relatively high degree of significance. In contrast, plan curvature, soil type (DOMSOIL), land use classification (LUCC), lithology, geomorphology, road density, and river proximity rank lower in importance. DEM and slope, as fundamental geographic and topographic indicators, are key in assessing landslide susceptibility, likely due to their influence on soil stability and water accumulation—factors closely associated with landslide potential. NDVI's notable role as an indicator of environmental stability and vegetation cover suggests that vegetation may play an essential role in supporting soil cohesion and mitigating landslide risk. While road density, river influence, lithology, and geomorphological form have relatively lower weights, their potential contributions to landslide susceptibility remain relevant, as landslides result from the complex interplay of multiple factors. These subtler features may demonstrate increased importance in specific contexts or when interacting with other influential variables.

The landslide susceptibility results of the optimal machine learning model

The results of the landslide susceptibility assessment for the area were derived after comprehensive training and validation of the model inputs. Utilizing the probability values generated by the Random Forest model, the classification methodology for landslide susceptibility zones delineates the ranges for various susceptibility levels by establishing thresholds based on the cumulative frequency distribution. Six susceptibility levels—low, relatively low, medium, moderately high, high, and extremely high—were defined in accordance with the probability distribution.

According to the classification strategy, the susceptibility categories are delineated as follows: low susceptibility is attributed to the lowest 10% of probability values, succeeded by moderately low susceptibility (20%), medium

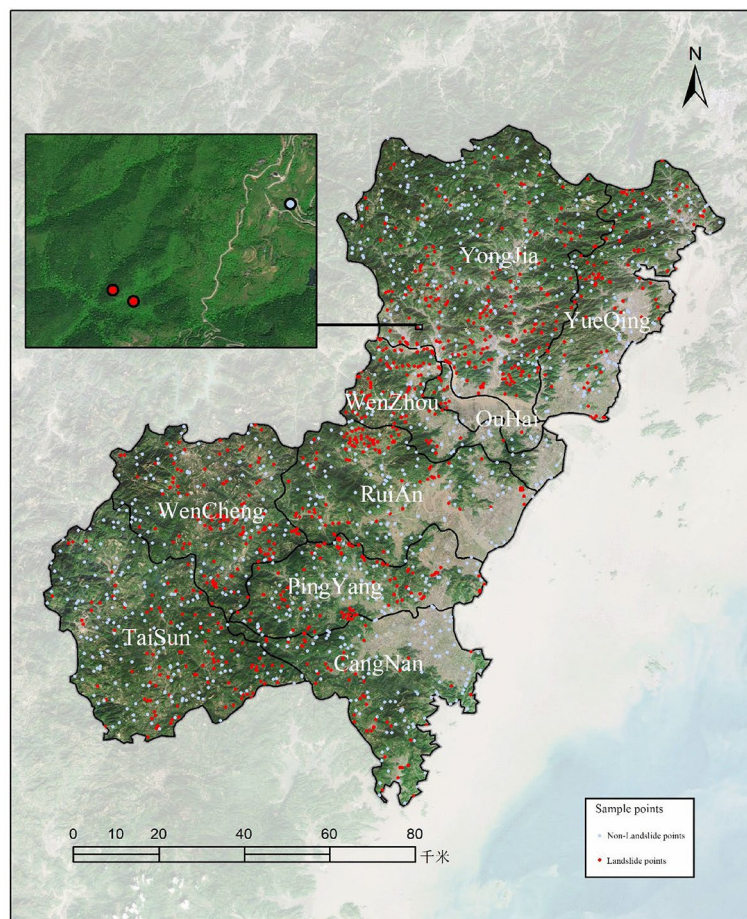


Fig. 6. Distribution map of landslide sample points. (Fig. 6 was created using ArcGIS 10.8.2 (<https://www.esri.com>), with satellite imagery provided by the software's default base map).

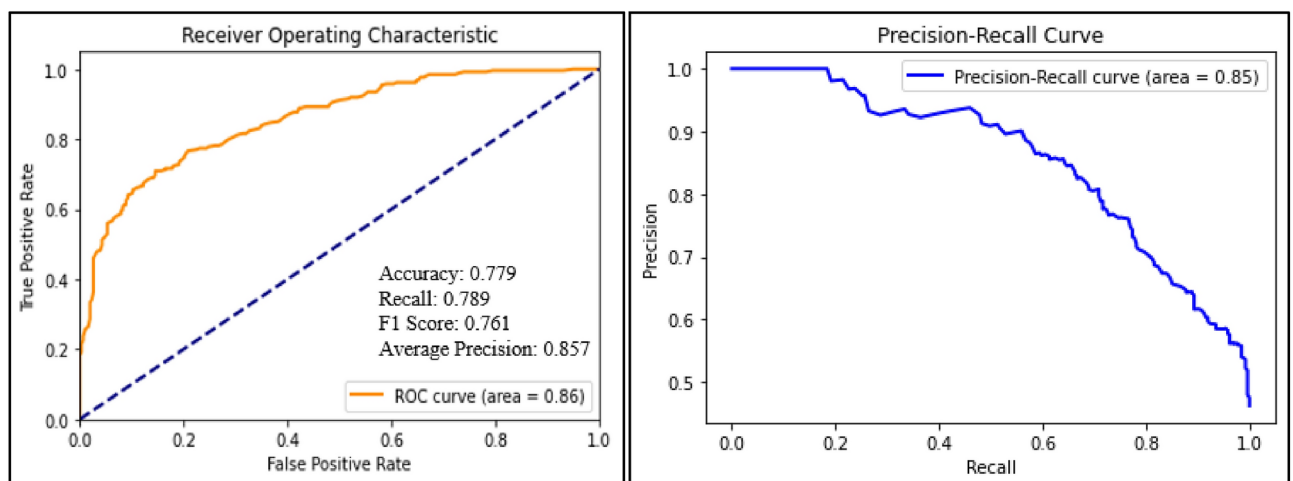


Fig. 7. Performance metrics of the Random Forest model under slope units.

susceptibility (30%), moderately high susceptibility (20%), high susceptibility (15%), and extremely high susceptibility (5%). [Figure 10](#) illustrates the results of the landslide vulnerability zoning across slope units.

As depicted in the figure, the northern and southwestern regions of Wenzhou predominantly contain the medium-high, high, and extremely high susceptibility zones, whereas the central and eastern plains are primarily characterized by low and moderately low susceptibility areas. A statistical table summarizing landslide

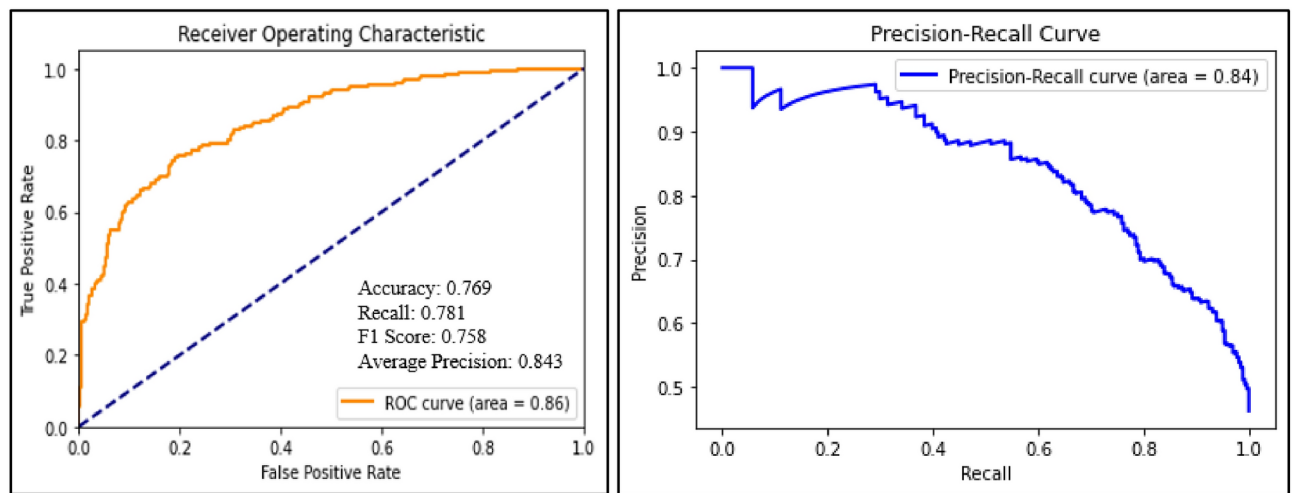


Fig. 8. Performance metrics of the XGBoost model under slope units.

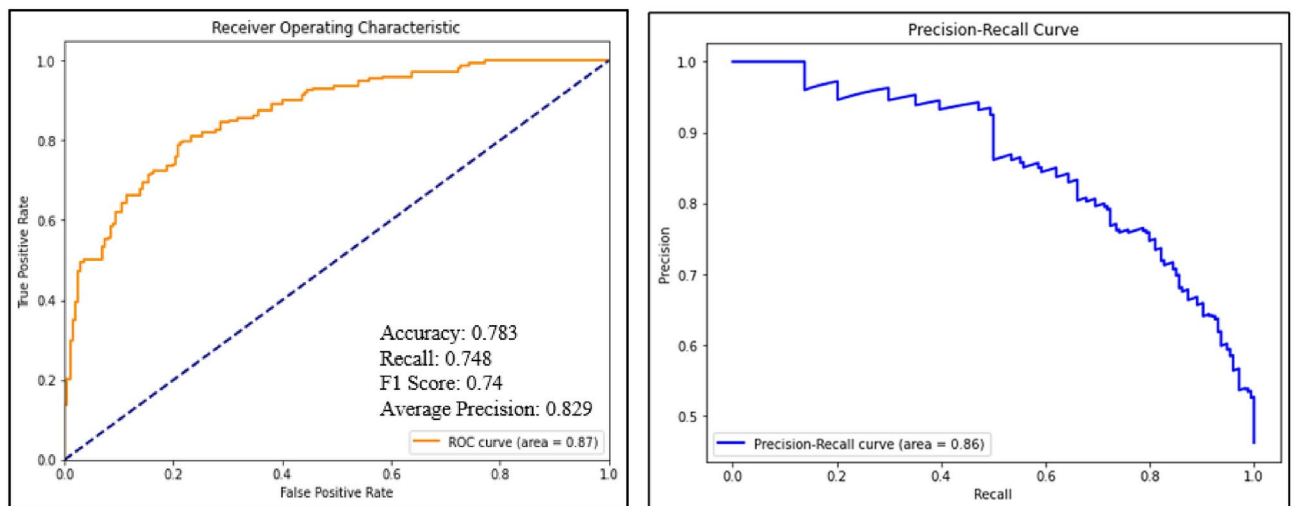


Fig. 9. Performance metrics of the LightGBM model under slope units.

Features	ROAD	RIVER	LITHOLOGY	LANDFORM	LUCC	DOMSOIL	PINGMIAN	POMIAN	NDVI	DEM	SLOPE
Relative importance	0.02	0.03	0.05	0.05	0.06	0.08	0.11	0.12	0.15	0.16	0.18

Table 3. Importance ranking of features in slope units.

occurrences within each susceptibility level was generated by analyzing the distribution of historical landslide events across the various susceptibility zones, as presented in Table 4.

The probability of landslide occurrence is calculated by dividing the number of landslides by the number of mapping units, which constitutes the methodology for determining landslide occurrence frequency in this study. This ratio serves as an indicator for assessing the relative risk or susceptibility to landslide disasters across different locations, as it reflects the frequency of landslides within a specified slope unit. The landslide density index is defined as the number of landslides divided by the area percentage.

The statistical data presented in the table clearly demonstrate a positive correlation between the landslide density index and the landslide occurrence frequency, both of which exhibit an increasing trend as the landslide susceptibility level rises. This pattern indicates that the susceptibility zoning results for slope units are consistent with observed conditions in the field.

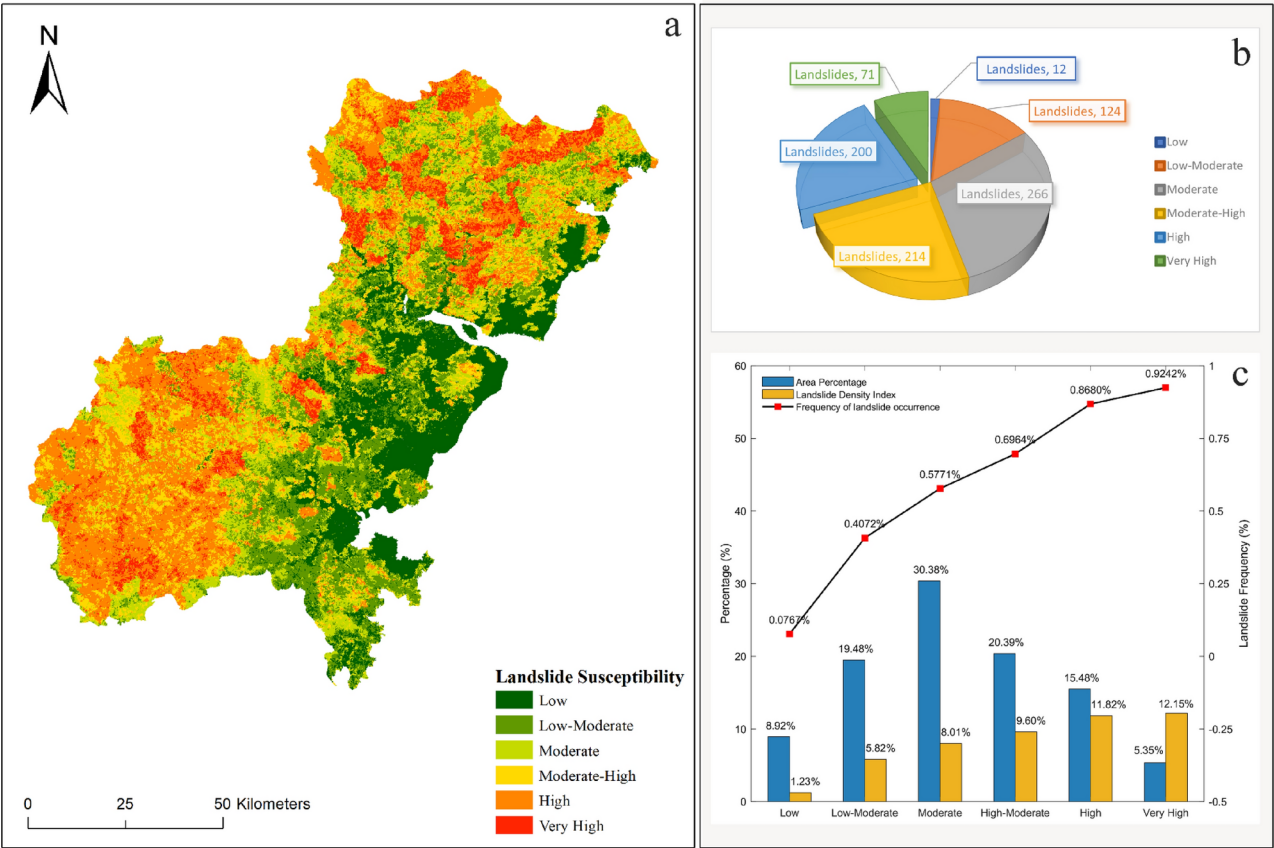


Fig. 10. (a) Susceptibility results of landslide hazards in slope units; (b) Proportion of landslides in the study area; (c) Statistical distribution of susceptibility across slope units. (Fig. 6 was created using ArcGIS 10.8.2 (<https://www.esri.com>)).

Landslide susceptibility	Slope units	Total area (km ²)	Landslides	Area ratio (%)	Landslide density index (%)	Landslide frequency (%)
Low	15,636	975.41	12	8.92	1.34	0.08
Low-moderate	30,455	2129.49	124	19.48	6.37	0.41
Moderate	46,091	3321.71	266	30.38	8.75	0.58
Moderate-high	30,731	2228.97	214	20.39	10.50	0.70
High	23,041	1692.48	200	15.48	12.92	0.87
Very high	7682	584.34	71	5.35	13.28	0.92

Table 4. Landslide statistics by slope unit.

Precipitation intensity in Wenzhou city under different recurrence periods

This study employed the Pearson Type III frequency curve method to evaluate rainfall intensity, facilitating a statistical analysis of the risk of landslide disasters induced by rainfall in Wenzhou across different return periods. This approach, which aims to predict potential future extreme rainfall events, is grounded in the highest annual rainfall data recorded in Wenzhou from 1961 to 2022. In the initial phase of the study, the maximum rainfall records for each year were extracted from climate data, and statistical analysis was performed to determine the sample mean (μ), sample standard deviation (σ), and skewness coefficient (γ) of these data. These statistical parameters provide foundational data for further frequency analysis.

According to the Pearson Type III distribution, the rainfall intensity (R_T) for each return period (T) can be calculated using the inverse of the cumulative distribution function. Specifically, for a given return period, the corresponding non-exceedance probability (P_T) can be computed, representing the probability that the rainfall will not exceed R_T in any given year. The non-exceedance probability P_T is given by the following formula:

$$P_T = 1 - \frac{1}{T}$$
$$Z_T = F^{-1}(P_T; \gamma)$$

where Z_T represents the return period (in years), F^{-1} is the inverse of the cumulative distribution function, P_T is the given non-exceedance probability, and γ is the sample skewness. Subsequently, the rainfall intensity for the return period can be expressed as:

$$R_T = \mu + Z_T \times \sigma$$

The Fig. 11 clearly indicates that the density of rainfall hotspots increases with the length of the return period. This trend suggests that the estimated probabilities of rainfall during extended return periods are more concentrated in these areas, as indicated by the maximum recorded rainfall values. The observed pattern underscores regions that are particularly susceptible to rainfall-induced landslides and emphasizes the increasing significance of extreme rainfall events associated with longer return periods.

Landslide hazard analysis

Analysis of relative risk results of landslides

By overlaying the landslide susceptibility with the rainfall values for different return periods, relative values of landslide hazard for each slope unit were obtained, as shown in Fig. 12.

The Fig. 13 illustrates a direct correlation between landslide hazard levels and excessive rainfall values. The impact of rainfall on landslide risk is particularly pronounced in high-slope regions with extended return periods. A hazard level heatmap for slope units was generated by mapping the quantity of slope units corresponding to different return periods, as depicted in Fig. 13. It is evident that, as the return period increases, the number of slope units classified as having high and extremely high hazard levels rises, while the number of slope units categorized as low and moderately low hazard decreases.

Table 5 and Fig. 12 presents a detailed analysis of the percentages of slope units categorized by various hazard levels across different return durations. The data reveal a significant reduction in low hazard slope units as return periods increase, with the proportion of low hazard slope units decreasing from 13.81% at a 5-year return period to 6.31% at a 100-year return period. Similarly, the percentages of moderately low and moderate hazard slope units also exhibit a gradual decline, dropping from 23.40 and 58.31% to 10.84 and 28.04%, respectively, as the return durations extend. Conversely, as return durations extend, there is a substantial increase in the percentages of high and extremely high hazard slope units. Specifically, high hazard units rise from 4.46 to 14.41%, while extremely high hazard units increase significantly from 0.02 to 40.39%. This trend indicates a decline in the overall stability of slope units in conjunction with rising rainfall intensity, leading to elevated hazard levels. Such findings underscore the critical need to focus on high and extremely high hazard slope units during periods of extreme rainfall.

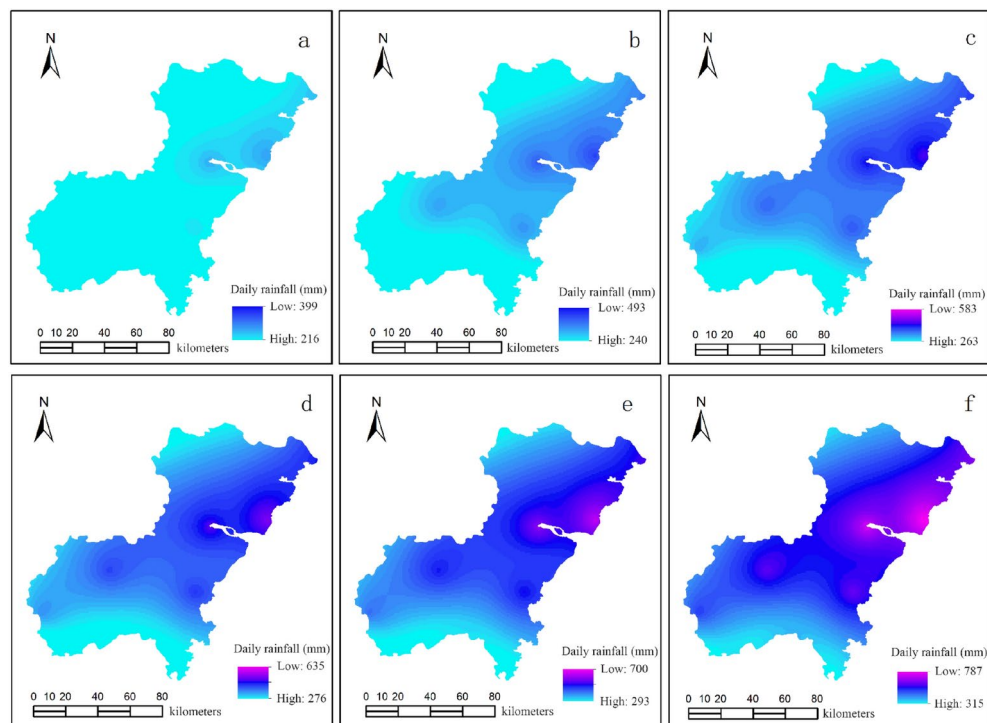


Fig. 11. Rainfall values for different return periods in Wenzhou. (a) 5-year return period; (b) 10-year return period; (c) 20-year return period; (d) 30-year return period; (e) 50-year return period; (f) 100-year return period. Figure 6 was created using ArcGIS 10.8.2 (<https://www.esri.com>).

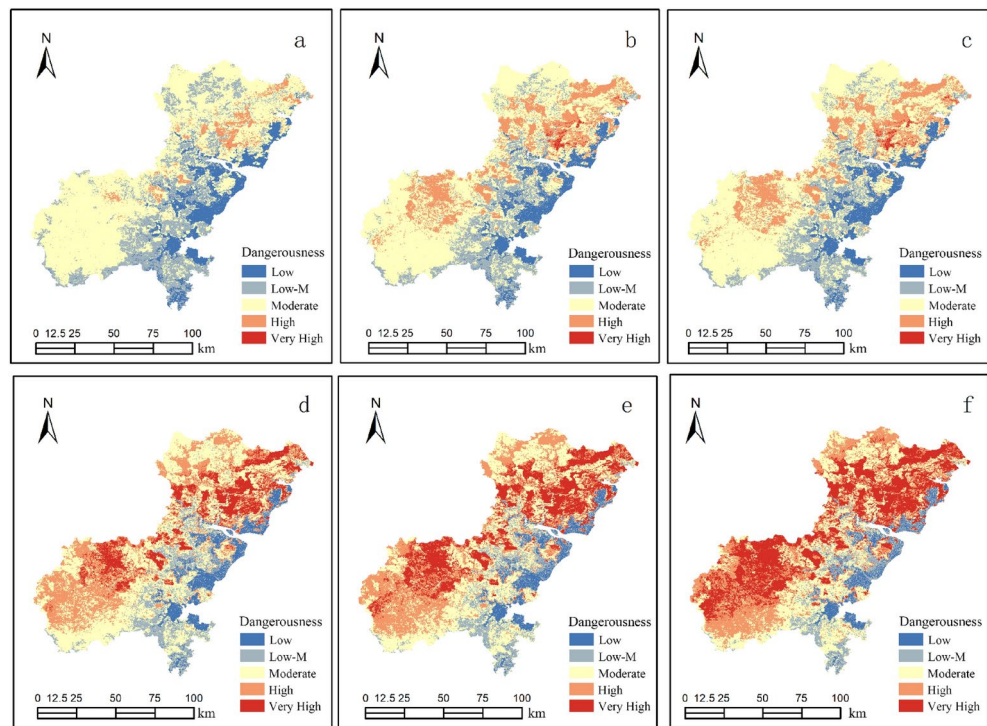


Fig. 12. Landslide hazard schematic diagrams under different rainfall return periods in Wenzhou. (a) 5-year return period; (b) 10-year return period; (c) 20-year return period; (d) 30-year return period; (e) 50-year return period; (f) 100-year return period. Figure 6 was created using ArcGIS 10.8.2 (<https://www.esri.com>).

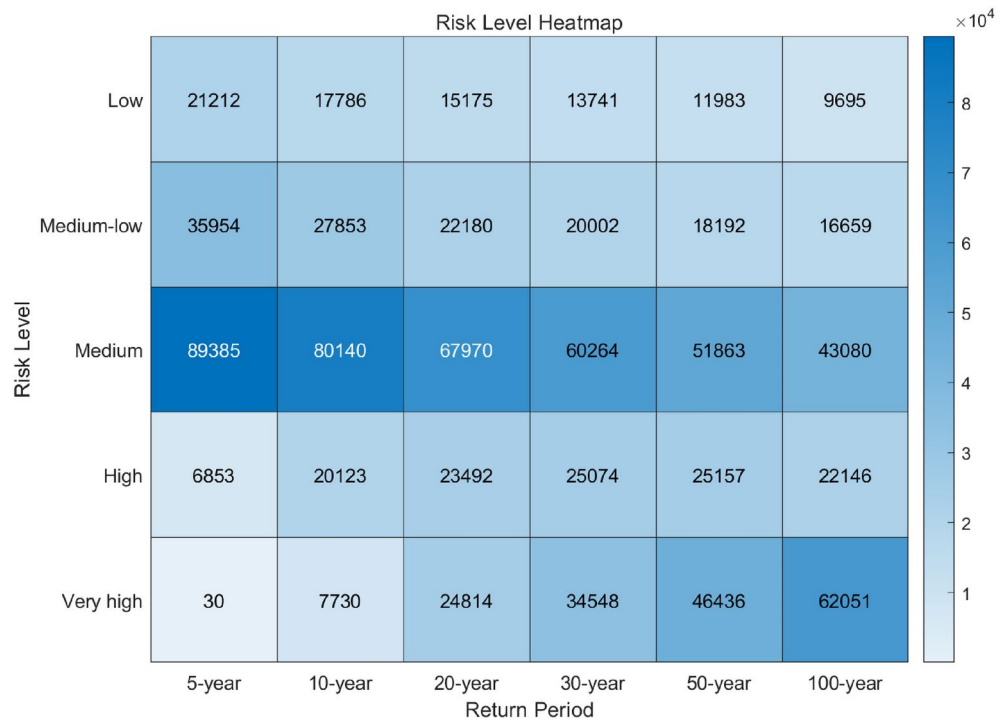


Fig. 13. Heatmap of hazard levels for slope units under different rainfall return periods.

Hazard level	5-year (%)	10-year (%)	20-year (%)	30-year (%)	50-year (%)	100-year (%)
Low	13.81	11.58	9.88	8.94	7.80	6.31
Medium-low	23.40	18.13	14.44	13.02	11.84	10.84
Medium	58.31	52.16	44.24	39.23	33.76	28.04
High	4.46	13.10	15.29	16.32	16.37	14.41
Very high	0.02	5.03	16.15	22.49	30.23	40.39

Table 5. Proportion of slope units by hazard level under different return periods.

Discussion
Results and limitations

The results of this study emphasize the advantages of using slope units as a finer and more effective analytical unit for landslide hazard assessment compared to traditional methods based on administrative or grid units. The Random Forest model, achieving an accuracy of 77.94%, demonstrated its robustness in susceptibility analysis, aligning with prior studies such as Zhao³¹, which highlighted the effectiveness of machine learning in landslide hazard zoning. The significant influence of factors such as slope, DEM, and NDVI supports established theories regarding the critical roles of topographic and vegetation parameters in landslide initiation^{3,12,35,36}. Validation through the Xuekou Village case study further reinforces the practical utility of these methods in disaster mitigation.

This study advances landslide risk assessment by refining slope unit delineation methods and incorporating high-resolution datasets. Enhancements to the r.slopeunits plugin and advanced preprocessing steps improved the accuracy of slope unit extraction, making the approach adaptable to diverse terrains. Under extreme rainfall scenarios, such as a 100-year return period, the observed increase in high-risk slope units underscores the dynamic nature of landslide hazards, consistent with studies¹⁶ linking extreme rainfall to increased geomorphological vulnerabilities. These findings provide actionable insights for disaster risk management, enabling targeted interventions like vegetation restoration and slope monitoring.

Despite these contributions, several limitations exist. First, the reliance on the Mohr–Coulomb failure criterion and static datasets restricts the model’s ability to capture dynamic interactions among geological, hydrological, and anthropogenic factors. Future research should integrate dynamic simulations and real-time meteorological data to improve predictive accuracy. While the Random Forest model performed well, exploring ensemble learning methods could enhance robustness and generalizability. Additionally, the absence of data on other triggers, such as seismic activities, limits the scope of this hazard assessment.

Addressing these limitations can help develop more comprehensive and adaptable frameworks for landslide risk assessment. This would strengthen disaster prevention strategies, offering better protection against geological hazards.

Analysis of slope stability and landslide hazard assessment in Xuekou village

The Mohr–Coulomb failure criterion was employed to evaluate the stress conditions leading to slope failure, with the safety factor used to predict and mitigate landslide risks. Soil shear strength, influenced significantly by pore water pressure, represents the threshold at which landslides are triggered. Increased soil moisture content raises pore water pressure, reducing soil shear strength. When the shear strength of soil in a slope unit drops below the shear stress, the slope nears the failure threshold, indicating a hazardous state^{35,37,38}. Other states are categorized as relatively safe. Based on safety factor values, slope stability was classified into five levels: low, medium-low, medium, medium-high, and high.

Each of the 100 slope units in Xuekou Village, selected as a case study, was systematically assigned a numerical designation. The categorization of landslide hazard levels for Xuekou Village was conducted based on the safety factor of the slope units once the soil moisture levels reached saturation. The resultant safety factor classification map for Xuekou Village, reflecting this saturation scenario, is presented in Fig. 14.

This study provides significant insights and addresses certain limitations to some extent. Through field investigations, we found that the high-risk results in our study align closely with existing local disaster prevention measures, which to some degree validates the applicability of our findings. However, future research should further integrate more detailed regional mitigation strategies and localized data to enhance the precision and reliability of risk assessments. Additionally, the calculation of the safety factor based on the Mohr–Coulomb criterion, which relies on constrained assumptions, may introduce uncertainties. Future studies should incorporate more detailed field data and dynamic simulations to refine these assessments further.

Dynamic and multi-source data, including real-time meteorological and updated remote sensing data, should also be integrated to enhance the timeliness and precision of landslide predictions, especially under rapidly changing environmental conditions. Employing ensemble learning methods, which combine multiple machine learning algorithms, could further improve prediction accuracy and generalizability. By addressing these challenges, future research can build on the foundations established in this study, advancing landslide risk assessments and disaster prevention strategies.

Conclusion

This study focuses on Wenzhou, a typical mountainous disaster-prone region in China, using historical landslide data as the basis. Terrain, geomorphological, and human activity-related data, including elevation, slope, land

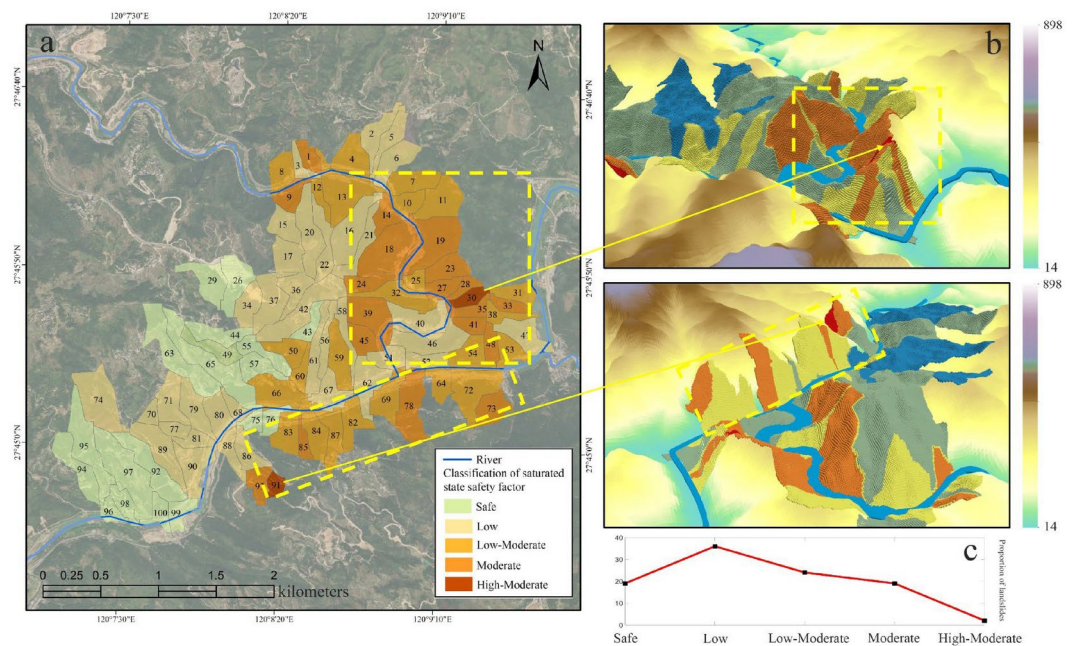


Fig. 14. (a): schematic diagram of safety factor classification in Xuekou village after soil water content reaches saturation; (b) 3d diagram of slope units in Xuekou village; (c) proportion of slope units by hazard level in Xuekou village. (Fig. 6 was created using ArcGIS 10.8.2 (<https://www.esri.com>)).

use, NDVI, lithology, soil type, geomorphology, and road density, were collected. Using ArcGIS software and machine learning models (Random Forest, XGBoost, and LightGBM), susceptibility analysis was performed for slope units across large-scale regions. While the research primarily takes Wenzhou as the study area, the methods and findings are applicable to other mountainous regions worldwide, particularly those prone to landslides. This adaptability highlights the broader significance of this study in improving disaster management and risk assessment in diverse geographical contexts. The study presents the following key findings:

- (1). **Optimization of Slope Unit Extraction:** An improved algorithm for slope unit extraction was proposed, enhancing extraction accuracy. Over 150,000 slope units were identified at a prefecture-level scale in Wenzhou, reducing issues like saw-tooth edges and unreasonable unit divisions, thus improving the precision of the extraction process. These improvements provide a robust framework that can be replicated in other regions to obtain accurate slope unit delineations.
- (2). **Impact of Extreme Rainfall on Hazard Levels:** The number of highly hazardous slope units increased significantly during extreme rainfall events. Rainfall intensity for different return periods was estimated using the Pearson Type III frequency curve, and the relative hazard was calculated by overlaying this data with susceptibility outcomes. The results showed that landslide risk increased with extreme rainfall, particularly in the mountainous and hilly regions, notably the southwestern and northeastern parts of Wenzhou. This relationship between rainfall intensity and landslide susceptibility is applicable to other mountainous regions, where similar geomorphological and climatic conditions exist.
- (3). **Case Study of Xuekou Village:** The feasibility of applying slope unit-based risk assessment results was demonstrated in Xuekou Village. A case study analyzed the distribution of hazard levels within its slope units and provided management recommendations for different hazard levels. These included targeted prevention for medium-to-high-risk units, regular monitoring, and enhancing vegetation coverage to mitigate landslide risk.

Despite focusing on a single prefecture in this study, computational constraints remain a primary limitation for expanding the scale of analysis. Future efforts will aim to optimize and parallelize computational processes, enabling susceptibility assessments at larger spatial scales, such as multiple counties or provinces. However, certain limitations persist. The reliance on the Mohr–Coulomb failure criterion and static datasets does not fully capture the dynamic interactions among geological, hydrological, and anthropogenic factors influencing landslide events. Furthermore, while the machine learning models employed demonstrated strong predictive performance, exploring ensemble methods that combine multiple algorithms could enhance prediction robustness and generalizability. To address these limitations, future research should integrate real-time meteorological data, high-resolution remote sensing imagery, and dynamic simulations, thereby improving the timeliness and accuracy of landslide susceptibility assessments. Expanding the application of the proposed framework to regions with diverse topographies and climatic conditions will validate its generalizability and adaptability. These advancements will contribute to the development of more comprehensive, data-driven approaches for

disaster prevention and control, ultimately supporting more resilient and sustainable communities in landslide-prone areas.

Data availability

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

Received: 14 December 2024; Accepted: 21 February 2025

Published online: 03 March 2025

References

- Jiang, W., Rao, P., Cao, R., Tang, Z. & Chen, K. Comparative evaluation of geological disaster susceptibility using multi-regression methods and spatial accuracy validation. *J. Geog. Sci.* **27**, 439–462 (2017).
- Wang, H., Wang, X., Zhang, C., Wang, C. & Li, S. Analysis on the susceptibility of environmental geological disasters considering regional sustainable development. *Environ. Sci. Pollut. Res.* **30**(4), 9749–9762 (2023).
- Cruden, D. A simple definition of a landslide. *Bull. Eng. Geol. Environ.* **43**(1), 27–29 (1991).
- Segoni, S. & Caleca, F. Definition of environmental indicators for a fast estimation of landslide risk at national scale. *Land* **10**(6), 621 (2021).
- Harp, E. L., Reid, M. E., McKenna, J. P. & Michael, J. A. Mapping of hazard from rainfall-triggered landslides in developing countries: examples from Honduras and Micronesia. *Eng. Geol.* **104**(3–4), 295–311 (2009).
- Guo, Y. et al. Study on the evaluation method and index system of major geological disasters based on AHP-taking the “6.24” landslide in Mao Xian county as an example. *City Disaster Reduct.* **2022**(3), 32–38 (2022).
- Higaki, D., Yoshida, K., Yoshimura, N. & Takahashi, T. Explanation of geological structure and movement of the Taninouchi landslide in relation to slope evolution processes. *Landslides* **32**(2), 1–9_1 (1995).
- Lian, B. et al. Formation mechanism analysis of irrigation-induced retrogressive loess landslides. *Catena* **195**, 104441 (2020).
- Sun, H. Y. et al. Analysis of deformation mechanism of landslide in complex geological conditions. *Bull. Eng. Geol. Environ.* **78**, 4311–4323 (2019).
- Li, B. et al. A multi-factor integrated method of calculation unit delineation for hydrological modeling in large mountainous basins. *J. Hydrol.* **597**, 126180 (2021).
- Sparrow, G. W. A. Some environmental factors in the formation of slopes. *Geogr. J.* **132**(3), 390–395 (1966).
- Larsen, M. C. Rainfall-triggered landslides, anthropogenic hazards, and mitigation strategies. *Adv. Geosci.* **14**, 147–153 (2008).
- Guo, M. Z., Gu, K. S. & Wang, C. Dynamic response and failure process of a counter-bedding rock slope under strong earthquake conditions. *Symmetry* **14**(1), 103 (2022).
- Dai, F. C., Lee, C. F. & Ngai, Y. Y. Landslide risk assessment and management: an overview. *Eng. Geol.* **64**(1), 65–87 (2002).
- Pan, D., Wang, J., Jia, H. & Zhao, J. Study of cartographic generalization on natural disaster risk mapping-taking point hazard-affected bodies as an example. *Geomat. Inf. Sci. Wuhan Univ.* **36**(1), 51–55 (2011).
- Alvioli, M., Marchesini, I., & Guzzetti, F. Nation-wide, general-purpose delineation of geomorphological slope units in Italy (No. e27066v1). *PeerJ* (2018).
- Huang, F. et al. Efficient and automatic extraction of slope units based on multi-scale segmentation method for landslide assessments. *Landslides* **18**, 3715–3731 (2021).
- Xiong, L., Li, S., Tang, G. & Strobl, J. Geomorphometry and terrain analysis: Data, methods, platforms and applications. *Earth Sci. Rev.* **233**, 104191 (2022).
- Kennelly, P. J. Terrain maps displaying hill-shading with curvature. *Geomorphology* **102**(3–4), 567–577 (2008).
- Minár, J., Evans, I. S. & Jenčo, M. A comprehensive system of definitions of land surface (topographic) curvatures, with implications for their application in geoscience modelling and prediction. *Earth Sci. Rev.* **211**, 103414 (2020).
- Park, S. J. & Van De Giesen, N. Soil-landscape delineation to define spatial sampling domains for hillslope hydrology. *J. Hydrol.* **295**(1–4), 28–46 (2004).
- Groenenendyk, D. G., Ferre, T. P., Thorp, K. R. & Rice, A. K. Hydrologic-process-based soil texture classifications for improved visualization of landscape function. *PloS ONE* **10**(6), e0131299 (2015).
- Krishnan, M. M. R., Choudhary, A., Chakraborty, C., Ray, A. K. & Paul, R. R. Texture based segmentation of epithelial layer from oral histological images. *Micron* **42**(6), 632–641 (2011).
- Alvioli, M. et al. Automatic delineation of geomorphological slope units with r. slopeunits v1.0 and their optimization for landslide susceptibility modeling. *Geosci. Model Dev.* **9**(11), 3975–3991 (2016).
- Alvioli, M., Guzzetti, F. & Marchesini, I. Parameter-free delineation of slope units and terrain subdivision of Italy. *Geomorphology* **358**, 107124 (2020).
- Zhao, Z., Liu, Z. Y. & Xu, C. Slope unit-based landslide susceptibility mapping using certainty factor, support vector machine, random forest, CF-SVM and CF-RF models. *Front. Earth Sci.* **9**, 589630 (2021).
- Wu, H. et al. Incorporating effects of slope units and sliding areas into seismically induced landslide risk modeling in tectonically active mountainous areas. *Remote Sens.* **16**(18), 3517 (2024).
- Carrara, M. et al. Gis techniques and statistical-models in evaluating landslide hazard. *Earth Surf. Process. Landf.* **16**(5), 427–445 (1991).
- Xie, M., Esaki, T. & Zhou, G. GIS method for slope-unit-based 3D landslide hazard evaluation. *Chin. J. Rock Mech. Eng.* **22**(6), 969–976 (2003).
- Zou, F. et al. Landslide hazard identification based on slope unit: A case study of shallow soil slope in Wanshan, Guizhou Province. *Chin. J. Geol. Hazard Control* **33**(3), 114–122 (2022).
- Zhao, X., Tan, S. & Li, Y. Risk assessment of geological hazards in Dongchuan District based on the methods of slope unit and combination weighting. *J. Yunnan Univ. Nat. Sci. Ed.* **43**(2), 299–305 (2021).
- Zhang, W., Wu, C., Tang, L., Gu, X. & Wang, L. Efficient time-variant reliability analysis of Bazimen landslide in the Three Gorges Reservoir Area using XGBoost and LightGBM algorithms. *Gondwana Res.* **123**, 41–53 (2023).
- Zhou, M., Shao, X. & Luo, M. Method and application of landslide geological hazard early-warning in Wenzhou city (in Chinese). *Chin. J. Geol. Hazard Control* **25**(2), 12–18 (2014).
- Zhao, X., Gao, X., Zhang, T., Zhou, M. & Xu, D. Potential degree index model of early warning system for geological hazards: A case study of Wenzhou in Zhejiang Province. *Geogr. Geo Inf. Sci.* **34**(5), 53–59 (2018).
- Li, C., Zhao, X., Xu, X. & Qu, X. Study on the differences between Hoek-Brown parameters and equivalent Mohr-Coulomb parameters in the calculation slope critical acceleration and permanent displacement. *Sci. Rep.* **14**(1), 15128 (2024).
- Li, S. Y. et al. Formation and failure mechanism of the landslide: a case study for Huaipa, Western Henan, China. *Environ. Earth Sci.* **80**, 1–12 (2021).
- Jeandet, L., Steer, P., Lague, D. & Davy, P. Coulomb mechanics and relief constraints explain landslide size distribution. *Geophys. Res. Lett.* **46**(8), 4258–4266 (2019).

38. Corominas, J. et al. Recommendations for the quantitative analysis of landslide risk. *Bull. Eng. Geol. Environ.* **73**, 209–263 (2014).

Acknowledgements

This research was funded by the National Key R&D Program of China (Grant No. 2022YFC3004404) and the State Grid Zhejiang Electric Power Co., Ltd. Technology Project (No.B311DS240008, Research on Rapid Assessment Technology for Comprehensive Risk of Typhoon Disaster in Large-Scale Overhead Transmission Lines). And research also was funded by the National Key Research and Development Program of China, Grant Numbers 2017YFA0604903 and 2017YFA0604904 which are funded by the Ministry of Science and Technology of the People's Republic of China.

Author contributions

K.D. and D.S. conceptualized the study and designed the methodology. K.D. and J.L. collected and processed the data. J.L. and C.W. performed the machine learning analyses. H.Z. contributed to the optimization of slope unit extraction methods. Z.W. and S.W. prepared the case study and statistical analyses. H.Z. wrote the main manuscript text and prepared figures. All authors reviewed and edited the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-91669-7>.

Correspondence and requests for materials should be addressed to Z.H.

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.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025