



## OPEN A systematic assessment of regional landslide risk under typhoon rainfall: a case study of Taishun, Zhejiang, China in September 2016

Chenchen Xie<sup>1,2,3,4</sup>, Chong Xu<sup>2,3,4,5</sup>✉, Xiwei Xu<sup>1</sup>, Yu Huang<sup>6</sup>, Yinke Li<sup>6</sup>, Kejie Chen<sup>2,7</sup>, Zhiwen Zheng<sup>8</sup>, Wei Zhang<sup>8</sup> & Saier Wu<sup>3,4</sup>

To clarify the occurrence patterns and risk characteristics of typhoon-induced rainfall landslides, this study assessed landslide susceptibility in Taishun County, Zhejiang Province, and surrounding areas affected by Typhoon Meranti in September 2016. Based on a detailed inventory of 4,102 landslides, a susceptibility framework incorporating key topographic and rainfall-related factors was established using an automated machine learning approach. The results show that landslides are predominantly distributed at elevations of 400–800 m and slope gradients of 25°–40°, with frequent occurrences associated with cumulative rainfall of 160–180 mm. The optimal ensemble model achieved high predictive performance and demonstrated strong spatial agreement between predicted high-risk zones and observed landslide distributions. Factor contribution analysis indicates that accumulated rainfall, elevation, and slope are the primary controlling factors, with their interactions playing a critical role in landslide susceptibility. Overall, this study provides an effective and interpretable framework for understanding typhoon-triggered landslide mechanisms and supporting regional early warning and disaster mitigation efforts.

**Keywords** Landslide inventory, Spatial distribution pattern, Automated machine learning, Hazard assessment, SHAP interpretation

The southeastern coastal region of China, characterized by a subtropical monsoon climate, frequently experiences typhoon-induced heavy rainfall. Combined with the widespread distribution of mountainous and hilly terrain and complex geological conditions, this region has become highly prone to rainfall-induced landslides<sup>1–4</sup>. Landslides in these areas are often sudden, extensive, and chain-reactive, resulting in casualties and significant property losses, and posing serious threats to regional ecological security and sustainable socio-economic development<sup>5–7</sup>. In recent years, previous studies have made notable progress in the assessment of typhoon-induced landslides<sup>8</sup>; Sajwan et al.<sup>4,7</sup>; Xie et al.<sup>9</sup>. Despite these advances, several critical challenges remain. First, the integration of multi-source heterogeneous data is still insufficient. Most existing studies focus on single-type datasets and fail to comprehensively integrate topographic, geological, and hydrological factors, leading to incomplete characterization of the controlling variables of landslides<sup>10–13</sup>. In particular, deficiencies exist in model construction and data foundations. Some studies rely on traditional models with subjectively assigned weights, or use incomplete and inaccurate landslide inventories, thereby limiting assessment accuracy (Gupta et al.<sup>4,6</sup>; Huang et al.<sup>14</sup>; Huang et al.<sup>15,16,17</sup>). Moreover, the interpretability of models remains limited. Although

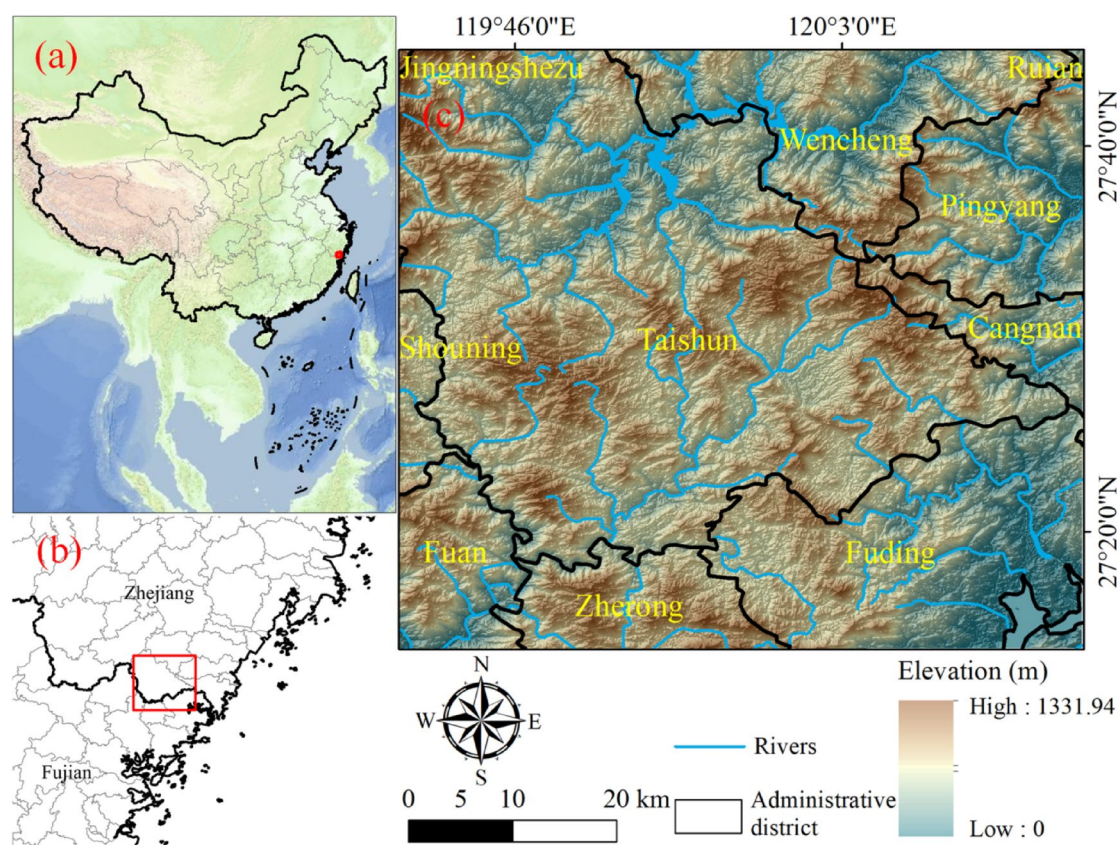
<sup>1</sup>School of Earth Sciences and Resources, China University of Geosciences (Beijing), Beijing 100083, China. <sup>2</sup>South China Field Scientific Observation and Research Station for Climate-Driven Landslide Risk, Ministry of Natural Resources, Guangzhou 510599, China. <sup>3</sup>National Institute of Natural Hazards, Ministry of Emergency Management of China, Beijing 100085, China. <sup>4</sup>Key Laboratory of Compound and Chained Natural Hazards Dynamics, Ministry of Emergency Management of China, Beijing 100085, China. <sup>5</sup>School of Geology and Mining Engineering, Xinjiang University, Urumqi 830017, China. <sup>6</sup>Department of Geotechnical Engineering, College of Civil Engineering, Tongji University, Shanghai 200092, China. <sup>7</sup>Department of Earth and Space Sciences, Southern University of Science and Technology, Shenzhen 518055, China. <sup>8</sup>Guangdong Geological Environment Monitoring Station, Guangzhou 510599, China. ✉email: chongxu@ninhm.ac.cn

advanced algorithms such as machine learning have improved predictive performance, their “black-box” nature often obscures the contribution mechanisms of different factors<sup>18</sup>; Xu et al.<sup>19</sup>; Zhang et al.<sup>2021</sup>. In particular, for the unique scenario of typhoon-induced rainfall, current research lacks in-depth analysis of the chain disaster processes, which constrains the precision and applicability of regional landslide hazard prevention and control strategies<sup>22–24</sup>.

To address these gaps, this study focuses on Taishun County, Zhejiang Province, and its surrounding areas affected by Typhoon Meranti in 2016. First, a high-quality landslide inventory was established through the integration of multi-source data and human–computer interactive interpretation, providing a robust and reliable sample basis for model training. Second, an ensemble learning approach was adopted to construct a regional landslide hazard assessment model, thereby enhancing predictive accuracy and model robustness. In addition, SHAP-based analysis was employed to quantitatively interpret the relative contributions of influencing factors and their interactions, improving the transparency and interpretability of the model outcomes. On this basis, the spatial distribution patterns and occurrence characteristics of landslides under typhoon-induced rainfall conditions were systematically investigated. The results not only support landslide early warning and hazard zoning in Taishun County and adjacent areas, but also provide a transferable methodological framework for landslide risk assessment in other typhoon-prone regions along the southeastern coast of China.

### Overview of the study area

The study area is located in the southwestern part of Zhejiang Province and the northeastern part of Fujian Province, encompassing Taishun County and its surrounding regions (Fig. 1). The terrain is highly complex, characterized by an interlaced pattern of mountains and hills. The delineation of the study area was based on the spatial concentration of landslides: interpretation was first conducted within the core area of high landslide density and then gradually extended outward. A buffer zone of approximately 3–5 km beyond the outer boundary was further examined, during which no additional landslides were identified. The final study area covers approximately 3,433.34 km<sup>2</sup> and lies in a coastal zone dominated by a subtropical monsoon climate, where abundant rainfall frequently triggers rainfall-induced geological hazards.

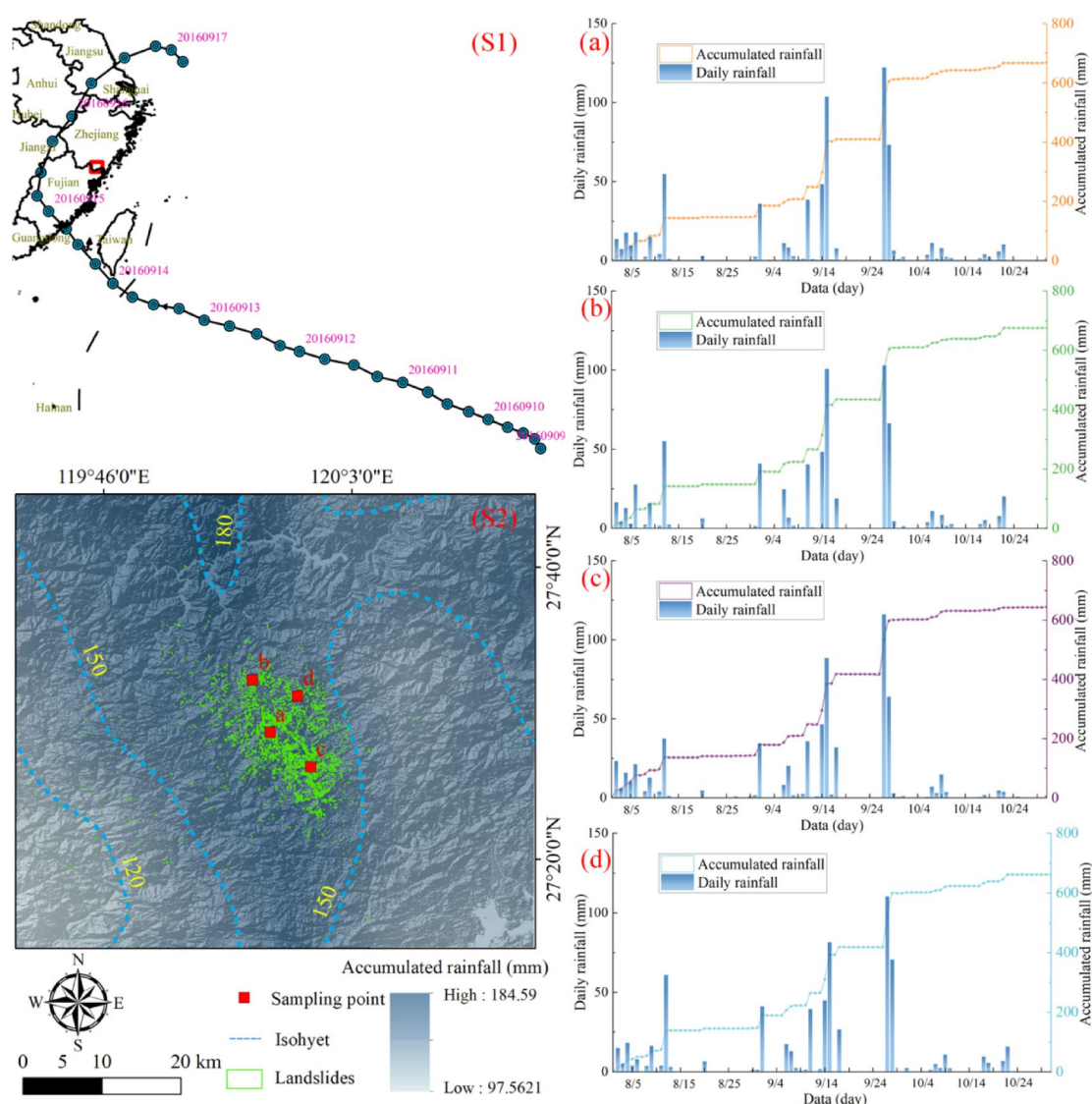


**Fig. 1.** Location of the study area and its related geographical information. (a) Location of the study area with respect to the national boundary; (b) Location of the study area with respect to provincial boundaries; (c) Geographical overview of the study area.

## Data and methods

### Rainfall data

The GPM (Global Precipitation Measurement) daily rainfall product provides comprehensive rainfall information and is widely used in studies of rainfall-induced landslides. The GPM data, available in grid format (<https://search.earthdata.nasa.gov>), have a spatial resolution of  $0.1^\circ \times 0.1^\circ$ . For subsequent analysis, the data were resampled to a 30 m resolution. Subfigure S1 in Fig. 2 illustrates the track and influence process of Typhoon Meranti, marking the relative position of the study area along the typhoon path. As shown, Meranti entered mainland China on 14 September 2016 and dissipated on 17 September, with the closest proximity to the study area occurring on 15–16 September. Landslide interpretation was conducted using remote sensing images acquired in mid-August and mid-October 2016. To better understand the rainfall conditions associated with this event, daily GPM rainfall data from 1 August to 31 October 2016 (a total of 92 days) were extracted. Four rainfall sampling points (a–d) were selected for detailed analysis, all located within zones of high landslide density. As shown in Fig. 2a and d, the cumulative rainfall at all four sites exhibited a steadily increasing trend, with a pronounced surge between 14 and 17 September. Because the post-event remote sensing imagery and the typhoon's influence period fall within this interval, the daily rainfall from these four days was summed to represent the cumulative rainfall for the event (Subfigure S2 in Fig. 2). Among the four sampling points, point c recorded the highest cumulative rainfall (168.7 mm), followed by point b (168.1 mm). The maximum daily rainfall was observed at point a on 15 September (103.8 mm), followed by point b on the same day (100.8 mm).

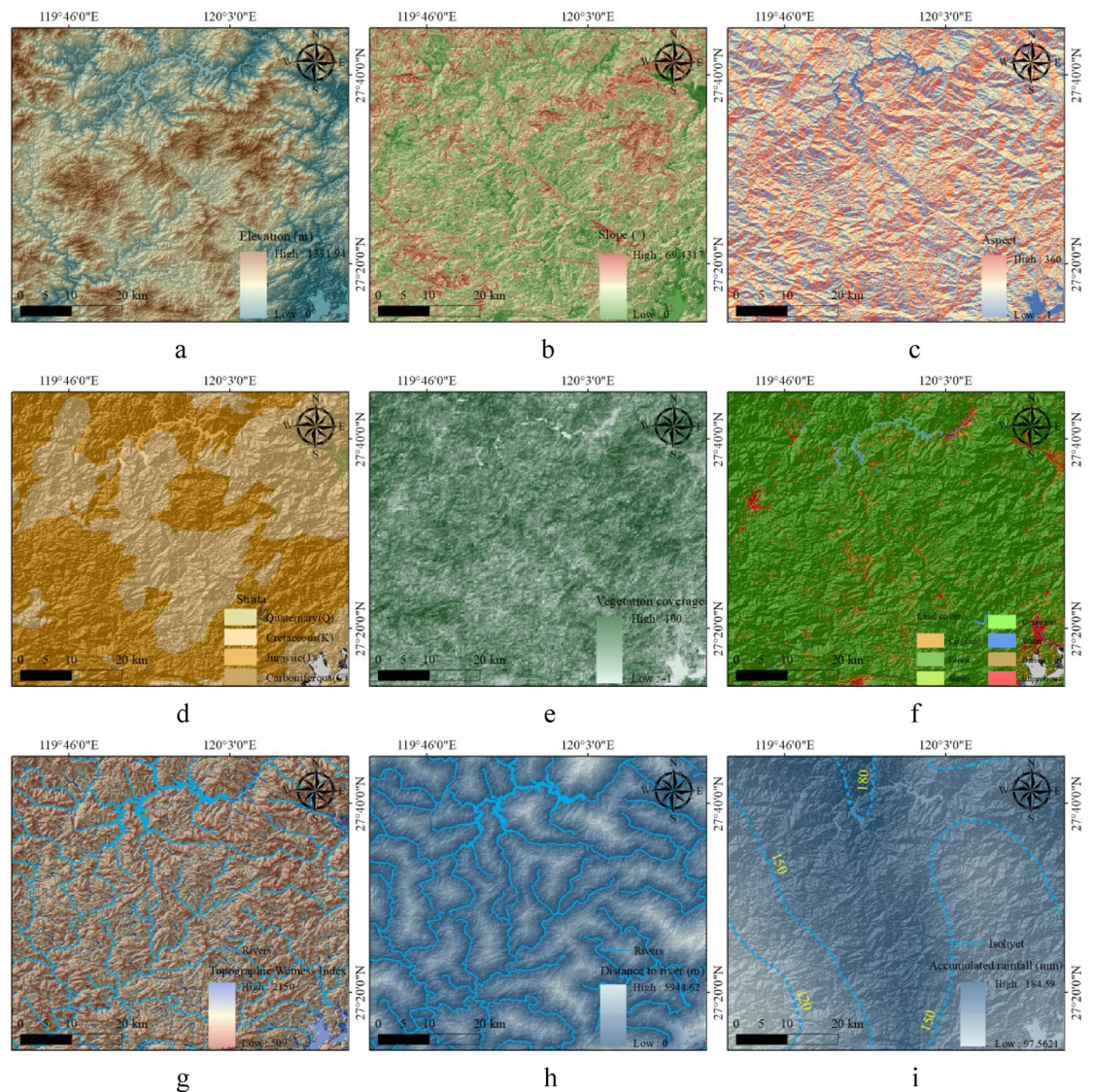


**Fig. 2.** Rainfall conditions in the study area during the typhoon. (S1) Typhoon influence process and the relative position of the study area; (S2) Cumulative rainfall distribution over four days and the locations of the four sampling points; (a) Daily and cumulative rainfall at sampling point a over 92 days; (b) Daily and cumulative rainfall at sampling point b over 92 days; (c) Daily and cumulative rainfall at sampling point c over 92 days; (d) Daily and cumulative rainfall at sampling point d over 92 days.

### Influence factor data

To systematically analyze the influence of topography, geological and hydrological conditions, and typhoon-induced rainfall on landslide occurrence, nine landslide influencing factors were selected in this study: elevation, slope, aspect, strata, vegetation coverage, land cover, topographic wetness index (TWI), distance to river, and accumulated rainfall (Fig. 3). These factors were used to explore the spatial distribution patterns and controlling mechanisms of landslides.

Elevation data were obtained from the Copernicus Digital Elevation Model (DEM) released by the European Space Agency (ESA), with a spatial resolution of 30 m (<https://panda.copernicus.eu>). Based on these data, three derived topographic factors, including slope, aspect, and the topographic wetness index (TWI), were generated. Strata data were sourced from the “National 1:200,000 Digital Geological Map Spatial Database” in MapGIS format<sup>25</sup>. After extraction and format conversion, the strata data were made suitable for analysis (<http://www.ngac.org.cn>). Vegetation coverage was represented as a percentage ranging from 0 to 100%<sup>26</sup>. Considering that landslides rarely occur in water bodies, vegetation coverage for water areas was assigned a value of -1. Land cover data were derived from China’s first annual Landsat-based land cover product (Yang et al.<sup>48</sup>), with nine categories: cropland, forest, shrub, grassland, water, snow/ice, barren, impervious, and wetland. Given that the event occurred in 2016, the 30 m resolution data from 2016 were selected. Hydrographic data for the study area were obtained from the “1:1,000,000 Public Version Basic Geographic Information Data” in the National Geographic Information Resource Directory Service System (<https://www.webmap.cn>). River network features were extracted and a distance-to-river raster was generated through spatial analysis. Accumulated rainfall data were derived from the GPM daily precipitation product. Using the temporal range corresponding to the typhoon



**Fig. 3.** Nine landslide-influencing factors used in this study. (a) Elevation; (b) Slope; (c) Aspect; (d) Strata; (e) Vegetation coverage; (f) Land cover; (g) Topographic Wetness Index (TWI); (h) Distance to river; (i) Accumulated rainfall.

event, daily rainfall data were cumulatively summed on the GeoScene platform and subsequently resampled to a 30 m resolution.

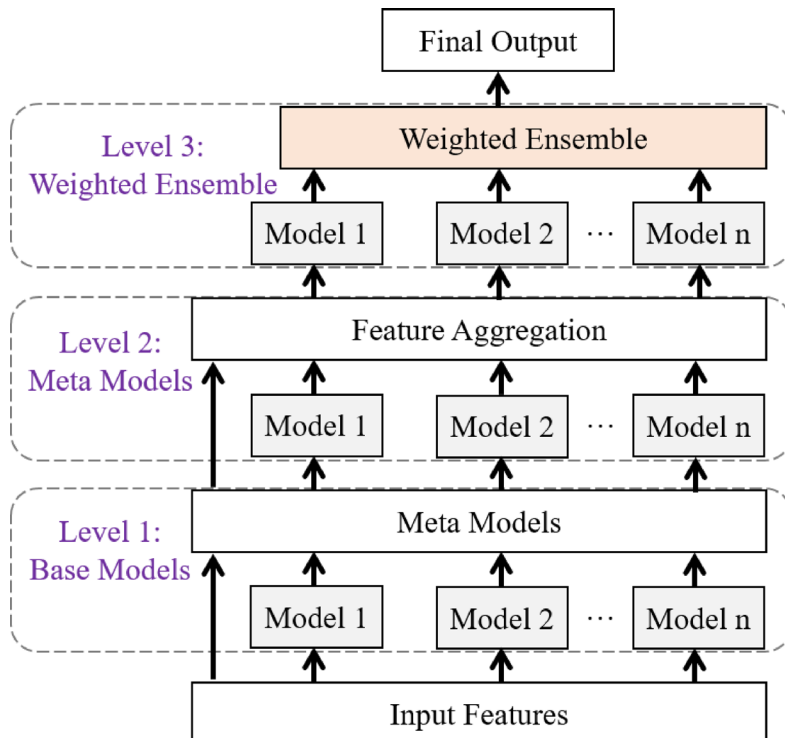
## Methods

In this study, a landslide inventory triggered by the rainfall event was systematically established using high-resolution optical remote sensing imagery combined with a human–computer interactive interpretation approach<sup>27–29</sup>. Landslide areas were identified by comparing pre- and post-event imagery, based on features such as vegetation loss, color changes, and flow morphology. The delineated landslide boundaries were digitized as vector data<sup>30,31</sup>. To enhance identification accuracy, high-resolution imagery and three-dimensional visualization available on the Google Earth platform were used for topographic verification, allowing the exclusion of disturbances caused by human activities such as building construction, road works, and mining. Field surveys and ground verification further confirmed the reliability of the landslide inventory. Additionally, by applying standardized interpretation criteria and digital human–computer interactive methods, geometric characteristics, color, texture, and spatial distribution of landslides were analyzed to ensure the inventory’s accuracy and practical utility.

The AutoGluon framework adopts a hierarchical ensemble learning architecture designed to enhance predictive performance through multi-level model integration. In this study, a three-level architecture is used to illustrate the AutoGluon framework (Fig. 4). At the base layer (Level 1), multiple base models are trained on the input features and generate initial predictions. In the second layer (Level 2), the outputs of these base models are aggregated by meta-models to further optimize the predictions. Finally, at the weighted ensemble layer (Level 3), the outputs of the meta-models are combined using weighted aggregation to produce the final prediction results. This hierarchical ensemble approach allows AutoGluon to automatically select and adjust the contributions of individual models, thereby improving prediction accuracy and model generalization. Table 1 summarizes the core workflow of the AutoGluon framework implemented in this study.

SHAP (Shapley Additive exPlanations) values provide a method for interpreting machine learning models by quantifying the contribution of each feature to the prediction outcome. Originating from cooperative game theory, SHAP values measure the marginal contribution of each participating feature within a collaborative setting. The calculation of SHAP values can be divided into three steps. Assume a model that depends on multiple features  $X = (x_1, x_2, \dots, x_n)$ . The SHAP value for each feature  $x_i$  is computed by considering its marginal contribution across all possible feature permutations. The calculation formula for the SHAP value of  $x_i$  is as follows:

$$\varphi_i(f) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)]$$



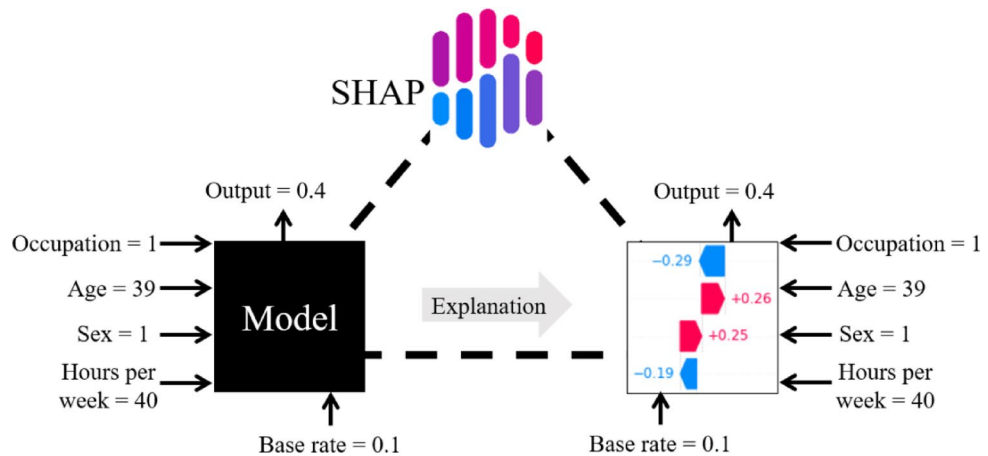
**Fig. 4.** Three-level architecture of the AutoGluon framework.

```

# Step 1: Import necessary libraries
from autogluon.tabular import TabularDataset, TabularPredictor
# Step 2: Load your dataset
train_data = TabularDataset('train.csv')
test_data = TabularDataset('test.csv')
# Step 3: Initialize the AutoML model and fit the model
predictor = TabularPredictor(label='target').fit(train_data)
# Step 4: Evaluate the model performance and predict new data
performance = predictor.evaluate(test_data)
predictions = predictor.predict(test_data)

```

**Table 1.** Example of the AutoGluon workflow code.



**Fig. 5.** Conceptual workflow of SHAP value calculation.

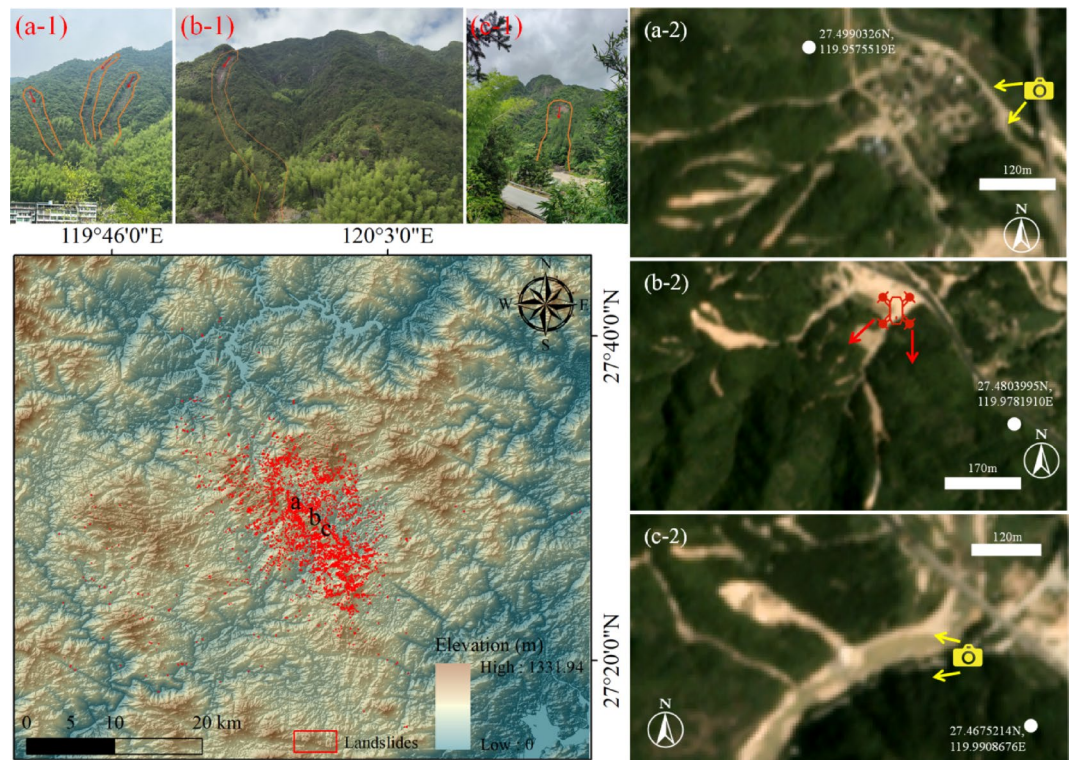
Here,  $N$  denotes the set of all features, i.e., the complete feature set.  $S$  represents a subset of features, with  $S \subseteq N \setminus \{i\}$ .  $f(S)$  is the model prediction based on the feature subset  $S$ , while  $f(S \cup \{i\})$  is the model prediction after adding feature  $x_i$  to subset  $S$ . The weight term  $\frac{|S|!(|N|-|S|-1)!}{|N|!}$  reflects the importance of feature  $i$  in a specific feature permutation within the SHAP framework. The calculation proceeds in three main steps. First, the marginal contribution of each feature is computed, i.e., the change in model prediction when feature  $x_i$  is added to the current subset of features. Second, this process is repeated across all possible feature subsets. Finally, the weighted average of the marginal contributions over all possible feature permutations is taken to obtain the final SHAP value. The computation steps and workflow are illustrated in Fig. 5.

## Results and analysis

### Landslide inventory

Using human–computer interactive interpretation, pre- and post-event remote sensing imagery were analyzed to identify rainfall-induced landslides. Within the study area of approximately 3,433.34 km<sup>2</sup>, a total of 4,102 landslides were delineated. The largest landslide covered an area of 21,334.97 m<sup>2</sup>, the total landslide area was approximately 6.09 km<sup>2</sup>, and the mean landslide area was 1,484.15 m<sup>2</sup>. To further verify the reliability of the landslide inventory, three representative landslide sites were selected for field investigation. As shown in Fig. 6, the landslides generally exhibit elongated shapes, consistent with the typical appearance of rainfall-induced landslides in southeastern China. The source zones are prominent, with deposits composed of loosely structured mixtures of soil and rock of varying sizes and irregular boundaries, clearly delineated from the surrounding stable slopes. Landslides frequently occur on steeper slopes, with uneven vegetation coverage in affected areas, showing evident displacement, damage, or even complete destruction. In contrast, the adjacent non-landslide areas are densely vegetated. In some cases, exposed bedrock prevents vegetation regrowth, leaving scars visible even years after the landslide occurrence.

A landslide density analysis was conducted using a search radius of 2 km, and Gaussian kernel density estimation was applied to calculate point density and area percentage. As shown in Fig. 7a and b, the maximum point density and area density are 48.84 km<sup>-2</sup> and 8.48%, respectively. Landslides are densely distributed along a northwest–southeast direction in the central part of the study area. Based on the landslide inventory, three metrics, including landslide count, landslide area, and frequency density, were analyzed. Figure 7c presents the distribution of landslide area versus landslide count. Results indicate that the cumulative number of



**Fig. 6.** Overall distribution of rainfall-induced landslides. (a-1) Field survey of landslide a; (b-1) UAV image of landslide b; (c-1) Field survey of landslide c; (a-2) Post-event remote sensing image of landslide a; (b-2) Post-event remote sensing image of landslide b; (c-2) Post-event remote sensing image of landslide c.

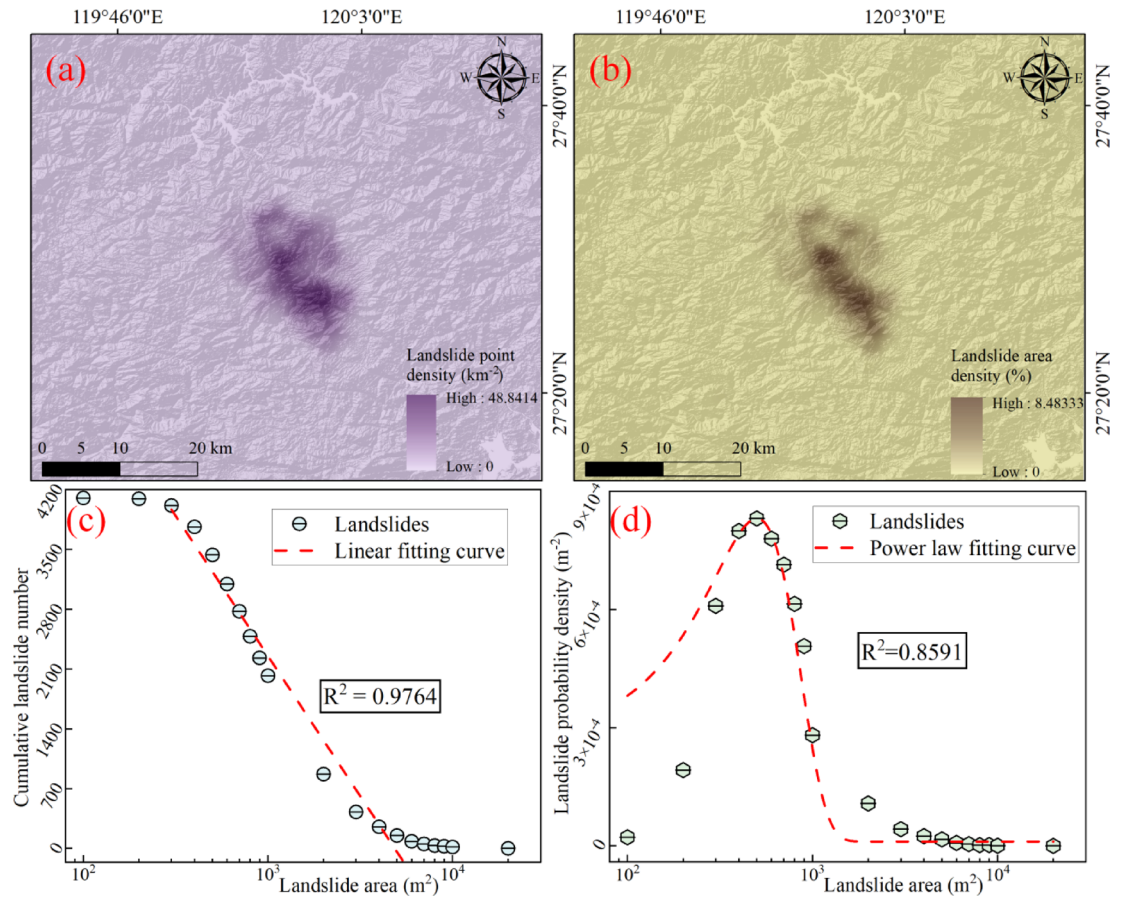
landslides decreases gradually with increasing landslide area, with the most pronounced reduction occurring in the 1,000–2,000 m<sup>2</sup> range. A linear fit was performed on 95.8% of the landslide data, yielding a correlation coefficient of 0.974, indicating a strong linear relationship. Figure 7d shows the distribution of landslide area versus frequency density, with the horizontal and vertical axes representing landslide area and frequency density, respectively. The fitted curve exhibits an initial increasing trend followed by a gradual decline. And the overall trend follows a power-law distribution with a correlation coefficient of 0.859. This result supports the completeness of the landslide inventory.

### Landslide mobility analysis

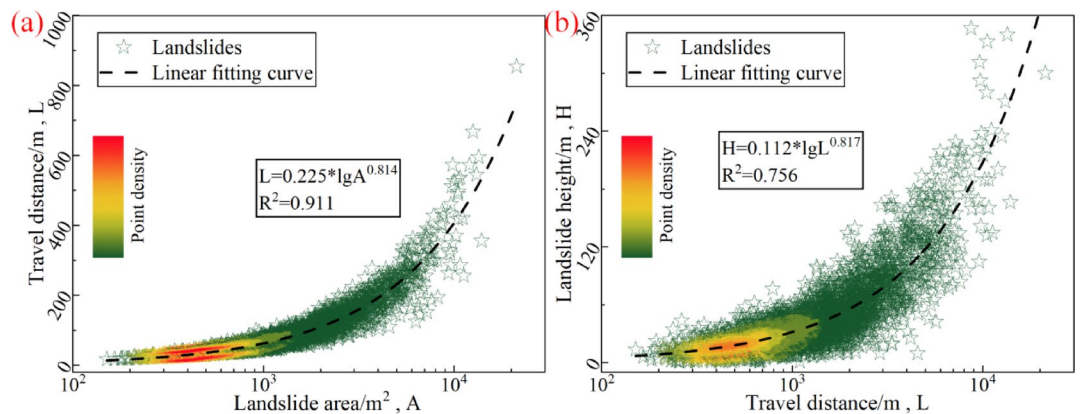
To gain deeper insights into landslide occurrence patterns, the mobility of landslides was statistically analyzed. As shown in Fig. 8a, landslide runout distance ( $L$ ) exhibits a power-law increase with landslide area ( $A$ ), with a correlation coefficient of 0.911. This trend indicates a strong relationship between landslide area and runout distance. Larger landslides imply greater source material; under rainfall-induced conditions, water infiltration saturates the soil–rock mass, reducing shear strength and promoting sliding. Moreover, the accumulated kinetic energy of large landslides is more significant, enabling longer movement under the action of gravity. Figure 8b shows that the vertical drop ( $H$ ) of landslides also follows a power-law relationship with runout distance ( $L$ ), with a correlation coefficient of 0.756. This suggests that the greater the vertical height of the landslide mass, the farther it can travel. Larger vertical drops generally indicate steeper initial slopes and more available gravitational potential energy. Higher efficiency in converting this potential energy into kinetic energy results in longer landslide runout distances.

### Landslide spatial distribution analysis

In this study, nine influencing factors were analyzed in relation to two statistical indicators, landslide count and area, to reveal the regularity between landslide occurrence and natural geographic conditions. Overall, landslide spatial distribution is governed by a combination of topography, geology, vegetation, hydrology, and rainfall. From the topographic perspective, as shown in Fig. 9a and b, elevation and slope exert the most significant control over landslide occurrence. Landslides are predominantly concentrated in areas of moderate elevation (approximately 400–800 m) and moderate to steep slopes (25°–40°). This pattern is closely related to the widespread mountainous and hilly terrain with pronounced relief in southeastern coastal regions, reflecting substantial gravitational potential energy in such landscapes. Runoff tends to concentrate in these areas, stresses on slopes are higher, and materials are looser, providing favorable conditions for landslide development. Conversely, regions at very low or very high elevations are either relatively flat or composed of more stable rock masses, and thus are less prone to landslides. In terms of aspect, as shown in Fig. 9c, landslide counts are notably higher on

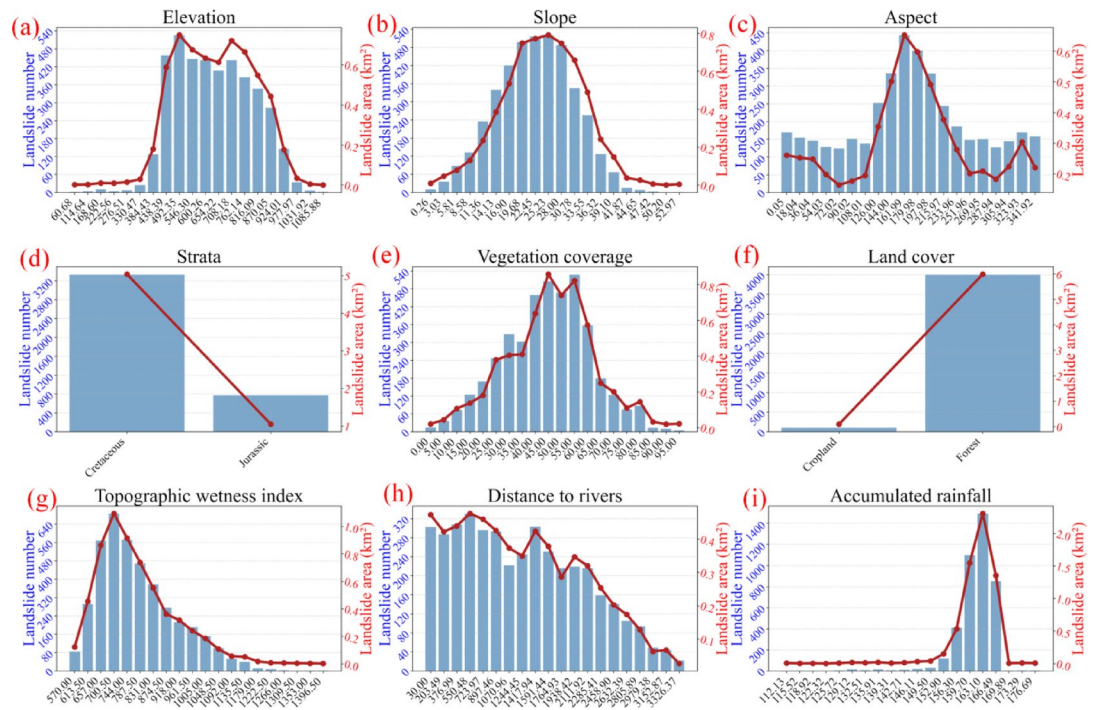


**Fig. 7.** Distribution and size statistics of rainfall-induced landslides. (a) Landslide area density and proportion of landslides in each administrative region; (b) Landslide point density and the number of landslides in each administrative region; (c) Variation of cumulative landslide count with landslide area; (d) Variation of landslide probability density with landslide area.



**Fig. 8.** Mobility analysis of rainfall-induced landslides. (a) Relationship between landslide runout distance and landslide area; (b) Relationship between landslide vertical drop and landslide runout distance.

south- and southeast-facing slopes, which corresponds to the prevailing rainfall directions in the region. The area is controlled by the southeast monsoon, so south- and southeast-facing slopes directly experience intense rainfall, leading to high infiltration and elevated pore-water pressure, which readily destabilizes the slopes. In addition, these slope aspects exhibit large temperature and moisture differences, and long-term weathering further loosens soil–rock structure, reducing slope stability. Regarding geological factors, Fig. 9d indicates that landslides are mainly distributed in Cretaceous strata. In southeastern coastal regions, Cretaceous formations



**Fig. 9.** Statistical analysis of nine influencing factors and landslides. (a) Elevation; (b) Slope; (c) Aspect; (d) Strata; (e) Vegetation coverage; (f) Land cover; (g) Topographic Wetness Index; (h) Distance to river; (i) Accumulated rainfall.

are generally composed of sandstone, mudstone, or conglomeratic sandstone. Well-developed joints and weak lithology, combined with intense chemical weathering under a tropical monsoon climate, facilitate the formation of thick residual deposits. In contrast, Jurassic or granitic formations are generally harder and more resistant, resulting in fewer landslides. Regarding vegetation conditions, Fig. 9e indicates that both landslide frequency and affected area peak under intermediate vegetation coverage (approximately 40%–70%). This pattern does not imply a direct causal relationship but reflects the combined effects of vegetation structure, soil properties, and hydrological processes. Areas with moderate vegetation coverage often correspond to forested or shrub-dominated slopes where root systems provide partial reinforcement but are insufficient to fully stabilize thick soil layers<sup>32,33</sup>. These areas typically exhibit higher soil depth and water retention capacity, which can increase susceptibility to shallow landslides under intense rainfall. During typhoon events, rapid soil saturation and concentrated surface runoff further exacerbate slope instability, making moderately vegetated slopes more prone to failure<sup>34,35</sup>. In contrast, areas with dense vegetation coverage generally possess well-developed root networks that enhance soil cohesion and slope stability, resulting in a lower frequency of landslides. Concerning land cover, Fig. 9f illustrates that landslide area and number are significantly higher in forests compared to croplands. This is likely due to forested areas being located on steeper slopes with thicker soil layers and concentrated rainfall, rather than forests directly inducing landslides. Cropland areas are usually on gentler terrain with more stable slopes, hence fewer landslides. For hydrology and rainfall factors, Fig. 9g shows that landslide counts initially increase and then decrease with Topographic Wetness Index (TWI). Southeastern coastal areas experience a humid monsoon climate with strong slope water cycles and frequent groundwater fluctuations. Under moderate moisture conditions, slopes are near a critical saturation state, with sufficient water to reduce friction but not fully saturated, facilitating shallow landslides. Excessively wet areas are typically low-lying with gentle slopes, which is unfavorable for landslide formation. As shown in Fig. 9h, proximity to rivers is positively correlated with landslide count and area, reflecting significant riverbank erosion, which weakens toe stability and forms high-incidence linear zones. Finally, Fig. 9i demonstrates that landslide count and area peak when accumulated rainfall is approximately 160 ~ 180 mm, indicating that intense rainfall is the direct triggering factor. Typhoon-induced heavy rainfall, characterized by short-duration high intensity and high cumulative amounts, rapidly increases soil moisture and pore-water pressure, reducing matric suction and facilitating the development of potential sliding surfaces. Prolonged multi-day rainfall or successive typhoon events can further trigger large-scale, clustered landslides.

### Landslide hazard assessment

In machine learning model training, when the ratio of landslide to non-landslide samples is balanced, the model can more equitably learn the core features of both classes, effectively avoiding bias caused by the dominance of majority-class samples. To ensure reasonable and reliable model fitting, negative samples were randomly selected from areas located more than 200 m away from landslide surfaces, with the number of negative samples equal to that of landslide samples, resulting in 4,102 non-landslide points. Combined with the 4,102 landslide

samples collected, a complete dataset of 8,204 samples was constructed for the landslide hazard assessment. Pearson correlation analysis, commonly used in landslide hazard assessment studies, was applied to quantify the relationships among the nine selected landslide-influencing factors. This method measures the degree of linear correlation between two variables, with correlation coefficients ranging from -1 to 1, providing an intuitive reflection of both the strength and direction of the linear association. During feature engineering and model construction, highly correlated variables may introduce information redundancy, potentially interfering with the model's identification of key features. Therefore, correlation analysis is essential to detect multicollinearity issues. As shown in Fig. 10, the absolute values of the Pearson correlation coefficients among most factors are below 0.4, indicating low correlation. This result demonstrates that the influencing factors are largely linearly independent, with no significant multicollinearity observed. All nine factors can be included in the landslide hazard model without removing strongly correlated variables. Hence, the selected influencing factors exhibit good independence and scientific validity, fully meeting the core requirements for machine learning model training.

To evaluate model performance in the classification task, Receiver Operating Characteristic (ROC) curves were plotted and the corresponding area under the curve (AUC) was calculated. The ROC curve presents the false positive rate (FPR) on the horizontal axis and the true positive rate (TPR) on the vertical axis, providing an intuitive visualization of model performance across different classification thresholds. The AUC quantifies the area under the curve, ranging from 0 to 1, with values closer to 1 indicating superior classification performance. Figure 11 shows the ROC curves and corresponding AUC values for 14 models on both the training and testing datasets. In the training set (Fig. 11a), all models exhibit high AUC values, with the WeightedEnsemble model performing the best, achieving an AUC of 0.9985. Other mainstream ensemble learning models, traditional k-nearest neighbors (KNN) models, and neural network models also demonstrate varying degrees of performance. In the testing set (Fig. 11b), AUC values for all models decrease to some extent, consistent with the expected performance degradation on unseen data, yet overall classification ability remains strong. The WeightedEnsemble model continues to perform best with an AUC of 0.9693. Ensemble learning models follow closely, whereas KNN performance declines more noticeably, and the generalization capability of neural network models does not surpass that of mainstream ensemble models. Among the models, the WeightedEnsemble further enhances the performance ceiling through model fusion. Mainstream ensemble learning models consistently demonstrate stable and superior classification performance across both training and testing datasets. The KNN model is more sensitive to data distribution and exhibits limited generalization. Neural network models do not achieve

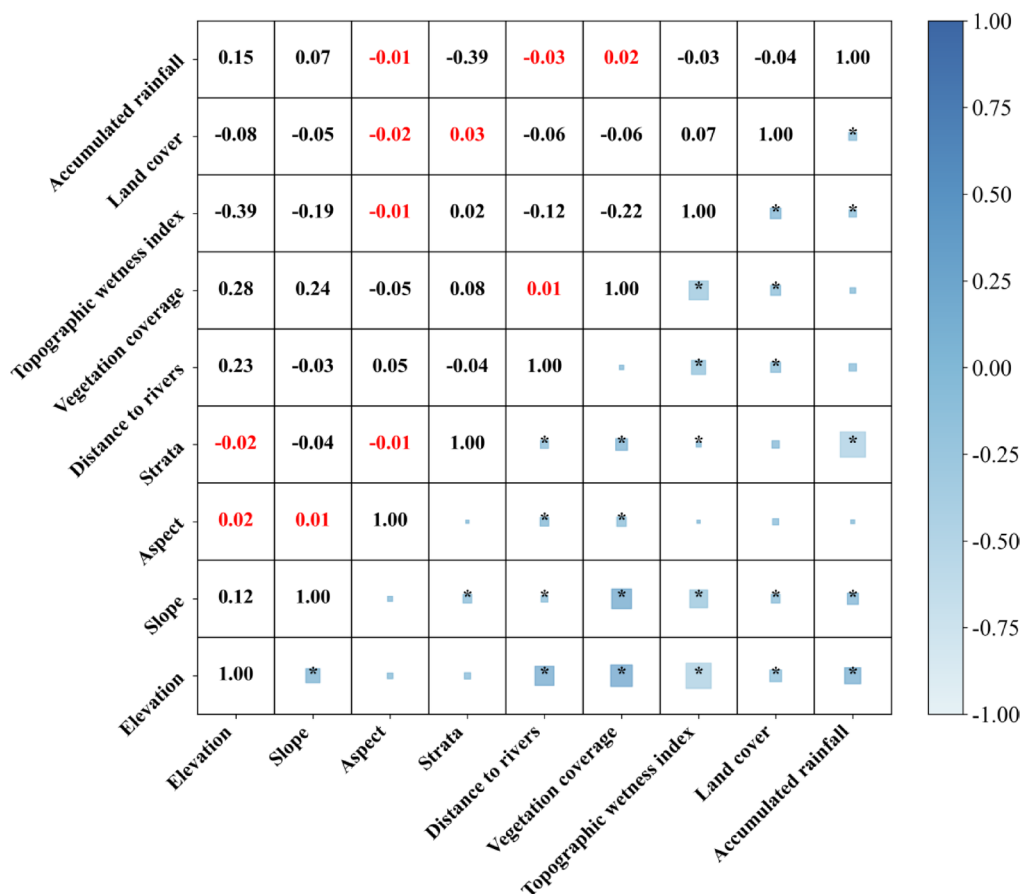
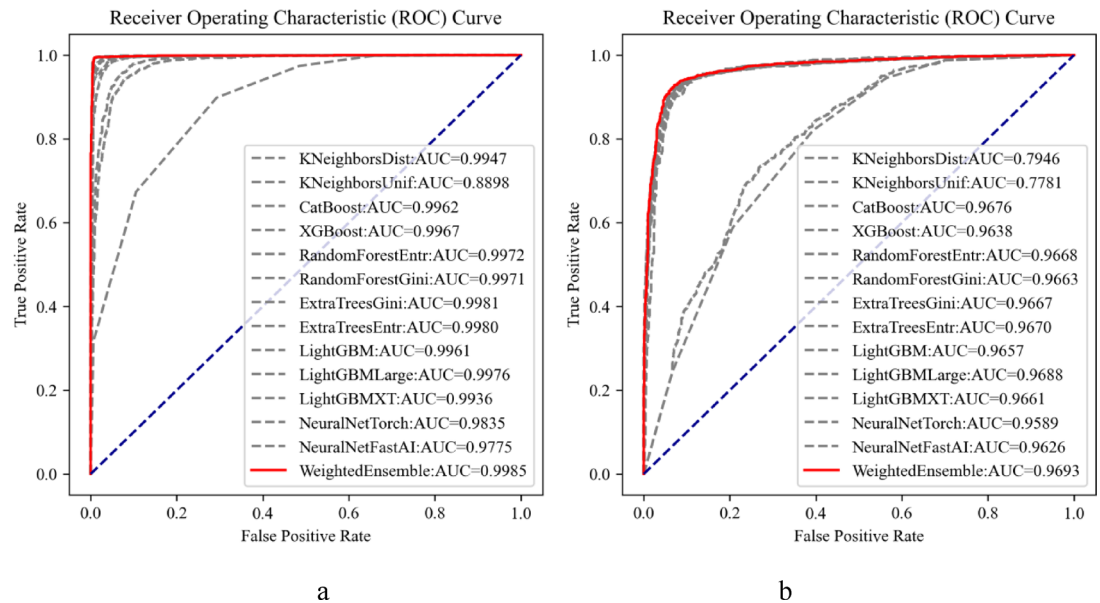


Fig. 10. Pearson correlation analysis between landslide and non-landslide samples.



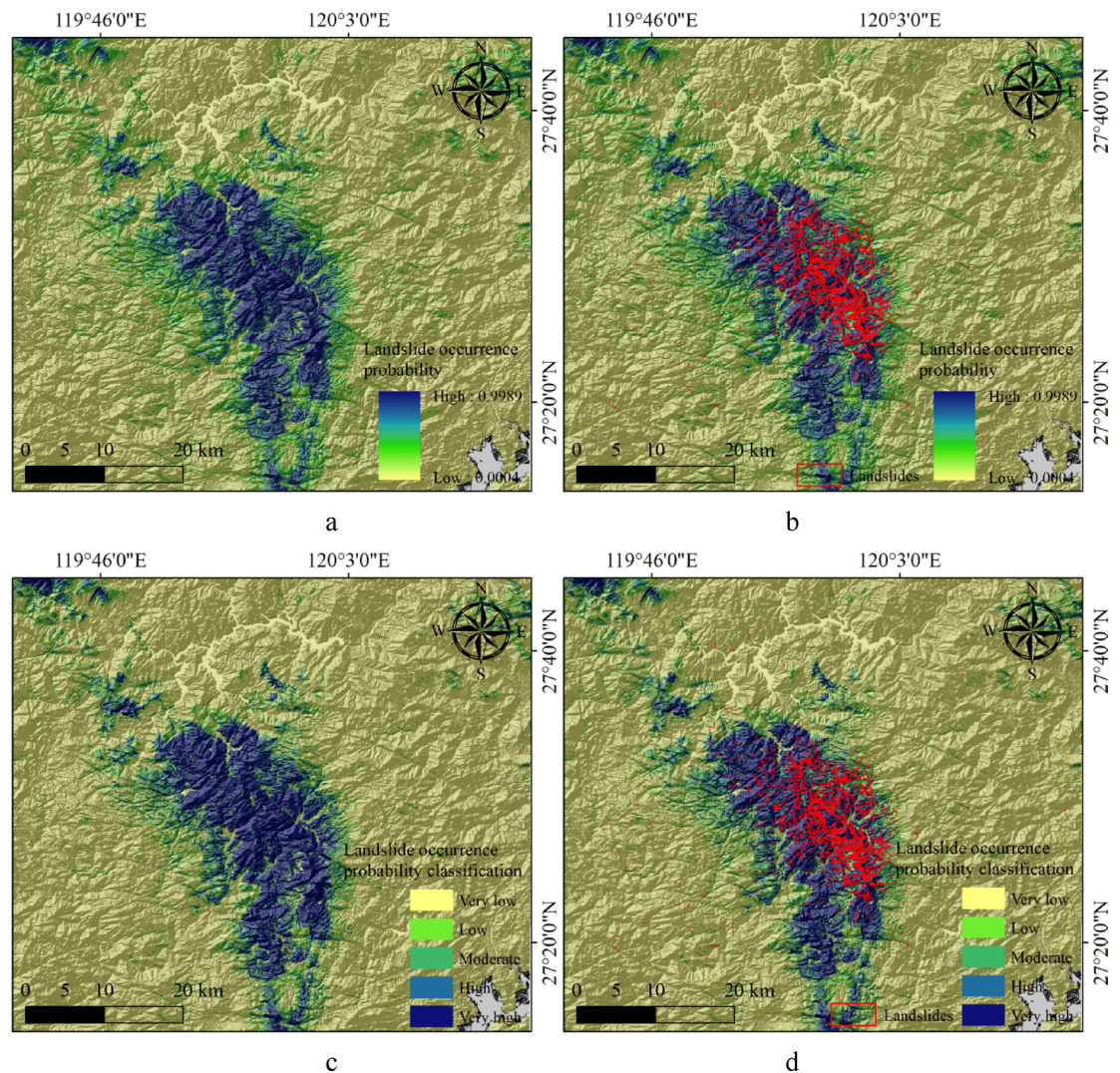
**Fig. 11.** ROC curves and AUC values of 14 models based on the sample dataset. (a) ROC curves and AUC values of 14 models on the training set; (b) ROC curves and AUC values of 14 models on the testing set.

the performance level of ensemble learning models in this task, potentially due to data characteristics or model architecture design.

Figure 12a illustrates the spatial distribution of rainfall-induced landslide hazard probabilities in the study area, with values ranging from 0.0004 to 0.9989. Probability differences are visually represented using a color gradient. Overall, the central mountainous region of the study area exhibits high-probability clustering, reflecting the significant potential risk of landslide occurrence in areas where terrain slope, soil–rock properties, and rainfall interactions are pronounced. To validate the probability model, actual landslide locations were overlaid on the probability map (Fig. 12b). It is evident that the vast majority of landslide points are concentrated in areas with higher probability values, indicating a strong spatial correspondence between predicted high-probability zones and actual landslide occurrences. This demonstrates the model's effectiveness in capturing potential hazard-prone areas for rainfall-induced landslides. The probability values were further classified into five hazard levels using an equal-interval method with an interval of 0.2. Figure 12c presents the spatial distribution of these hazard levels, distinguished by different colors. Very high-risk zones are primarily concentrated in the core mountainous region of the study area, high-risk zones form strip-like belts surrounding the very high-risk areas, and moderate- to low-risk zones occupy the peripheral regions. When actual landslide occurrences are overlaid on the classified hazard map (Fig. 12d), landslide points are predominantly located within the “Very High” and “High” risk zones, further confirming the rationality of the classification results. Only a very small fraction of landslide points appears in moderate- or low-risk areas, which does not compromise the overall effectiveness of the classification model.

### SHAP value analysis

In this study, SHAP analysis was conducted on the test dataset using the XGBoost model within the AutoGluon framework. As shown in Fig. 13a, Accumulated Rainfall contributes most significantly to the landslide model, followed by Elevation and Slope. The bee swarm plot illustrates how different values of each factor influence the model output. High values of accumulated rainfall exhibit a strong positive contribution, indicating that heavy rainfall is a primary trigger for landslides. Overall, among the categories of terrain, anthropogenic, hydrological, and geological factors, hydrological factors (Accumulated Rainfall, Distance to River) and terrain factors (Elevation, Slope, etc.) are identified as the key drivers. Figure 13b depicts the influence paths of factor combinations on model output for individual samples, with each line representing one sample. The convergence and divergence of lines visually demonstrate how multiple factors jointly affect the predicted landslide probability. For instance, the combination of Accumulated Rainfall and Elevation amplifies the model output range, clearly revealing the mechanism by which multiple factors synergistically influence landslide hazard. Figure 13c presents detailed SHAP value variations for individual factors. Taking Accumulated Rainfall as an example, SHAP values increase sharply once rainfall exceeds a significant threshold, indicating a critical point beyond which rainfall strongly triggers landslides. Elevation shows positive and fluctuating SHAP values in the mid-to-high range, reflecting the uncertainty of landslide risk in mid-to-high altitude areas. The trends in SHAP values are consistent with previous landslide pattern analyses, indirectly validating the model's reliability and alignment with realistic landslide development processes. Notably, the relationship between SHAP values and factor values is often nonlinear. For example, the Topographic Wetness Index positively contributes in the mid-range but shows reduced influence at very low or high values. Vegetation coverage exhibits both positive and



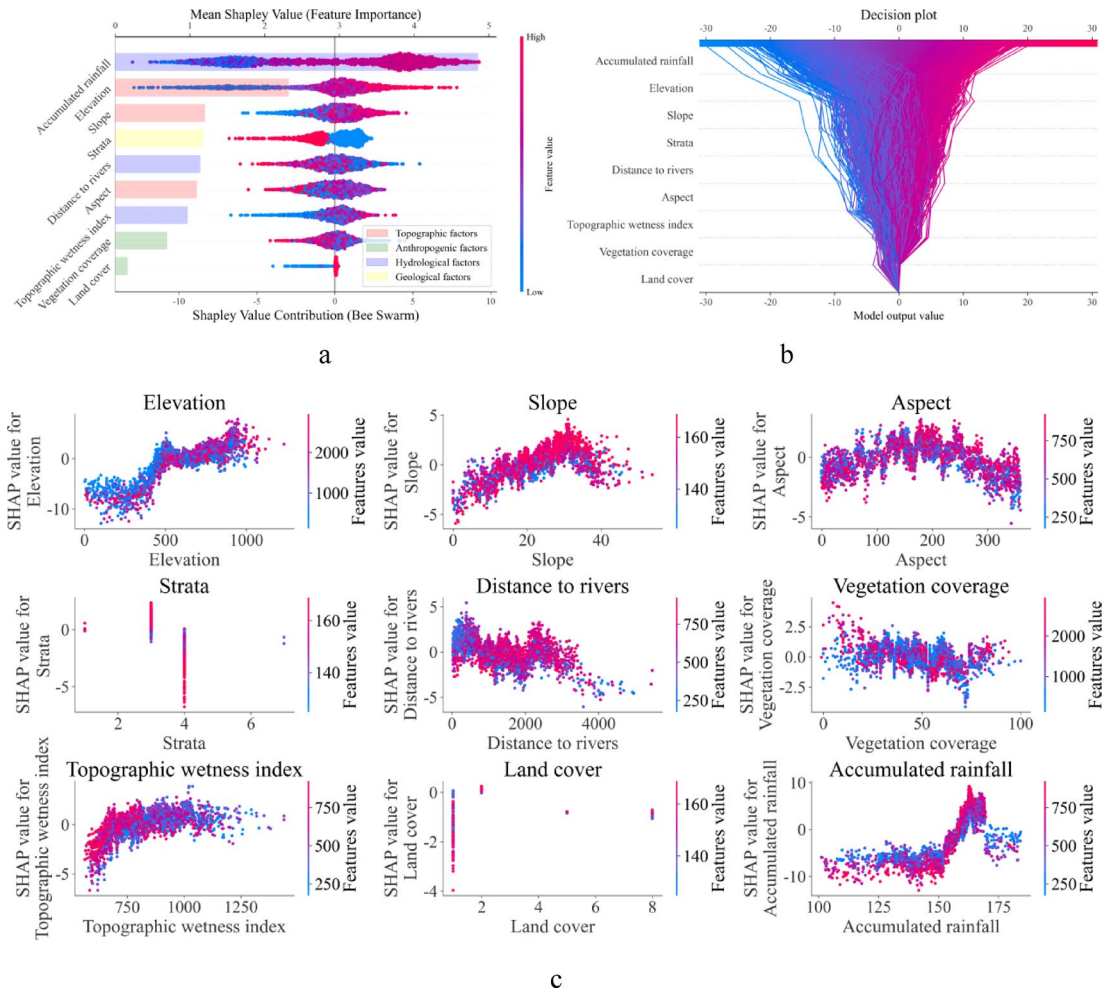
**Fig. 12.** Probability and classified maps of rainfall-induced landslide hazard. (a) Hazard probability map of rainfall-induced landslides; (b) Hazard probability map overlaid with landslide occurrences; (c) Classified hazard map of rainfall-induced landslides; (d) Classified hazard map overlaid with landslide occurrences.

negative contributions, indicating that the stabilizing or destabilizing effect of vegetation on slopes depends on the degree of coverage and its interaction with regional environmental conditions.

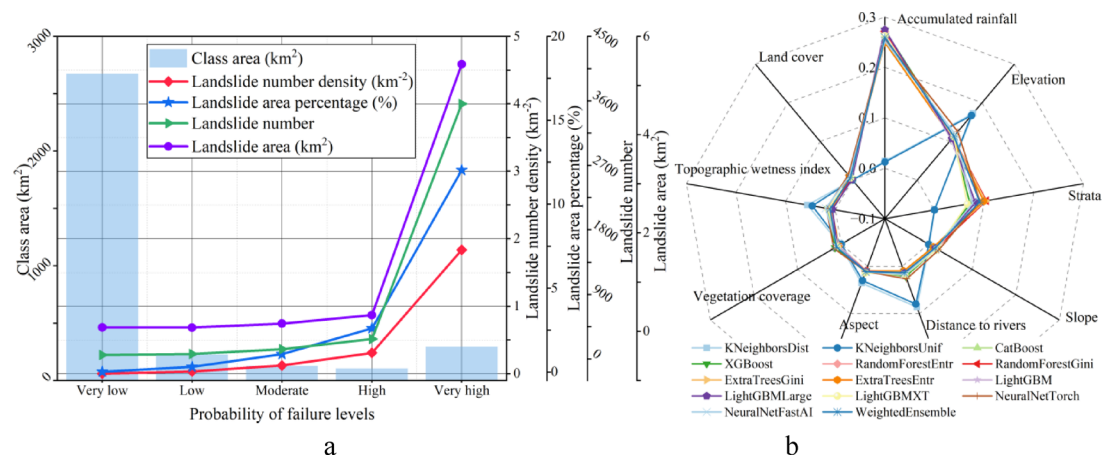
## Discussion

### Analysis of assessment results

In this study, the distribution patterns of landslide hazard across different classification levels were statistically analyzed, including class area, landslide number density (LND), landslide area percentage (LAP), landslide number, and landslide area. As shown in Fig. 14a, the areas corresponding to the “Very Low,” “Low,” and “Moderate” classes account for the majority of the study area, whereas the “High” and “Very High” classes are spatially concentrated, indicating that most regions exhibit relatively low landslide hazard. As the hazard level increases, both LND and LAP progressively rise. In the “Very High” class, all five statistical indicators reach their maximum values: 296.04 km<sup>2</sup> (class area), 1.84 km<sup>-2</sup> (LND), 12.02% (LAP), 3,559 (landslide number), and 5.44 km<sup>2</sup> (landslide area). In the “High” class, LND and LAP attain the second-highest values at 106.71 km<sup>2</sup>, 0.31 km<sup>-2</sup>, 2.61%, 279, and 0.33 km<sup>2</sup>, respectively. These results indicate that the core areas of frequent landslides identified in the model align well with actual hazard zones, confirming the effectiveness of the hazard assessment model. Figure 14b presents the feature importance ranking obtained from an ensemble analysis of 14 models on the AutoGluon platform. The cross-validation of multiple models effectively reduces the bias inherent in single-model evaluations. Among the models, 12 assign a weight greater than 0.25 to Accumulated Rainfall, and all 14 models assign a weight greater than 0.1 to Elevation. The importance of Accumulated Rainfall, Elevation, and Slope consistently ranks high across multiple models, indicating that hydrological and terrain factors are the primary drivers of rainfall-induced landslides. Although Strata, Vegetation Coverage, and Land Cover rank lower in importance, they still play a non-negligible role in influencing landslide occurrence.



**Fig. 13.** SHAP analysis of the XGBoost model trained within the AutoGluon framework. (a) Bee swarm plot showing mean Shapley values and individual Shapley value contributions; (b) Decision plot illustrating the influence paths of feature combinations on model output; (c) Subplots showing the relationship between SHAP values and feature values for each factor.



**Fig. 14.** Distribution patterns of different hazard classification levels and feature importance ranking. (a) Distribution patterns of landslide hazard across classification levels; (b) Feature importance ranking of nine landslide-influencing factors based on 14 models in AutoGluon.

## Application prospects and limitations

Landslides triggered by typhoon-induced rainfall often exhibit sudden onset, rapid disaster formation, and pronounced chain effects. Systematic assessment of regional landslide hazards under typhoon rainfall can identify spatial patterns of high-frequency and high-risk zones. Prior to a typhoon, hazard assessment results can be used to delineate potential danger areas in advance. During the typhoon event, combining real-time monitoring data with hazard assessment models enables rapid identification of high-risk landslide zones, improving emergency response efficiency and reducing potential losses<sup>36,37</sup>. Coupling landslide hazard assessment results with the spatial distribution and value density of exposed elements allows for a direct understanding of the risk levels faced by different areas and types of assets, providing a scientific basis for targeted mitigation measures<sup>38,39</sup>. Based on the risk levels of exposed elements, preventive engineering measures can be strategically deployed, such as installing landslide warning signs and constructing check dams near mountain settlements, as well as reinforcing slopes along key transportation corridors<sup>40,41</sup>. With the development of remote sensing, the Internet of Things (IoT), and meteorological radar technologies, the volume of landslide-related data has grown exponentially. Large-scale models can efficiently integrate these multi-source heterogeneous datasets, uncover complex interrelations, and enhance the accuracy and timeliness of landslide hazard assessment<sup>42,43</sup>. Furthermore, such models can be deeply integrated with geographic information systems (GIS) and decision support systems (DSS), not only generating hazard assessment results but also automatically producing multiple mitigation and disaster response plans based on the assessment outcomes (Wang et al.<sup>44</sup>; Xie et al.<sup>45</sup>). The effectiveness of these plans can be simulated and predicted, assisting decision-makers in selecting optimal strategies and advancing landslide hazard assessment from single predictions toward intelligent decision-making. However, at the current stage, landslide risk assessment frameworks still exhibit several critical limitations. The accuracy of assessment outcomes is largely constrained by the spatial and temporal resolution of rainfall data, particularly during extreme typhoon-induced localized precipitation events, where existing monitoring systems often fail to fully and precisely capture rainfall dynamics. Moreover, the completeness and quality of data on key exposed elements, including infrastructure distribution and socio-economic assets, constitute additional sources of uncertainty that may undermine the reliability of risk assessment results.

## Conclusion

This study focuses on Taishun County, Zhejiang Province, and its surrounding areas affected by Typhoon Meranti in September 2016. By integrating multi-source datasets, constructing a detailed landslide inventory, applying machine learning-based modeling, and interpreting model outputs using SHAP values, a systematic assessment of regional landslide risk under typhoon-induced rainfall conditions was conducted. The main conclusions are summarized as follows:

(1) Spatial clustering and size distribution of landslides. A total of 4,102 rainfall-induced landslides were identified using human-computer interactive interpretation, covering a total area of 6.09 km<sup>2</sup> with an average area of 1,484.15 m<sup>2</sup>. Landslides are densely distributed in the northwestern to southeastern mountainous belt of the central study area. Landslide runout distance shows a power-law relationship with both landslide area and elevation difference, with correlation coefficients of 0.911 and 0.756, respectively, indicating a close association between landslide size and mobility.

(2) Multi-factor control and identification of key drivers. Landslide occurrence is governed by multiple factors, with Accumulated Rainfall (high-frequency interval: 160~180 mm), Elevation (400~800 m), and Slope (25°~40°) identified as the most critical drivers. Secondary factors include strata type, distance to rivers, and others. South- and southeast-facing slopes, moderate vegetation coverage (40~70%), Cretaceous strata, and medium-humidity topographic conditions exhibit significantly elevated landslide probabilities, highlighting the synergistic influence of topography, geology, hydrology, and rainfall.

(3) Performance of ensemble learning models in hazard assessment. Ensemble learning models demonstrated excellent capability for landslide hazard evaluation. Within the AutoGluon framework, the WeightedEnsemble model achieved the best performance, with AUC values of 0.9985 and 0.9693 for the training and testing datasets, respectively. High-risk zones are concentrated in the central mountainous area, showing good spatial correspondence with actual landslide occurrences, which validates the model's reliability and practical applicability.

(4) Insights from SHAP analysis on factor mechanisms and critical thresholds. SHAP analysis revealed the contribution of each factor to model output, with Accumulated Rainfall exerting the largest influence. High rainfall values trigger landslides by reducing the shear strength of soil and rock. SHAP results further confirmed that Accumulated Rainfall, Elevation, and Slope are the core drivers of landslide occurrence. Many factors exhibit non-linear relationships with landslide hazard and possess distinct critical thresholds; for example, SHAP values increase sharply when accumulated rainfall exceeds a specific threshold, providing quantitative guidance for landslide early-warning criteria.

(5) Although this study focuses on a typical typhoon-affected region in southeastern China, the proposed landslide hazard assessment framework is not region-specific. By integrating multi-source environmental data, ensemble learning models, and interpretable analysis techniques, the framework can be readily adapted to other typhoon-prone mountainous regions with similar climatic and geomorphological settings. With appropriate adjustments to input data and local conditioning factors, the framework is scalable across different spatial extents and transferable to other regions experiencing frequent extreme rainfall events, thereby offering broad potential for regional landslide risk assessment and disaster mitigation.

## Data availability

The datasets used or analysed during the current study are available from the corresponding author upon reasonable request.

Received: 10 November 2025; Accepted: 24 March 2026

Published online: 27 March 2026

## References

- Cui, Y., Yang, L., Xu, C. & Zheng, J. Spatial distribution of shallow landslides caused by Typhoon Lekima in 2019 in Zhejiang Province, China. *J. Mt. Sci.* **21** (5), 1564–1580. <https://doi.org/10.1007/s11629-023-8377-y> (2024).
- Zeng, T. et al. Double-index rainfall warning and probabilistic physically based model for fast-moving landslide hazard analysis in subtropical-typhoon area. *Landslides* **21** (4), 753–773. <https://doi.org/10.1007/s10346-023-02187-4> (2024).
- Zhuang, Y. et al. Typhoon, rainfall and trees jointly cause landslides in coastal regions. *Eng. Geol.* **298**, 106561. <https://doi.org/10.1016/j.enggeo.2022.106561> (2022).
- Zhuang, Y., Xing, A., Sun, Q. & Jiang, Y. Insights into initiation of typhoon-induced deep-seated landslides in Southeast coastal China. *Nat. Hazards*. **119** (1), 721–749. <https://doi.org/10.1007/s11069-023-06138-z> (2023).
- Shao, X. et al. Landslides triggered by the 2022 Ms. 6.8 Luding strike-slip earthquake: An update. *Eng. Geol.* **335**, 107536. <https://doi.org/10.1016/j.enggeo.2024.107536> (2024).
- Wang, F. et al. An improved monitoring strategy for shallow rainfall-induced landslides under critical site-specific recognition. *Landslides* **21** (10), 2569–2581. <https://doi.org/10.1007/s10346-024-02298-6> (2024b).
- Zhang, M. et al. Field model experiments to determine mechanisms of rainstorm-induced shallow landslides in the Feiyunjiang River basin, China. *Eng. Geol.* **262**, 105348. <https://doi.org/10.1016/j.enggeo.2019.105348> (2019).
- Nie, W., Tian, C., Song, D., Liu, X. & Wang, E. Disaster process and multisource information monitoring and warning method for rainfall-triggered landslide: a case study in the southeastern coastal area of China. *Nat. Hazards*. **121** (3), 2535–2564. <https://doi.org/10.1007/s11069-024-06897-3> (2025).
- Xie, C. et al. Advances in the study of natural disasters induced by the 23.7 extreme rainfall event in North China. *Nat. Hazards Res.* **5** (1), 1–13. <https://doi.org/10.1016/j.nhres.2025.01.003> (2025c).
- Chen, W. et al. Exploring advanced machine learning techniques for landslide susceptibility mapping in Yanchuan County, China. *Earth Sci. Inf.* **17** (6), 5385–5402. <https://doi.org/10.1007/s12145-024-01455-8> (2024).
- Chowdhuri, I. et al. Ensemble approach to develop landslide susceptibility map in landslide dominated Sikkim Himalayan region, India. *Environ. Earth Sci.* **79** (20), 476. <https://doi.org/10.1007/s12665-020-09227-5> (2020).
- Xie, C., Huang, Y., Li, L., Li, T. & Xu, C. Detailed inventory and spatial distribution analysis of rainfall-induced landslides in Jiexi County, Guangdong Province, China in August 2018. *Sustainability* **15** (18), 13930. <https://doi.org/10.3390/su151813930> (2023).
- Zhang, Y. et al. Effects of combination of influencing factors on earthquake-induced landslide susceptibility assessments. *Environ. Earth Sci.* **84** (7), 189. <https://doi.org/10.1007/s12665-025-12203-6> (2025).
- Huang, F. et al. Uncertainties in landslide susceptibility prediction modeling: A review on the incompleteness of landslide inventory and its influence rules. *Geosci. Front.* **15** (6), 101886. <https://doi.org/10.1016/j.gsf.2024.101886> (2024a).
- Huang, F. et al. Modelling landslide susceptibility prediction: A review and construction of semi-supervised imbalanced theory. *Earth Sci. Rev.* **250**, 104700. <https://doi.org/10.1016/j.earscirev.2024.104700> (2024b).
- Lin, Q. et al. National-scale data-driven rainfall induced landslide susceptibility mapping for China by accounting for incomplete landslide data. *Geosci. Front.* **12** (6), 101248. <https://doi.org/10.1016/j.gsf.2021.101248> (2021).
- Xu, C., Dai, F., Xu, X. & Lee, Y. H. GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China. *Geomorphology* **145**, 70–80. <https://doi.org/10.1016/j.geomorph.2011.12.040> (2012a).
- Wang, J., Zhu, H. H., Zhang, W., Tan, D. Y. & Pasuto, A. Enhancing landslide displacement prediction using a spatio-temporal deep learning model with interpretable features. *J. Geophys. Research: Mach. Learn. Comput.* **2** (3), e2025JH000592. <https://doi.org/10.1029/2025JH000592> (2025).
- Xu, C., Xu, X., Dai, F. & Saraf, A. K. Comparison of different models for susceptibility mapping of earthquake triggered landslides related with the 2008 Wenchuan earthquake in China. *Comput. Geosci.* **46**, 317–329. <https://doi.org/10.1016/j.cageo.2012.01.002> (2012b).
- Zhang, J. et al. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. *J. Environ. Manage.* **332**, 117357. <https://doi.org/10.1016/j.jenvman.2023.117357> (2023a).
- Zhou, X., Wen, H., Li, Z., Zhang, H. & Zhang, W. An interpretable model for the susceptibility of rainfall-induced shallow landslides based on SHAP and XGBoost. *Geocarto Int.* **37** (26), 13419–13450. <https://doi.org/10.1080/10106049.2022.2076928> (2022).
- Fu, B., Liang, X., Wen, X., Zhang, S. & Cheng, Q. Construction of an urban collaborative emergency response network based on a disaster chain: an example of typhoon disaster in Fujian Province. *Nat. Hazards*. **121**, 19329–19354. <https://doi.org/10.1007/s11069-025-07575-8> (2025).
- Yang, X., Qin, X., Zhou, X., Chen, Y. & Gao, L. Assessment of disaster mitigation capability oriented to typhoon disaster chains: A case study of Fujian Province, China. *Ecol. Ind.* **167**, 112621. <https://doi.org/10.1016/j.ecolind.2024.112621> (2024).
- Yang, X. et al. Risk of Compound Typhoon Disaster Chains: Insights from Southeastern China. *Int. J. Disaster Risk Sci.* **16**, 870–887. <https://doi.org/10.1007/s13753-025-00674-x> (2025).
- Li, C. et al. China National Digital Geological Map (Public Version at 1:200 000 Scale) Spatial Database. *Geol. CHINA*. **46** (S1), 1–10. <https://doi.org/10.12029/gc2019Z101> (2019).
- Tateishi, R. et al. Production of global land cover data—GLCNMO. *Int. J. Digit. Earth.* **4** (1), 22–49. <https://doi.org/10.1080/17538941003777521> (2011).
- Xie, C. et al. Detailed inventory and initial analysis of landslides triggered by extreme rainfall in the northern Huaiji County, Guangdong Province, China, from June 6 to 9, 2020. *Geoenvironmental Disasters*. **12** (1), 7. <https://doi.org/10.1186/s40677-025-00311-1> (2025b).
- Xu, C. Preparation of earthquake-triggered landslide inventory maps using remote sensing and GIS technologies: Principles and case studies. *Geosci. Front.* **6** (6), 825–836. <https://doi.org/10.1016/j.gsf.2014.03.004> (2015).
- Xu, C., Xu, X., Yao, X. & Dai, F. Three (nearly) complete inventories of landslides triggered by the May 12, 2008 Wenchuan Mw 7.9 earthquake of China and their spatial distribution statistical analysis. *Landslides* **11** (3), 441–461. <https://doi.org/10.1007/s10346-013-0404-6> (2014).
- Xie, C., Huang, Y., Xu, C., Dai, K. & Xu, X. Over 100,000 landslides triggered by typhoon-induced rainfall in North China in July 2023 1–20 (Landslides, 2026). <https://doi.org/10.1007/s10346-026-02698-w>
- Xie, C., Xu, C., Xu, X., Huang, Y. & Gao, H. Analysis of a comprehensive inventory of rainfall-induced landslides and multi-factor coupling mechanisms in Eastern Guangdong, China, in August 2018. *Environmental Earth Sciences*, **84**(24), 707. (2025d). <https://doi.org/10.1007/s12665-025-12726-y>
- Li, X., Zhang, Y., Ji, X., Strauss, P. & Zhang, Z. Effects of shrub-grass cover on the hillslope overland flow and soil erosion under simulated rainfall. *Environ. Res.* **214**, 113774. <https://doi.org/10.1016/j.envres.2022.113774> (2022).

33. Wang, F. et al. Effects of rainfall patterns, vegetation cover types and antecedent soil moisture on run-off and soil loss of typical Luvisol in southern China. *Earth. Surf. Proc. Land.* **49** (10), 2998–3012. <https://doi.org/10.1002/esp.5871> (2024c).
34. Jiang, H. et al. Probability analysis of shallow landslides in varying vegetation zones with random soil grain-size distribution. *Environ. Model. Softw.* **183**, 106267. <https://doi.org/10.1016/j.envsoft.2024.106267> (2025).
35. Zhu, J. et al. Soil moisture and hysteresis affect both magnitude and efficiency of root reinforcement. *Catena* **219**, 106574. <https://doi.org/10.1016/j.catena.2022.106574> (2022).
36. Dai, X., Chen, J., Zhang, T. & Xue, C. Integrated Landslide Risk Assessment via a Landslide Susceptibility Model Based on Intelligent Optimization Algorithms. *Remote Sens.* **17** (3), 545. <https://doi.org/10.3390/rs17030545> (2025).
37. Marín-Rodríguez, N. J., Vega, J., Zanabria, O. B., González-Ruiz, J. D. & Botero, S. Towards an understanding of landslide risk assessment and its economic losses: a scientometric analysis. *Landslides* **21** (8), 1865–1881. <https://doi.org/10.1007/s10346-024-02272-2> (2024).
38. Xie, C. et al. Comprehensive analysis and assessment of extreme rainfall-induced clustered landslides: a case study of southern Qingyuan city, Guangdong province, China, in June 2020. *Georisk: Assess. Manage. Risk Eng. Syst. Geohazards.* **1**, 1–23. <https://doi.org/10.1080/17499518.2025.2563759> (2025).
39. Yin, D. B., Zheng, Q., Zhou, A. & Shen, S. L. Enhancing landslide hazard prevention: Mapping vulnerability via considering the effects of human factors. *Int. J. Disaster Risk Reduct.* **108**, 104509. <https://doi.org/10.1016/j.ijdrr.2024.104509> (2024).
40. Singh, G., Kumar, S., Karmakar, R. & Mishra, A. K. Rapid assessment of landslide exposure to elements at risk for decision support in regional landslide forecasting. *Discover Geoscience.* **3** (1), 181. <https://doi.org/10.1007/s44288-025-00299-z> (2025).
41. Zhang, Y., Ayyub, B. M., Gong, W. & Tang, H. Risk assessment of roadway networks exposed to landslides in mountainous regions—A case study in Fengjie County, China. *Landslides* **20** (7), 1419–1431. <https://doi.org/10.1007/s10346-023-02045-3> (2023b).
42. Xu, Y. et al. Cas landslide dataset: A large-scale and multisensor dataset for deep learning-based landslide detection. *Sci. Data.* **11**(1), 12. doi:<https://doi.org/10.1038/s41597-023-02847-z> (2024).
43. Zhan, L., Li, W. & Min, W. FA-ResNet: Feature affine residual network for large-scale point cloud segmentation. *Int. J. Appl. Earth Obs. Geoinf.* **118**, 103259. <https://doi.org/10.1016/j.jag.2023.103259> (2023).
44. Wang, C. et al. Near-real-time earthquake-induced fatality estimation using crowdsourced data and large-language models. *Int. J. Disaster Risk Reduct.* **111**, 104680. <https://doi.org/10.1016/j.ijdrr.2024.104680> (2024a).
45. Xie, C. et al. Leveraging the DeepSeek large model: A framework for AI-assisted disaster prevention, mitigation, and emergency response systems. *Earthq. Res. Adv.* **5** (4), 100378. <https://doi.org/10.1016/j.eqrea.2025.100378> (2025a).
46. Gupta, S. K. & Shukla, D. P. Handling data imbalance in machine learning based landslide susceptibility mapping: a case study of Mandakini River Basin, North-Western Himalayas. *Landslides* **20** (5), 933–949. <https://doi.org/10.1007/s10346-022-01998-1> (2023).
47. Sajwan, A. & Ramana, G. Integrating geoinformatics and numerical modelling for landslide back-analysis and forecasting: a proactive mitigation study of the Shiv Bawri landslide. *Landslides* **21** (10), 2481–2499. <https://doi.org/10.1007/s10346-024-02321-w> (2024).
48. Yang, J. & Huang, X. 30 m annual land cover and its dynamics in China from 1990 to 2019. *Earth Syst. Sci. Data Discuss.* **13** (8), 3907–3925. <https://doi.org/10.5194/essd-13-3907-2021> (2021).

## Author contributions

Chenchen Xie: Data curation; Writing—original draft; Writing—review and editing; Validation. Chong Xu: Data curation; Methodology; Conceptualization; Writing—review and editing; Supervision; Formal analysis. Xiwei Xu: Conceptualization; Data curation; Supervision. Yu Huang: Data curation. Yinke Li: Data curation. Kejie Chen: Data curation. Zhiwen Zheng: Data curation. Wei Zhang: Data curation. Saier Wu: Data curation.

## Funding

This research was funded by the South China Field Scientific Observation and Research Station for Climate-Driven Landslide Risk, Ministry of Natural Resources (2025HNF003) and the National Key Research and Development Program of China (2024YFC3012604).

## Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to C.X.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://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) 2026