



# OPEN A new framework for landslide susceptibility mapping in contiguous impoverished areas using machine learning and catastrophe theory

Wei Zhou<sup>1</sup>, Yingzhi Zhou<sup>3,5</sup>, Shuneng Liang<sup>4</sup>✉, Chengnian Zhang<sup>1</sup>, Hongzhou Dai<sup>3</sup> & Xiaofei Sun<sup>2,3</sup>✉

Landslides are among the most frequent and dangerous geological disasters worldwide, making accurate landslide susceptibility mapping (LSM) crucial for effective disaster prevention. This study introduces a novel LSM framework by integrating random forest (RF), support vector machine (SVM), and catastrophe theory (CT), and applies it to the contiguous impoverished areas of Liangshan, Sichuan. First, we selected 12 factors representing both internal environmental and external triggering conditions to assess landslide susceptibility. The frequency ratio method was used to assess the correlation between historical landslides and these factors. Second, CT was integrated into the RF- and SVM-based LSM models, resulting in two integrated models (RF-CT and SVM-CT) for generating LSM in the region. Finally, the receiver operating characteristic curve was used to evaluate and compare the accuracy of the methods. The results show that the RF-CT and SVM-CT frameworks performed well, with a 10% improvement in the success rate (0.899 for RF-CT and 0.873 for SVM-CT), and a 5% improvement in the prediction rate (0.783 for RF-CT and 0.775 for SVM-CT) compared with the individual RF and SVM models. These findings provide valuable insights for disaster prevention, poverty alleviation, and sustainable development in the study area.

**Keywords** Machine learning, Cusp catastrophe model, Spatial modeling, Landslide susceptibility, ROC curve

A landslide is the downward movement of rocks, soil, and organic material along a slope under the influence of gravity<sup>1</sup>. As a destructive natural disaster, landslides cause extensive damage to vegetation, infrastructure, and property, often resulting in loss of life<sup>2</sup>. China experiences one of the highest frequencies of landslides globally, particularly in mountainous regions where they hinder resource development and economic growth<sup>3,4</sup>. The contiguous impoverished areas of Liangshan, Sichuan (CIALS), serve as the core of the Jinsha River Hydropower Base and are a critical region for geological disaster early warning in the Yangtze River Basin. Since the reservoir's construction and filling, significant changes in the area's environmental and geological conditions have led to an increase in both the scale and frequency of landslides. These frequent events threaten homes, public infrastructure, and agricultural production, causing casualties and substantial property losses<sup>5</sup>. Post-disaster recovery often requires significant financial resources, but the region's fragile economic foundation hampers effective reconstruction, exacerbating poverty in the CIALS. High-precision landslide susceptibility mapping (LSM) is essential for prioritizing governance and prevention measures in high-risk areas. It can help reduce the impact of landslides, enhance disaster resilience, and promote sustainable development<sup>6</sup>. Accurately LSM has thus become a critical priority for disaster management and regional economic development in CIALS.

<sup>1</sup>The Fourth Geological Brigade of Jiangxi Geological Bureau, Pingxiang 337000, China. <sup>2</sup>State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu 610059, China. <sup>3</sup>College of Geography and Planning, Chengdu University of Technology, Chengdu 610059, China. <sup>4</sup>Land Satellite Remote Sensing Application Center, Ministry of Natural Resources of China, Beijing 100048, China. <sup>5</sup>Forest and Grassland Fire Monitoring Center of Sichuan Province, Sichuan Forestry and Grassland Bureau, Chengdu 610081, China. ✉email: liangsn83@163.com; sunxiaofei@cdut.edu.cn

LSM is an effective approach for reducing and mitigating future landslide-related losses<sup>7</sup>. It estimates the spatial probability of landslide occurrences, helping to identify high-risk areas. This knowledge is essential for understanding the causes of landslides, predicting future events, and providing a scientific basis for land use planning, major project site selection, and disaster prevention<sup>8</sup>. With advancements in Geographic information system (GIS) and Remote sensing (RS) technologies, large and complex spatial data can now be processed systematically and efficiently, providing valuable support for large-scale LSM<sup>9,10</sup>. Researchers have employed various data and methods for LSM, which generally fall into two categories: subjective weighting methods and mathematical model evaluations. Subjective weighting methods, while useful in specific regions, rely heavily on researchers' knowledge, making them less generalizable<sup>11</sup>. These methods also struggle to handle high-dimensional, non-normal, and nonlinear data, limiting their adaptability<sup>12</sup>. Examples include the Delphi method and the Analytic hierarchy process. In contrast, mathematical model-based approaches, particularly those using machine learning (ML), are more objective, replicable, and adaptable to complex data. This has led to their increasing use in LSM. For example, Ngo et al.<sup>13</sup> applied Recurrent neural networks and Convolutional neural networks to create a national-scale LSM for Iran, identifying key factors such as slope, geology, land use, and proximity to faults as primary drivers of landslides. Youssef et al.<sup>14</sup> compared several advanced ML methods for LSM and found that Random forest (RF) and Linear discriminant analysis outperformed others. Zhang et al.<sup>15</sup> developed an LSM model for mountainous areas using the SHapley additive interpretation and eXtreme gradient boosting algorithm, demonstrating better generalization for regions with similar characteristics. These studies have provided scientifically grounded insights into the spatial probability and dynamics of landslide occurrences, aiding prevention efforts. However, the choice of mapping method should be based on the specific characteristics of the study area and the selected evaluation criteria, as the generalization abilities of models can vary regionally.

Recently, mixed and integrated ML models have gained favor among researchers for their superior performance<sup>16–18</sup>. Support vector machine (SVM) is well-suited for handling small sample sizes and providing robust explanations, particularly due to its capacity to process high-dimensional data, generalize effectively, and tackle non-linear problems<sup>19</sup>. RF is advantageous because it incorporates multiple input factors without excluding any, leading to improved classification outcomes and higher predictive accuracy<sup>20</sup>. As a branch of nonlinear theory, catastrophe theory (CT) provides an objective approach to address discontinuous changes in natural and social systems, reducing human bias in decision-making and enhancing model generalization<sup>21</sup>. CT focuses on the phenomena and laws of transition from one stable state to another stable state. It effectively explains sudden phenomena in nature and has wide applications in physics, chemistry, biology, and engineering<sup>22</sup>. CT has been extensively applied in groundwater potential mapping<sup>23</sup>, environmental carrying capacity analysis<sup>24</sup>, and flood and drought risk assessments<sup>25</sup>. This demonstrates the feasibility of applying CT to landslide susceptibility mapping. Therefore, this study introduced CT when using SVM and RF models for LSM, proposing a new LSM framework and conducting a comparative analysis of its performance. Ultimately, the study aims to generate a reliable landslide occurrence probability distribution map for the CIALS. This map will provide a scientific foundation for comprehensive geological hazard prevention and control, while also supporting the region's targeted poverty alleviation and ecological protection strategies of the Yangtze River Economic Belt.

## Study area

The CIALS is located at 102°12'–103°53' E and 26°50'–28°53' N, covering approximately 7,800 square kilometers (Fig. 1). Situated in the transitional zone between the Sichuan Basin and the Yunnan-Guizhou Plateau, the region is deeply cut and eroded by the Jinsha River system. The topography slopes from west to east, with elevations ranging from 4,756 m to 301 m. The region experiences an average annual temperature of 11 °C and receives 500–1000 millimeters of rainfall annually. Its geological structure is complex, and the lithology primarily consists of soft strata and interlayered sandy mudstone from the Cretaceous, Jurassic, and Triassic periods of the Mesozoic Era. According to detailed geological hazard surveys in Sichuan Province, the CIALS contains 471 landslides, including 2 extra-large, 21 large, 112 medium, and 336 small ones.

## Data source and conditioning factors

### Data preparation and data processing

The data collected for this study include historical landslide inventories, Landsat images, DEM, rainfall data, geological maps, population distribution, and other relevant datasets (Table 1).

The data collection and processing methods are as follows: (1) In 2016, the Sichuan Institute of Geological Survey conducted a comprehensive investigation of landslides in the CIALS. The survey included remote sensing interpretation, field investigation, and data compilation. During the remote sensing phase, high-resolution images and DEMs were used to identify landslides based on key geomorphological features such as flow tracks, head scarps on open slopes, and exposed soil in forests. In the field, landslide data were collected and verified using handheld GPS devices and drones. These landslides were then digitized in ArcGIS software, with attributes such as location and information on threatened buildings, populations, and property. (2) Using the Google Earth Engine platform, Landsat images are employed to obtain the Normalized Difference Vegetation Index (NDVI) of the CIALS. To minimize the impacts of snow, shadows, and clouds, the maximum value synthesis method<sup>26</sup> is applied, synthesizing the NDVI data from 2022 to 2024 and yielding the maximum NDVI values during this period. Additionally, land use data is derived from Landsat images using the SVM model, providing information on the distribution of human engineering activities in the CIALS. (3) DEM data is used to extract elevation, slope, plan curvature, relief degree, and water system data. (4) Rainfall data is interpolated into spatial data using Kriging interpolation<sup>27</sup>. (5) Population density is estimated based on night light data, land use data, and township population statistics<sup>28</sup>. (6) Peak ground acceleration (PGA) from earthquakes and geological data

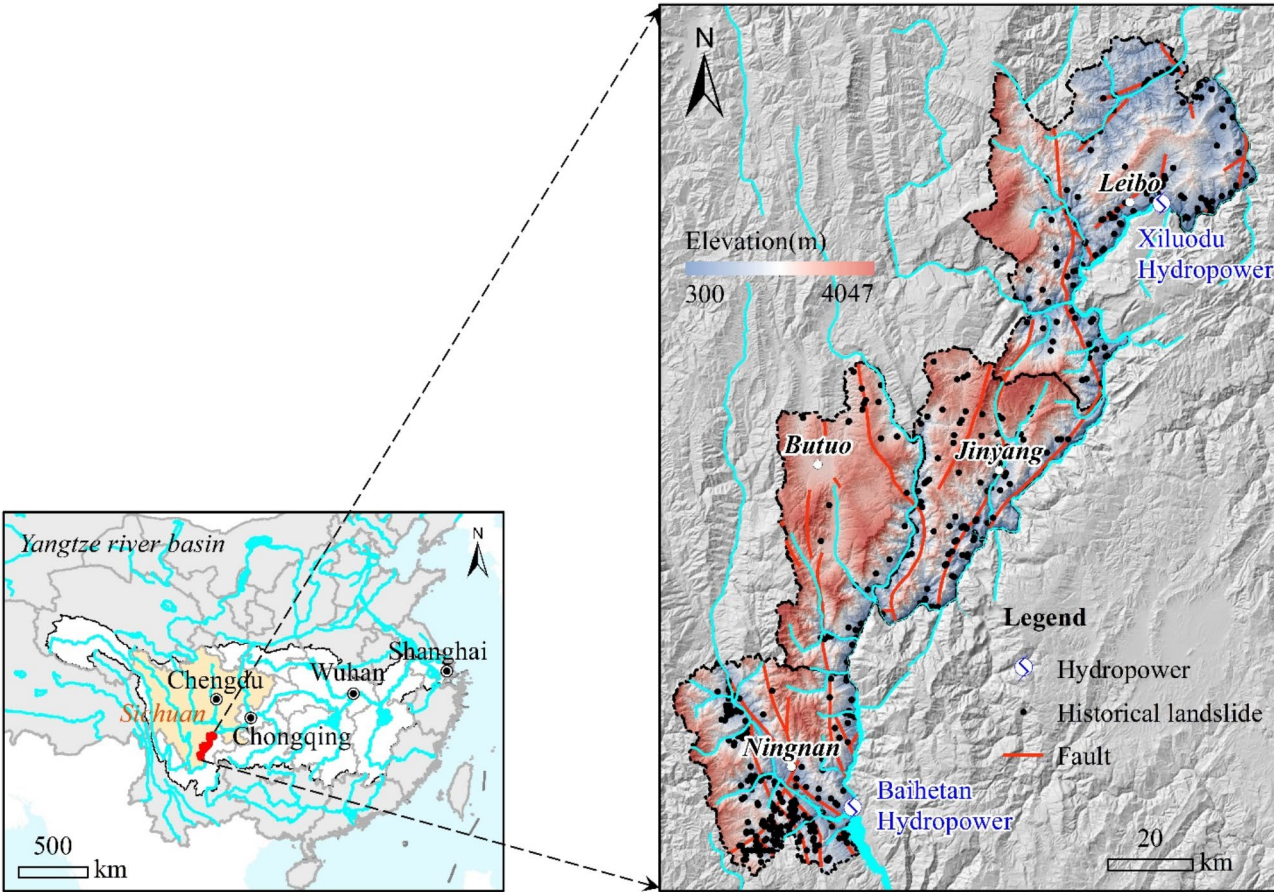


Fig. 1. Location map of the study area.

| Data type                      | Source                                                              | annotations                                                                      |
|--------------------------------|---------------------------------------------------------------------|----------------------------------------------------------------------------------|
| Landsat                        | Google Earth Engine                                                 | Select images from 2022 to 2024 and extract the vegetation coverage of the CIAIS |
| DEM                            | <a href="http://www.gscloud.cn/">http://www.gscloud.cn/</a>         | SRTM DEM, with a spatial resolution of 30 m. A total of 5 scenes are involved    |
| Rainfall                       | <a href="http://data.cma.cn/">http://data.cma.cn/</a>               | The average annual precipitation year by year from 2015 to 2020                  |
| Geological data                | <a href="http://www.ngac.org.cn/">www.ngac.org.cn/</a>              | Geological map on a scale of 1:500,000                                           |
| PAG                            | <a href="https://www.cea.gov.cn/">https://www.cea.gov.cn/</a>       | Seismic ground motion peak acceleration zoning map of China                      |
| Demographic data               | <a href="https://www.hongheiku.com/">https://www.hongheiku.com/</a> | Data of the seventh national census of China                                     |
| Nightlight data                | <a href="https://www.ngdc.noaa.gov/">https://www.ngdc.noaa.gov/</a> | NPP-VIIRS nighttime light data                                                   |
| Historical landslide inventory | Detailed survey results of geological hazards in Sichuan Province   | Provided by Sichuan Institute of Geological Survey.                              |

Table 1. Data and sources required for this study.

are processed as vector data, requiring only coordinate transformation. To ensure spatial alignment, all data is projected using the Gauss-Kruger system with a spatial resolution of 100 m × 100 m.

Landslide conditioning factors

Landslides result from the interaction of multiple factors, including geology, topography, meteorology, and human activities<sup>29</sup>. Currently, there is no unified standard for selecting conditioning factors in LSM. Therefore, based on existing research<sup>30–33</sup> and available data, this study selects 12 conditioning factors from two categories: internal environmental factors and external triggering factors. 8 internal environmental factors were selected based on the following criteria: First, due to the significant altitude difference and deep-cut terrain of CIAIS, four topographical factors, namely altitude, slope angle, slope aspect, and plane curvature, were chosen. Second, given the complex and fragmented lithology of the area and the development of geological structures, which are major contributors to landslides, lithology and distance to faults were selected as geological factors. Third, because the area is severely eroded by the Jinsha River system and shows significant spatial variation in vegetation within the dry hot valley, distance to rivers and NDVI were selected. Additionally, 4 external triggering factors

were considered: First, frequent earthquakes in the area and surrounding areas are major causes of landslides, so PGA was selected to represent earthquake impact. Second, the concentrated and intense rainfall in the area is a key factor in triggering landslides. Third, there is a large population, most of whom live in river valleys. Large-scale human activities, such as the “Village-to-Village Project” and hydropower development, have induced numerous landslides. Therefore, population density and distance to engineering activities were selected.

When conducting LSM, it is of major importance to recognize the conditioning factors that can cause the landslide and the process that could trigger the movement. The frequency ratio (FR) method relies on the relationship between historical landslide distributions and various conditioning factors, revealing the correlation between landslide locations and these factors<sup>34</sup>. In this study, the FR method was used to analyze the spatial relationship between historical landslide locations and 12 conditioning factors, providing a basis for LSM modeling (Table 2).

#### *Altitude*

As gravity weakens with increasing altitude, loose materials at lower altitudes are more prone to collapse under gravitational force. Therefore, altitude plays a significant role in the distribution of geological hazards<sup>35</sup>. In this study, altitude was reclassified into 7 categories (Fig. 2a), each with 500-meter intervals, and it was found that the probability of landslides is highest below 2,000 m above sea level.

#### *Slope angle*

Slope angle plays a critical role in the formation, development, and susceptibility to landslides, and is widely recognized as a key input parameter in LSM by numerous studies<sup>36</sup>. In the CIALS, most landslides occur on slopes ranging between 10° and 30° (Fig. 2b).

#### *Slope aspect*

Slope aspect indirectly influences landslides and is a crucial factor in LSM<sup>37</sup>. In this study, slope aspect was reclassified into 9 categories (Fig. 2c), with most landslides concentrated in the east (67.5°–112.5°) and southeast (112.5°–157.5°) directions in the CIALS.

#### *Plan curvature*

Hillsides with concave plan curvature are more susceptible to landslides compared to those with convex curvature<sup>38</sup>. In this study, plan curvature values range from −49.89 to 53.78 and have been reclassified into 6 categories (Fig. 2d).

#### *Lithology*

Lithology significantly influences slope stability, serving as a key factor in LSM<sup>39</sup>. In the CIALS, lithology is classified into 5 categories: hard rocks, harder rocks, softer rocks, soft rocks, and loose deposits (Fig. 2e). Landslides are predominantly found in regions characterized by softer rocks.

#### *Distance to faults*

Fault zones can significantly compromise the integrity and continuity of rock masses, thereby diminishing slope stability<sup>40</sup>. Additionally, these zones often contain abundant groundwater and exhibit high levels of geological activity, which can further reduce the shear strength of the surrounding rock. Consequently, the fault buffer zone is categorized into 5 classifications (Fig. 2f), and a distance to faults map is created with intervals of 1,000 m.

#### *Distance to rivers*

River down cutting creates steep walls or cliffs, making these regions susceptible to landslides<sup>41</sup>. Therefore, this study establishes 5 river buffer zone categories at intervals of 200 m (Fig. 2g). The analysis reveals that the probability of landslides is significantly higher within the first 200 m from the river compared to other regions.

#### *NDVI*

Vegetation plays a crucial role in slope stability by influencing the hydrological and mechanical properties of slopes, thereby impacting landslide occurrences<sup>42</sup>. In the CIALS, the NDVI values range from −0.48 to 0.86 and are reclassified into 5 categories using the natural breaks method (Fig. 2h).

#### *PGA*

The CIALS is located at the junction of the Longnan-Mabian seismic belt and the Anning River Valley seismic belt in China. The PGA of earthquakes can reflect the seismic load (Fig. 2i), so it is the main external triggering factor<sup>43</sup>.

#### *10 rainfall*

Rainfall is a critical triggering factor for landslides, exerting both direct and indirect effects on their occurrence<sup>44</sup>. In this study, the annual average precipitation from 15 meteorological stations is utilized to create a rainfall map (Fig. 2j), which is then reclassified into 6 categories at 30-millimeter intervals.

#### *11 Population density*

Population density is a significant factor influencing the landslide occurrence<sup>45</sup>. Higher population density can impact the landscape and environment, potentially increasing landslide risk. Analysis of population density statistics and existing landslide distributions in the CIALS reveals a linear correlation between population density and the probability of landslide occurrence (Fig. 2k).

| Conditioning factors            |                        | Classes        | Percentage of landslides (%) | Number of landslides | Percentage of domain (%) | Area (km) | FR    |
|---------------------------------|------------------------|----------------|------------------------------|----------------------|--------------------------|-----------|-------|
| Intrinsic environmental factors | Altitude (m)           | < 1000         | 0.213                        | 70                   | 0.072                    | 562.85    | 2.942 |
|                                 |                        | 1000–1500      | 0.350                        | 115                  | 0.178                    | 1387.89   | 1.960 |
|                                 |                        | 1500–2000      | 0.286                        | 94                   | 0.214                    | 1664.69   | 1.336 |
|                                 |                        | 2000–2500      | 0.100                        | 33                   | 0.207                    | 1614.68   | 0.483 |
|                                 |                        | 2500–3000      | 0.049                        | 16                   | 0.206                    | 1605.63   | 0.236 |
|                                 |                        | 3000–3500      | 0.003                        | 1                    | 0.103                    | 799.34    | 0.030 |
|                                 |                        | > 3500         | 0.000                        | 0                    | 0.019                    | 147.30    | 0.000 |
|                                 | Slope angle (°)        | < 10           | 0.085                        | 28                   | 0.090                    | 700.5     | 0.946 |
|                                 |                        | 10–20          | 0.343                        | 113                  | 0.243                    | 1890.19   | 1.414 |
|                                 |                        | 20–30          | 0.313                        | 103                  | 0.291                    | 2266.96   | 1.075 |
|                                 |                        | 30–40          | 0.158                        | 52                   | 0.239                    | 1863.82   | 0.660 |
|                                 |                        | 40–50          | 0.070                        | 23                   | 0.104                    | 811.25    | 0.671 |
|                                 |                        | > 50           | 0.030                        | 10                   | 0.032                    | 249.66    | 0.947 |
|                                 | Plan curvature         | < -3           | 0.024                        | 8                    | 0.044                    | 339.14    | 0.558 |
|                                 |                        | -3 ~ -1.5      | 0.070                        | 23                   | 0.083                    | 645.46    | 0.843 |
|                                 |                        | -1.5 ~ 0       | 0.389                        | 128                  | 0.394                    | 3064.73   | 0.988 |
|                                 |                        | 0 ~ 1.5        | 0.401                        | 132                  | 0.354                    | 2755.92   | 1.133 |
|                                 |                        | 1.5 ~ 3        | 0.088                        | 29                   | 0.086                    | 668.51    | 1.026 |
|                                 |                        | > 3            | 0.027                        | 9                    | 0.040                    | 308.63    | 0.690 |
|                                 | Slope aspect           | Flat           | 0.000                        | 0                    | 0.001                    | 5.26      | 0.000 |
|                                 |                        | North          | 0.082                        | 27                   | 0.109                    | 845.39    | 0.755 |
|                                 |                        | Northeast      | 0.143                        | 47                   | 0.135                    | 1050.71   | 1.058 |
|                                 |                        | East           | 0.195                        | 64                   | 0.178                    | 1386.87   | 1.092 |
|                                 |                        | Southeast      | 0.182                        | 60                   | 0.157                    | 1222.66   | 1.161 |
|                                 |                        | South          | 0.094                        | 31                   | 0.110                    | 852.26    | 0.860 |
|                                 |                        | Southwest      | 0.109                        | 36                   | 0.108                    | 838.07    | 1.016 |
|                                 |                        | West           | 0.103                        | 34                   | 0.105                    | 815.83    | 0.986 |
|                                 |                        | Northwest      | 0.091                        | 30                   | 0.098                    | 765.35    | 0.927 |
|                                 | Lithology              | Hard rocks     | 0.015                        | 5                    | 0.014                    | 111.81    | 1.058 |
|                                 |                        | Harder rocks   | 0.289                        | 95                   | 0.497                    | 3870.86   | 0.581 |
|                                 |                        | Softer rocks   | 0.611                        | 201                  | 0.440                    | 3424.27   | 1.388 |
|                                 |                        | Soft rocks     | 0.082                        | 27                   | 0.037                    | 291.82    | 2.189 |
|                                 |                        | Loose deposits | 0.003                        | 1                    | 0.011                    | 83.62     | 0.283 |
|                                 | Distance to faults (m) | < 1000         | 0.310                        | 102                  | 0.176                    | 1371.46   | 1.759 |
|                                 |                        | 1000–2000      | 0.240                        | 79                   | 0.162                    | 1262.52   | 1.480 |
|                                 |                        | 2000–3000      | 0.173                        | 57                   | 0.136                    | 1059.02   | 1.273 |
|                                 |                        | 3000–4000      | 0.094                        | 31                   | 0.119                    | 923.32    | 0.794 |
|                                 |                        | > 4000         | 0.182                        | 60                   | 0.407                    | 3166.06   | 0.448 |
|                                 | Distance to rivers (m) | < 200          | 0.204                        | 67                   | 0.107                    | 830.60    | 1.908 |
|                                 |                        | 200–400        | 0.134                        | 44                   | 0.100                    | 776.54    | 1.340 |
|                                 |                        | 400–600        | 0.143                        | 47                   | 0.096                    | 745.11    | 1.492 |
|                                 |                        | 600–800        | 0.094                        | 31                   | 0.092                    | 714.53    | 1.026 |
|                                 |                        | > 800          | 0.426                        | 140                  | 0.606                    | 4715.60   | 0.702 |
|                                 | NDVI                   | < 0.2          | 0.012                        | 4                    | 0.018                    | 141.34    | 0.669 |
|                                 |                        | 0.2–0.35       | 0.094                        | 31                   | 0.095                    | 737.96    | 0.994 |
|                                 |                        | 0.35–0.50      | 0.495                        | 163                  | 0.380                    | 2959.59   | 1.303 |
|                                 |                        | 0.50–0.65      | 0.313                        | 103                  | 0.340                    | 2649.35   | 0.920 |
|                                 |                        | > 0.65         | 0.085                        | 28                   | 0.166                    | 1294.15   | 0.512 |
| Continued                       |                        |                |                              |                      |                          |           |       |

| Conditioning factors        |                                        | Classes   | Percentage of landslides (%) | Number of landslides | Percentage of domain (%) | Area (km) | FR    |
|-----------------------------|----------------------------------------|-----------|------------------------------|----------------------|--------------------------|-----------|-------|
| External triggering factors | PGA (g)                                | 0.15      | 0.289                        | 95                   | 0.379                    | 2950.98   | 0.762 |
|                             |                                        | 0.2       | 0.666                        | 219                  | 0.520                    | 4045.95   | 1.280 |
|                             |                                        | 0.3       | 0.046                        | 15                   | 0.101                    | 785.46    | 0.452 |
|                             | Rainfall (mm/yr)                       | < 570     | 0.076                        | 25                   | 0.067                    | 521.91    | 1.133 |
|                             |                                        | 570–600   | 0.055                        | 18                   | 0.064                    | 501.89    | 0.848 |
|                             |                                        | 600–630   | 0.049                        | 16                   | 0.085                    | 661.39    | 0.572 |
|                             |                                        | 630–660   | 0.073                        | 24                   | 0.103                    | 804.44    | 0.706 |
|                             |                                        | 660–690   | 0.602                        | 198                  | 0.408                    | 3178.34   | 1.474 |
|                             |                                        | > 690     | 0.146                        | 48                   | 0.272                    | 2114.41   | 0.537 |
|                             | Population density (per)               | < 1       | 0.748                        | 246                  | 0.875                    | 6806.13   | 0.855 |
|                             |                                        | 1–5       | 0.179                        | 59                   | 0.090                    | 697.68    | 2.000 |
|                             |                                        | 5–10      | 0.040                        | 13                   | 0.023                    | 175.53    | 1.752 |
|                             |                                        | 10–15     | 0.021                        | 7                    | 0.008                    | 58.95     | 2.809 |
|                             |                                        | 15–20     | 0.006                        | 2                    | 0.003                    | 24.00     | 1.972 |
|                             |                                        | > 20      | 0.006                        | 2                    | 0.003                    | 20.10     | 2.353 |
|                             | Distance to engineering activities (m) | < 500     | 0.462                        | 152                  | 0.194                    | 1509.69   | 2.382 |
|                             |                                        | 500–1000  | 0.167                        | 55                   | 0.132                    | 1030.87   | 1.262 |
|                             |                                        | 1000–1500 | 0.116                        | 38                   | 0.107                    | 829.95    | 1.083 |
|                             |                                        | 1500–2000 | 0.067                        | 22                   | 0.116                    | 901.83    | 0.577 |
|                             |                                        | > 2000    | 0.188                        | 62                   | 0.451                    | 3510.05   | 0.418 |

**Table 2.** Spatial relationship between landslide frequency and landslide conditioning factors.

### 12 Distance to human engineering activities

Human engineering activities in the CIALS, including road construction, hydropower development, and urban expansion, have intensified, leading to increased damage to mountain slopes<sup>46</sup>. This trend contributes to slope instability, triggering numerous landslides and reactivating ancient ones. In this study, 5 buffer zone categories are established at intervals of 500 m (Fig. 2l), with landslides predominantly occurring within a distance of less than 500 m from these activities.

## Methodology

The technical approach of this study comprises 4 key steps (Fig. 3): (1) Sample generation: 10 groups of random samples were created from the historical landslide inventory, with each group divided into a training dataset (70%) and a validating dataset (30%). (2) Conditioning factors selection: Relevant landslide conditioning factors specific to the study area were identified. (3) Model integration: CT was integrated with both SVM and RF models through coordinate transformation and GIS technology to generate LSM. (4) Accuracy assessment: The accuracy and stability of the four methods (RF, SVM, RF-CT, and SVM-CT) were assessed using the Receiver operating characteristic (ROC) curve across the different random samples.

### Support vector machine

SVM is a ML technique widely used for classification, regression, and anomaly detection, particularly effective with linearly inseparable and high-dimensional datasets<sup>47</sup>. It learns from labeled training data and aims to find the optimal hyperplane that best separates different classes (Fig. 4a). For linearly separable data, SVM identifies the maximum margin hyperplane, ensuring the greatest separation between classes. When the data is not linearly separable, SVM uses kernel functions to map the data into a higher-dimensional feature space, where it then finds the optimal hyperplane. Given a set of samples  $(x_i, y_i)$ , the process for determining the optimal hyperplane in SVM is as follows:

$$\begin{cases} \min_{\phi, \theta, \delta} \left( \frac{1}{2} \phi^T \phi + C \sum_{i=1}^n \delta_i \right) \\ y_i (\phi^T \phi (x_i) + \theta) \geq 1 - \delta_i \\ \delta_i \geq 0, \quad i = 1, 2, \dots, n \end{cases} \quad (1)$$

In Eq. (1),  $\phi$  represents the weight vector defining the hyperplane, while  $\delta_i$  is the slack variable that permits misclassifications. The condition  $\delta_i \geq 0$  indicates that each point is either on the correct side of the margin or within an acceptable error range.  $C$  is the penalty parameter.  $\theta$  denotes the offset of the hyperplane from the origin, and  $y_i$  is the label of the training data. The function  $\phi(x_i)$  maps the input data  $x_i$  into a high-dimensional feature space, facilitating nonlinear classification.

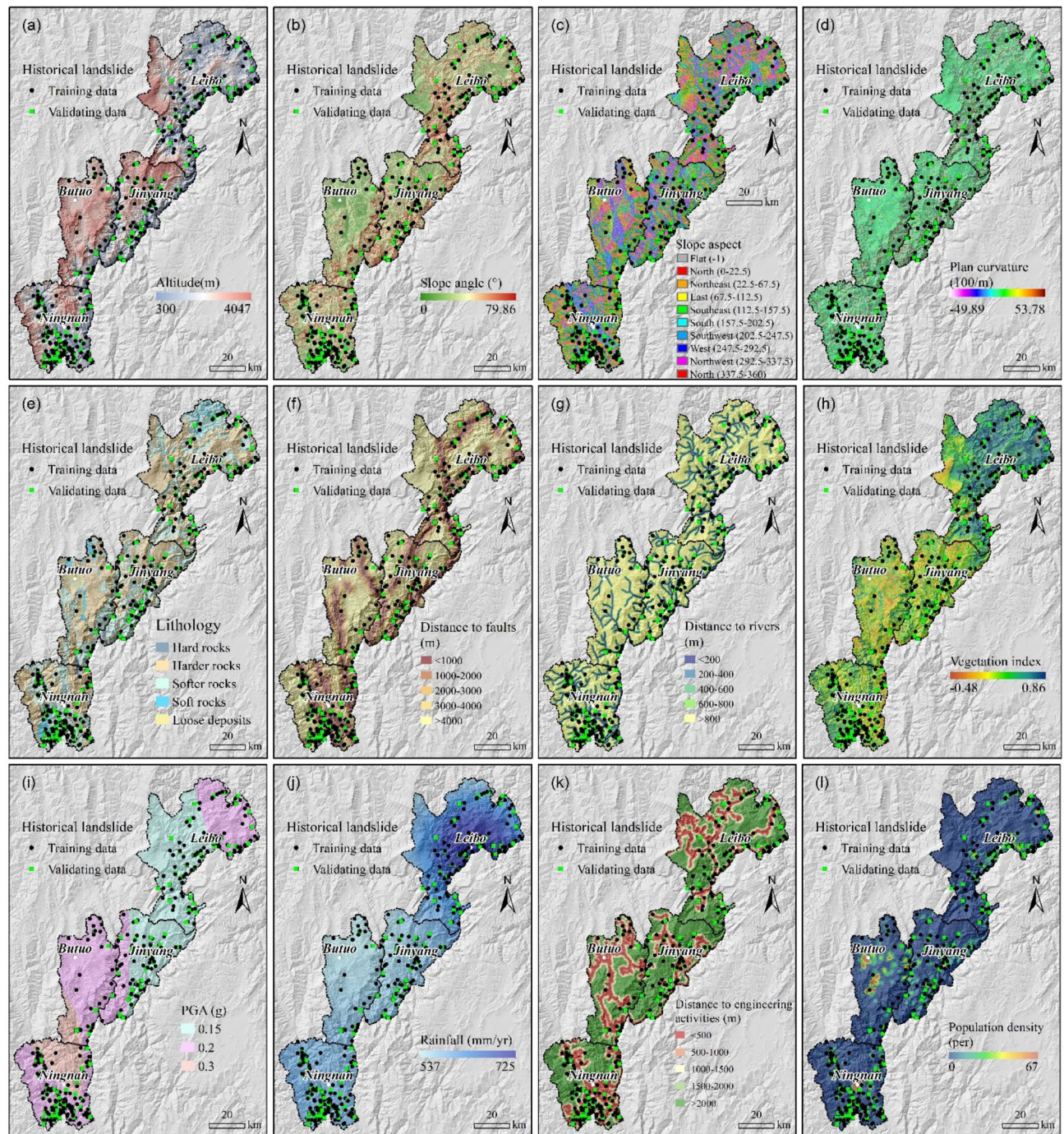


**Fig. 2.** Spatial distribution of conditioning factors in the CIALS; (a–l) represent Altitude, Slope angle, Plan curvature, Slope aspect, Lithology, Distance to faults, Distance to rivers, NDVI, PGA, Rainfall, Distance to engineering activities, Population density, respectively.

### Random forest

RF is an ensemble learning algorithm based on decision trees and is widely applied in tasks such as classification, regression, and feature selection<sup>48</sup>. The process of constructing an RF model involves several steps: First, the Bootstrap method is used to randomly sample with replacement from the original training dataset to create  $n$  new sample sets, each used to generate  $n$  decision trees. Second, at each resampling, a subset of features is randomly selected to build the decision tree. Finally, these decision trees are combined to form a “forest” (Fig. 4b). For classifying new data, the final classification is determined by the majority vote of all decision trees. RF models are highly capable of handling missing values and noisy data, and they are generally resistant to overfitting. They are also known for their robustness and accuracy when managing complex data<sup>49</sup>. Given these advantages, the RF model was chosen for this study to generate LSM for the CIALS.

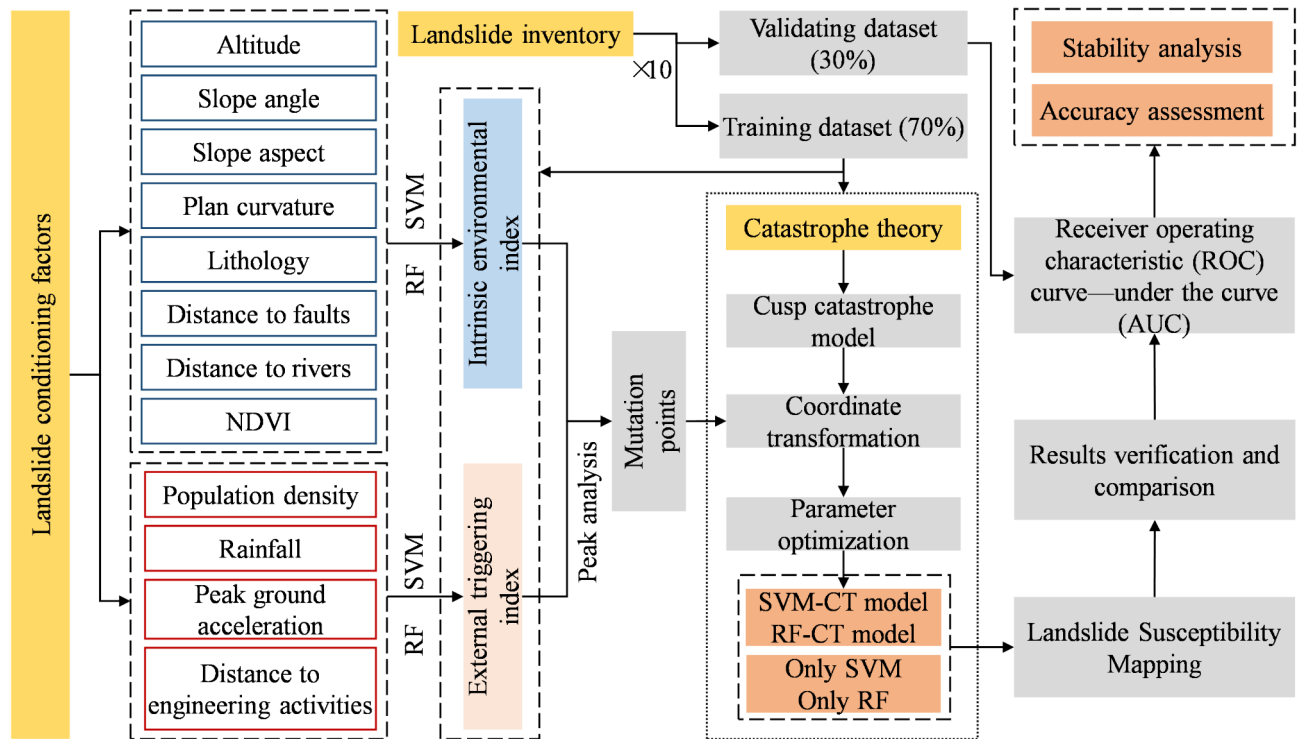


Fig. 3. Technical process of this study.

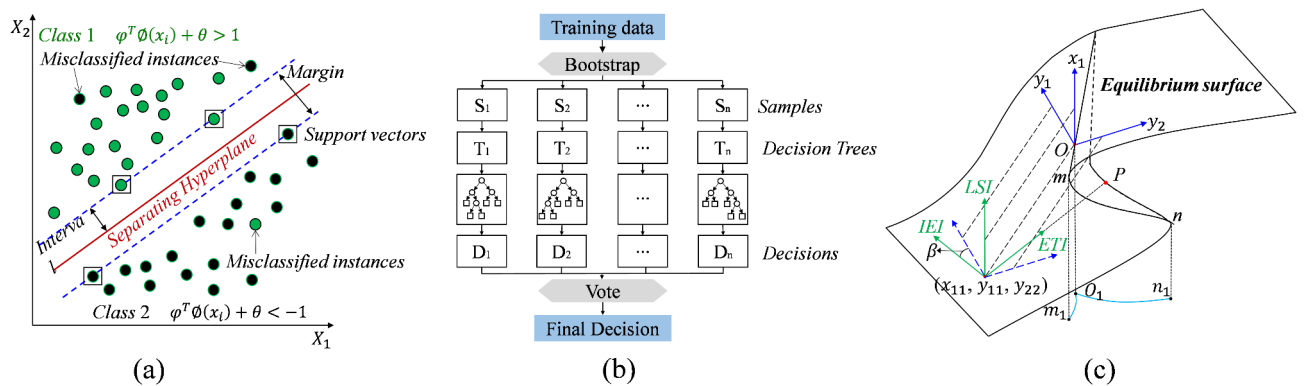


Fig. 4. (a) is the classification principle and hyperplane of SVM; (b) is the schematic diagram of the RF model structure; (c) represents the CCM equilibrium surface and coordinate conversion process, where  $o_1m_1$  and  $o_1n_1$  are bifurcation curves.

### LSM framework

In this study, intrinsic environmental factors and external triggering factors are calculated using ML (SVM and RF) algorithms to derive the intrinsic environmental index (IEI) and external triggering index (ETI), respectively. These indexes are used as control variables for constructing a landslide susceptibility framework based on the cusp catastrophe model (CCM) within CT<sup>50</sup>. The core task in constructing the landslide susceptibility framework is to establish the surface using the IEI, ETI and landslide susceptibility index (LSI). The correspondence between the equilibrium surface  $(x_1, y_1, y_2)$  of the CCM and the surface  $(LSI, IEI, ETI)$  is determined through a coordinate transformation. This transformation ensures that the function with LSI, IEI and ETI meets the conditions of the cusp mutation type. Finally, based on the property that the mutations of CCM should occur in the cusp-shaped folding zone, the landslide susceptibility assessment formula is derived. The specific steps for constructing the framework are as follows:

### Equation of the equilibrium surface

To accurately simulate the relationship between the control variables IEI and ETI and the state variable LSI, an equilibrium surface equation is established, referring to the equilibrium surface  $(x_1, y_1, y_2)$  as determined by the CCM. The equation of the equilibrium surface is as follows:

$$ax_1^3 + bx_1y_1 = y_2 \quad (2)$$

In Eq. (2), all coefficients are non-zero constants. Let  $x_1 = x_0$  be a fixed value, then  $y_1$  and  $y_2$  exhibit a linear relationship, and the equilibrium surface consists of straight lines parallel to the  $y_1oy_2$  plane. As shown in Fig. 4c,  $y_2$  has a positive correlation with  $x_1$ ; however, when  $y_2$  reaches a certain range, a slight increase in  $y_2$  causes a sharp and discontinuous rise in  $x_1$ .

### Coordinate conversion of the equilibrium surface

In this study, let the spatial translation amount of equilibrium surface  $(x_1, y_1, y_2)$  to  $(LSI, IEI, ETI)$  is denoted as  $(x_{11}, y_{11}, y_{22})$ , and the rotation angle around the LSI-axis is  $\beta$ , the transformation formula is given by:

$$\begin{cases} LSI + x_{11} = x_1 \\ ETI \sin \beta + IEI \cos \beta + y_{11} = y_1 \\ ETI \cos \beta - IEI \sin \beta + y_{22} = y_2 \end{cases} \quad (3)$$

According to Fig. 4c and Eq. (3), it is evident that the origin coordinates  $(x_{11}, y_{11}, y_{22})$  of the equilibrium surface  $(LSI, IEI, ETI)$  are on the equilibrium surface  $(x_1, y_1, y_2)$ . Therefore, the following equation holds:

$$ax_{11}^3 + bx_{11}y_{11} = y_{22} \quad (4)$$

Let  $ax_{11}^3 + bx_{11}y_{11} = y_{22}$  represent the IEI-axis. This line serves as the intersection between the plane passing through the origin of the new coordinate system and the equilibrium surface  $(x_1, y_1, y_2)$ . The new coordinate system is obtained by translating the original system and rotating by  $\beta$  degrees around the LSI-axis. Thus, the following equation applies:

$$\arcsin \left( \frac{bx_{11}}{\sqrt{1 + bx_{11}^2}} \right) = \arccos \left( \frac{1}{\sqrt{1 + bx_{11}^2}} \right) = \beta \quad (5)$$

### New coordinate system

Perform peak analysis on IEI and ETI to obtain the mutation points  $IEI_1, IEI_2, ETI_1$ , and  $ETI_2$ . Since  $(LSI_1, IEI_1, ETI_1)$  is the critical point where landslide susceptibility undergoes a mutation, and the origin of the coordinate system  $(x_1, y_1, y_2)$  is also a critical point where a mutation exists. Therefore, for the point  $(LSI_1, IEI_1, ETI_1)$  in the new coordinate system  $(LSI, IEI, ETI)$ , make a plane perpendicular to the IEI-axis. This plane also passes through the origin of the coordinate system  $(x_1, y_1, y_2)$ . Then the mutation point  $(LSI_1, IEI_1, ETI_1)$  in the coordinate system  $(x_1, y_1, y_2)$  is  $(LSI_1 + x_{11}, ETI_1 \sin \beta + IEI_1 \cos \beta + y_{11}, ETI_1 \cos \beta - IEI_1 \sin \beta + y_{22})$ , and the direction vector of the LSI-axis is  $(0, \cos \beta, \sin \beta)$ . Therefore, the equation of this plane is:

$$0 = y_1 \cos \beta - y_2 \sin \beta \quad (6)$$

Substitute the mutation point  $(LSI_1 + x_{11}, ETI_1 \sin \beta + IEI_1 \cos \beta + y_{11}, ETI_1 \cos \beta - IEI_1 \sin \beta + y_{22})$  into Eq. (6), and solve it simultaneously with Eq. (5) to obtain:

$$ETI_1 + \frac{y_{11}}{\sqrt{1 + bx_{11}^2}} - \frac{by_{22}x_{11}}{\sqrt{1 + bx_{11}^2}} = 0 \quad (7)$$

In the coordinate system  $(LSI, IEI, ETI)$ , a mutation will occur again at the point  $(LSI_2, IEI_2, ETI_2)$ , that is, point P in Fig. 4c. The coordinate point  $P(LSI_2 + x_{11}, ETI_2 \sin \beta + IEI_2 \cos \beta + y_{11}, ETI_2 \cos \beta - IEI_2 \sin \beta + y_{22})$  is the intersection and mutation point of the ETI-axis and the singular point set of the equilibrium surface. Therefore, there exist three roots  $(x_{11}, x_{12}$  and  $x_{13})$  for the equation  $ETI_2 \cos \beta - IEI_2 \sin \beta + y_{22} = ax_1^3 + bx_1(ETI_2 \sin \beta + IEI_2 \cos \beta + y_{11})$ , where two roots are equal  $(x_{11} = x_{13})$ , and satisfy the equation  $(x_1 - x_{11})(x_1 - x_{12})(x_1 - x_{13}) = 0$ . Solving this equation gives:

$$x_1^3 - (2x_{11} + x_{12})x_1^2 + (x_{11}^2 + 2x_{11}x_{12})x_1 - x_{11}^2x_{12} = 0 \quad (8)$$

Set the coefficients of Eq. (8) equal to those of  $ETI_2 \cos \beta - IEI_2 \sin \beta + y_{22} = ax_1^3 + bx_1(ETI_2 \sin \beta + IEI_2 \cos \beta + y_{11})$ , and substitute Eq. (5) to obtain the following equation:

$$\begin{cases} x_{12} + 2x_{11} = 0 \\ x_{11}^2 + 2x_{11}x_{12} = \frac{b}{a} \left( \frac{IEI_2}{\sqrt{1+bx_{11}^2}} + \frac{ETI_2bx_{11}}{\sqrt{1+bx_{11}^2}} + y_{11} \right) \\ ay_{22}x_{11} = \frac{ETI_2}{\sqrt{1+bx_{11}^2}} - \frac{IEI_2bx_{11}}{\sqrt{1+bx_{11}^2}} + y_{22} \end{cases} \quad (9)$$

Since landslide susceptibility gradually increases as the control variable coefficients increase within a certain range around point  $P$ , a transition in susceptibility occurs at point  $P$ . The magnitude of the transition from  $x_{11}$  to  $x_{12}$  is given by:

$$\begin{cases} x_{11} - x_{12} = \sqrt{-3IEI} = \sqrt[3]{27ETI} \\ Max(x_{11} - x_{12}) \end{cases} \quad (10)$$

#### Landslide susceptibility assessment formula

By solving Eqs. (5), (3) and (2), obtain the expression of the new equilibrium surface ( $LSI$ ,  $IEI$ ,  $ETI$ ). And through the *Cardano* formula, obtain the landslide susceptibility assessment formula established:

$$L = \frac{1}{4} \left[ \frac{1}{a} \left( \frac{IEIbx_{11}}{\sqrt{1+bx_{11}^2}} - \frac{ETI}{\sqrt{1+bx_{11}^2}} - y_{22} \right) \right]^2 + \frac{1}{27} \left[ \frac{b}{a} \left( \frac{IEI}{\sqrt{1+bx_{11}^2}} + \frac{ETIbx_{11}}{\sqrt{1+bx_{11}^2}} + y_{11} \right) \right]^3 \quad (11)$$

#### Parameter solving

- (1) Assign a random non-zero constant to  $b$ , and use Eqs. (7), (9), and (10) to calculate the remaining 5 parameters ( $a, b, x_{11}, x_{12}, y_{11}, y_{22}$ ).
- (2) Using the 6 initially determined coefficients, derive the bifurcation curve equation ( $o_1m_1, o_1n_1$ ). The equations for curves  $o_1m_1$  and  $o_1n_1$  are provided by Eqs. (12) and (13).
- (3) If more than 90% of the validating datasets lies outside the bifurcation curve ( $o_1m_1, o_1n_1$ ), then the chosen value of  $b$  satisfies the solution requirements.

$$a \left( \sqrt{\frac{-by_1}{3a}} \right)^3 + by_1 \left( \sqrt{\frac{-by_1}{3a}} \right) = y_2 \quad (12)$$

$$a \left( -\sqrt{\frac{-by_1}{3a}} \right)^3 - by_1 \left( \sqrt{\frac{-by_1}{3a}} \right) = y_2 \quad (13)$$

According to the distribution of  $IEI$  and  $ETI$  values, their mutation positions are identified through peak analysis (Fig. 5a,b). Following the parameter solution steps, it is observed that as the absolute value of  $b$  increases, the absolute values of  $a$ ,  $y_{11}$ , and  $y_{22}$  initially rise, then decline, and eventually stabilize, while the values of  $x_{11}$  and  $x_{12}$  remain constant (Fig. 5c). When more than 90% of the landslide verification datasets are located far from the bifurcation curve, the range of acceptable values for  $b$  can be established (Fig. 5d). Furthermore, since the origin coordinates lie on the initial equilibrium surface, the parameters  $x_{11}$ ,  $y_{11}$ , and  $y_{22}$  also satisfy Eq. (4). Consequently, the final value of  $b$  can be determined (Fig. 5e), allowing for the calculation of other parameter values for the landslide susceptibility framework (RF-CT and SVM-CT).

#### Verification and comparison

The ROC curve is a graphical representation of sensitivity, widely used in various fields, including LSM<sup>51,52</sup>. In the ROC curve, the Area under the curve (AUC) quantifies the uncertainty of model predictions while accounting for detection bias. This study utilizes the ROC curve, along with the spatial distribution of historical landslide datasets, to validate the results of the LSM.

This study utilize training and validating datasets to assess the framework's performance. The training dataset is used to calculate the success rate, while the validating dataset assesses the prediction rate. Additionally, separate SVM and RF models are employed to generate landslide susceptibility maps based on landslide conditioning factors, enabling a comparative analysis with the framework developed in this study.

## Result

### Landslide susceptibility mapping

By substituting the  $IEI$ ,  $ETI$ , and framework parameters into Eq. (11), we can derive the  $LSI$  for the CIALS. The natural breaks method categorizes the  $LSI$  into 5 classifications: Very Low, Low, Moderate, High, and Very High (Fig. 6). As illustrated in Fig. 7, the area classified as the Very high susceptibility zone is 644.57 km<sup>2</sup>, representing 8.28% of the total area. The High susceptibility zone covers 1337.66 km<sup>2</sup>, accounting for 17.19%. These two categories are primarily located in Ningnan County, in the southern part of the CIALS, as well as along the Jinsha River and its tributary valleys. The Moderate susceptibility zone spans 1937.16 km<sup>2</sup> (24.89%), the Low susceptibility zone covers 2342.62 km<sup>2</sup> (30.10%), and the Very low susceptibility zone encompasses 1520.37 km<sup>2</sup> (19.54%). These three categories are mainly found in mountainous areas with abundant vegetation, situated far from human activities. Overall, the spatial distribution patterns of the landslide susceptibility maps generated by the four methods are relatively consistent.

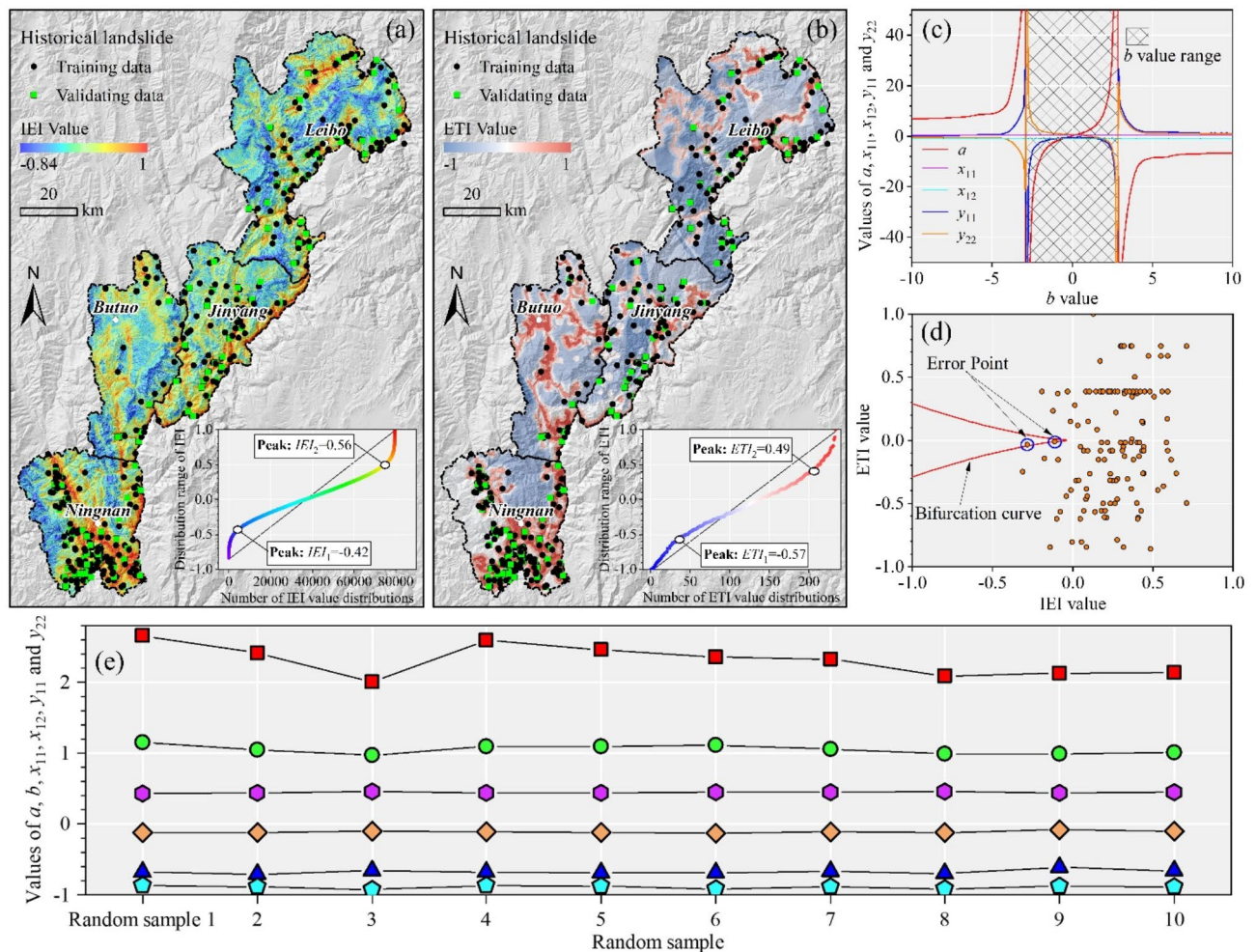


**Fig. 5.** Framework parameter solving process.

### Model verification

According to the verification results of the ROC curve (Fig. 8), the LSM obtained using RF-CT shows the highest average success rate, with an AUC value of 0.899 and the lowest average standard error of 0.014. It also achieves the highest average prediction rate (AUC = 0.783) along with the lowest average standard error of 0.068. Overall, the LSMs derived from the four methods demonstrate relatively high accuracy. The average mapping success rates, from highest to lowest, are: RF-CT (0.899), SVM-CT (0.873), RF (0.761), and SVM (0.758). Similarly, the average prediction rates, from highest to lowest, are: RF-CT (0.783), SVM-CT (0.775), SVM (0.737), and RF (0.733). Notably, the average success rates of RF-CT and SVM-CT are approximately 10% higher than those of RF and SVM, while the average prediction rates exceed those of RF and SVM by about 5%.

### Discussion

#### Accuracy analysis

The RF, SVM, RF-CT, and SVM-CT models are compared based on the generated LSMs. Qualitatively, the overall spatial distribution of Very high and High susceptibility zones shows a strong correlation with historical landslide locations. For a typical area (Fig. 6), while most landslides are concentrated in Very high and High susceptibility zones, the RF-CT and SVM-CT models generate smaller Very high and High susceptibility zones, and offering more detailed descriptions. Quantitatively, the average area of the classification categorizes of the 10 results was analyzed. The proportion of historical landslides in Very low and Low susceptibility zones for RF-CT and SVM-CT is 7.64% and 8.49%, respectively—significantly lower than 12.95% for RF and 16.77% for SVM. In Very high and High susceptibility zones, historical landslides account for 75.37% in RF-CT, 71.55% in SVM-CT, 67.30% in RF, and 63.69% in SVM. Similarly, the landslide density is higher for RF-CT (0.18 per km<sup>2</sup>) and SVM-CT (0.17 per km<sup>2</sup>) than for RF and SVM (both 0.16 per km<sup>2</sup>) in Very high and High susceptibility zones. Both qualitative and quantitative analyses show that RF-CT and SVM-CT outperform RF and SVM.

As shown in Fig. 7, regardless of the changes in the training and validating datasets, the changes in landslide susceptibility categories for RF-CT and SVM-CT are significantly smaller than those for the RF and SVM models. Figure 8c,d further confirms this result. From the training dataset, within the 95% confidence interval, there are significant differences in the AUC value ranges of the 10 sample groups. Specifically, the AUC value

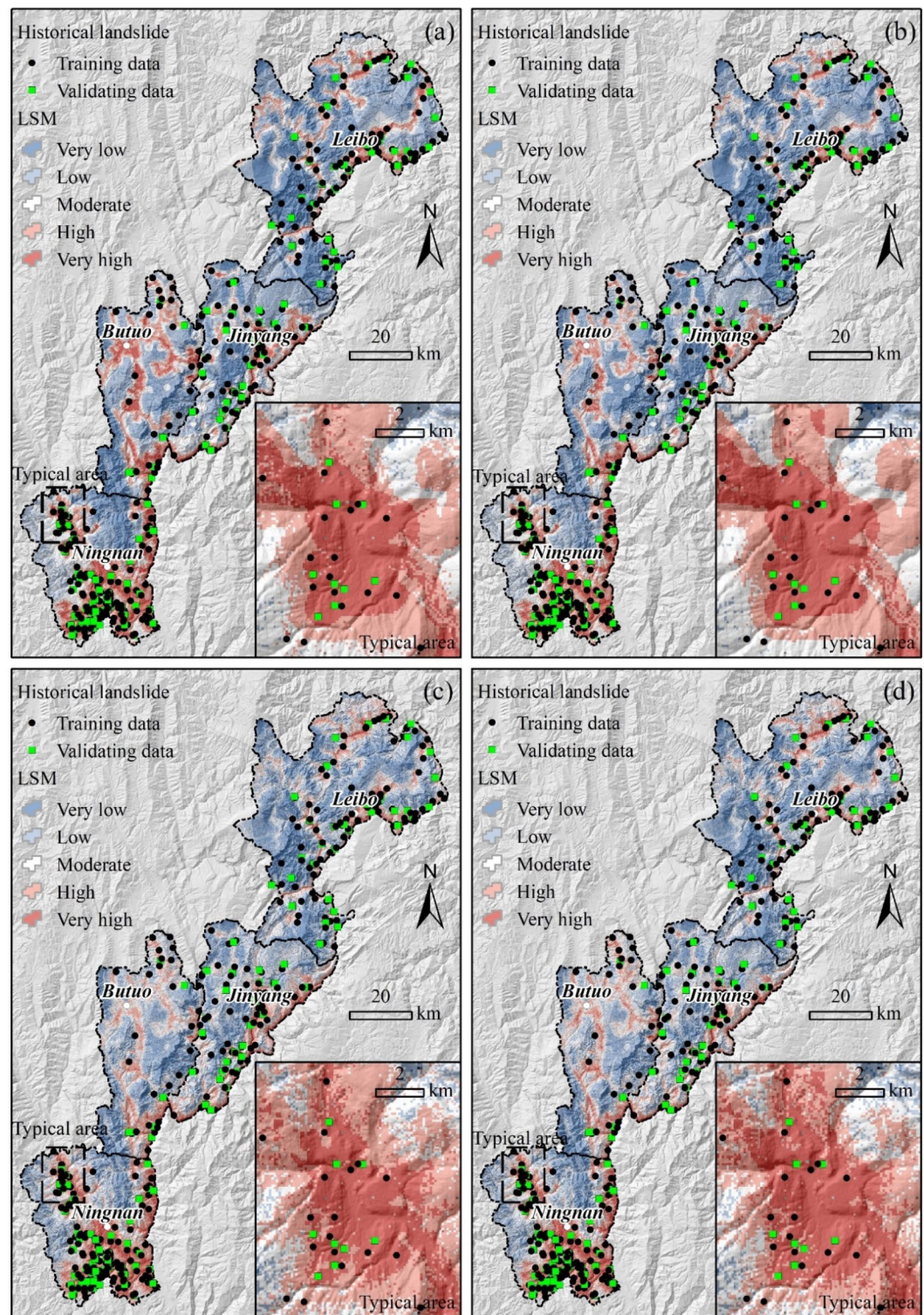


**Fig. 6.** LSM results of the study area. (a) RF; (b) SVM; (c) RF-CT; (d) SVM-CT.

ranges for RF, SVM, RF-CT, and SVM-CT are 0.724–0.796 (range difference of 0.072), 0.723–0.792 (0.069), 0.879–0.918 (0.039), and 0.846–0.895 (0.049), respectively. On the validating dataset, the AUC values for RF, SVM, RF-CT, and SVM-CT range from 0.698 to 0.768 (0.070), 0.700–0.775 (0.075), 0.749–0.817 (0.061), and 0.745–0.806 (0.068), respectively. Regarding the change in AUC value ranges, on the training dataset, the AUC range for RF-CT is 45% narrower than that for RF, and for SVM-CT, it is 29% narrower than that for SVM. On the validating dataset, the AUC range for RF-CT is 11% narrower than that for RF, and for SVM-CT, it is 10% narrower than that for SVM.

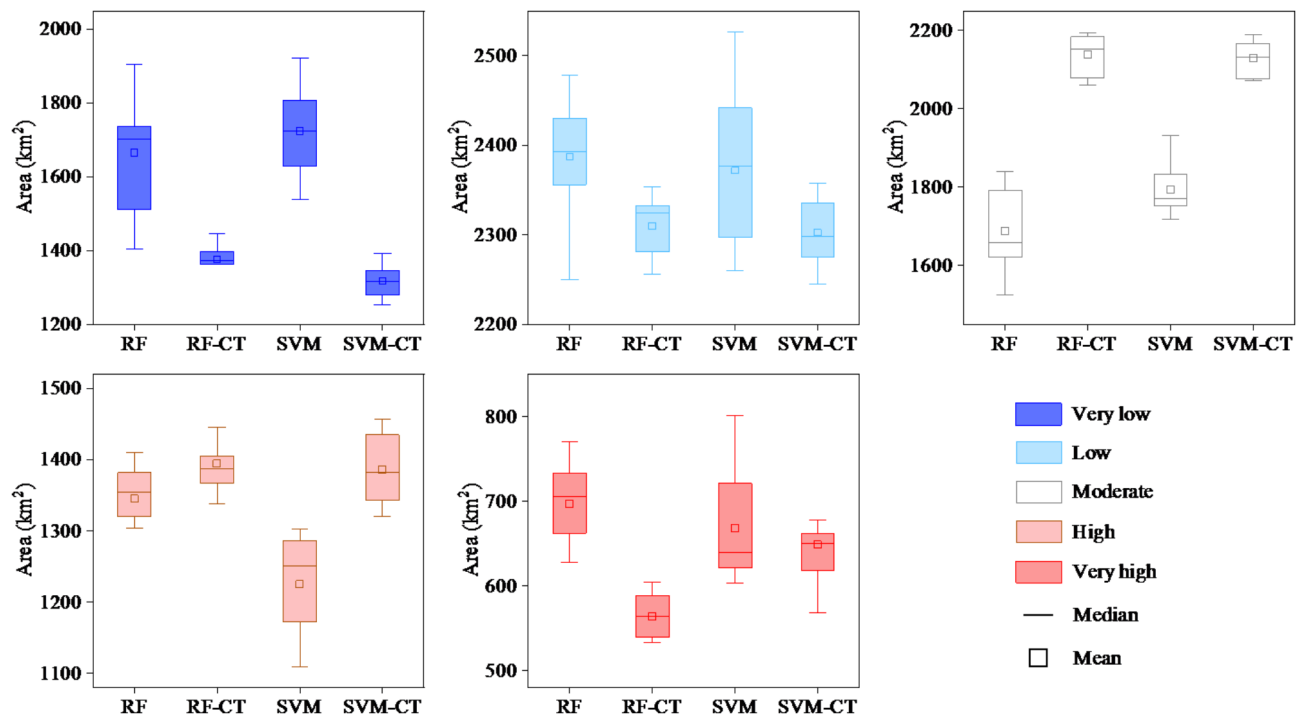


**Fig. 7.** Area of landslide susceptibility classification based on 10 groups of random samples.

These results indicate that integrating CT into RF- and SVM-based LSM models is not only feasible but also significantly improves the accuracy, stability and generalization of landslide susceptibility predictions.

### The roles of RF, SVM and CT in the new framework

Existing studies indicate that hybrid and integrated ML algorithms enhance the accuracy of LSM<sup>16,53–55</sup>. The LSM framework developed in this study exhibits this characteristic, demonstrating significantly better performance than individual RF and SVM models. Although the mapping accuracy of RF is slightly higher than that of SVM, this aligns with current research findings<sup>56,57</sup>.

The integration of CT with RF and SVM offers a more robust prediction framework for assessing landslide occurrence probabilities. First, based on CT characteristics, conditional factors are categorized into internal environmental factors and external triggering factors, which are comprehensively analyzed by RF or SVM. This approach mitigates the uncertainties associated with traditional CT application methods (Catastrophe fuzzy membership functions, CFMFs) when determining the relative importance of conditional factors<sup>58</sup>. Second, RF's inherent randomness reduces correlations among trees<sup>59</sup>, while SVM maximizes the margin between samples and the decision boundary<sup>60</sup>, addressing the dependency of CFMFs on the independence of conditional factors. Third, while RF generally lacks high generalization capability, the combined application of CT and RF enhances model training stability across different areas by stabilizing critical points using three variables<sup>61</sup>. Additionally, SVM can be resource-intensive when handling large datasets<sup>62</sup>, and categorizing factors into internal environmental factors and external triggering factors can improve SVM training efficiency.

In summary, landslide occurrence is a highly complex nonlinear process influenced by multiple environmental factors. SVM and RF are effective in addressing this complexity, while CT reveals the mutation characteristics of the process. CT provides a rational and mathematical framework that enables the integration of SVM and RF output results with historical landslide data (validating dataset) to identify the “mutation point” of landslide occurrence, thus deriving a landslide susceptibility assessment formula. This integration results in a more reliable and robust LSM technique compared to using SVM or RF alone. In addition, most LSMs rely on model parameter tuning, which often leads to strong performance in some regions but poor results in others<sup>63</sup>. This study uses 10 different random sample datasets to demonstrate that the proposed framework not only ensures high accuracy in LSMs but also offers stability, making it a universal model applicable to any area with relevant conditioning factors and historical landslide inventory data.

### The limitations of the new framework

Although the new framework demonstrates “superior” performance, it also has certain limitations. First, to ensure that all data transition at point  $P$  during the framework construction, we selected the maximum value across all assessment units when setting the transition amount from  $x_{11}$  to  $x_{12}$  (Eq. 10), without considering the spatial characteristics, as the control variable is spatial data. Second, the conditioning factors form the basis of LSM, and any bias in these factors can affect the LSM results. For instance, the spatial interpolation method was used to convert annual precipitation data from meteorological stations into raster data. This method assumes

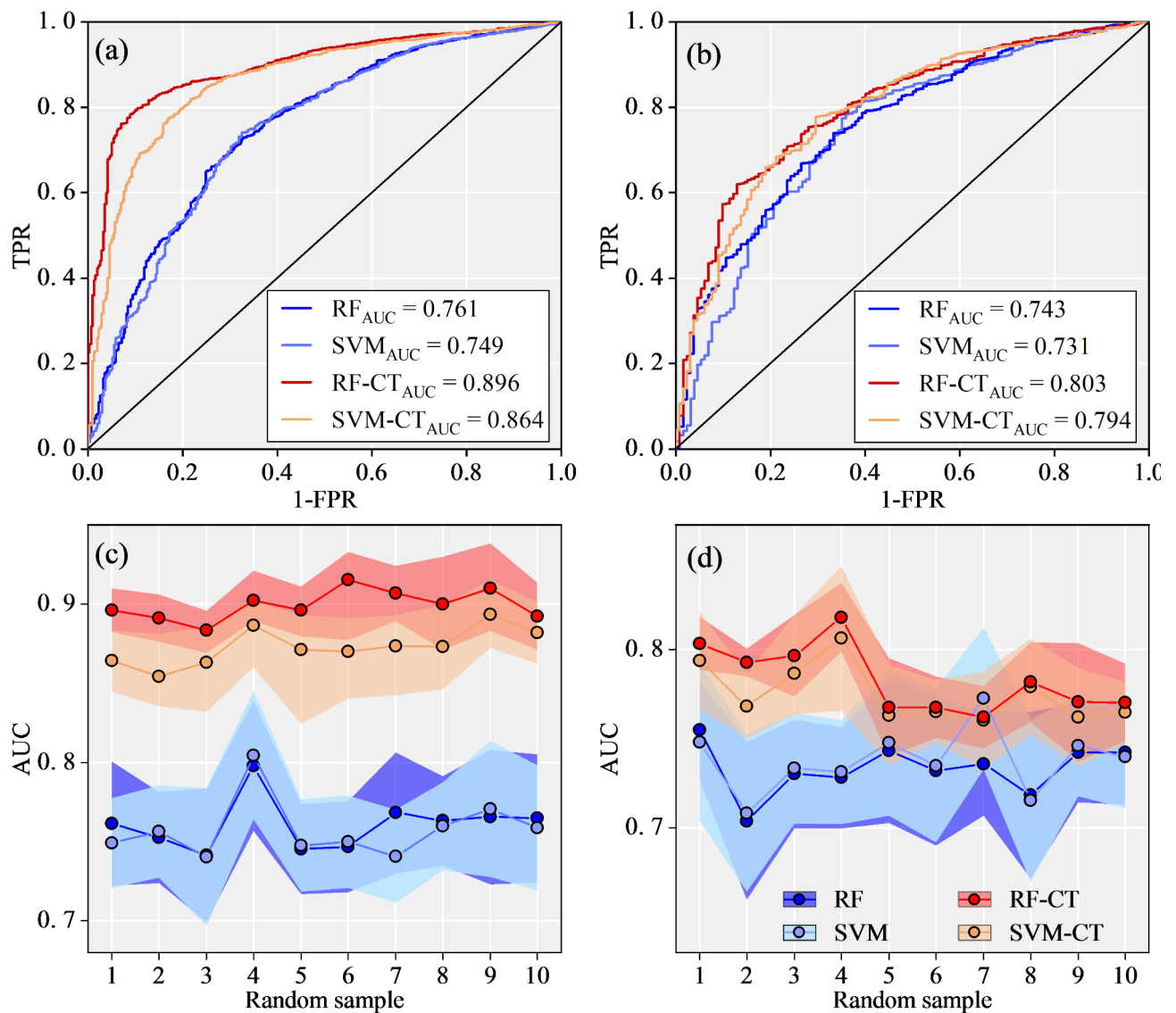


**Fig. 8.** Verification results of 10 groups of random samples. (a) and (b) are the ROC curves of the success rate and prediction rate of a certain group of random samples; (c) is the success rate of 10 groups of random samples; (d) is the prediction rate of 10 groups of random samples.

that points closer in space are more likely to share similar characteristics<sup>50</sup>. However, using known point data to estimate unknown point data can affect the accuracy of LSM. Finally, when selecting conditional factors, we overlooked certain factors that are difficult to quantify or collect. For example, we did not account for the impact of policies, management, and prevention measures on landslide susceptibility.

## Conclusions

In this study, we developed a new framework for predicting landslide occurrence probability by integrating RF, SVM, and CT. A typical area (CIALS) was selected for landslide susceptibility predicting, verification, and comparative analysis. The main conclusions are as follows:

- (1) *Feasibility of CT integration:* Introducing CT in conjunction with RF and SVM for LSM is effective. Both RF-CT and SVM-CT demonstrate excellent mapping performance, with success rates approximately 10% higher and prediction rates about 5% higher than those of RF and SVM.
- (2) *Spatial distribution of susceptibility zones:* In CIALS, the Very high susceptibility zone covers 644.57 km<sup>2</sup> (8.28% of the total area), the High susceptibility zone spans 1337.66 km<sup>2</sup> (17.19%), the Moderate susceptibility zone encompasses 1937.16 km<sup>2</sup> (24.89%), the low susceptibility zone is 2342.62 km<sup>2</sup> (30.10%), and the Very low susceptibility zone covers 1520.37 km<sup>2</sup> (19.54%). The spatial distribution patterns of landslide susceptibility derived from RF, SVM, RF-CT, and SVM-CT are relatively consistent and strongly correlate with the distribution of historical landslides.

- (3) **Practical Implications:** The research findings provide valuable references for disaster managers, poverty alleviation workers, and decision-makers in the CIALS. Governments and relevant organizations can utilize these results to avoid developing or conducting geological investigations in areas at risk of landslides. This proactive approach is one of the most effective strategies for reducing future losses of life and property, alleviating poverty, and enhancing sustainability.

## Data availability

The data supporting the findings of this study are provided in Table 1 of the main text. The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 28 October 2024; Accepted: 23 January 2025

Published online: 27 March 2025

## References

- Xu, Q. et al. Remote sensing for landslide investigations: A progress report from China. *Eng. Geol.* **321**, 107156 (2023).
- Casagli, N., Intrieri, E., Tofani, V., Gigli, G. & Raspini, F. Landslide detection, monitoring and prediction with remote-sensing techniques. *Nat. Rev. Earth Environ.* **4** (1), 51–64 (2023).
- Deng, X., Zeng, M., Xu, D. & Qi, Y. Why do landslides impact farmland abandonment? Evidence from hilly and mountainous areas of rural China. *Nat. Hazards*. **113** (1), 699–718 (2022).
- Zhang, S. et al. Fatal landslides in China from 1940 to 2020: Occurrences and vulnerabilities. *Landslides* **20** (6), 1243–1264 (2023).
- Xiao, S. et al. Landslide damming threats along the Jinsha River, China. *Engineering*, (2024).
- Shrestha, M., Noppradit, P., Shrestha, R. P. & Dahal, R. K. Understanding livelihood insecurity due to landslides in the mid-hill of Nepal: A case study of Bahrabise Municipality. *Int. J. Disaster Risk Reduct.* **104**, 104399 (2024).
- Sun, X. et al. A novel landslide susceptibility optimization framework to assess landslide occurrence probability at the regional scale for environmental management. *J. Environ. Manage.* **322**, 116108 (2022).
- Loche, M., Alvioli, M., Marchesini, I., Bakka, H. & Lombardo, L. Landslide susceptibility maps of Italy: Lesson learnt from dealing with multiple landslide types and the uneven spatial distribution of the national inventory. *Earth Sci. Rev.* **232**, 104125 (2022).
- Sonker, I. & Tripathi, J. N. Remote sensing and GIS-based landslide susceptibility mapping using frequency ratio method in Sikkim Himalaya. *Quat. Sci. Adv.* **8**, 100067 (2022).
- Patil, A. S. & Panhalkar, S. S. Remote sensing and GIS-based landslide susceptibility mapping using LNRF method in part of western ghats of India. *Quat. Sci. Adv.* **11**, 100095 (2023).
- Hodasová, K. & Bednarik, M. Effect of using various weighting methods in a process of landslide susceptibility assessment. *Nat. Hazards*. **105**, 481–499 (2021).
- Nhu, V. H. et al. Comparison of support vector machine, bayesian logistic regression, and alternating decision tree algorithms for shallow landslide susceptibility mapping along a mountainous road in the west of Iran. *Appl. Sci.* **10** (15), 5047 (2020).
- Ngo, P. T. T. et al. Evaluation of deep learning algorithms for national scale landslide susceptibility mapping of Iran. *Geosci. Front.* **12** (2), 505–519 (2021).
- Youssef, A. M. & Pourghasemi, H. R. Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia. *Geosci. Front.* **12** (2), 639–655 (2021).
- Zhang, J. et al. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. *J. Environ. Manage.* **332**, 117357 (2023).
- Chen, Y., Ming, D., Ling, X., Lv, X. & Zhou, C. Landslide susceptibility mapping using feature fusion-based CPCNN-ML in Lantau Island, Hong Kong. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* **14**, 3625–3639 (2021).
- Ado, M. et al. Landslide susceptibility mapping using machine learning: A literature survey. *Remote Sens.* **14** (13), 3029 (2022).
- Ma, J. et al. Automated machine learning-based landslide susceptibility mapping for the three gorges reservoir area, China. *Math. Geosci.* **56** (5), 975–1010 (2024).
- Huang, Y. & Zhao, L. Review on landslide susceptibility mapping using support vector machines. *Catena* **165**, 520–529 (2018).
- Arabameri, A., Pradhan, B., Rezaei, K. & Lee, C. W. Assessment of landslide susceptibility using statistical and artificial intelligence-based FR-RF integrated model and multiresolution DEMs. *Remote Sens.* **11** (9), 999 (2019).
- Sun, X. et al. Tracking sustainable development in mining towns: A novel framework integrating socioeconomic and eco-environmental perspectives through coupling coordination degree. *Environ. Impact Assess. Rev.* **109**, 107641 (2024).
- Chen, H. T., He, J., Wang, W. C. & Chen, X. N. Simulation of maize drought degree in Xi'an City based on cusp catastrophe model. *Water Sci. Eng.* **14** (1), 28–35 (2021).
- Singh, L. K., Jha, M. K. & Chowdary, V. M. Assessing the accuracy of GIS-based multi-criteria decision analysis approaches for mapping groundwater potential. *Ecol. Ind.* **91**, 24–37 (2018).
- Wang, Y., Wang, Y., Su, X., Qi, L. & Liu, M. Evaluation of the comprehensive carrying capacity of interprovincial water resources in China and the spatial effect. *J. Hydrol.* **575**, 794–809 (2019).
- Ziarh, G. F., Asaduzzaman, M., Dewan, A., Nashwan, M. S. & Shahid, S. Integration of catastrophe and entropy theories for flood risk mapping in peninsular Malaysia. *J. Flood Risk Manag.* **14** (1), e12686 (2021).
- Sun, X. F. et al. Spatiotemporal change of vegetation coverage recovery and its driving factors in the Wenchuan earthquake-hit areas. *J. Mt. Sci.* **18** (11), 2854–2869 (2021a).
- Pellicone, G., Caloiero, T., Modica, G. & Guagliardi, I. Application of several spatial interpolation techniques to monthly rainfall data in the Calabria region (southern Italy). *Int. J. Climatol.* **38** (9), 3651–3666 (2018).
- Bagan, H. & Yamagata, Y. Analysis of urban growth and estimating population density using satellite images of nighttime lights and land-use and population data. *GIScience Remote Sens.* **52**(6), 765–780 (2015).
- Lacroix, P., Handwerker, A. L. & Bièvre, G. Life and death of slow-moving landslides. *Nat. Rev. Earth Environ.* **1** (8), 404–419 (2020).
- Vakhshoori, V. & Zare, M. Is the ROC curve a reliable tool to compare the validity of landslide susceptibility maps? *Geomatics Nat. Hazards Risk*. **9** (1), 249–266 (2018).
- Chowdhuri, I. et al. Ensemble approach to develop landslide susceptibility map in landslide dominated Sikkim Himalayan region, India. *Environ. Earth Sci.* **79**, 1–28 (2020).
- Di Napoli, M. et al. Machine learning ensemble modelling as a tool to improve landslide susceptibility mapping reliability. *Landslides* **17** (8), 1897–1914 (2020).
- Agboola, G., Beni, L. H., Elbayoumi, T. & Thompson, G. Optimizing landslide susceptibility mapping using machine learning and geospatial techniques. *Ecol. Inf.* **81**, 102583 (2024).
- Li, L. et al. A modified frequency ratio method for landslide susceptibility assessment. *Landslides* **14**, 727–741 (2017).

35. Nakileza, B. R. & Nedala, S. Topographic influence on landslides characteristics and implication for risk management in upper Manafwa catchment. *Mt. Elgon Uganda Geoenviron. Disasters*. **7**, 1–13 (2020).
36. Intrieri, E., Carlà, T. & Gigli, G. Forecasting the time of failure of landslides at slope-scale: A literature review. *Earth Sci. Rev.* **193**, 333–349 (2019).
37. Celtek, S. The effect of aspect on landslide and its relationship with other parameters. In *Landslides*. IntechOpen (2021).
38. Ohlmacher, G. C. Plan curvature and landslide probability in regions dominated by earth flows and earth slides. *Eng. Geol.* **91** (2–4), 117–134 (2007).
39. Henriques, C., Zêzere, J. L. & Marques, F. The role of the lithological setting on the landslide pattern and distribution. *Eng. Geol.* **189**, 17–31 (2015).
40. Qin, Y. et al. Spatial distribution of near-fault landslides along Litang fault zones, eastern tibetan Plateau. *Geomorphology* **455**, 109189 (2024).
41. Zou, Y. et al. Factors controlling the spatial distribution of coseismic landslides triggered by the mw 6.1 Ludian earthquake in China. *Eng. Geol.* **296**, 106477 (2022).
42. Sun, H. et al. Influence of spatial heterogeneity on landslide susceptibility in the transboundary area of the Himalayas. *Geomorphology* **433**, 108723 (2023).
43. Shi, Y., Zhang, Z., Xue, C. & Feng, Y. Machine learning prediction of Co-seismic Landslide with Distance and Azimuth instead of Peak Ground Acceleration. *Sustainability* **16** (19), 8332 (2024).
44. Kirschbaum, D., Kapnick, S. B., Stanley, T. & Pascale, S. Changes in extreme precipitation and landslides over high mountain Asia. *Geophys. Res. Lett.* **47**(4), e2019GL085347 (2020).
45. Pourghasemi, H. R., Moradi, H. R. & Fatemi Aghda, S. M. Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. *Nat. Hazards*. **69**, 749–779 (2013).
46. Li, Y., Wang, X. & Mao, H. Influence of human activity on landslide susceptibility development in the three gorges area. *Nat. Hazards*. **104**, 2115–2151 (2020).
47. Abdullah, D. M. & Abdulazeez, A. M. Machine learning applications based on SVM classification a review. *Qubahan Acad. J.* **1** (2), 81–90 (2021).
48. Izquierdo-Verdiguier, E. & Zurita-Milla, R. An evaluation of guided regularized Random Forest for classification and regression tasks in remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* **88**, 102051 (2020).
49. Shirvani, Z. A holistic analysis for landslide susceptibility mapping applying geographic object-based random forest: A comparison between protected and non-protected forests. *Remote Sens.* **12** (3), 434 (2020).
50. Sun, X. et al. Integrated decision-making model for groundwater potential evaluation in mining areas using the cusp catastrophe model and principal component analysis. *J. Hydrology: Reg. Stud.* **37**, 100891 (2021).
51. Cantarino, I., Carrion, M. A., Goerlich, F. & Martinez Ibañez, V. A ROC analysis-based classification method for landslide susceptibility maps. *Landslides* **16**, 265–282 (2019).
52. Wang, Y. et al. Utilizing deep learning approach to develop landslide susceptibility mapping considering landslide types. *Bull. Eng. Geol. Environ.* **83** (11), 1–19 (2024).
53. Pham, B. T., Bui, D. T., Prakash, I. & Dholakia, M. B. Hybrid integration of Multilayer Perceptron neural networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *Catena* **149**, 52–63 (2017).
54. Fang, Z., Wang, Y., Peng, L. & Hong, H. Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping. *Comput. Geosci.* **139**, 104470 (2020).
55. Elkhachy, I. et al. Landslide susceptibility mapping and management in Western Serbia: An analysis of ANFIS-and SVM-based hybrid models. *Front. Environ. Sci.* **11**, 1218954 (2023).
56. Agrawal, N. & Dixit, J. GIS-based landslide susceptibility mapping of the Meghalaya-Shillong Plateau region using machine learning algorithms. *Bull. Eng. Geol. Environ.* **82** (5), 170 (2023).
57. Miao, F. et al. Landslide dynamic susceptibility mapping base on machine learning and the PS-InSAR Coupling Model. *Remote Sens.* **15** (22), 5427 (2023).
58. Kaur, L., Rishi, M. S., Singh, G. & Thakur, S. N. Groundwater potential assessment of an alluvial aquifer in Yamuna sub-basin (Panipat region) using remote sensing and GIS techniques in conjunction with analytical hierarchy process (AHP) and catastrophe theory (CT). *Ecol. Ind.* **110**, 105850 (2020).
59. Hu, J. & Szymczak, S. A review on longitudinal data analysis with random forest. *Brief. Bioinform.* **24** (2), bbad002 (2023).
60. Pisner, D. A. & Schnyer, D. M. Support vector machine. In *Machine Learning* (eds Pisner, D. A. & Schnyer, D. M.) 101–121 (Academic, 2020).
61. Papacharalampous, A. E. & Vlahogianni, E. I. Modeling microscopic freeway traffic using cusp catastrophe theory. *IEEE Intell. Transp. Syst. Mag.* **6** (1), 6–16 (2014).
62. Gaye, B., Zhang, D. & Wulamu, A. Improvement of support vector machine algorithm in big data background. *Math. Probl. Eng.* **2021** (1), 5594899 (2021).
63. Sun, D., Xu, J., Wen, H. & Wang, D. Assessment of landslide susceptibility mapping based on bayesian hyperparameter optimization: A comparison between logistic regression and random forest. *Eng. Geol.* **281**, 105972 (2021).

## Author contributions

W.Z. and Y.Z. are co-first authors of the article. Conceptualization, W.Z., Y.Z. and S.L.; methodology, W.Z. and Y.Z.; software, W.Z. and S.L.; validation, W.Z. and Y.Z.; formal analysis, Y.Z.; investigation, W.Z. and Y.Z.; data curation, W.Z.; writing—original draft preparation, Y.Z.; writing—review and editing, W.Z., Y.Z., X.S. and H.D.; visualization, Y.Z.; supervision, X.S. and S.L.; funding acquisition, X.S. All authors reviewed the manuscript.

## Funding

This work was supported by the National Natural Science Foundation of China (42401387), Department of Science and Technology of Sichuan Province (2020YFS0530), and Natural Science Foundation of Sichuan Province (2024NSFSC0786).

## Declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to S.L. or X.S.

**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) 2025