sustainability

MDPI

Article

Towards Sustainable Development: Landslide Susceptibility Assessment with Sample Optimization in Guiyang County, China

Yuzhong Kong [1,2,†], Kangcheng Zhu [2,3,†], Hua Wu [1,2,*], Chong Xu [4,5], Ze Meng [1,2], Hui Kong [6], Wen Tan [6], Xiangyun Kong [1,2], Xingwang Chen [1,2], Linna Chen [1,2] and Tong Xu [7]

1 School of Engineering, Xi Zang University, Lhasa 850032, China; kongyuzhong@stu.utibet.edu.cn (Y.K.)
2 Bomi Geological Hazards Field Scientific Observation and Research Station of the Ministry of Education, Bomê, Nyingchi 860300, China
3 School of Ecology and Environment, Xizang University, Lhasa 850032, China
4 National Institute of Natural Hazards, Ministry of Emergency Management of China, Beijing 100085, China
5 Key Laboratory of Compound and Chained Natural Hazards Dynamics, Ministry of Emergency Management of China, Beijing 100085, China
6 Geospatial Survey and Monitoring Institute of Hunan Province, Changsha 410129, China
7 College of Earth and Planetary Sciences, Chengdu University of Technology, Chengdu 610059, China
* Correspondence: xzwhua@utibet.edu.cn
† These authors contributed equally to this work.

Abstract

Here we present a high-resolution landslide susceptibility model for Guiyang County, China, developed to support sustainable disaster risk management. Our approach couples optimized positive and negative training samples with an ensemble of machine-learning algorithms to maximize predictive fidelity. We compiled a georeferenced inventory of 146 landslides by integrating historical records with systematic field validation. Sample optimization was central to our methodology: landslide presence points were refined via buffer-based dilution, and four classifiers—SVM, LDA, RF, and ET—were trained with identical covariate sets to ensure comparability. Three strategies for selecting pseudo-absences—buffering, low-slope filtering, and coupling with the IOE—were benchmarked. The Slope-IOE-O model, which synergizes low-gradient screening with entropy-weighted sampling, yielded the highest predictive capacity (AUC = 0.965). SHAP-based interpretability revealed that slope, monthly maximum rainfall, surface roughness, and elevation collectively dominate susceptibility, with pronounced non-linearities and interactions. Slope contribution peaks at 20–30°, monthly maximum rainfall exhibits a critical threshold near 225 mm, and the synergy between high roughness and road density amplifies landslide risk. Spatially, susceptibility follows a pronounced north–south gradient, with high-hazard corridors aligned along northern and southern mountain belts and the urban core of southern Guiyang County. By integrating rigorously curated training data with robust machine-learning workflows, this study provides a transferable framework for proactive landslide risk assessment, offering scientific support for sustainable land-use planning and resilient development in mountainous regions.

Keywords: landslide susceptibility assessment; sample optimization; machine learning; Guiyang County; SHAP; sustainable disaster-risk reduction

check for updates

Academic Editor: Yunfeng Ge

Received: 4 October 2025
Revised: 20 October 2025
Accepted: 23 October 2025
Published: 28 October 2025

Citation: Kong, Y.; Zhu, K.; Wu, H.; Xu, C.; Meng, Z.; Kong, H.; Tan, W.; Kong, X.; Chen, X.; Chen, L.; et al. Towards Sustainable Development: Landslide Susceptibility Assessment with Sample Optimization in Guiyang County, China. *Sustainability* **2025**, *17*, 9575. https://doi.org/10.3390/su17219575

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Landslides are a type of geological disaster, characterized by their high incidence, diversity, and widespread impact on a global scale [1,2]. They not only cause significant

destruction to the natural environment but also exert profound impacts on human society. With the intensification of global climate change and the advancement of human engineering projects, the risk of landslide disasters is further exacerbated [3,4]. In 2024, a total of 5719 geological disasters occurred in China, of which 3316 were landslides, accounting for 58% of all geological disasters [5]. While landslides are influenced by both natural factors and human activities, their inherent natural processes are uncontrollable and subject to long-term dynamics, making complete elimination of landslide hazards unattainable [6,7]. However, we can predict and mitigate the impacts of landslides through scientific methods and technological means, thereby minimizing the losses caused by these disasters. Conducting landslide susceptibility assessments is an effective approach to addressing this issue.

Today, landslide-susceptibility studies rely chiefly on three approaches: expert rules, physical simulation, and data-driven algorithms [8,9]. The extensive dependence on expert knowledge in experience-driven models makes them susceptible to subjectivity, which constrains their generalizability and practical application [10]. Physics-based models demand exhaustive hydrological and geotechnical inputs, so they are rarely scaled to regional mapping [11]. Because data-driven methods are objective and easily transferable, they have become the preferred choice for most researchers [12]. These empirical techniques split into classical statistical tests and an expanding family of machine-learning algorithms [13]. Statistical inference retrospectively dissects the parameter constellations that precipitated historical landslides [14]; by contrast, machine-learning algorithms autonomously learn and quantify non-linear, high-dimensional coupling patterns among predisposing factors, yielding accurate extrapolation of landslide susceptibility to unmapped terrain.

Machine learning algorithms suitable for landslide susceptibility mapping include both single models and ensemble models, with single models comprising linear regression [15], logistic regression [16,17], support vector machines (SVM) [18], and others, and ensemble models including random forests (RF) [19], gradient boosting trees [20], and others. In recent years, researchers have developed additional models by employing techniques such as boosting, bagging, stacking, voting, and blending of different models, such as XGBoost [21], LightGBM [22], Simple Stacking [23], Weighted Voting [24], Basic Blending [25], and others. However, different algorithm types vary in their- data processing methods and applicability. Currently, there is no consensus on a standardized methodology for landslide susceptibility assessment. We benchmarked two single models, SVM and linear discriminant analysis (LDA)—against two ensemble models—RF and extra trees (ET)—to systematically evaluate predictive performance.

A primary determinant of predictive accuracy in landslide susceptibility modeling is the quality, balance, and representativeness of the training samples, as they serve as the foundational basis for achieving reliable model performance [26]. For reliable landslide-susceptibility mapping, samples must capture the full spectrum of settings and traits tied to slope failure. Landslide samples, typically acquired through field surveys or remote sensing interpretation, are highly accurate. However, their spatial extent is significantly smaller than that of non-landslide areas, resulting in a substantial sample imbalance [27]. Landslides are episodic and sparse, especially in certain regions; this shortage skews machine-learning performance. Balancing and augmenting the landslide set is therefore imperative.

Currently, the two most used methods for optimizing landslide samples are oversampling and synthetic sample generation [28]. Oversampling offers a simple way to improve model performance on the minority class by augmenting its samples. Nevertheless, the potential for introduced redundancy necessitates careful consideration during implementation [29]. Synthetic sample generation is more complex in its implementation, generating new synthetic samples by interpolating between minority class samples, which

increases sample diversity and reduces the risk of overfitting, However, the generated synthetic samples may not fully conform to the actual distribution, particularly when the data distribution is complex, or may result in an excessive number of synthetic samples being generated in dense minority class regions, leading to overfitting in these areas [30]. Therefore, both optimization methods for landslide samples have their respective advantages and disadvantages. To address the scarcity and imbalance between landslide and non-landslide samples, this study employed a straightforward oversampling approach. Non-landslide samples are generally not directly obtainable and are typically acquired indirectly by avoiding potential landslide points using various methods. Currently, No consensus yet exists on how non-landslide points should be chosen; the three prevailing strategies are the buffer-distance, factor-constrained, and coupled-model methods [31]. The buffer size in the buffer distance method is difficult to determine, and both excessively large and small buffer sizes can affect the classification of landslide susceptibility [32]. Caution must be exercised in buffer zone sizing for non-landslide sample selection. An insufficient buffer can incorporate pseudo non-landslide samples—areas geomorphologically similar to landslide sites—which dilutes the model's predictive capability and causes an underestimation of susceptibility. Conversely, an excessively large buffer size may cause non-landslide samples to be overly confined in the environmental feature space, resulting in poor global representativeness of the non-landslide sample set and consequently an overestimation of landslide hazard [33]. Besides the factor-constrained selection method—which incorporates multiple factors, notably slope, for effective non-landslide sample selection [34,35]—coupled models (e.g., information value with frequency ratio) have also achieved commendable results [36,37]. Despite their ability to enhance model accuracy, these methods carry a risk: the selected non-landslide samples may fall within geomorphologically susceptible areas, ultimately compromising the predictive performance of the machine learning models. It is widely held that higher purity in non-landslide samples enables a prediction model to extract more accurate classification features from the conditioning factors, thereby improving its performance [38]. Therefore, this study employs these three methods and their three combined approaches to optimize non-landslide samples, aiming to identify the optimal method for optimizing non-landslide samples and thereby enhance the accuracy of machine learning.

This study develops a landslide susceptibility assessment method based on sample optimization and machine learning for areas with limited landslide data. First, both landslide and non-landslide samples were optimized to identify the most effective strategy. Then, the optimal hyperparameters for four models (SVM, LDA, RF, and ET) were determined using a grid search, and their susceptibility maps were generated. Finally, the models were comprehensively evaluated using the AUC value, Accuracy, Precision, Recall, F1-score, and landslide frequency ratio [39,40].

2. Materials and Methods

2.1. Study Area

2.1.1. Topographical and Geological Conditions

Guiyang County is in the southwest of Chenzhou, in southeastern Hunan Province, with longitude ranging from 112°13′26″ E to 112°55′46″ E and latitude from 25°27′15″ N to 26°13′30″ N. As the largest and most populous county in Chenzhou City, Guiyang County covers a total area of 2958.61 km^2 [41] (Figure 1). Guiyang County is situated on the northern side of the Nanling Mountains, with the Tashan and Dayishan Mountains in the north, the northern foothills of Qitianling in the south, and extensive hilly and upland areas in the middle. The regional topography is characterized by elevated northern and southern parts and a depressed central area, resulting in a saddle-shaped relief. The landscape

comprises rugged ridges and deeply incised gullies, rising from 59 m to 1400 m within a relief of 1341 m and locally exceeding 70°. This extreme steepness substantially undermines slope stability and precipitates a sharp rise in landslide probability [41].

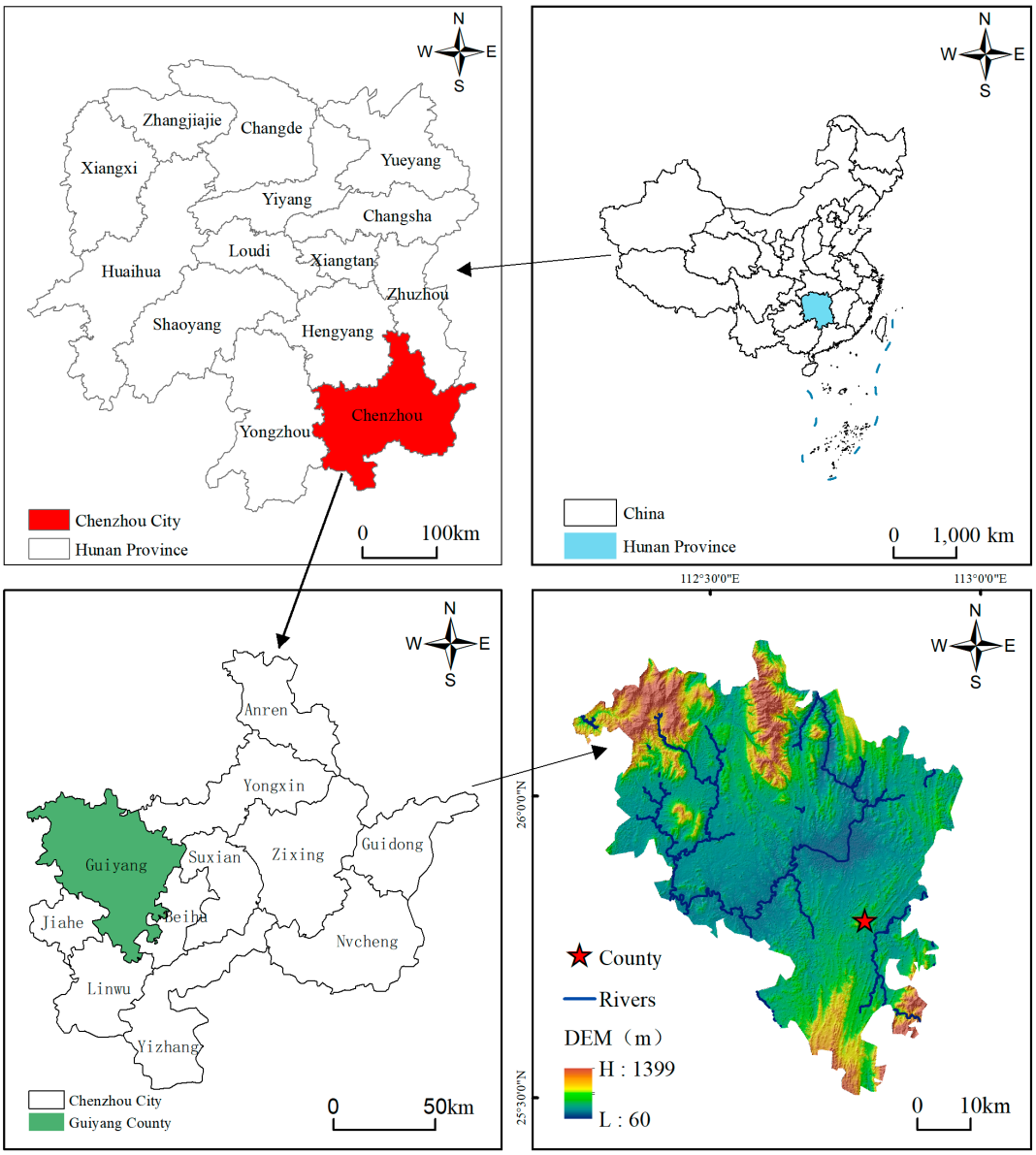


Figure 1. Location of the study area.

Guiyang County lies at the intersection of radial Leiyang–Linwu and NE-trending Xinhua-Xia structures. Faults and folds strike ~20°, especially in Huangshaping mine; the NE corner belongs to the Yong–Chen fold belt. Outcrops are Upper Paleozoic mudstone, shale, siltstone, limestone and dolomite, capped by thin Quaternary deposits. Under humid subtropical weathering, soft mudstone/shale hydrate and argillite, creating weak bedding planes; preferential erosion undercuts competent limestone, promoting toppling and planar slides. Karstic limestone is dissected by joints and faults that channel groundwater, further lowering shear strength. Intensive extraction of more than 60 minerals has damaged rock mass and hydrology [41]. Together, tectonic weak layers, karst groundwater and mining disturbance make slopes highly unstable and landslides frequent.

Guiyang County experiences a humid subtropical monsoon climate, receiving 1400–1800 mm of precipitation annually, 55% of which is concentrated between April

and June when 3-h intensities can exceed 80 mm; these extreme rainfall events constitute a primary trigger of landslides. These mass-movements pose direct threats to critical linear infrastructure—including the Gui-Xin Expressway and the Ganzhou–Chenzhou Railway—and to strategic hydraulic assets such as the Ouyang-Hai Reservoir irrigation scheme. Consequently, quantitative landslide-susceptibility mapping is essential for delineating high-priority zones, optimizing corridor alignment, and designing slope-stabilization strategies that safeguard regional water security and underpin sustainable socio-economic development.

2.1.2. Hydrological and Climatic Conditions

Guiyang County has a well-developed river system, with three primary rivers—Chongling River, Baishui River, and Yi River—56 secondary rivers, 22 tertiary rivers, and 8 quaternary rivers, with a total river length of 1363 km, a river network density of 0.46 km/km^2, and an annual runoff volume of 2.03 billion cubic meters [42]. The dense river network may lead to soil saturation during heavy rainfall, further reducing slope stability. Guiyang County sits in a subtropical monsoon zone, characterized by concentrated precipitation and frequent heavy rainstorms. The annual precipitation is unevenly distributed, with significant concentration in the summer and frequent heavy rainstorms. Heavy rainfall is a primary trigger for landslides, as it raises pore water pressure and weakens the shear strength of slopes, thereby increasing disaster frequency.

2.2. Data Sources

2.2.1. Landslide Relic Data Sources

The analysis of landslide relic data is essential for understanding their spatial distribution and predicting future occurrences, thus holding vital importance for landslide hazard mitigation studies. The sources of landslide relic data mainly fall into two categories: First, data can be obtained and collected through field surveys, which yield accurate and reliable information but are time-consuming and labor-intensive, especially in remote or inaccessible areas. Second, data can be derived from the interpretation of high-resolution satellite images, which offers broad coverage and allows for rapid acquisition of landslide relic data over large areas at relatively low cost. However, the accuracy of interpretation is influenced by image resolution, weather conditions, and interpretation techniques, making it difficult to identify landslide relics in complex terrains or areas with dense vegetation [43,44]. Given that Guiyang County is in a subtropical monsoon climate zone, the vegetation throughout the area is extremely lush, which significantly affects landslide interpretation. Therefore, this study employs field surveys to obtain landslide relic data, with a total of 146 landslide relics identified in study area (Table 1). The landslide points were mapped onto the Digital Elevation Model (DEM) using ArcGIS 10.8 (Figure 2).

Table 1. Landslide data table for Guiyang County.

No.	Date	Type	V. (m^3)	No.	Date	Type	V. (m^3)	No.	Date	Type	V. (m^3)	No.	Date	Type	V. (m^3)
1	2001/5/	Soil	4500	38	2006/7/	Soil	24,000	75	2014/6/	Soil	1200	112	2005/5/	Soil	14,000
2	2004/5/	Soil	2250	39	2014/3/	Rock	5200	76	2001/5/	Soil	28,800	113	2013/5/	Soil	1500
3	2013/5/	Soil	4500	40	2002/8/	Soil	24,000	77	2008/6/	Soil	8000	114	2014/5/	Soil	100
4	2012/7/	Soil	6000	41	2002/8/	Soil	12,000	78	2004/6/	Soil	12,800	115	2013/6/	Soil	200
5	2008/6/	Soil	8000	42	2014/6/	Rock	1440	79	2014/5/	Soil	120	116	2012/6/	Soil	50
6	2013/6/	Soil	108	43	2014/3/	Soil	3000	80	1996/5/	Soil	4440	117	2013/5/	Soil	320
7	2014/4/	Soil	720	44	2014/6/	Soil	1950	81	2010/4/	Soil	1200	118	2013/5/	Soil	150
8	2006/7/	Soil	12,000	45	2014/5/	Soil	2550	82	2014/3/	Soil	150	119	2013/5/	Soil	8000
9	2007/7/	Soil	720	46	2002/5/	Soil	5400	83	2013/3/	Soil	300	120	2007/3/	Soil	300
10	2004/5/	Soil	3600	47	2001/6/	Soil	20,000	84	2002/6/	Soil	42,970	121	2013/6/	Soil	9600
11	2006/7/	Soil	3000	48	2002/6/	Soil	24,000	85	2003/6/	Soil	1200	122	2013/6/	Soil	2400
12	2006/7/	Soil	5000	49	2002/4/	Soil	28,800	86	2003/6/	Soil	18,000	123	2010/6/	Soil	19,200

Table 1. *Cont.*

No.	Date	Type	V. (m^3)	No.	Date	Type	V. (m^3)	No.	Date	Type	V. (m^3)	No.	Date	Type	V. (m^3)
13	2000/6/	Soil	1000	50	2006/7/	Soil	5400	87	2010/6/	Soil	9600	124	2002/4/	Soil	8000
14	2008/2/	Soil	1360	51	2002/6/	Soil	22,000	88	2005/5/	Soil	4935	125	2001/3/	Soil	21,600
15	1998/5/	Soil	225	52	2012/3/	Soil	2400	89	2014/4/	Soil	8000	126	2003/5/	Soil	28,000
16	2005/5/	Soil	400	53	2010/6/	Soil	4800	90	2010/6/	Soil	4500	127	2013/4/	Rock	2100
17	1997/5/	Soil	1800	54	2002/6/	Soil	40,000	91	2012/6/	Soil	16,000	128	2012/6/	Soil	16,800
18	2005/5/	Soil	5250	55	2013/6/	Soil	3600	92	2006/7/	Soil	147,450	129	2001/6/	Soil	12,800
19	2003/7/	Rock	1000	56	2002/8/	Soil	30,600	93	2005/6/	Soil	9600	130	2006/7/	Soil	3400
20	2001/5/	Soil	3600	57	1992/4/	Soil	3200	94	1996/8/	Soil	15,000	131	1998/6/	Soil	12,000
21	2004/5/	Soil	6240	58	2013/6/	Soil	3120	95	1998/5/	Soil	1800	132	2002/6/	Soil	8820
22	2012/4/	Soil	18,800	59	2005/5/	Soil	7500	96	2006/7/	Soil	3600	133	2002/6/	Soil	14,640
23	2014/3/	Soil	400	60	2002/6/	Soil	19,200	97	1997/8/	Soil	3000	134	2002/6/	Soil	3200
24	2014/4/	Soil	28,710	61	2002/6/	Soil	460	98	2003/5/	Soil	16,000	135	2002/6/	Soil	14,000
25	2014/3/	Soil	140	62	2011/4/	Soil	720	99	1994/6/	Soil	10,000	136	1998/6/	Soil	1720
26	2014/4/	Soil	24,000	63	1999/7/	Soil	3080	100	2007/6/	Soil	12,800	137	2002/6/	Soil	28,000
27	2014/3/	Rock	1050	64	1992/5/	Soil	11,200	101	2007/5/	Soil	400	138	2002/6/	Soil	11,200
28	2014/4/	Soil	7050	65	2002/5/	Soil	600	102	1992/7/	Soil	1360	139	1998/5/	Soil	14,400
29	2014/5/	Soil	3000	66	1987/5/	Soil	1200	103	2004/6/	Soil	3360	140	1998/6/	Soil	7200
30	2014/5/	Soil	2100	67	2004/6/	Soil	1600	104	2006/6/	Soil	13,200	141	2002/6/	Soil	8000
31	2014/6/	Soil	1950	68	1979/6/	Soil	2400	105	1998/5/	Soil	1400	142	2002/6/	Soil	12,000
32	2003/12/	Soil	8160	69	1980/7/	Soil	1050	106	1992/4/	Soil	4560	143	2002/6/	Soil	4000
33	2006/7/	Soil	10,800	70	2012/6/	Soil	1080	107	1996/6/	Soil	2500	144	1998/6/	Soil	9600
34	2006/7/	Soil	8800	71	1997/6/	Soil	1500	108	2014/5/	Soil	3200	145	2014/5/	Soil	338
35	2006/7/	Soil	19,200	72	2002/7/	Soil	10,000	109	1998/3/	Soil	1,920,000	146	2013/9/	Soil	2280
36	1996/8/	Soil	4000	73	1972/5/	Soil	60,000	110	1991/5/	Soil	9600				
37	2006/7/	Soil	8000	74	2004/6/	Soil	16,000	111	2012/4/	Soil	2000				

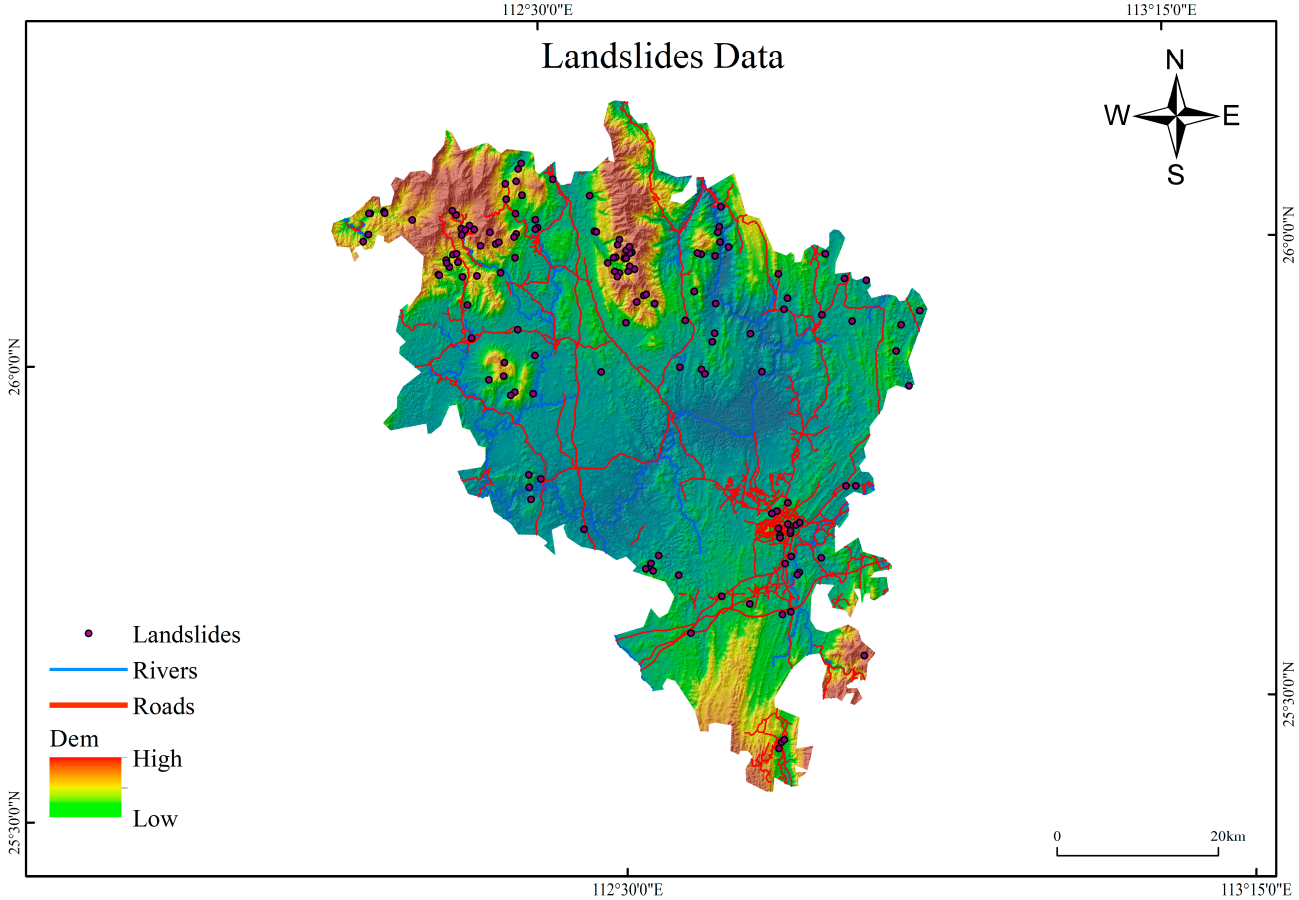


Figure 2. Landslide relic dataset.

As shown in Figure 2, landslides in Guiyang County are primarily distributed in the relatively high-altitude northern and southern mountainous regions. In addition, the southern area surrounding the county town, which has a higher road density, also

experiences a higher frequency of landslides, while the central region, characterized by lower altitude and relatively flat terrain, has fewer landslides.

2.2.2. Data Sources of Landslide Conditioning Factors

Landslide conditioning factors can primarily be categorized into four major groups: Terrain, Hydrology, Land cover, and Lithology, with most studies selecting around 10 easily accessible factors [45,46]. Different conditioning factors may vary in their contributions across different study areas or models, and some researchers have summarized the usage frequency of each conditioning factor [27]. For a comprehensive representation of landslide characteristics, 18 conditioning factors with high and moderate usage frequencies were selected to establish a landslide susceptibility assessment system for Guiyang County. The corresponding data sources are detailed in Table 2.

Table 2. Data sources of landslides impact factors.

Conditioning Factor	Name of the Data	Resolution/Scale	Data Type	Data Source
Density of fault (DOF), Lithology Type	Geological map of China	1:50,000	Vector	https://www.ngac.cn/
Elevation, Slope, Aspect, Profile Curvature, Plane Curvature, Terrain Wetness Index (TWI), Stream Power Index (SPI): Roughness, Cutting-Depth, Relief, Elevation Coefficient of Variation (CV)	Spatial resolution DEM data for China	30 m	Raster	http://www.gscloud.cn/
Density of road (DOR), Density of steam (DOS)	Basic geographic data of river systems, roads, and administrative boundaries	1:1,000,000	Vector	http://www.gscloud.cn/
Rainfall	The dataset of precipitation in China from 1991 to 2020	30 m	Raster	http://www.gisrs.cn/
Normalized Difference Vegetation Index (NDVI)	NDVI data for China in 2020	30 m	Raster	http://www.nesdc.org.cn/
Land use	Land cover data for China in 2020	30 m	Raster	https://www.ncdc.ac.cn/

2.3. Mapping Units

In landslide susceptibility prediction, slope units and grid units are two commonly used types of mapping units, Slope units are typically delineated using hydrological information from ridge lines and valley lines. This method can incorporate certain topographic and geomorphological information into the slope units, but there is no unified standard for the parameters used during the generation process, and the process is highly complex, particularly when manual corrections to the slope units are required, which can be very time-consuming [47,48]. Thus, grid cells are adopted as the mapping units in this study. The selection of grid cell size is crucial for both model accuracy and computational efficiency. A grid cell size that is too small can lead to excessive computational load, while a size that is too large may fail to accurately capture the spatial distribution characteristics of landslides. An established empirical formula exists for determining the appropriate grid cell size [49].

$$G_s = 7.49 + 6 \times 10^4 S - 2 \times 10^{-9} S^2 + 2.9 \times 10^{-15} S^3 \quad (1)$$

where G_s denotes the appropriate grid size, and S is the denominator of the map scale.

This study employs a survey scale of 1:50,000. Using Equation (1), the theoretical grid size (G_s) is calculated as 32.853 m; however, a practical size of 30 m × 30 m was adopted

for the evaluation unit to facilitate data analysis. This resulted in a total of 3,289,874 grid cells for Guiyang County.

2.4. Landslide Conditioning Factors

A uniform coordinate system was applied to all landslide conditioning factors, using a Universal Transverse Mercator (UTM) projection referenced to zone 49N (based on the 3° division scheme). The classification was performed as follows: discrete variables (i.e., lithology, land use, aspect) were categorized based on observational criteria; continuous variables were classified employing the natural breaks method (Figure 3).

Elevation: Elevation is one of the fundamental characteristics of terrain, Elevation itself does not directly trigger landslides but indirectly alters the types and frequencies of landslides by influencing climate, hydrology, vegetation, and human activities. The elevation data for Guiyang County was obtained from a national DEM with a spatial resolution of 30 m. The elevation, ranging from 59 to 1400 m, was categorized into five classes: 59–249 m, 249–393 m, 393–584 m, 584–826 m, and 826–1400 m. The first two classes (below 393 m) were found to occupy 75% of the study area.

Slope: Slope measures the steepness of a terrain. As the slope angle increases, so does the downslope gravitational force, making it easier to overcome the shear resistance of the material and consequently raising the probability of a landslide. Using a 30-m DEM, slopes were categorized into five classes: 0–6.86°, 6.86–13.73°, 13.73–21.79°, 21.79–32.53°, and 32.53–76.10°. The classification scheme resulted in the first two gentle slope classes (0–13.73°) collectively occupying 65% of the study area.

Aspect: Aspect refers to the geographical direction that a slope faces, which influences solar radiation, precipitation distribution, vegetation types, and weathering processes, thereby indirectly affecting slope stability, and it is generally believed that sun-facing slopes (aspect) are more prone to landslides than shaded slopes. Aspect was derived from a 30-m resolution DEM and classified into eight directional classes: north, northeast, east, southeast, south, southwest, west, and northwest.

Profile Curvature: Profile Curvature refers to the terrain curvature in the direction of slope gradient, which describes the concave and convex shapes of the slope in the vertical direction. Using a 30-m DEM, profile curvature was categorized into five classes: −37.08 to −4.65, −4.65 to −1.44, −1.44 to 0.81, 0.81 to 3.70, and 3.70 to 44.81, where the range from −1.44 to 3.70 accounts for 80%.

Plane Curvature: Plane Curvature refers to the terrain curvature in the direction of contour lines, which describes the bending shape of the slope in the horizontal direction. Using a 30-m DEM, plane curvature was categorized into five classes: −25.78 to −2.96, −2.96 to −0.98, −0.98 to 0.33, 0.33 to 2.09, and 2.09 to 30.18, where the range from −0.98 to 2.09 accounts for 83%.

Lithology Type: Lithology type refers to the material composition and engineering geological properties of surface or subsurface rocks and soils, which directly influences key stability parameters of slopes, such as shear strength, permeability, and weathering rate. The lithology types were interpreted from the 1:50,000 geological map of Hunan Province. Following the Chinese Engineering Rock Mass Classification Standard (GB/T 50218-2014 [50]), the lithologies were classified into five groups: Harder Rock (e.g., quartzite, basalt, granite), Hard Rock (e.g., slate, limestone), Weak Rock (e.g., dolomite, conglomerate), Weaker Rock (e.g., argillaceous limestone, sandstone, siltstone), and Loose Rock (e.g., silty clay). Collectively, Hard Rock and Weak Rock account for 88% of the study area.

DOF: DOF is a structural metric representing the total length of faults per unit area (km/km^2), thus quantifying the regional intensity of tectonic fragmentation. The higher the fault density, the more intense the tectonic influence on the rock mass, and the poorer the

stability is generally. The fault density data were derived from the 1:50,000 geological map of Hunan Province, with fault information extracted, and fault density was calculated using the fishnet method, divided into five categories: 0–0.25, 0.25–0.76, 0.76–1.23, 1.23–1.76, and 1.76–3.16.

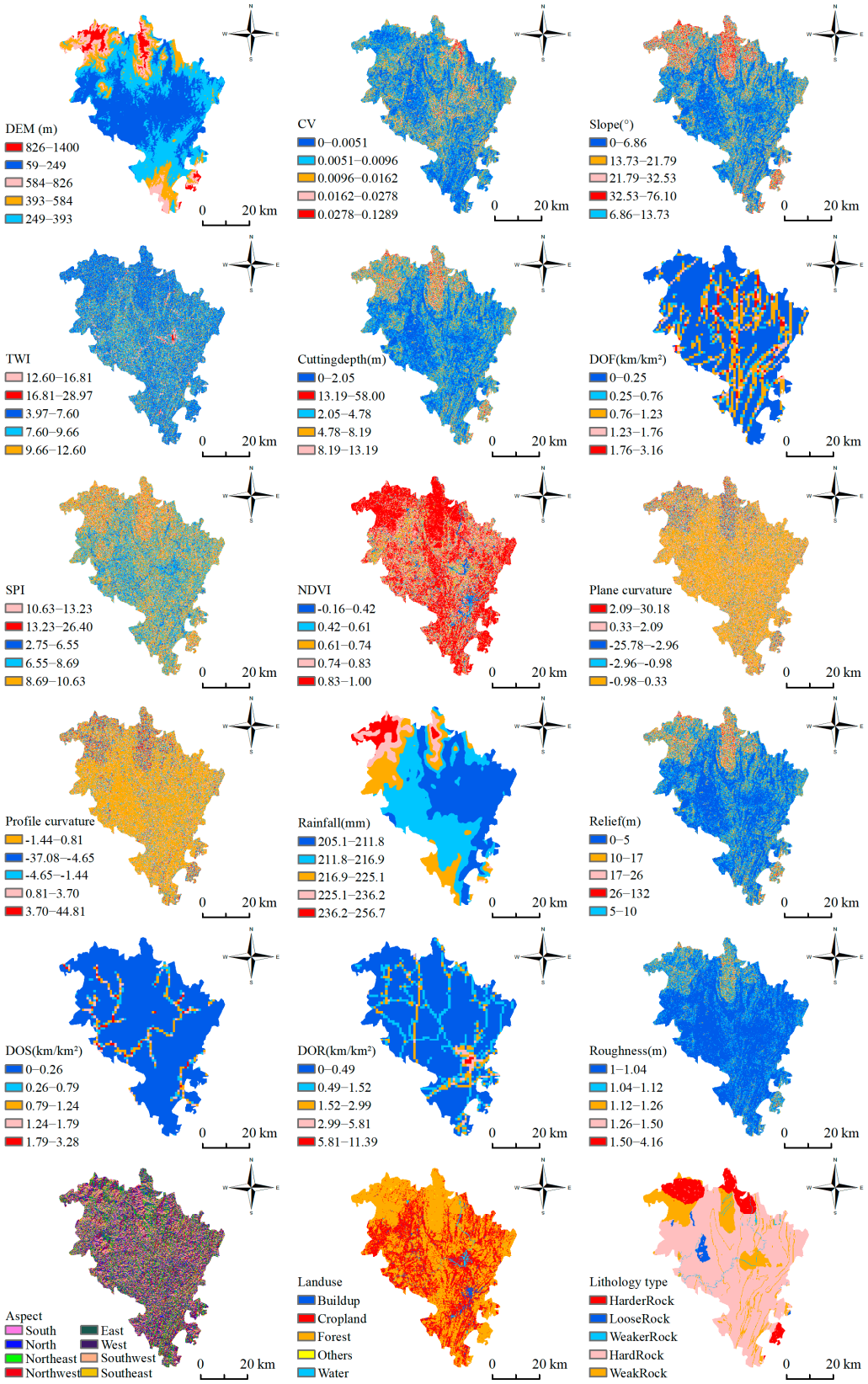


Figure 3. Grading diagram of conditioning factors.

Rainfall: Monthly Maximum Rainfall (unit: mm) is defined as the intensity of extreme precipitation events and is a major landslide trigger [51]. It functions through the following mechanism: heavy rainfall saturates the slope, increasing pore water pressure and reducing both effective stress and shear strength, which ultimately causes landslides. Based on the collected landslide dataset, Guiyang County experiences the highest number of landslides in June, which also coincides with the month of the highest monthly rainfall in the county. Therefore, this study utilizes the national monthly maximum rainfall data from 1991 to 2020, divided into five categories: 205.1–211.8, 211.8–216.9, 216.9–225.1, 225.1–236.2, and 236.2–256.7, with the range of 205.1–216.9 accounting for 73% of the area.

DOS: DOS is a hydrological metric representing the total length of rivers per unit area (km/km^2), thus reflecting the degree of surface runoff development and the regional erosive capacity of the river network. Areas with high DOS typically exhibit strong terrain dissection and active hydrological activity, which significantly affect slope stability. This study extracted river vector data from basic geographical data in the study area, and calculated river density using the fishnet method, divided into five categories: 0–0.26, 0.26–0.79, 0.79–1.24, 1.24–1.79, and 1.79–3.28.

NDVI: NDVI reflects vegetation status and cover density, and it exerts a dual influence on landslide activity. Positively, root systems reinforce soil shear strength, while canopy interception reduces rainfall infiltration and runoff. Plant transpiration also lowers soil moisture, suppressing pore water pressure buildup and thus slope instability. Conversely, the weight of large trees can increase slope load, and decaying roots may create preferential flow paths for water, potentially triggering deep-seated landslides. Using a 30-m DEM, NDVI were categorized into five classes: −0.16 to 0.42, 0.42 to 0.61, 0.61 to 0.74, 0.74 to 0.83, and 0.83 to 1.00, with the range of 0.74 to 1.00 accounting for 77% of the area.

TWI: The potential for water accumulation and soil moisture was assessed using the TWI. This index was computed from a 30-m resolution DEM and subsequently categorized into five classes: 3.97–7.60, 7.60–9.66, 9.66–12.60, 12.60–16.81, and 16.81–28.97. The combined area of the two lowest TWI classes (3.97–9.66) constitutes 75% of the total area.

Land Use Type: Dense forest cover significantly increases soil shear strength, thereby stabilizing slopes. Conversely, sparse vegetation—such as shrubs, grassland, and bare land—provides weak root reinforcement, resulting in low shear strength and a higher propensity for landslides. The land cover data were obtained from the 30-m annual dataset of China provided by the National Cryosphere Desert Data Center. A reclassification was performed, consolidating the original types into five categories: arable land, forest, water bodies, urban land, and other land uses (shrubs, grasslands, and bare land). Forests and arable land together constitute 97% of the total area.

DOR: DOR is a spatial metric representing the total length of the road network per unit area (km/km^2), thus quantifying the intensity of human engineering activities and associated surface modification. This study extracted road vector data from basic geographical data in the study area, and calculated road density using the fishnet method, divided into five categories: 0–0.49, 0.49–1.52, 1.52–2.99, 2.99–5.81, and 5.81–11.39.

SPI: SPI is defined as the rate of kinetic energy loss per unit flow width, which reflects the erosive power of surface runoff. This makes it a useful measure for estimating the potential for slope erosion and sediment transport. The SPI, extracted from a 30-m resolution DEM, was classified into five intervals: 2.75–6.55, 6.55–8.69, 8.69–10.63, 10.63–13.23, and 13.23–26.40. The classification resulted in the middle two classes (6.55–10.63) occupying 71% of the study area.

CV: The dispersion of elevation was assessed using the CV, defined as the ratio of the standard deviation to the mean. This index was computed from a 30-m resolution DEM. The resulting CV values were categorized into five intervals: 0–0.0051, 0.0051–0.0096,

0.0096–0.0162, 0.0162–0.0278, and 0.0278–0.1289. The two lowest categories (0–0.0096) were found to occupy 71% of the study area, indicating predominantly uniform terrain.

Roughness: Topographic roughness is a parameter that characterizes the micro-scale undulations of the terrain, reflecting the complexity of local topography, which is a key indicator for identifying potential landslide surfaces (unit: m). Calculated from a 30-m DEM, the topographic roughness was divided into five intervals (1–1.04, 1.04–1.12, 1.12–1.26, 1.26–1.50, 1.50–4.16), with the range of 1–1.12 covering 92% of the study area.

Cutting-Depth: The cutting depth, defined as the vertical distance from the current surface to the pristine geomorphic datum, was computed from a 30-m DEM. This metric, expressed in meters, serves as a core indicator of surface modification intensity for landslide risk evaluation in high mountain gullies. The computed values were categorized into five intervals: 0–2.05, 2.05–4.78, 4.78–8.19, 8.19–13.19, and 13.19–58. The combined area of the first three intervals (0–8.19 m) constitutes 92% of the region.

Relief: The relief amplitude, which quantifies the maximum elevation difference within a local analysis window, was computed from a 30-m resolution DEM. This metric serves as a core indicator for characterizing terrain dissection and for disaster mitigation planning in mountainous regions (units: meters). The calculated values were categorized into five intervals: 0–5, 5–10, 10–17, 17–26, and 26–132. The combined area of the low to moderate amplitude categories (0–17 m) constitutes 93% of the total area.

2.5. Data Correlation Analysis Method

Given that landslide susceptibility assessment can be affected by collinearity among conditioning factors, which reduces model reliability, a two-step screening process was utilized. First, multicollinearity was assessed using the Variance Inflation Factor (VIF > 5 indicates issues [52]). Second, pairwise linear correlations were examined using the Pearson coefficient ($|r| > 0.7$ suggests a strong relationship [53]). Factors flagged by either criterion were considered for removal to minimize information redundancy.

2.6. Landslide Susceptibility Assessment Method

2.6.1. IOE Model

The IOE model is a statistical model designed to assess the significance of various landslide conditioning factors. It achieves this by measuring the degree of disorder (entropy) within the data, which objectively reflects each factor's importance and its relative contribution to the overall assessment.

The following calculations and formulas are derived from the methodology presented in reference [40]:

$$FR_{ij} = \frac{a}{b} \tag{2}$$

$$P_{ij} = \frac{FR_{ij}}{\sum_{j=1}^{s} FR_{ij}} \tag{3}$$

$$H_i = -\sum_{j=1}^{s} P_{ij} \times log_2 P_{ij} \tag{4}$$

$$H_{i,max} = \log_2 S \tag{5}$$

$$I_i = \frac{H_{i,max} - H_i}{H_{i,max}} \tag{6}$$

$$P_i = \frac{1}{S} \sum_{j=1}^{s} FR_{ij} \tag{7}$$

$$W_i = I_i \times P_i \tag{8}$$

where the frequency ratio FR_{ij} refers to the landslide density within class j of factor i compared to the overall study area. Terms a and b signify the ratio of landslide points and the area ratio for that specific class, respectively, while P_{ij} is its probability density. The variable s denotes the total number of classes within the factor. Furthermore, H_i and $H_{i,max}$ represents the calculated entropy and the maximum possible entropy for factor i, respectively, from which the information coefficient I_i is derived. Finally, W_i is the comprehensive weight resulting from this calculation for each factor.

$$LSI = \sum_{i=1}^{n} FR \times W_i \tag{9}$$

where LSI represents the Landslide Susceptibility Index.

2.6.2. SVM Model

The SVM model [54] is a powerful supervised learning algorithm. Its core idea is to find the optimal hyperplane that maximizes the margin between different classes, thereby enhancing generalization ability. This maximum-margin hyperplane is determined by the closest data points, known as support vectors. The SVM can handle both linear and non-linear problems by employing kernel functions to map data into higher-dimensional spaces.

2.6.3. LDA Model

The LDA model [55] is a supervised dimensionality reduction and classification technique. It aims to find a linear projection that maximizes the separation between classes while minimizing the dispersion within each class. This is achieved by maximizing the ratio of between-class variance to within-class variance, under the assumption that all classes share a common covariance matrix and are normally distributed. LDA is particularly effective for datasets with high-dimensional features and a limited number of samples.

2.6.4. RF Model

RF model [56] is an ensemble learning algorithm, which enhances model accuracy and robustness by constructing multiple decision trees. It effectively handles high-dimensional data, is robust to missing values and outliers, and can assess feature importance. RF helps reduce the risk of overfitting, enhances model generalization capability, and is easily parallelizable, speeding up the training process. This method can handle both classification and regression problems while effectively modeling nonlinear data patterns. RF typically requires no complex parameter tuning, and performs well even with default settings, making it an ideal choice for solving complex data problems.

2.6.5. ET Model

ET model [57] is an ensemble method that performs classification or regression by building many randomized decision trees and aggregating their predictions. Unlike RF, the ET model introduces more randomness in tree construction, such as randomly selecting features and split points at each decision node. This extreme randomization makes the model more robust, enabling it to capture complex patterns and interactions in the data. The ET model is typically used for handling high-dimensional data, and it has some robustness against outliers and noise. Due to its simple structure and fast training speed, and its good performance on many datasets, the ET model has been widely used in practical applications.

2.6.6. Landslide Sample Optimization Methods

A landslide sample optimization procedure was performed to mitigate class imbalance and enhance model accuracy. First, a 30-m buffer zone was created around the original landslide points. Subsequently, synthetic samples, numbering three times the original landslide count, were randomly generated within this buffer. The final dataset, comprising both original and synthetic samples, contained a total of 584 landslide samples.

2.6.7. Non-Landslide Sample Optimization Methods

Six methodologies for non-landslide sample optimization were implemented in a comparative framework to determine the most effective approach (Figure 4). The methodologies are delineated as follows:

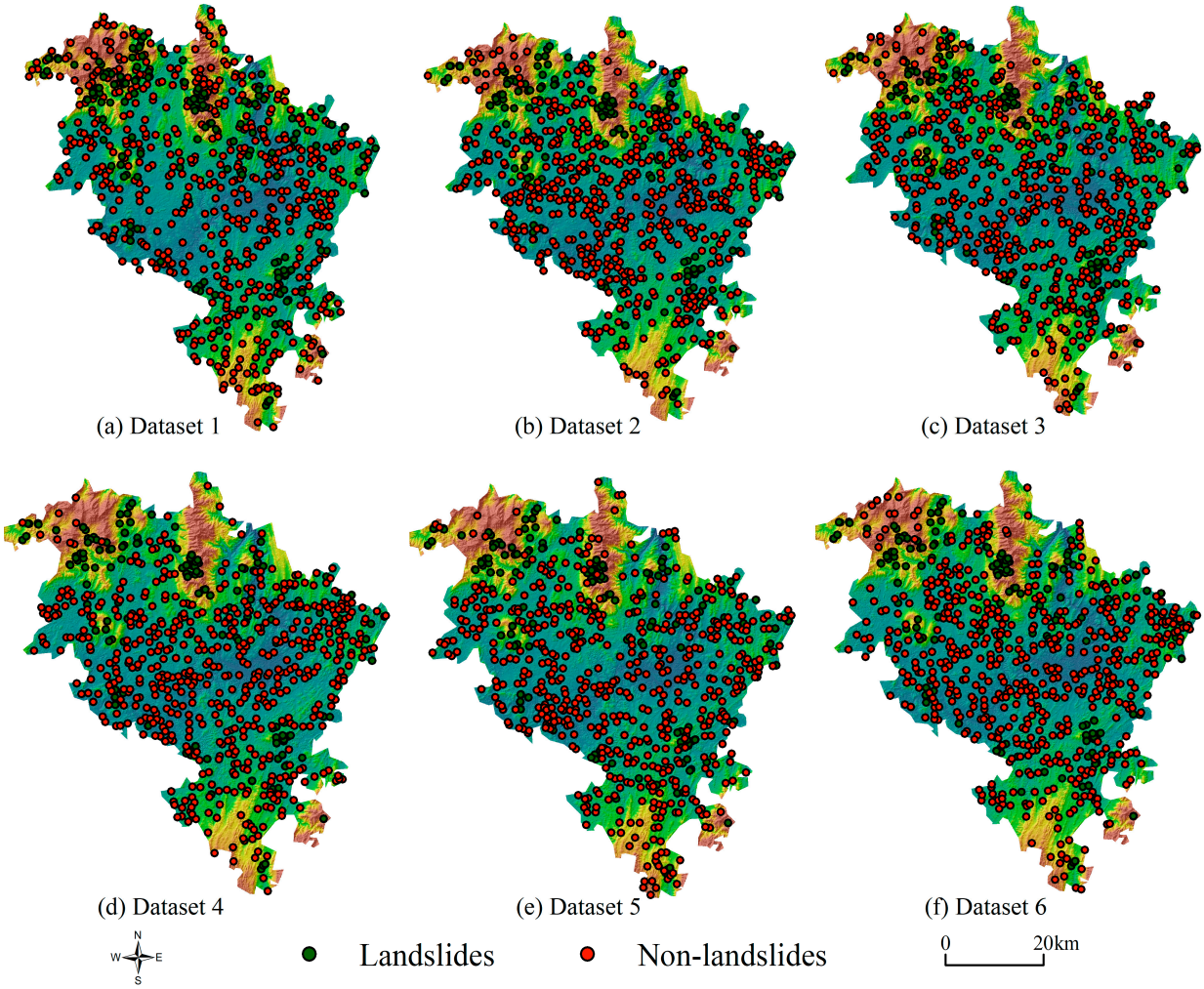


Figure 4. The datasets of different sample optimization methods.

Buf-O (Dataset 1): Non-landslide samples were randomly selected from beyond a 500-m buffer surrounding landslide inventories [32].

IOE-O (Dataset 2): Sampling was conducted within zones classified as very low or low susceptibility by the Index of Entropy (IOE) model.

Slope-O (Dataset 3): Samples were extracted from terrain characterized by low slope angles.

Slope-IOE-O (Dataset 4): Sampling was confined to the spatial overlap between low-slope areas and the IOE model's very low/low susceptibility zones.

Buf-Slope-O (Dataset 5): Samples were derived from the intersection of low-slope areas and regions external to the 500-m landslide buffer.

Buf-IOE-O (Dataset 6): Samples were selected from the intersection of the IOE model's very low/low susceptibility zones and the area outside the 500-m landslide buffer.

In all cases, the selected non-landslide samples were merged with the optimized landslide samples to form the final dataset for model training.

2.6.8. SHAP Feature Interpretation Method

The "black box" nature of many machine learning models often obscures their internal prediction processes. To mitigate this, SHapley Additive exPlanations (SHAP)—a method rooted in the game theory of Lloyd Shapley [58] and introduced to machine learning by Lundberg [59]—provides a powerful solution for interpretability. SHAP improves transparency and trust by quantifying the contribution of each input feature to the model's output. Its core strength lies in facilitating both global interpretation, which assesses overall feature importance across the dataset, and local interpretation, which deconstructs the prediction for a single sample. The formula for calculating Shapley values is as follows:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_x(S \cup \{i\}) - f_x(S)] \tag{10}$$

where ϕ_i denotes the Shapley value of the i-th feature, F represents the total number of features, $f_x(S \cup \{i\})$ represents the model prediction value for sample x when feature i is added to the feature subset S, $f_x(S)$ represents the model prediction value when only the feature subset S is used.

2.6.9. Validation Metrics

The classification models were evaluated based on the Receiver Operating Characteristic (ROC) curve and a confusion matrix. The ROC curve's position is related to the top-left corner and its corresponding Area Under the Curve (AUC) value were used to gauge model performance, with higher values indicating greater robustness and classification accuracy. Additionally, a confusion matrix was utilized to compare predicted classifications with actual values, enabling the computation of Accuracy (Acc), Precision, and the F1-score—the harmonic mean of precision and recall. The formulas for these metrics are as follows [40]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

In the confusion matrix, TP (True Positive) and TN (True Negative) represent the counts of correctly classified positive and negative instances, respectively. Conversely, FP (False Positive) denotes the count of negative instances misclassified as positive, while FN (False Negative) refers to the count of positive instances misclassified as negative.

2.7. Landslide Susceptibility Assessment Workflow

2.7.1. Assessment Steps

(a) Landslide Inventory Compilation: A landslide inventory is constructed by integrating multiple data sources, including historical records, remote sensing interpretation, and field investigation data.

(b) Construction of Conditioning Factor System: A set of conditioning factors was established, covering topographic, hydrological, geological, and environmental attributes (e.g., slope, rainfall, lithology, NDVI).

(c) Screen conditioning factors: Use multicollinearity detection, Pearson correlation coefficients, and collinearity diagnostics to select key factors affecting landslides and eliminate highly correlated factors such as cutting depth and relief amplitude.

(d) Construct the model dataset: Expand the landslide database by increasing the number of landslide points through a 30-m buffer zone and construct the non-landslide dataset using random sampling and non-landslide sample selection methods.

(e) Model Evaluation and Selection: The four machine learning models (SVM, LDA, RF, and ET) were evaluated and compared. The evaluation was based on a suite of indicators, including ROC curves, AUC values, and confusion matrices, to facilitate the selection of the superior model and dataset.

(f) Susceptibility Mapping and Accuracy Assessment: A landslide susceptibility map was generated using the selected optimal model. The accuracy of the prediction was assessed by comparing the statistical outcomes of the susceptibility zoning.

2.7.2. Technical Approach

Figure 5 depicts the technical workflow adopted in this study.

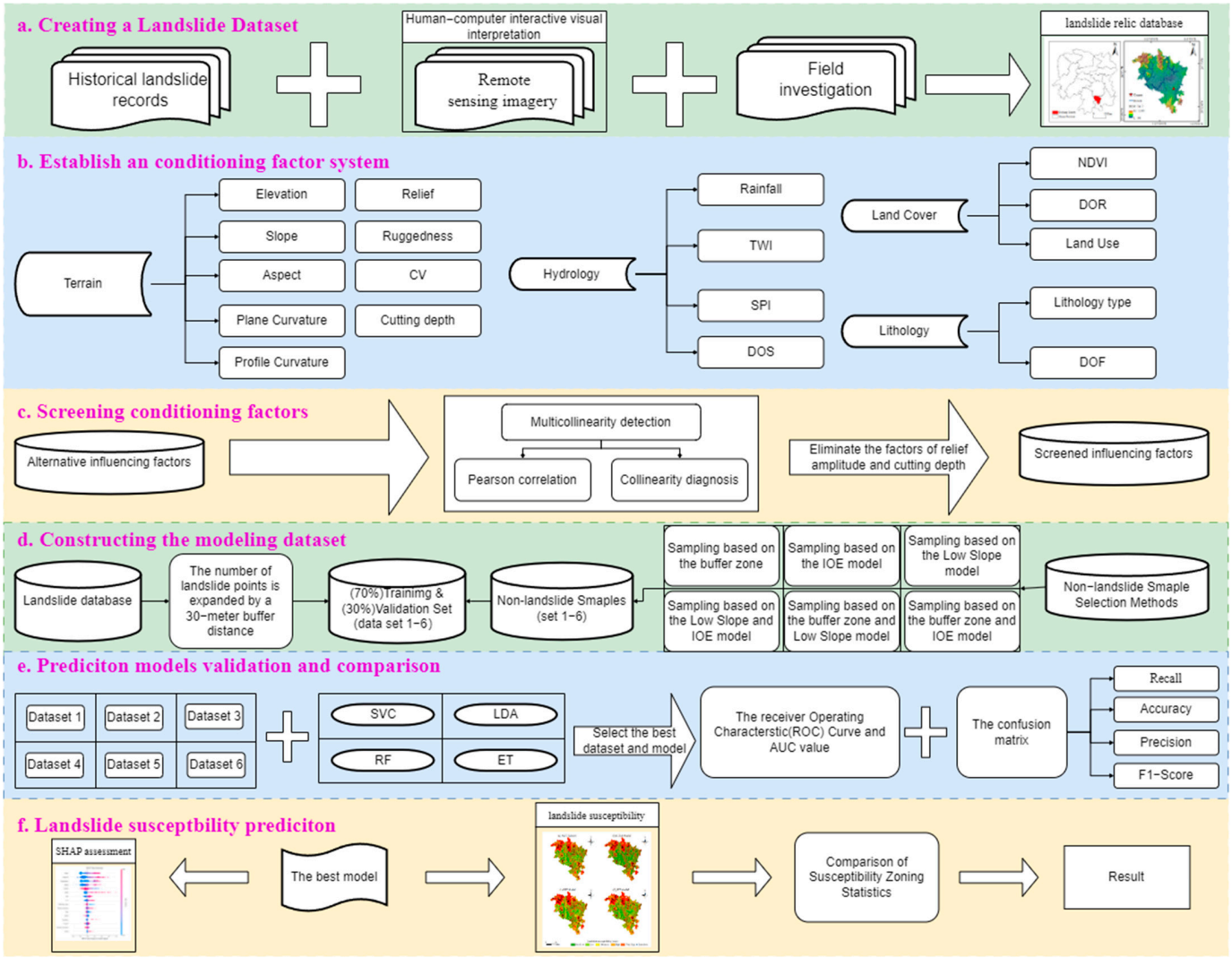


Figure 5. Flow chart of technical research.

3. Result

3.1. Multicollinearity Diagnosis and Pearson Correlation Analysis

A collinearity diagnosis was conducted for all conditioning factors using the linear regression module in SPSS 27.0, with the results (VIF) summarized in Table 3. The Pearson correlation analysis between each pair of factors was conducted using the correlation analysis module in SPSS 27.0 (Figure 6).

Table 3. Multi-collinearity test by tolerance and VIF.

Factors	Tolerances	VIF	Factors	Tolerances	VIF	Factors	Tolerances	VIF
Aspect	0.940	1.063	DOS	0.968	1.033	Lithology type	0.580	1.723
CV	0.583	1.715	DOR	0.570	1.755	Relief	0.250	3.993
DOF	0.968	1.033	Roughness	0.602	1.660	Cutting depth	0.236	4.244
DEM	0.611	1.636	Slope	0.246	4.062	Plane curvature	0.812	1.231
Landuse	0.608	1.644	SPI	0.709	1.411	Profile curvature	0.928	1.077
NDVI	0.749	1.335	TWI	0.708	1.412	Rainfall	0.590	1.695

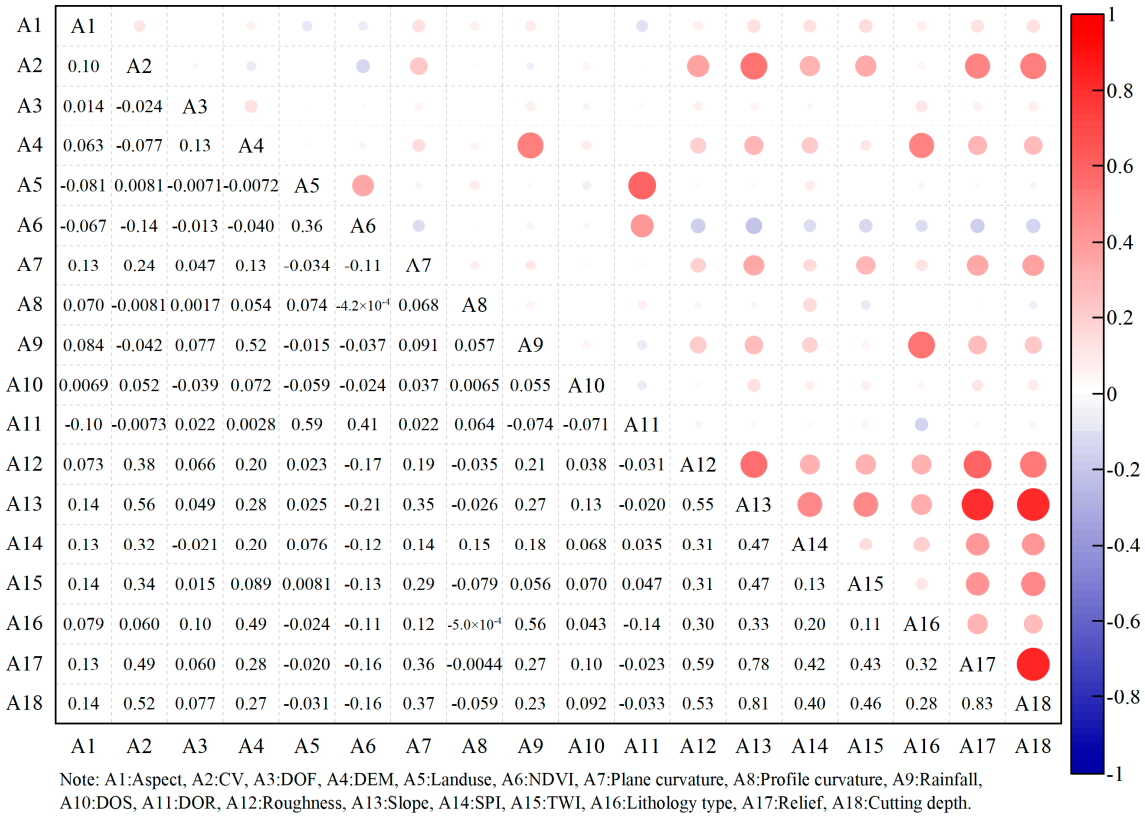


Figure 6. Correlation coefficient matrix.

Diagnostic results indicated that all 18 conditioning factors exhibited tolerance values exceeding 0.2 and VIF values under the threshold of 5, which confirms that significant multicollinearity was not present among the factors. Notably, the VIF values for incision depth, relief, and slope were approximately 4, and their Pearson correlation coefficients surpassed 0.7. The slope factor was deemed essential for the assessment due to its fundamental influence on landslide mechanisms and was therefore retained. Conversely, the factors of incision depth and relief amplitude were consequently removed from the analysis. Consequently, the remaining 16 hazardous factors exhibited neither collinearity nor strong correlation.

3.2. Relationship Between Conditioning Factors and Landslide Relics

The landslide inventory points were assigned to the Frequency Ratio value of their respective factor classes via the "Extract Multi Values to Points" function in ArcGIS 10.8. The Frequency Ratio for a class is calculated as: (Number of landslide points in the class/Total landslide points)/(Area of the class/Total study area). Subsequently, the number of disaster points and grid cells for each level of the conditioning factors were tallied. Finally, the values of a, b, FR_{ij}*, P_{ij}, H_i, $H_{i,max}$, I_i, P_i, and W_i* were calculated using Excel spreadsheets according to Equations (2)–(8) (Table 4).

Table 4. IOE model parametrical collation table.

Factors	Classes	No. of Landslides	No. of Raster	a	b	FR_{ij}*	P_{ij}	H_i	$H_{i,max}$	I_i	P_i	W_i	W_i*
Slope	0–6.86	23	1,147,291	0.1575	0.3487	0.4518	0.0716	0.2723	2.3219	0.0543	1.2624	0.0686	0.0277
	6.86–13.73	46	979,935	0.3151	0.2979	1.0579	0.1676	0.4319					
	13.73–21.79	42	647,322	0.2877	0.1968	1.4621	0.2316	0.4888					
	21.79–32.53	23	377,973	0.1575	0.1149	1.3713	0.2173	0.4785					
	32.53–76.10	12	137,353	0.0822	0.0418	1.9688	0.3119	0.5243					
Plane Curvature	−25.78–−2.96	3	56,199	0.0205	0.0171	1.2030	0.2016	0.4658	2.3219	0.0130	1.1934	0.0156	0.0063
	−2.96–−0.98	24	366,651	0.1644	0.1114	1.4751	0.2472	0.4984					
	0.98–0.33	61	1,795,141	0.4178	0.5457	0.7658	0.1283	0.3801					
	0.33–2.09	50	935,136	0.3425	0.2842	1.2049	0.2019	0.4661					
	2.09–30.18	8	136,747	0.0548	0.0416	1.3184	0.2209	0.4813					
DOS	0–0.26	128	2,889,251	0.8767	0.8782	0.9984	0.2100	0.4729	2.3219	0.0194	0.9507	0.0184	0.0074
	0.26–0.79	5	120,621	0.0342	0.0367	0.9342	0.1965	0.4613					
	0.79–1.24	7	141,948	0.0479	0.0431	1.1113	0.2338	0.4902					
	1.24–1.79	5	95,776	0.0342	0.0291	1.1765	0.2475	0.4986					
	1.79–3.28	1	42,278	0.0068	0.0129	0.5331	0.1121	0.3540					
DOF	0–0.25	98	2,127,293	0.6712	0.6466	1.0382	0.2091	0.4721	2.3219	0.0187	0.9928	0.0185	0.0075
	0.25–0.76	17	265,940	0.1164	0.0808	1.4405	0.2902	0.5180					
	0.76–1.23	22	655,917	0.1507	0.1994	0.7559	0.1523	0.4135					
	1.23–1.76	5	147,171	0.0342	0.0447	0.7657	0.1542	0.4160					
	1.76–3.16	4	93,553	0.0274	0.0284	0.9635	0.1941	0.4591					
DOR	0–0.49	97	2,477,520	0.6644	0.7531	0.8823	0.0712	0.2714	2.3219	0.1561	2.4788	0.3869	0.1562
	0.49–1.52	17	493,559	0.1164	0.1500	0.7762	0.0626	0.2503					
	1.52–2.99	17	237,857	0.1164	0.0723	1.6106	0.1300	0.3826					
	2.99–5.81	13	72,027	0.0890	0.0219	4.0671	0.3282	0.5275					
	5.81–11.39	2	8911	0.0137	0.0027	5.0575	0.4081	0.5277					
Rainfall	2051–2118	40	1,210,437	0.2740	0.3679	0.7447	0.1167	0.3616	2.3219	0.2059	1.2765	0.2628	0.1061
	2118–2169	36	1,190,378	0.2466	0.3618	0.6816	0.1068	0.3446					
	2169–2251	38	532,952	0.2603	0.1620	1.6068	0.2517	0.5010					
	2251–2362	31	219,354	0.2123	0.0667	3.1846	0.4990	0.5005					
	2362–2567	1	136,753	0.0068	0.0416	0.1649	0.0258	0.1363					
TWI	3.97–7.60	73	1,305,235	0.5000	0.3967	1.2604	0.2853	0.5162	2.3219	0.0324	0.8834	0.0286	0.0116
	7.60–9.65	45	1,171,820	0.3082	0.3562	0.8654	0.1959	0.4607					
	9.65–12.60	21	491,509	0.1438	0.1494	0.9629	0.2180	0.4791					
	12.60–16.81	5	272,014	0.0342	0.0827	0.4143	0.0938	0.3202					
	16.81–28.97	2	49,296	0.0137	0.0150	0.9143	0.2070	0.4704					
NDVI	−1649–4246	12	90,541	0.0822	0.0275	2.9866	0.3535	0.5303	2.3219	0.0878	1.6897	0.1483	0.0599
	4246–6115	20	199,141	0.1370	0.0605	2.2632	0.2679	0.5091					
	6115–7378	35	472,658	0.2397	0.1437	1.6687	0.1975	0.4622					
	7378–8260	40	915,188	0.2740	0.2782	0.9850	0.1166	0.3615					
	8260–9999	39	1,612,346	0.2671	0.4901	0.5451	0.0645	0.2551					
DEM	59–249	35	1,316,776	0.2397	0.4003	0.5990	0.1061	0.3435	2.3219	0.1055	1.1287	0.1190	0.0481
	249–393	49	1,147,025	0.3356	0.3487	0.9627	0.1706	0.4352					
	393–584	41	426,851	0.2808	0.1297	2.1645	0.3835	0.5303					
	584–826	19	276,005	0.1301	0.0839	1.5513	0.2749	0.5121					
	826–1400	2	123,217	0.0137	0.0375	0.3659	0.0648	0.2559					
Lithology type	Harder Rock	22	312,971	0.1507	0.0951	1.5841	0.2786	0.5136	2.3219	0.1710	1.1373	0.1944	0.0785
	Hard Rock	76	2,359,103	0.5205	0.7171	0.7260	0.1277	0.3791					
	Weak Rock	46	538,857	0.3151	0.1638	1.9237	0.3383	0.5290					
	Weaker Rock	2	31,022	0.0137	0.0094	1.4528	0.2555	0.5030					
	Loose Rock	0	47,909	0.0000	0.0146	0.0001	0.0000	0.0003					

Table 4. *Cont.*

Factors	Classes	No. of Landslides	No. of Raster	*a*	*b*	FR_{ij}*	P_{ij}	H_i	$H_{i,max}$	I_i	P_i	W_i	W_i*
Land use	Cropland	46	1,220,474	0.3151	0.3710	0.8494	0.0865	0.3054	2.3219	0.2723	1.9646	0.5350	0.2160
	Forest	88	1,955,766	0.6027	0.5945	1.0140	0.1032	0.3382					
	Others	1	4998	0.0068	0.0015	4.5086	0.4590	0.5157					
	Water	0	36,807	0.0000	0.0112	0.0001	0.0000	0.0002					
	Buildup	11	71,829	0.0753	0.0218	3.4509	0.3513	0.5302					
Aspect	North	12	459,635	0.0822	0.1397	0.5884	0.0725	0.2745	3.0000	0.0109	1.0141	0.0111	0.0045
	Northeast	17	334,051	0.1164	0.1015	1.1468	0.1414	0.3990					
	East	18	466,401	0.1233	0.1418	0.8697	0.1072	0.3454					
	Southeast	21	401,912	0.1438	0.1222	1.1775	0.1451	0.4041					
	South	22	410,405	0.1507	0.1247	1.2080	0.1489	0.4091					
	Southwest	20	367,835	0.1370	0.1118	1.2253	0.1510	0.4119					
	West	21	466,432	0.1438	0.1418	1.0146	0.1251	0.3751					
	Northwest	15	383,203	0.1027	0.1165	0.8821	0.1087	0.3481					
SPI	2.75–6.55	1	263,540	0.0068	0.0801	0.0856	0.0187	0.1072	2.3219	0.1106	0.9174	0.1015	0.0410
	6.55–8.68	37	1,104,319	0.2534	0.3357	0.7551	0.1646	0.4285					
	8.68–10.63	67	1,231,618	0.4589	0.3744	1.2259	0.2672	0.5088					
	10.63–13.23	35	571,961	0.2397	0.1739	1.3790	0.3006	0.5213					
	13.23–26.40	6	118,436	0.0411	0.0360	1.1416	0.2489	0.4994					
Roughness	1–1.04	81	2,292,192	0.5548	0.6967	0.7964	0.1159	0.3604	2.3219	0.1909	1.3738	0.2623	0.1059
	10.4–1.12	45	722,696	0.3082	0.2197	1.4032	0.2043	0.4681					
	1.12–1.26	19	210,869	0.1301	0.0641	2.0304	0.2956	0.5197					
	1.26–1.50	0	55,578	0.0000	0.0169	0.0001	0.0000	0.0002					
	1.50–4.16	1	8539	0.0068	0.0026	2.6390	0.3842	0.5302					
Cutting depth	0–2.05	24	1,272,842	0.1644	0.3869	0.4250	0.0744	0.2788	2.3219	0.0522	1.1427	0.0597	0.0241
	2.05–4.78	65	1,139,489	0.4452	0.3464	1.2855	0.2250	0.4842					
	4.78–8.19	38	599,351	0.2603	0.1822	1.4288	0.2501	0.5000					
	8.19–13.19	17	227,622	0.1164	0.0692	1.6830	0.2946	0.5194					
	13.19–58	2	50,570	0.0137	0.0154	0.8913	0.1560	0.4181					
Relief	0–5	32	1,608,874	0.2192	0.4890	0.4483	0.0751	0.2804	2.3219	0.0848	1.1945	0.1013	0.0409
	5–10	60	937,863	0.4110	0.2851	1.4417	0.2414	0.4950					
	10–17	38	524,499	0.2603	0.1594	1.6326	0.2733	0.5115					
	17–26	15	176,589	0.1027	0.0537	1.9142	0.3205	0.5261					
	26–132	1	42,049	0.0068	0.0128	0.5360	0.0897	0.3121					
Profile Curvature	−37.08–−4.65	3	86,895	0.0205	0.0264	0.7781	0.1667	0.4309	2.3219	0.0243	0.9335	0.0227	0.0092
	−4.65–−1.44	18	415,162	0.1233	0.1262	0.9771	0.2093	0.4723					
	−1.44–0.81	81	1,991,942	0.5548	0.6055	0.9164	0.1963	0.4611					
	0.81–3.70	40	645,598	0.2740	0.1962	1.3962	0.2991	0.5208					
	3.70–44.81	4	150,277	0.0274	0.0457	0.5999	0.1285	0.3804					
CV	0–0.0051	37	1,224,790	0.2534	0.3723	0.6808	0.0954	0.3233	2.3219	0.0853	1.4279	0.1218	0.0492
	0.0051–0.0096	56	1,119,905	0.3836	0.3404	1.1269	0.1578	0.4204					
	0.0096–0.0162	38	689,186	0.2603	0.2095	1.2425	0.1740	0.4390					
	0.0162–0.0278	11	225,863	0.0753	0.0687	1.0975	0.1537	0.4153					
	0.0278–0.1289	4	30,130	0.0274	0.0092	2.9916	0.4190	0.5258					

The entropy index weight (W_i) serves as a metric for the influence of each conditioning factor on landslide development, thereby facilitating an objective analysis of their relative contributions. The weight assigned to a factor is directly proportional to its contribution to landslide susceptibility. The quantitative analysis ranks the conditioning factors in the following descending order of contribution: land type > road density > rainfall > roughness > lithology type > NDVI > CV > DEM > SPI > relief > slope > incision depth > TWI > profile curvature > fault density > stream density > plan curvature > aspect. Among them, the dominant factors such as land type, road density, and rainfall have smaller entropy values but larger weights, indicating their significant impact on landslide development. Factors such as plan curvature and aspect display higher entropy values yet lower weights, signifying their relatively limited role in landslide development.

3.3. Analysis of Different Sample Optimization Methods

We constructed SVM, LDA, RF, and ET models using the scikit-learn library in Python 3.11. All six datasets (1–6) were used to train and test these models with a split ratio of 7:3 for training to testing. The precision of the models was assessed using ROC curves, shown in Figure 7.

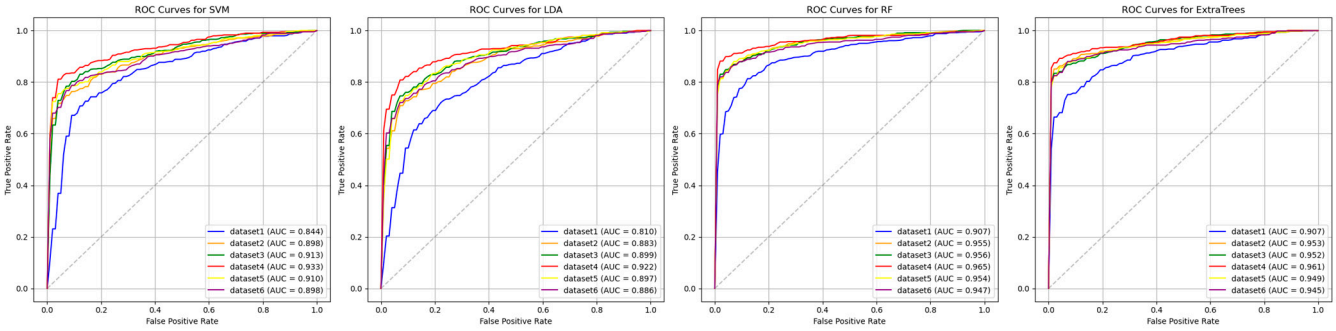


Figure 7. AUC for different non-landslide sample optimization methods.

Hyperparameters are configuration variables that are fixed before learning commences and remain unchanged while the model optimises its parameters on data. These variables govern model capacity, convergence speed and generalisation; Their selection directly determines predictive performance. Identifying optimal hyperparameters is thus critical to maximising model efficacy. Systematic tuning—whether through grid search [60], Bayesian optimisation [61], evolutionary strategies [62] or meta-learning [23]—is now regarded as an indispensable component of every machine-learning pipeline. Owing to the low-dimensional hyper-parameter space of the model under study, exhaustive grid search offers the most direct route to the global optimum.

The optimal parameters for each model, finalized after multiple tuning iterations, are listed in Table 5. To mitigate the bias from imbalanced data distribution and improve training precision, we employed 10-fold cross-validation using the StratifiedKFold function with the parameters: n_splits = 10, shuffle = True, and random_state = 42.

Table 5. Optimal hyperparameters for different models.

No.	Model	Parameters
1	SVM	*C* = 10, *gamma* = scale, and *kernel* = rbf (all other parameters as default).
2	LDA	*shrinkage* = None, and *solver* = lsqr (all other parameters as default).
3	RF	*max_depth* = 20, *min_samples_leaf* = 1, *min_samples_split* = 2, and *n_estimators* = 200 (all other parameters as default).
4	ET	*max_depth* = 20, *min_samples_leaf* = 1, *min_samples_split* = 2, and *n_estimators* = 100 (all other parameters as default).

Figure 7 shows that all models achieve robust classification across datasets 1–6, with AUC values consistently above 0.8. RF and ET outperform SVM and LDA, exhibiting markedly higher AUC values. Systematic variation of non-landslide sampling strategies substantially enhances predictive performance. Specifically, AUC improves by 0.089 for SVM (0.844 → 0.933), 0.112 for LDA (0.810 → 0.922), 0.058 for RF (0.907 → 0.965) and 0.054 for ET (0.907 → 0.961). For the best-performing RF model, the strategies yield AUCs in descending order: Slope-IOE-O > Slope-O > IOE-O > Buf-Slope-O > Buf-IOE-O > Buf-O. Within single-objective optimizations, performance decreases as Slope-O > IOE-O > Buf-O; under hybrid optimizations, the hierarchy becomes Slope-IOE-O > Buf-Slope-O > Buf-

IOE-O. Overall, Slope-IOE-O delivers the highest predictive accuracy across all evaluated sampling strategies.

3.4. Analysis of Model Performance and Effectiveness

3.4.1. Analysis of Model Performance

Based on the results of different sample optimization methods, the Slope-IOE-O method was found to be the most effective. Therefore, we utilized this dataset (dataset4) to apply the SVM, LDA, RF, and ET models with their optimal parameters for analysis, and the confidence intervals were presented in Table 6.

Table 6. Confidence Intervals of AUC, Accuracy, Precision, F1, and Recall for the Different Models.

Model	AUC		ACC		Precision		F1		Recall	
	Value	CI.	Value	CI.	Value	CI.	Value	CI.	Value	CI.
SVM	0.933	[0.871, 0.946]	0.872	[0.784, 0.885]	0.908	[0.773, 0.964]	0.861	[0.781, 0.880]	0.818	[0.681, 0.833]
LDA	0.922	[0.862, 0.958]	0.840	[0.768, 0.900]	0.875	[0.765, 0.994]	0.826	[0.743, 0.893]	0.782	[0.622, 0.827]
RF	0.965	[0.912, 0.983]	0.906	[0.858, 0.943]	0.925	[0.848, 0.994]	0.900	[0.859, 0.942]	0.876	[0.783, 0.923]
ET	0.961	[0.923, 0.979]	0.909	[0.843, 0.936]	0.916	[0.820, 0.994]	0.905	[0.840, 0.937]	0.894	[0.749, 0.928]

Table 6 shows that all models exhibit narrow 95% confidence intervals, signifying robust performance estimates. RF and ET surpass SVM and LDA across nearly all metrics, and their tighter intervals denote both reduced variance and heightened stability. Specifically, RF marginally excels in AUC and precision, whereas ET demonstrates superior accuracy, F1-score and recall.

3.4.2. Analysis of Model Effectiveness

A landslide susceptibility map for Guiyang County was generated using Python 3.11, Dataset 2, and four models (SVM, LDA, RF, and ET). This map was subsequently categorized in ArcGIS 10.8 into five susceptibility levels (very high, high, moderate, low, and very low) through the natural breaks' classification method, as presented in Figure 8.

Model validation was performed by overlaying the original landslide points onto the susceptibility map. The number of landslides and grid cells were counted for each susceptibility level, enabling the computation of landslide density and frequency ratio. The frequency ratio for a given level was computed using the formula: (Percentage of Landslides in the Level)/(Percentage of Total Area occupied by the Level). The results of this analysis are presented in Table 7.

As evaluated by the landslide frequency ratio in very high and high susceptibility zones (Table 6), the models are ranked in descending order of performance as: RF, ET, SVM, and LDA, with RF and ET models producing nearly identical areal proportions in these critical zones. RF and ET yield nearly identical areal proportions for very high and high susceptibility zones. Collectively, these zones cover merely 42% of the study region but encompass 96% of the inventoried landslides.

Table 7. Classification performance of different coupling models.

Model	Landslide Susceptibility Zone Levels	Number of Grid Cells	Area Proportion	Number of Landslides	Landslide Number Ratio	Landslide Frequency Ratio
ET	Very low	982,144	0.30	2	0.01	0.05
	low	638,882	0.19	2	0.01	0.07
	middle	292,052	0.09	6	0.04	0.46
	high	321,124	0.10	11	0.08	0.77
	Very high	1,055,672	0.32	125	0.86	2.67

Table 7. *Cont.*

Model	Landslide Susceptibility Zone Levels	Number of Grid Cells	Area Proportion	Number of Landslides	Landslide Number Ratio	Landslide Frequency Ratio
RF	Very low	912,176	0.28	2	0.01	0.05
	low	660,851	0.20	1	0.01	0.03
	middle	334,600	0.10	3	0.02	0.20
	high	224,453	0.07	23	0.16	2.31
	Very high	1,157,794	0.35	117	0.80	2.28
SVM	Very low	1,124,243	0.34	8	0.05	0.16
	low	499,757	0.15	11	0.08	0.50
	middle	237,673	0.07	8	0.05	0.76
	high	382,477	0.12	7	0.05	0.41
	Very high	1,045,724	0.32	112	0.77	2.41
LDA	Very low	1,550,851	0.47	17	0.12	0.25
	low	272,606	0.08	10	0.07	0.83
	middle	163,123	0.05	8	0.05	1.11
	high	161,655	0.05	4	0.03	0.56
	Very high	1,141,639	0.35	107	0.73	2.11

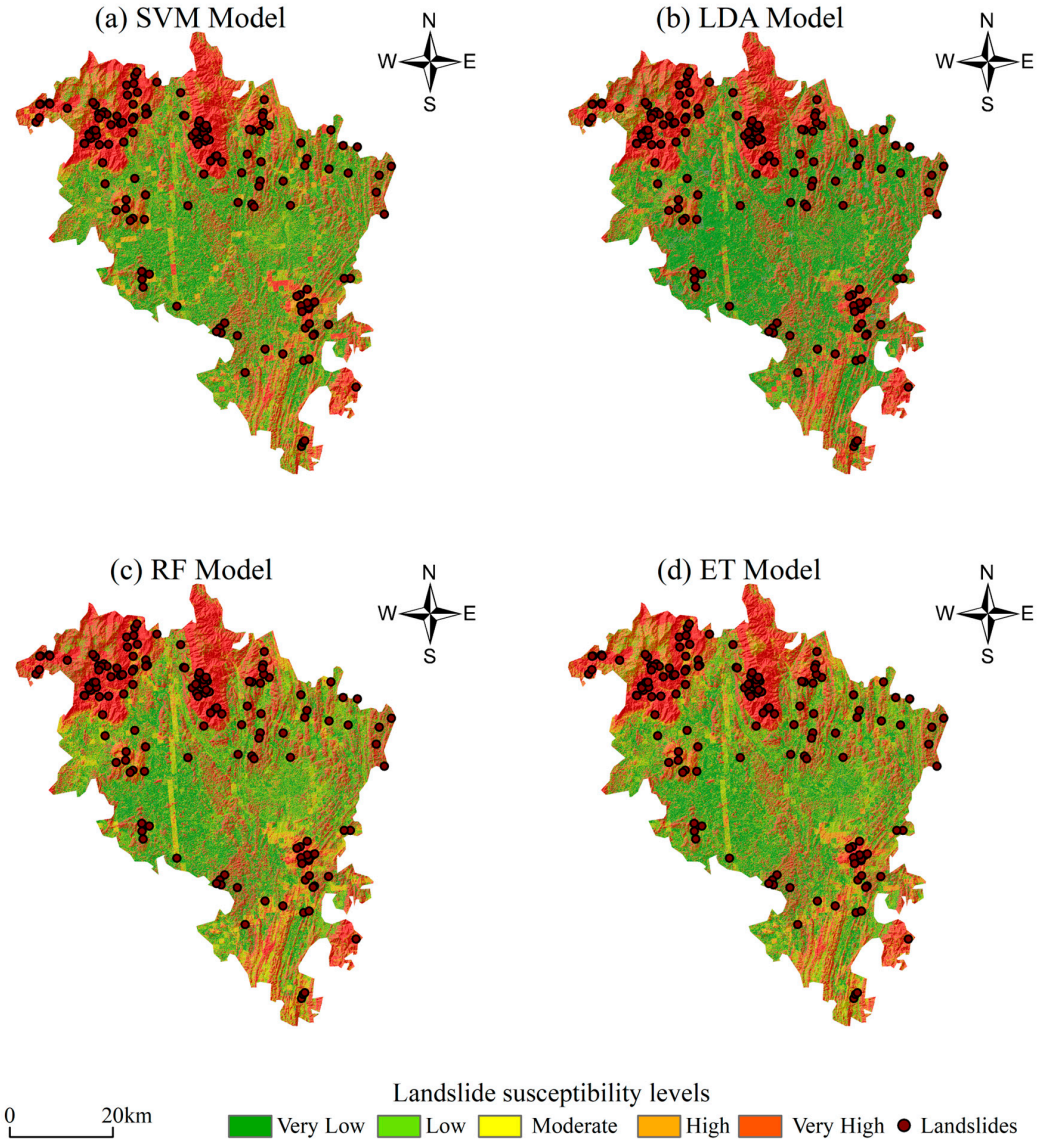


Figure 8. Landslide susceptibility mapping premised on different coupling models.

Overall, the RF and ET models demonstrate superior classification performance and effectiveness compared to the SVM and LDA models, with RF being marginally superior to ET and thus exhibiting the best overall performance.

3.5. Model Interpretability via SHAP

Machine-learning models are inherently black boxes, rendering their decision logic opaque; therefore, interpretability frameworks are essential for trustworthy evaluation. We focus on Slope-IOE-O (Dataset 4)—the optimal sampling strategy—and the top-performing RF model, leveraging SHAP for global and local interpretation. Using Tree Explainer (SHAP v0.44, Python 3.11), we quantify global feature importance and instance-level Shapley values, elucidate dominant landslide controls and factor interdependencies, and thereby dissect the model's internal decision mechanism.

3.5.1. Global Interpretation

The SHAP summary plot is displayed in Figure 9. The SHAP value for each predictor, plotted on the horizontal axis, is interpreted as follows: a positive value signifies that the feature elevates the model's landslide susceptibility prediction, whereas a negative value corresponds to a decrease in the predicted risk. Each point corresponds to an individual sample, with colour indicating the factor's magnitude (red = high, blue = low). Factors exhibiting broad horizontal dispersion exert greater influence, whereas tight clustering around zero indicates limited predictive relevance. Transparent bars denote global feature importance, with bar length proportional to each factor's relative contribution.

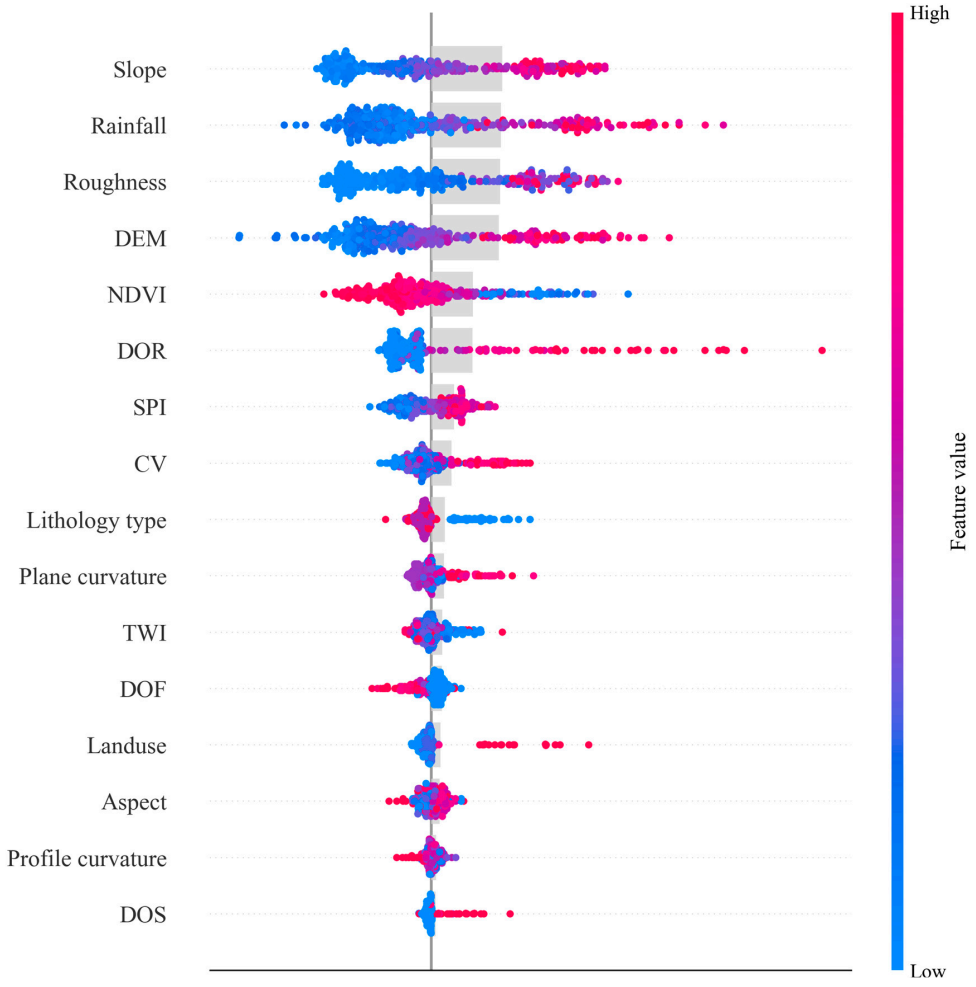


Figure 9. SHAP value summary.

Transparent bars in the SHAP summary plot identify slope, maximum rainfall, surface roughness and elevation as the four dominant landslide drivers among the sixteen predictors, as evidenced by their greatest bar lengths. Conversely, profile curvature and stream density contribute minimally, exerting negligible influence on landslide occurrence. Slope, Rainfall, Roughness and DEM exhibit wide, positively skewed SHAP distributions, indicating that higher values amplify landslide risk: red (high-value) points align with positive SHAP scores, whereas green (low-value) points cluster in the negative domain, denoting risk suppression. Low NDVI and DOF values map predominantly to positive SHAP scores, implying that reduced levels enhance landslide susceptibility. The dispersed SHAP values for DOR and land-use variables suggest strong interaction effects with other predictors. Aspect and profile curvature exhibit SHAP values tightly centred on zero, underscoring their negligible impact on landslide risk.

3.5.2. Local Interpretation

Focusing locally on the four dominant landslide drivers—slope, maximum rainfall, surface roughness and elevation—Figure 10a–d present univariate SHAP dependence plots in which the *x*-axis records each predictor's raw value and the *y*-axis quantifies its direct, monotonic or non-monotonic, contribution to landslide probability, with every point representing an individual observation whose vertical position signals the magnitude and direction of influence on model output. Extending to interactions, Figure 10e–h display bivariate SHAP dependence landscapes where the x- and y-axes give the raw values of the primary driver and its strongest covariate, the colour scale encodes their joint SHAP interaction value, and warmer (cooler) hues reveal progressively stronger positive (negative) synergistic modulation of landslide susceptibility.

The SHAP dependence plot for slope shows low or negative SHAP values at gentle inclinations, indicating limited landslide susceptibility. Between 0° and 20°, SHAP values rise monotonically with slope, reflecting an increasingly positive contribution to landslide probability. Between 20° and 30°, SHAP values plateau at their maximum, signifying the strongest slope-driven amplification of risk. Above 30°, SHAP values decline, consistent with reduced landslide likelihood in steep terrains underlain by competent bedrock and devoid of loose regolith, which collectively enhance slope stability. The bivariate SHAP interaction plot with road density—the dominant covariate—shows that high-slope pixels coincide with sparse road networks, reflecting the county's hilly terrain where roads cluster in gentle valleys and plains rather than steep uplands. Moreover, low-slope samples paired with high road density yield elevated SHAP values, evidencing that intensive road construction can trigger landslides even on gentle gradients.

The SHAP dependence plot for monthly maximum rainfall reveals low or negative SHAP values below 215 mm, denoting subdued landslide susceptibility. At 225 mm, SHAP values increase linearly and plateau at their maximum, signifying peak rainfall-driven landslide amplification. The interaction plot shows that, for identical rainfall, high-elevation samples (red) exhibit greater SHAP values than low-elevation samples (blue), evidencing elevation-dependent risk amplification. Between 225 mm and 235 mm, pronounced fluctuations and non-linearity suggest soil-moisture saturation and heightened modulation by covariates. Above 235 mm, SHAP values decrease; the interaction plot shows that most samples exceed 600 m, indicating that excessive rainfall attenuates landslide risk at high elevations.

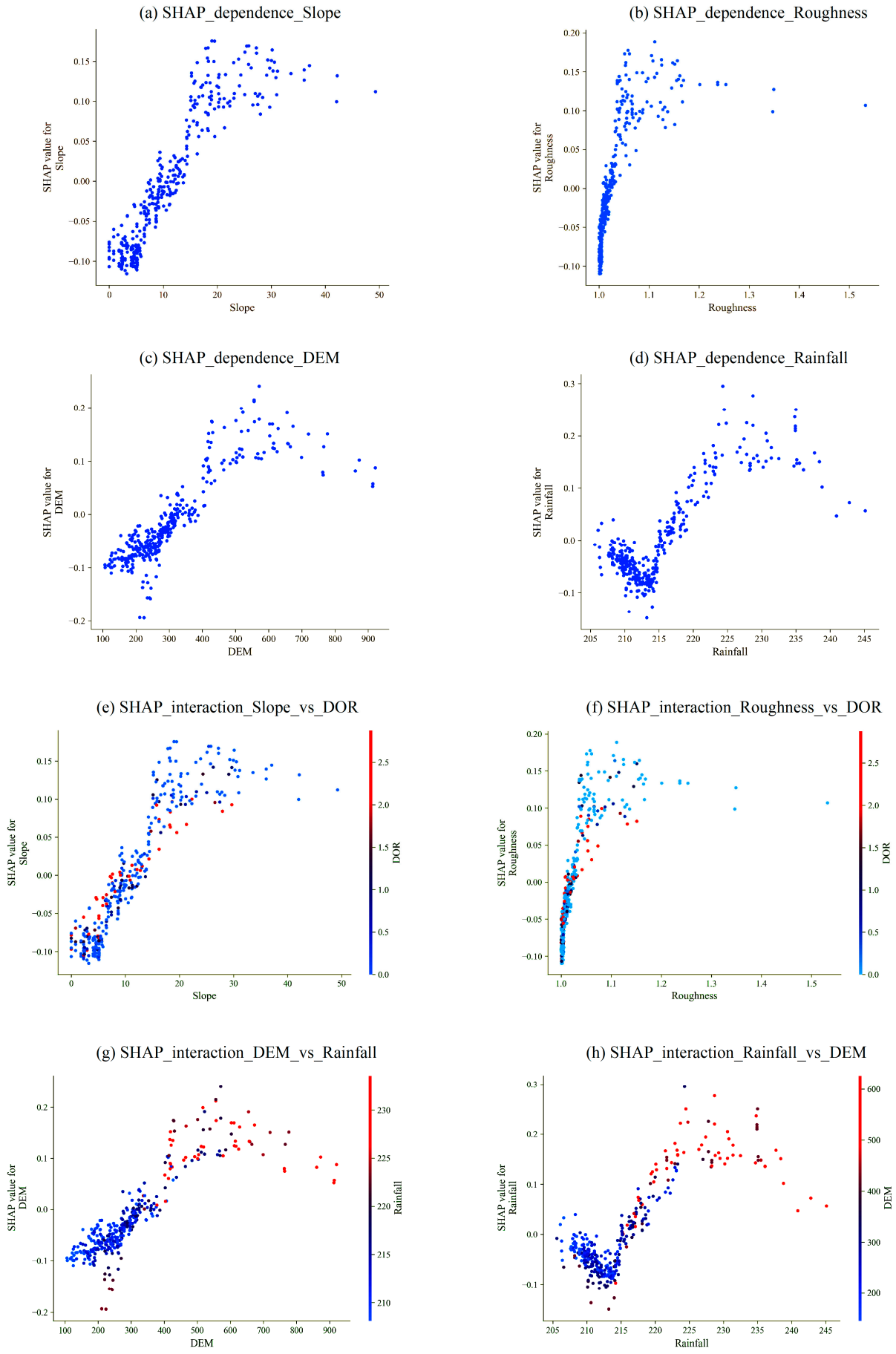


Figure 10. SHAP single-factor and two-factor interaction dependence plots.

Below a roughness index of 1.04, SHAP values remain low, and samples form a dense, linear cloud, indicating minimal landslide susceptibility and limited covariate influence. Between 1.04 and 1.2, SHAP values rise steeply, signaling a pronounced escalation in landslide risk. Beyond 1.2, SHAP values plateau and sample dispersion intensifies, indicating diminishing marginal risk and heightened interaction effects. Road construction markedly amplifies the landslide hazard associated with surface roughness when the roughness index exceeds 1.04 and road density (DOR) surpasses 1.0, a conjunction that demands heightened vigilance.

SHAP dependence plots for elevation reveal low or negative SHAP values below 300 m, reflecting stable fluvial plains and valley floors whose terrain exerts negligible, or even suppressive, effects on landslide initiation. Between 100 m and 450 m, SHAP values increase monotonically, mirroring escalating landslide susceptibility. Within the 450–700 m band, SHAP values plateau at elevated levels yet exhibit pronounced variability, denoting high landslide risk strongly modulated by covariates. Beyond 700 m, SHAP values decline modestly, with interaction plots revealing that landslide likelihood is instead dictated by the synergistic effects of high elevation and extreme rainfall.

4. Discussion

4.1. Hybrid Optimisation of Non-Landslide Samples

We present a systematic framework for sample optimisation. To mitigate the paucity (n = 146) and spatial bias of landslide inventories in Guiyang County, landslide records are augmented via oversampling and complemented by a novel non-landslide selection strategy—Slope-IOE-O. Low-slope constraints (Slope-O) are intersected—via a spatial overlay—with zones of minimal susceptibility delineated by the entropy-weighted information-value model (IOE-O), establishing a coupled physical–statistical framework for purifying non-landslide samples. The method implements a dual-guarantee mechanism: low-slope areas inherently exhibit geotechnical stability and low failure probability, while IOE-O statistically isolates zones of minimal susceptibility. Their spatial concordance yields negatives that are simultaneously topographically and statistically robust, elevating sample purity and minimising mislabelled pseudo-landslides. Relative to conventional buffer-based selection—prone to mislabeling proximal hazard zones—Slope-IOE-O exploits spatial concordance to sharpen the environmental distinction between positives and negatives, elevating RF AUC from 0.907 to 0.965. Accordingly, Slope-IOE-O is advocated as the primary optimisation strategy, with single-criterion alternatives reserved for data-scarce scenarios that preclude hybrid implementation. Landslide inventories derive from field surveys and archival records, ensuring veracity. Eighteen multi-source covariates—encompassing topography, hydrology, geology and anthropogenic drivers—were compiled; collinearity and correlation screening distilled the set to sixteen predictors, eliminating redundancy and furnishing high-fidelity model inputs. The integrated optimisation of samples and predictors yields a transferable protocol for susceptibility assessment in geologically complex, data-limited settings.

4.2. Mechanistic Attribution of Landslide Susceptibility Using SHAP

Using 146 inventoried landslides in Guiyang County—compiled from literature and field surveys—we model susceptibility via LDA, RF and ET. To elucidate factor contributions and interactions, we conduct a systematic SHAP analysis of model outputs. SHAP identifies slope, monthly maximum rainfall, surface roughness and elevation as the four dominant drivers, quantifies their importance in descending order, and reveals non-linear pathways and interdependencies. Dependence plots reveal that slope contributions peak at 20–30°, declining beyond 30° where exposed bedrock reduces susceptibility. Monthly

rainfall exhibits a critical threshold (225 mm); beyond this threshold, rising pore-water pressure sharply increases failure probability. Interaction SHAP values further demonstrate significant synergy between surface roughness and road density, and between rainfall and elevation, jointly modulating landslide spatial patterns. By providing quantitative attribution, SHAP overcomes the black-box limitation, corroborates the physical plausibility of predictions, and enhances interpretability in geomechanically and hydrological contexts, thereby offering a robust visual and quantitative toolkit for landslide-susceptibility modelling in complex terrains [59,63].

4.3. Spatial Heterogeneity of Landslide Susceptibility and Its Primary Drivers

Landslide susceptibility in Guiyang County exhibits a pronounced north–south high-risk belt that sandwiches a low-lying central corridor, revealing strong spatial heterogeneity. Very-high susceptibility is confined to three hotspots: (i) the northern uplands (>584 m elevation, >21.79° slope, topographic roughness > 1.12); (ii) the steep southern highlands; and (iii) the built-up county seat and adjacent peri-urban fringe. Conversely, the central tract—comprising low-relief hills, terraces, plains and open water—is assigned to the very-low susceptibility class.

Excessive landslide probability in the northern mountains arises from synergistic topographic–hydrological forcing. High elevations and pronounced roughness focus runoff, sustaining chronic soil saturation [64]; steep terrain amplifies gravitational loading, and vigorous fluvial incision undermines mechanical stability [65], whereas intense monthly rainfall and elevated SPI values promote failure via surface erosion and elevated pore-water pressures [66]. Within the southern county seat, landslide susceptibility is exacerbated by the confluence of dense anthropogenic disturbance, lateral river undercutting and predisposed lithological structures. Compacted urban fabric (low NDVI) and pervasive cut slopes unload slope toes and relax rock mass [67]; lateral migration of the Xi River and attendant stage fluctuations weaken bank materials and scour slope bases; dense fault networks, intensely fractured strata and groundwater flow along discontinuities collectively diminish shear resistance. The central tract's very-low susceptibility reflects subdued relief, gentle gradients, limited runoff convergence and negligible erosion, which together preclude landslide initiation.

RF, ET, SVM and LDA unanimously reproduce the north–south high-risk belt flanking a low-risk core, a pattern congruent with elevation and slope, corroborating model robustness. We therefore recommend focusing mitigation on (i) drainage and bio-engineered slope protection in the northern highlands, (ii) strict regulation of cut slopes, riparian armory and revegetation in the southern county seat, and (iii) preservation of the central low-risk tract's intact, low-relief topography.

4.4. Comparison with Other Studies

Leveraging 18 conditioning factors, we delineated landslide susceptibility across Guiyang County, revealing slope, monthly maximum rainfall, surface roughness and elevation as the four dominant drivers. Liu Leilei employed a deterministic factor model in Hunan's red-bed terrain, identified land-use type, elevation and topographic relief as principal predictors [68]. Liu Ruiyang coupling SHAP with RF, found that lithology, fault proximity, elevation, road proximity and river proximity dictated typhoon-triggered landslides in Zixing City [69]. Divergences arise from (1) distinct geological contexts and landslide sample characteristics, and (2) differing predictor suites and model architectures that reorder factor importance and modulate causal pathways [70]. To enhance interpretability, we employed SHAP, which quantifies each feature's Shapley contribution within a game-theoretic framework, ensuring transparent and traceable decisions. Global SHAP

metrics identify dominant drivers, whereas local explanations quantify site-specific contributions and elucidate underlying mechanisms, thereby validating causal links between predictors and landslide occurrence.

Four landslide susceptibility maps for Guiyang County were produced using SVM, LDA, RF, and ET models. High spatial concordance among the models was observed, with the very-high and very-low susceptibility zones exhibiting consistent spatial patterns, thereby attesting to the robustness of the modeling framework. Susceptibility exhibits a north- and south-high, central-low gradient, with very-high zones concentrated in the northern mountains, the vicinity of the southern county seat and selected villages of Taihe Town, and very-low zones confined to the central lowlands. This pattern accords with the regional zonation of Xu Zhaojun [71], derived from an index-based method. However, Xu Zhaojun incorporated ground subsidence and rockfalls, potentially underestimating landslide susceptibility in some areas, and their coarse-resolution mapping limited spatial fidelity. Our machine-learning workflow delivers 30 m × 30 m resolution, markedly enhancing both accuracy and precision.

4.5. Future Research Directions

Comparative evaluation of non-landslide sampling protocols identifies an optimal strategy that substantially enhances classifier performance and partially alleviates data scarcity in landslide inventories. Landslide samples derive from rigorously validated historical records and field surveys, warranting the use of buffer-based augmentation. When inventories are suspected, buffer methods risk amplifying noise and introducing distributional bias and are therefore discouraged. Future research should integrate spatial-autocorrelation metrics and topographic constraints within sampling protocols or harness multi-source data fusion to enhance the fidelity and representativeness of both landslide and non-landslide inventories.

Landslide footprints are spatially restricted relative to stable terrain. Although preserving this imbalance can mirror natural prevalence, a universal landslide-to-non-landslide ratio remains elusive. Ratios ranging from 1:160 [27] to 1:1 [72] have been reported, with intermediate prescriptions of 1:2 [73] and 1:10 [74]. The optimal ratio appears contingent on both spatial extent and the representativeness of predictor space. Future work could integrate cost-sensitive weighting into grid search or Bayesian optimization frameworks to self-adaptively identify the optimal class ratio for maximizing predictive accuracy.

The prevalence of rainfall-induced landslides, which constitute 90% of events in China [75], mandates focused research. The risk is particularly acute in Southern China, where more than 60% of the annual precipitation is delivered between June and September. This concentration, coupled with a rising trend in extreme rainfall events, underscores the critical importance of developing accurate rainfall-threshold models for landslide early warning. Future efforts should assimilate high-resolution rainfall data with temporally explicit landslide inventories to dissect the coupled response of slope stability to rainfall intensity and duration. Machine-learning models can be trained to link rainfall metrics to landslide response, with k-fold cross-validation and Bayesian optimisation used to derive optimal thresholds. Accounting for spatial–temporal heterogeneity, dynamic threshold models should be developed to enhance predictive accuracy and to inform regional disaster-risk reduction.

5. Conclusions

Landslide susceptibility in Guiyang County, Hunan Province, is systematically assessed through an integrated framework that couples sample-optimization strategies with machine-learning models and interpretability analyses; key conclusions are summarised below.

(1) To counteract the paucity and spatial bias of landslide inventories, we devise and validate the Slope-IOE-O hybrid sampling protocol, integrating low-slope thresholds with IOE-delineated extremely low-susceptibility zones to maximise the purity and representativeness of non-landslide samples. The strategy elevates the RF AUC from 0.907 to 0.965, markedly surpassing both traditional buffer techniques and single-criterion alternatives.

(2) Across SVM, LDA, RF and ET, RF delivers the highest performance across AUC, accuracy, precision, F1 and recall (AUC = 0.965), attesting to its reliability and stability for landslide prediction in Guiyang County.

(3) SHAP interpretability identifies slope, monthly maximum rainfall, surface roughness and elevation as the four dominant drivers of regional landslide susceptibility. Each exhibits pronounced non-linear effects and interactions: slope contributions peak at 20–30°, monthly rainfall displays a threshold at 225 mm, and the synergistic effect of high roughness and road density markedly amplifies risk.

(4) Landslide susceptibility exhibits a north–south-high, central-low pattern; high-susceptibility areas are mainly distributed in the northern and southern mountains, the southern urban core and their surrounding areas—regions closely linked to topography, hydrology and anthropogenic forcing. Very-low susceptibility zones occupy the central hills and plains, characterised by gentle topography and high geotechnical stability.

(5) The proposed sample-optimization framework and modelling pipeline provide a transferable protocol for landslide assessment in geologically complex, data-scarce regions, especially in the heavy-rainfall zones of southern China.

Future research should integrate multi-source data, dynamic-threshold modelling and automated sample-ratio optimisation to enhance predictive performance under extreme climatic conditions. Real-time monitoring fused with deep-learning architectures will further advance intelligent, fine-resolution landslide early-warning systems.

Author Contributions: Conceptualization, Y.K.; Methodology, Y.K.; Software, Y.K., K.Z., Z.M., X.C., L.C. and T.X.; Validation, Y.K.; Formal analysis, Y.K.; Investigation, Y.K.; Resources, Y.K.; Data curation, Y.K., C.X., H.K., W.T. and X.K.; Writing—original draft, Y.K. and Z.M.; Writing—review & editing, Y.K.; Visualization, Y.K.; Supervision, Y.K.; Project administration, Y.K.; Funding acquisition, Y.K. and H.W. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China (Grant No. U23A2047 awarded to H.W.), the Xizang Autonomous Region Science and Technology Department (Grant Nos. XZ202401YD0028, XZ202402ZD0001, and XZ202401ZY0057 awarded to H.W.), and Xizang University (Grant No. 2025-GSP-S011 awarded to Y.K.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in mendeley at DOI:10.17632/jprswjdvz4.2. [mendeley].

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

References

1. Seçkin, F.; Hakan, T.; Abdullah, A.; Luigi, L.; Petley, D.N.; Tolga, G. Understanding Fatal Landslides at Global Scales: A Summary of Topographic, Climatic, and Anthropogenic Perspectives. *Nat. Hazards* **2024**, *120*, 6437–6455. [CrossRef]
2. Wen, H.; Li, W.; Xu, C.; Daimaru, H. Landslides in Forests around the World: Causes and Mitigation. *Forests* **2023**, *14*, 629. [CrossRef]

3. Capobianco, V.; Choi, C.E.; Crosta, G.; Hutchinson, D.J.; Jaboyedoff, M.; Lacasse, S.; Nadim, F.; Reeves, H. Effective Landslide Risk Management in Era of Climate Change, Demographic Change, and Evolving Societal Priorities. *Landslides* **2025**, *22*, 2915–2933. [CrossRef]
4. Mateja, J.A.; Nejc, B.; Ela, Š.; Peter, F.; Luigi, G.S.; Anže, M.; Tina, P. Climate Change Increases the Number of Landslides at the Juncture of the Alpine, Pannonian and Mediterranean Regions. *Sci. Rep.* **2023**, *13*, 23085. [CrossRef] [PubMed]
5. Ministry of Natural Resources. *Natural Resources Bulletin of China, 2024*; China Natural Resources News; Ministry of Natural Resources: Beijing, China, 2025.
6. Tao, Z.; Luo, S.; Zhu, C.; He, M. Dynamie Mechanical Monitoring of Landslide and Case Analysis of Failure Process. *J. Eng. Geol.* **2022**, *30*, 177–186. [CrossRef]
7. Zhang, Z.; Huang, X.; Cai, Y.; Fu, J.; Yue, Z.; Yang, R.; Han, C. The Evolution Pattern and Influence of Human Activities of Landslide Driving Factors in Wulong Section of the Three Gorges Reservoir Area. *Chin. J. Geol. Hazard Control* **2022**, *33*, 39–50. [CrossRef]
8. Somogyvári, M.; Chicas, S.D.; Li, H.; Mizoue, N.; Ota, T.; Du, Y. Landslide Susceptibility Mapping Core-Base Factors and Models' Performance Variability: A Systematic Review. *Nat. Hazards* **2024**, *120*, 1–21. [CrossRef]
9. Jia, Z.; Cheng, Z.; Chang, Z.; Li, Q.; Peng, Y.; Jiang, B.; Huang, F. Modeling and Uncertainty in Landslide Susceptibility Prediction Considering the Coupling Mode of Landslide Types. *Earth Sci.* **2025**, *50*, 2311–2329. [CrossRef]
10. Zhang, L.; Jiang, S. Data Driven Weight Model for Reqional Landslide Susceptibility Assessment and Its Application. *Hydrogeol. Eng. Geol.* **2004**, *6*, 33–36. Available online: https://www.zhangqiaokeyan.com/academic-journal-cn_hydrogeology-engineering-geology_thesis/0201254217118 (accessed on 3 October 2025).
11. Al-Najjar, H.A.H.; Pradhan, B.; He, X.; Sheng, D.; Alamri, A.; Gite, S.; Park, H.-J. Integrating Physical and Machine Learning Models for Enhanced Landslide Prediction in Data-Scarce Environments. *Earth Syst. Environ.* **2024**. [CrossRef]
12. Wang, J.; Wang, Y.; Li, Y.; Wei, S.; Li, C.; Wang, Y.; Qi, H. Landslide Susceptibility Assessment Based on Weighted Information Value Model: A Case Study of Chongqing City. *Sci. Soil Water Conserv.* **2023**, *21*, 53–62. [CrossRef]
13. Lu, Y.; Xu, H.; Wang, C.; Yan, G.; Huo, Z.; Peng, Z.; Liu, B.; Xu, C. A Novel Strategy Coupling Optimised Sampling with Heterogeneous Ensemble Machine-Learning to Predict Landslide Susceptibility. *Remote Sens.* **2024**, *16*, 3663. [CrossRef]
14. Marzini, L.; D'Addario, E.; Papasidero, M.P.; Chianucci, F.; Disperati, L. Influence of Root Reinforcement on Shallow Landslide Distribution: A Case Study in Garfagnana (Northern Tuscany, Italy). *Geosciences* **2023**, *13*, 326. [CrossRef]
15. Vanani, A.A.G.; Shoaei, G.; Zare, M. Landslide Susceptibility Mapping in North Tehran, Iran: Linear Regression, Neural Networks, and Fuzzy Logic Approaches. *Geotech. Geol. Eng.* **2024**, *42*, 7159–7186. [CrossRef]
16. Xu, C.; Xu, X. Logistic Regression Model and Its Validation for Hazardmapping of Landslides Triggered by Yushu Earthquake. *J. Eng. Geol.* **2012**, *20*, 326–333. Available online: http://www.gcdz.org/en/article/id/11136 (accessed on 3 October 2025).
17. Xu, C.; Dai, F.; Xu, S.; Xu, X.; He, H.; Wu, X.; Shi, F. Application of Logistic Regression Model on the Wenchuan Earthquaketriggered Landslide Hazard Mapping and Its Validation. *Hydrogeol. Eng. Geol.* **2013**, *40*, 98–104. [CrossRef]
18. Kavzoglu, T.; Sahin, E.K.; Colkesen, I. Landslide Susceptibility Mapping Using GIS-Based Multi-Criteria Decision Analysis, Support Vector Machines, and Logistic Regression. *Landslides* **2014**, *11*, 425–439. [CrossRef]
19. Li, M.X.; Wang, H.Y.; Chen, J.L.; Zheng, K. Assessing Landslide Susceptibility Based on the Random Forest Model and Multi-Source Heterogeneous Data. *Ecol. Indic.* **2024**, *158*, 111600. [CrossRef]
20. Yang, K.; Niu, R.; Song, Y.; Dong, J.; Zhang, H.; Chen, J. Dynamic Hazard Assessment of Rainfall Induced Landslides Using Gradient Boosting Decision Tree with Google Earth Engine in Three Gorges Reservoir Area, China. *Water* **2024**, *16*, 1638. [CrossRef]
21. Wen, H.; Liu, B.; Di, M.; Li, J.; Zhou, X. A SHAP-Enhanced XGBoost Model for Interpretable Prediction of Coseismic Landslides. *Adv. Space Res.* **2024**, *74*, 3826–3854. [CrossRef]
22. Sun, D.; Wu, X.; Wen, H.; Gu, Q. A LightGBM-Based Landslide Susceptibility Model Considering the Uncertainty of Non-Landslide Samples. *Geomat. Nat. Hazards Risk* **2023**, *14*, 2213807. [CrossRef]
23. Song, Y.; Song, Y.; Wang, C.; Wu, L.; Wu, W.; Li, Y.; Li, S.; Chen, A. Landslide Susceptibility Assessment through Multi-Model Stacking and Meta-Learning in Poyang County, China. *Geomat. Nat. Hazards Risk* **2024**, *15*, 2354499. [CrossRef]
24. Shruti, S.; Tarunpreet, B.; Verma, A.K. A Novel Voting Ensemble Model for Spatial Prediction of Landslides Using GIS. *Int. J. Remote Sens.* **2020**, *41*, 929–952. [CrossRef]
25. Zhang, R.; Guan, Y. Application of CNN-LSTM Hybrid Model in Predicting Surface Displacement of Accumula Ted Landslide Sites. *North China Farthquake Sci.* **2025**, *43*, 1–8. [CrossRef]
26. Wang, Y.; Wu, X.; Zhou, K.; Lin, G.; Peng, B.; Zhice, F. Integrating a Multi-Dimensional Deep Convolutional Neural Network with Optimized Sample Selection for Landslide Susceptibility Assessment. *Geo-Spat. Inf. Sci.* **2025**, *15*, 1–21. [CrossRef]
27. Huang, F.; Xiong, H.; Jiang, S.H.; Yao, C.; Fan, X.; Catani, F.; Chang, Z.; Zhou, X.; Huang, J.; Liu, K. Modelling Landslide Susceptibility Prediction: A Review and Construction of Semi-Supervised Imbalanced Theory. *Earth-Sci. Rev.* **2024**, *250*, 104700. [CrossRef]

28. Wu, H.Y.; Zhou, C.; Liang, X.; Wang, Y.; Yuan, P.C.; Wu, L.X. Evaluation of landslide susceptilbility based on sample optimization strategy research. *Geomat. Inf. Sci. Wuhan Univ.* **2023**, *49*, 1–15. [CrossRef]
29. Liu, Y.; Chen, C.; He, Q.; Li, K. Landslide Susceptibility Evaluation Considering Positive and Negative Sample Optimization. *Acta Geod. Cartogr. Sin.* **2025**, *54*, 308–320. [CrossRef]
30. Ge, Q.; Li, J.; Lacasse, S.; Sun, H.; Liu, Z. Data-Augmented Landslide Displacement Prediction Using Generative Adversarial Network. *J. Rock Mech. Geotech. Eng.* **2024**, *16*, 4017–4033. [CrossRef]
31. Liu, M.M. Landslide Susceptibility Analysis Method Considering Sample Optimization and Spatial Characteristics. Ph.D. Thesis, Liaoning Technical University, Fuxin, China, 2024.
32. Miao, Y.; Zhu, A.; Yang, L.; Bai, S.; Liu, J.; Deng, Y. Sensitivity of BCS for Sampling Landslide Absence Datain Andslide Susceptibility Assessment. *Mt. Res.* **2016**, *34*, 432–441. [CrossRef]
33. Yao, X.; Tham, L.G.; Dai, F.C. Landslide Susceptibility Mapping Based on Support Vector Machine: A Case Study on Natural Slopes of Hong Kong, China. *Geomorphology* **2008**, *101*, 572–582. [CrossRef]
34. Miao, Y.; Zhu, A.; Yang, L.; Bai, S.; Zeng, C. A New Method of Pseudo Absence Data Generation in Landslide Susceptibility Mapping. *Geogr. Geo-Inf. Sci.* **2016**, *32*, 61–67+127. [CrossRef]
35. Cui, Y.; Zhu, L.; Xu, M.; Miao, H. Optimizing TSES Method Based on the Environmental Factors to Select Negative Samples and Its Application in Landslide Susceptibility Evaluation. *Bull. Geol. Sci. Technol.* **2024**, *43*, 192–199. [CrossRef]
36. Guo, Y.; Dou, J.; Xiang, Z.; Ma, H.; Dong, A.; Luo, W. Susceptibility Evaluation of Wenchuan Coseismic Landslides by Gradient-boosting Decision Tree and Random Forest Based on Optimal Negative Samplesampling Strategies. *Bull. Geol. Sci. Technol.* **2024**, *43*, 251–265. [CrossRef]
37. Zhou, X.; Huang, F.; Wu, W.; Zhou, C.; Zeng, S.; Pan, L. Regional Landslide Susceptibility Prediction Based on Negative Sample Selected by Coupling Information Value Method. *Adv. Eng. Sci.* **2022**, *54*, 25–35. [CrossRef]
38. Fu, Y.; Fan, Z.; Li, X.; Wang, P.; Sun, X.; Ren, Y.; Cao, W. The Influence of Non-Landslide Sample Selection Methods on Landslide Susceptibility Prediction. *Land* **2025**, *14*, 722. [CrossRef]
39. Gu, T.; Duan, P.; Wang, M.; Li, J.; Zhang, Y. Effects of Non-Landslide Sampling Strategies on Machine Learning Models in Landslide Susceptibility Mapping. *Sci. Rep.* **2024**, *14*, 7201. [CrossRef] [PubMed]
40. Kong, Y.; Wu, H.; Xu, C.; Sun, J.; Zhu, K.; Zhang, C.; Zhou, J.; Xu, T.; Su, T.; Zhang, Z.; et al. Landslide Susceptibility Mapping Using an Entropy Index-Based Negative Sample Selection Strategy: A Case Study of Luolong County. *PLoS ONE* **2025**, *20*, e0322566. [CrossRef]
41. Zhang, L. Research on the Mode of Mineland Reclamation in the County of Guiyang. Master's Thesis, Hunan Normal University, Changsha, China, 2013.
42. Wang, P. Research on Guiyang County HV Distribution Network Planning. Master's Thesis, Hunan University, Changsha, China, 2016.
43. Zhang, Y.; Ming, D.; Zhao, W.; Xu, L.; Zhao, Z.; Liu, R. The Extraction and Analysis of Luding Earthquake—Induced Landslide Based on High- Resolution Optical Satellite Images. *Remote Sens. Nat. Resour.* **2023**, *35*, 161–170. [CrossRef]
44. Liu, P.; Wei, Y.; Wang, Q.; Chen, Y.; Xie, J. Research on Post-Earthquake Landslide Extraction Algorithm Based on Improved U-Net Model. *Remote Sens.* **2020**, *12*, 894. [CrossRef]
45. Zhu, Y.; Sun, D.; Wen, H.; Zhang, Q.; Ji, Q.; Li, C.; Zhou, P.; Zhao, J. Considering the Effect of Non-Landslide Sample Selection on Landslide Susceptibility Assessment. *Geomat. Nat. Hazards Risk* **2024**, *15*, 2392778. [CrossRef]
46. Cui, Y.L.; Yang, W.H.; Xu, C.; Wu, S. Distribution of Ancient Landslides and Landslide Hazard Assessment in the Western Himalayan Syntaxis Area. *Front. Earth Sci.* **2023**, *11*, 1135018. [CrossRef]
47. Lu, C.; Bo, Z. A New Slope Unit Extraction Method Based on Improved Marked Watershed. *Matec Web Conf.* **2018**, *232*, 04070. [CrossRef]
48. Yu, C.; Chen, J. Application of a GIS-Based Slope Unit Method for Landslide Susceptibility Mapping in Helong City: Comparative Assessment of ICM, AHP, and RF Model. *Symmetry* **2020**, *12*, 1848. [CrossRef]
49. Liu, S.; Zhu, J.; Yang, D.; Ma, B. Comparative Study of Geological Hazard Evaluation Systems Using Grid Units and Slope Units under Different Rainfall Conditions. *Sustainability* **2022**, *14*, 16153. [CrossRef]
50. *GB/T 50218-2014*; Standard for Engineering Classification of Rock Mass. China Planning Press: Beijing, China, 2014.
51. Riaz, M.T.; Basharat, M.; Ahmed, K.S.; Sirfraz, Y.; Shahzad, A.; Shah, N.A. Failure Mechanism of a Massive Fault–Controlled Rainfall–Triggered Landslide in Northern Pakistan. *Landslides* **2024**, *21*, 2741–2767. [CrossRef]
52. Yu, L.B.; Wang, Y.; Pradhan, B. Enhancing Landslide Susceptibility Mapping Incorporating Landslide Typology via Stacking Ensemble Machine Learning in Three Gorges Reservoir, China. *Geosci. Front.* **2024**, *15*, 101802. [CrossRef]
53. Zhao, P.; Wen, G.; He, Z.; Wang, G.; Chen, L.; Shen, X.; Wang, K.; Tang, H. Shallow Landslide Susceptibility Assessment in Jinsha River Basin Based on Machine Learning Models. *Water Resour. Hydropower Eng.* **2024**, *55*, 53–70. [CrossRef]
54. Kumar, D.; Thakur, M.; Dubey, C.S.; Shukla, D.P. Landslide Susceptibility Mapping & Prediction Using Support Vector Machine for Mandakini River Basin, Garhwal Himalaya, India. *Geomorphology* **2017**, *295*, 115–125. [CrossRef]

55. Alfonso, M.R.-C.; Luis, F.P.-S.; Mario, G.T.-V.; Juan, P.M.; Ana, C.S.-R. Linear Discriminant Analysis to Describe the Relationship between Rainfall and Landslides in Bogota, Colombia. *Landslides* **2016**, *13*, 671–681. [CrossRef]
56. Du, P.; Chen, N.S.; Wu, K.N.; Li, Z.; Zhang, Y.Y.L. Evaluation of landslide susceptibility in southeast tibet based on a random forest model. *J. Chengdu Univ. Technol. (Sci. Technol. Ed.)* **2024**, *51*, 328–344. [CrossRef]
57. Halder, K.; Srivastava, A.K.; Ghosh, A.; Das, S.; Banerjee, S.; Pal, S.C.; Chatterjee, U.; Bisai, D.; Ewert, F.; Gaiser, T. Improving Landslide Susceptibility Prediction through Ensemble Recursive Feature Elimination and Meta-Learning Framework. *Sci. Rep.* **2025**, *15*, 5170. [CrossRef]
58. Shapley, L.S. *A Value for N-Person Games*; RAND Corporation: Santa Monica, CA, USA, 1952.
59. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Auckland, New Zealand, 2–6 December 2024; Curran Associates Inc.: Red Hook, NY, USA, 2017; pp. 4768–4777.
60. Kanwar, M.; Pokharel, B.; Lim, S. A New Random Forest Method for Landslide Susceptibility Mapping Using Hyperparameter Optimization and Grid Search Techniques. *Int. J. Environ. Sci. Technol.* **2025**, *22*, 10635–10650. [CrossRef]
61. Xu, C.; Dai, F.C.; Yao, X.; Chen, J.; Tu, X.B.; Sun, Y.; Wang, Z.Y. Gis-based landslide susceptibility assessment using analytical hierarchy process in wenchuan earthquake region. *Chin. Joural Rock Mech. Eng.* **2009**, *28*, 3978–3985. [CrossRef]
62. Rahman, A.S.A.; A'kif, A.F.; Mohamed, K.K.; Nouh, M.A.; Rida, A.A. Spatial Mapping of Landslide Susceptibility in Jerash Governorate of Jordan Using Genetic Algorithm-Based Wrapper Feature Selection and Bagging-Based Ensemble Model. *Geomat. Nat. Hazards Risk* **2022**, *13*, 2252–2282. [CrossRef]
63. Zheng, D.; Li, Y.; Yan, C.; Wu, H.; Yamashiki, Y.A.; Gao, B.; Nian, T. Landslide Susceptibility Assessment Using AutoML-SHAP Method in the Southern Foothills of Changbai Mountain, China. *Landslides* **2025**, *22*, 1855–1875. [CrossRef]
64. Zhang, T.; Li, L.; Liu, F.; Hong, Z.; Qian, F.; Hu, B.; Zhang, M. Evaluation of Loess Landslide Susceptibility Based on Optimised Max Ent Model: A Case Study of Wuqi County in Shaanxi Province. *Northwestern Geol.* **2025**, *58*, 172–185. [CrossRef]
65. Cheng, J.; Xu, C.; Xu, X.; Zhang, S.; Zhu, P. Modeling Seismic Hazard and Landslide Occurrence Probabilities in Northwestern Yunnan, China: Exploring Complex Fault Systems with Multi-Segment Rupturing in a Block Rotational Tectonic Zone. *Nat. Hazards Earth Syst. Sci.* **2025**, *25*, 857–877. [CrossRef]
66. Liu, Y.; Chen, C. Landslide Susceptibility Evaluation Method Considering Spatial Heterogeneity and Feature Selection. *Acta Geod. Artographica Sin.* **2024**, *53*, 1417–1428. [CrossRef]
67. Wang, S.; Zhuang, J.; Fan, H.; Niu, P.; Jia, K.; Wang, J. Evaluation of Landslide Suseeplibilitly Based on Frequeney Raio and Ensemble Leaming Taking theBalang-Dege Seion in the Upstream of Jinsha River as an Example. *J. Eng. Geol.* **2022**, *30*, 817–828. [CrossRef]
68. Liu, L.; Xiao, H.; Wang, C.; Yao, T. Landslide Analysis Based on Susceptibility to Factors Causing Geological Disasters in Red Beds Area of Hunan Province. *Min. Metall. Eng.* **2024**, *44*, 169–174. [CrossRef]
69. Liu, R.; Xu, Q.; Pu, C.; Xu, F.; Wang, X.; Zhao, H.; Zhu, X.; He, N. Characteristics of Landslides Induced by Typhoon "GaeMi" in Zixing, Hunan, July 2024, and Their Geological Control Factors. *Geomat. Inf. Sci. Wuhan Univ.* **2025**, 1–22. [CrossRef]
70. Yang, J. Uncertainty Analysis of Rainfall-Induced Landslide Susceptibility Prediction and Risk Assessment Modeling. Master's Thesis, Nanchang University, Nanchang, China, 2022.
71. Xu, Z.; Xiao, N.; Liu, Z.; Li, Y. Research on Geological Disaster-Prone Area Basedon Susceptibility Index Method in Guiyang County. *J. Chang. Inst. Technol. (Nat. Sci. Ed.)* **2012**, *13*, 54–59. [CrossRef]
72. Hong, H.; Miao, Y.; Liu, J.; Zhu, A.-X. Exploring the Effects of the Design and Quantity of Absence Data on the Performance of Random Forest-Based Landslide Susceptibility Mapping. *Catena* **2019**, *176*, 45–64. [CrossRef]
73. Reza, P.H.; Aiding, K.; Norman, K.; Farzin, S. Investigating the Effects of Different Landslide Positioning Techniques, Landslide Partitioning Approaches, and Presence-Absence Balances on Landslide Susceptibility Mapping. *Catena* **2020**, *187*, 104364. [CrossRef]
74. Sun, D.; Wen, H.; Wang, D.; Xu, J. A Random Forest Model of Landslide Susceptibility Mapping Based on Hyperparameter Optimization Using Bayes Algorithm. *Geomorphology* **2020**, *362*, 107201. [CrossRef]
75. Li, Y. Method for the Warning of Precipitation Induced Landslides. Ph.D. Thesis, China University of Geosciences (Beijing), Beijing, China, 2005.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.