https://doi.org/10.5194/egusphere-2026-4550
Preprint. Discussion started: 6 August 2026

A Statistical-Physical Prior-Guided Framework for Multi-Class Landslide Susceptibility Mapping

Haixia Feng[a], Qingwu Hu[b]*, Yizhou Lan[c], Daoyuan Zheng[a], Yuxiang Xu[a], Jiayuan Li[a], Shunli Wang[a], Pengcheng Zhao[a]
[a]*School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China*
[b]*School of Remote Sensing and Information Engineering, Hubei Luojia Laboratory, Wuhan University, Wuhan, China*
[c]*School of Earth and Space Sciences, Peking University, Beijing China*

Correspondence:
Qingwu Hu
huqw@whu.edu.cn
18971070362
School of Remote Sensing & Information Engineering, Hubei Luojia Laboratory, Wuhan University, No. 129 Luoyu Road, Hongshan District, Wuhan 430072, China

Abstract: Regional landslide susceptibility mapping generally treats landslide/non-landslide discrimination as a binary learning task and subsequently divides predicted probabilities into several qualitative classes. This two-stage strategy does not directly represent the susceptibility gradient and may therefore reduce the spatial concentration of historical landslides within the very high susceptibility zone. This study proposes a statistical-physical prior-guided framework for direct five-class landslide susceptibility mapping along the Ancient Qin-Shu Roads in the Qinling-Daba Mountains. A landslide inventory containing 1,407 records and 14 conditioning factors was compiled. First, certainty factor (CF) analysis was used to generate a initial susceptibility map. The Shallow Landsliding Stability model (SHALSTAB) was then introduced to provide a physically meaningful stability mask with which the zonation was corrected and samples of very high, high, moderate, low, and very low susceptibility were constructed. Logistic regression (LR), support vector machine (SVM), extreme gradient boosting (XGBoost), random forest (RF), Feature Tokenizer Transformer (FT-Transformer), and Tabular Prior-Data Fitted Network (TabPFN) were compared under binary and multi-class settings. The results show that RF performed well in both tasks, attaining an area under the receiver operating characteristic curve (AUC) of 91.08% for binary classification and 92.53% for multi-class classification. In the binary result, the landslide-prone zone occupied 19.3% of the study area and contained 87.3% of the historical landslides, yielding a Hit Optimization Index (HOI) of 1.68. In the corresponding multi-class result, the very high susceptibility zone occupied only 13.5% of the study area while containing 91.8% of the historical landslides, with an HOI of 1.78. These findings indicate that multi-class learning provides greater historical-landslide coverage and spatial concentration. Compared with labels derived from CF alone, the SHALSTAB-constrained scheme increased accuracy from 65.05% to 68.60%, AUC from 91.30% to 92.53%, and HOI from 1.72 to 1.78, demonstrating that physical constraints can improve susceptibility zonation. Shapley additive explanations (SHAP) showed that elevation, land use and land cover (LULC), and distance to roads were the three principal controls on identification of very high susceptibility zones.
Keywords: Landslide susceptibility; Multi-class classification; Statistical-physical prior; Tabular learning

1. Introduction

Landslide susceptibility assessment addresses the question of where landslides are more likely to occur under given topographic, geological, hydrological, and anthropogenic conditions. It provides a basis for subsequent hazard and risk assessment, territorial spatial planning, and disaster prevention along transportation corridors (Fell et al., 2008; Reichenbach et al., 2018). With advances in high-resolution digital elevation models (DEM), multi-source remote sensing, interferometric synthetic aperture radar (InSAR), geological-hazard databases, and machine learning, landslide susceptibility assessment has evolved from early expert judgment and statistical overlay into a diversified research system encompassing statistical learning, machine learning, process-based modeling, and physics-informed integration (Merghadi et al., 2020; Feng et al., 2022). Data-driven models are now the mainstream approach. Goetz et al. (2015) compared generalized linear model (GLM), generalized additive model (GAM), weights of evidence (WoE), support vector machine (SVM), and random forest (RF), and concluded that no single model is optimal for all regions; model selection should jointly consider predictive accuracy, spatial transferability, and cartographic plausibility. Jiang et al. (2026) systematically reviewed machine-learning, ensemble-learning, and hybrid approaches to landslide susceptibility mapping. Yang et al. (2026) applied machine learning to the Tibetan Plateau and identified key drivers under contrasting topographic, geological, and climatic conditions.

Beyond purely data-driven models, increasing attention has been paid to coupling statistical learning with geological, hydrological, and geotechnical properties. Accordingly, landslide susceptibility research has shifted from simple comparisons of classifier accuracy toward sample quality, model interpretability, and the integration of physical information. Zhang et al. (2023) combined Shapley additive explanations (SHAP) with extreme gradient boosting (XGBoost) to analyze the spatial heterogeneity of landslide conditioning factors and improve model interpretability. Edrich et al. (2024) developed an RF framework for shallow-landslide susceptibility mapping that follows the Findable, Accessible, Interoperable, and Reusable (FAIR) principles. By systematically examining factor combinations, positive-to-negative sample ratios, and the spatial representativeness of training samples, they showed that directly extrapolating a local landslide inventory to an entire study area can introduce substantial bias.

https://doi.org/10.5194/egusphere-2026-4550
Preprint. Discussion started: 6 August 2026

Yang et al. (2024) compared the analytic hierarchy process (AHP), frequency ratio (FR), logistic regression (LR), deep learning, XGBoost, and Light Gradient-Boosting Machine (LightGBM) across different sample sizes. They found XGBoost and LightGBM relatively robust to changes in sample size and used SHAP to identify the importance of anthropogenic factors such as roads and land use. Le et al. (2024) further integrated Bayesian optimization, bootstrap resampling, Monte Carlo simulation, and SHAP to jointly evaluate the accuracy, predictive uncertainty, and factor contributions of multiple machine-learning models, helping shift susceptibility research from selecting the best model to determining whether and why a prediction is reliable. Liu et al. (2024) proposed a physics-informed optimization approach for data-driven susceptibility assessment. This approach partly reduced the uncertainty associated with random negative sampling and moved sample construction from empirical sampling toward stability-constrained sampling. Cui et al. (2025) combined a physical probabilistic model with a convolutional neural network (CNN) to develop a hybrid model for rainfall-induced landslide susceptibility. Li et al. (2025) coupled geographically weighted regression (GWR) with tabular neural network (TabNet) and used spatial regression relationships to select negative samples with low-risk semantics, demonstrating that both negative-sample construction and the choice of tabular model affect mapping stability. Other studies have begun to examine the effects of dynamic environmental variables and sample construction. Using California as a case study, Han and Semnani (2025) applied XGBoost to future-climate landslide susceptibility and identified class imbalance, conditioning-factor selection, and climatic extrapolation beyond the training-data range as key sources of uncertainty. Mejia-Manrique et al. (2025) incorporated rainfall forecasts, real-time soil moisture, and saturation into an attention-based CNN to produce dynamic susceptibility maps for extreme rainfall, indicating a transition from static environmental analysis toward event-driven prediction. Cifci et al. (2026) optimized artificial neural networks using evolutionary algorithms, further demonstrating that ensemble optimization remains an important route to improving model stability. On the Central Yunnan Plateau, Chen et al. (2026) optimized classifications of rainfall, lithology, and land use using point-biserial correlation and decision trees and integrated LR, SVM, and RF results. They found that 24-h rainfall, 30-day cumulative rainfall, loose rock and soil, carbonate rocks, and distances of less than 1,000 m from roads were important regional controls. On the loess plateau of western Shanxi, Zhou et al. (2026) incorporated cohesion, internal friction angle, permeability, and collapsibility coefficient into SVM, RF, and CNN models and proposed a weighted multi-model SHAP method. Although geotechnical parameters were not always the most important variables, they enhanced model generalization and physical interpretability. In the Maiji District of the western Qinling Mountains, Sang et al. (2026) stratified non-landslide samples by engineering-geological rock group and combined CatBoost, XGBoost, and LightGBM with several hyperparameter-optimization algorithms, showing that both the geological representativeness of negative samples and the model-optimization strategy affect susceptibility zonation.

Most existing regional studies formulate landslide inventories and environmental factors as a binary landslide/non-landslide problem and use LR, SVM, RF, or gradient-boosted trees to estimate landslide occurrence probabilities (Goetz et al., 2015; Merghadi et al., 2020). Multi-class classification can incorporate susceptibility gradients directly into the learning objective, allowing a model to learn differences among classes during training. Its central difficulty, however, is the absence of reliable multi-level labels that cover the entire region. Labels constructed solely from historical landslide density or statistical associations may amplify inventory omissions and fluctuations in sparse classes. Conversely, reliance on physical models alone is constrained by the spatial heterogeneity of geotechnical parameters, steady-state hydrological assumptions, and computational scale. Previous studies have shown that combining empirical statistical relationships with slope stability models can balance regional data availability against process plausibility (Dietrich et al., 2001; Goetz et al., 2011). This approach is also consistent with the core principle of theory-guided data science, in which domain knowledge constrains data-driven models (Karpatne et al., 2017). Meanwhile, the Feature Tokenizer Transformer (FT-Transformer) and Tabular Prior-Data Fitted Network (TabPFN) offer new options for geoscientific classification with small to medium-sized datasets. FT-Transformer learns complex interactions among variables through feature-wise attention, whereas TabPFN uses pretraining on synthetic tabular tasks for rapid small-sample inference (Gorishniy et al., 2021; Hollmann et al., 2025). The application of TabNet by Li et al. (2025) demonstrates the potential of deep models designed for structured geoscientific data, although tree-based models generally remain highly competitive. The suitability of FT-Transformer and TabPFN for direct five-class landslide susceptibility learning, and whether they provide stable gains over conventional ensemble models such as RF and XGBoost, therefore require evaluation using identical samples, data partitions, and metrics.

Against this background, this study proposes a statistical-physical prior-guided multi-class framework for direct five-class regional landslide susceptibility mapping. The specific objectives are to: (1) use the certainty factor (CF) to generate a statistical prior covering the entire study area and introduce the Shallow Landsliding Stability model (SHALSTAB) stability results to correct the initial susceptibility map, thereby constructing a five-class sample set informed by both statistical associations and physical constraints; (2) compare the landslides model performance and spatial mapping results of LR, SVM, XGBoost, RF, FT-Transformer, and TabPFN under binary and multi-class settings using identical samples and factors; (3) use the Hit Optimization Index (HOI) to evaluate the historical-landslide coverage and spatial concentration of the resulting susceptibility maps and compare them with binary baselines constructed from the same factors; and (4) interpret the principal controls on very high susceptibility zones using SHAP.

2. Study area and materials

https://doi.org/10.5194/egusphere-2026-4550
Preprint. Discussion started: 6 August 2026

2.1. Study area

The study area is located in the Qinling-Daba Mountains and adjacent regions, between 105°40′-110°30′ E and 32°10′-35°30′ N, as shown in Fig. 1. It includes Longnan in Gansu Province and Baoji, Xi'an, Hanzhong, and Shangluo in Shaanxi Province (Feng et al., 2024; Zheng et al., 2025). Buffers were established along four north-south Ancient Qin-Shu Roads, producing several elongated and spatially separated corridor areas. This delineation preserves the pronounced environmental gradient from the margin of the Guanzhong Plain, across the main Qinling ridge, to the northern margin of the Hanzhong Plain. The terrain is dominated by moderately and highly rugged mountains, with interspersed ridges, gullies, footslopes, and river terraces; local elevations exceed 3,000 m. Fault structures and stratigraphic lithologies are complex, and Quaternary unconsolidated deposits, slate, phyllite, mudstone, sandstone, carbonate rocks, and granitic bodies are all present. Alternating weakly bedded rock groups and competent rock masses, together with fault-related fracturing, river incision, and road-slope excavation, provide the material and topographic conditions for shallow landslides. Located in the Qinling Mountains and their northern and southern transition zones, the area has a climate jointly controlled by orographic uplift, monsoon circulation, and mountain moisture transport. Mean annual rainfall is approximately 668-1,035 mm and generally increases from north to south. The drainage network is well developed, and rivers and gullies commonly follow structural lineaments and topographic lows.

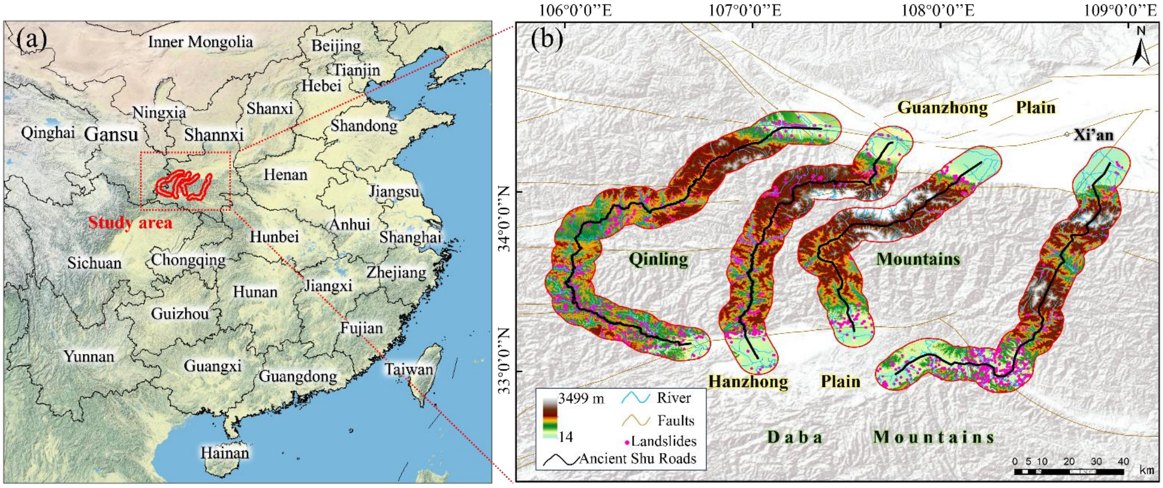

Fig. 1. Geographical location and detailed distribution of the study area. (a) Geographical location of the study area; (b) Elevation, rivers, faults and landslide distribution of the study area.

2.2. Materials

Historical hazard data were compiled from expert-verified landslides recorded in national geological-hazard investigations and censuses, together with disaster reports from the Shaanxi Environmental Monitoring Center and relevant county and district authorities. The final inventory comprised 1,407 positive landslide samples, with fields describing hazard type, occurrence time, location, coordinates, scale, triggering factors, and impacts. Based on the landslide characteristics of the study area, data availability, and previous studies (Chang et al., 2023; Liu et al., 2023; Wei et al., 2024; Huang et al., 2026; Ma et al., 2026), 14 core conditioning factors were selected from four categories: topography, geological setting, vegetation and hydrology, and human activity. These factors were elevation, aspect, slope, relief amplitude, lithology, distance to faults, distance to earthquake epicenters, soil type, normalized difference vegetation index (NDVI), distance to rivers, topographic wetness index (TWI), rainfall, land use and land cover (LULC), and distance to roads, as shown in Table 1. The topographic factors were elevation, aspect, slope, and relief amplitude. Elevation represents the vertical geomorphic position of a slope and is closely related to regional hydrothermal conditions, vegetation development, weathering and erosion intensity, and human activity, thereby indirectly affecting slope materials and the environment in which landslides develop. Aspect controls incoming solar radiation, and resulting differences in thermal, hydrological, and weathering processes affect slope stability. Relief amplitude characterizes land-surface form and ruggedness and reflects spatial differences in flow convergence and divergence. Slope directly affects internal stress distributions and surface runoff; steeper slopes generally experience larger downslope forces and stronger runoff erosion, which reduce stability. Relief amplitude characterizes land-surface form and ruggedness and reflects spatial differences in flow convergence and divergence. Geological factors included lithology, distance to faults, distance to epicenters, and soil type. Lithology is a basic geological control because variations in rock and soil type produce differences in strength, weathering, and the availability of loose material, thereby governing landslide distribution and scale. Fault activity and earthquakes are two major external controls on regional landsliding; slopes closer to active faults and earthquake epicenters are more strongly affected by tectonic compression and seismic shaking and are therefore more susceptible to failure. Soil type influences infiltration and regional hydrological cycling

https://doi.org/10.5194/egusphere-2026-4550
Preprint. Discussion started: 6 August 2026

and, together with lithology, determines the mechanical strength and water status of shallow slope materials. Vegetation and hydrological factors included NDVI, distance to rivers, TWI, and rainfall. Vegetation exerts a dual control on slope stability: dense roots can reinforce soil and inhibit shallow landslides, whereas the added weight of vegetation may intensify creep deformation in thick, deep-seated landslides. River incision reshapes regional topography; undercutting at a slope toe can create a free face and steepen the slope, directly triggering instability. NDVI represents surface vegetation cover, distance to rivers indicates the spatial influence of water bodies, and TWI reflects surface-water accumulation and soil moisture. The human-activity factors were LULC and distance to roads. Landslides seriously threaten people and property, while engineering activities can substantially alter the original stability of slopes. As socioeconomic development, population growth, and urbanization intensify disturbance of the geological environment, construction and road excavation increasingly trigger landslides, rockfalls, and related geological hazards worldwide.

Table 1
Landslide conditioning factors and data sources.

Category	No.	Conditioning factor	Description	Resolution	Data source
Topographic	1	DEM	Surface elevation	30 m	National 30 m DEM Dataset
Topographic	2	Aspect	Slope aspect	30 m	Derived from DEM
Topographic	3	Slope	Slope gradient	30 m	Derived from DEM
Topographic	4	Relief amplitude	Terrain relief	30 m	Resource and Environment Science and Data Center
Geological	5	Lithology	Lithological units	1:2,500,000	China Geological Cloud Platform
Geological	6	Distance to faults	Distance to the nearest fault	—	China Geological Cloud Platform
Geological	7	Distance to epicenters	Distance to the nearest earthquake epicenter	—	China Earthquake Data Center
Geological	8	Soil type	Surface soil classification	1:1,000,000	Resource and Environment Science and Data Center
Vegetation & hydrological	9	NDVI	Vegetation coverage	1 km	NASA Remote Sensing Data Platform
Vegetation & hydrological	10	Distance to rivers	Distance to the nearest river	—	Global Geographic Data Download Platform
Vegetation & hydrological	11	TWI	Topographic Wetness Index	30 m	Derived from DEM
Vegetation & hydrological	12	Rainfall	Annual cumulative precipitation	0.05°	Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS)
Human activity	13	LULC	Land use and land cover	30 m	Global Land Cover Dataset
Human activity	14	Distance to roads	Distance to the nearest road	—	Global Geographic Data Download Platform

3. Methodology

3.1. Overall framework

The overall workflow is shown in Fig. 2 and comprises three stages: data acquisition and preprocessing, sample construction, and landslide susceptibility modeling and evaluation. First, CF was used to generate a statistical susceptibility map, which was corrected using the SHALSTAB stability results to construct a five-class training dataset. Second, the landslide inventory and conditioning factors were subjected to consistent projection, resampling, and spatial registration. LR, FT-Transformer, XGBoost, SVM, TabPFN, and RF were trained for multi-class classification using identical samples and factors, with binary classification serving as the control. Finally, Recall, F1 score, accuracy, area under the receiver operating characteristic curve (AUC), and HOI were used to evaluate sample-level performance and spatial mapping, and SHAP was used to interpret the key controls identified by the best-performing model.

https://doi.org/10.5194/egusphere-2026-4550
Preprint. Discussion started: 6 August 2026

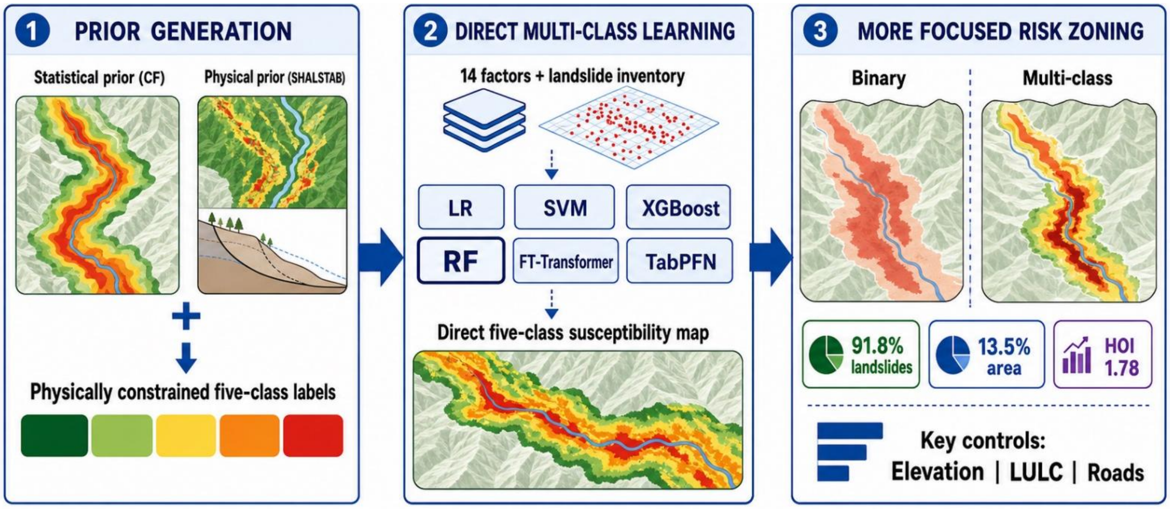

Fig. 2. Framework of the multi-class landslide susceptibility assessment.

3.2. Initial landslide susceptibility zonation map

The CF quantifies the statistical relationship between landslide occurrence and the classes of individual conditioning factors. By comparing the landslide probability within each factor class with the regional mean probability, it characterizes the positive or negative contribution of that class to landslide development (Qin et al., 2024). Let N be the total number of valid grid cells in the study area and L the number of landslides. If class i of factor j contains N_{ij} valid cells and L_{ij} landslides, the regional prior probability of landsliding, P_s, and the conditional landslide probability within that class, $P_{a,ij}$, are respectively defined as:

$$P_s = \frac{L}{N}$$
$$P_{a,ij} = \frac{L_{ij}}{N_{ij}} \tag{1}$$

The CF value for a class is calculated from the difference between its conditional probability and the regional prior probability:

$$CF_{ij} = \frac{P_{a,ij} - P_s}{P_{a,ij}(1 - P_s)}, P_{a,ij} \geq P_s$$
$$CF_{ij} = \frac{P_{a,ij} - P_s}{P_s(1 - P_{a,ij})}, P_{a,ij} < P_s \tag{2}$$

CF ranges from -1 to 1. A value greater than 0 indicates that the landslide probability within the factor class exceeds the regional average and is positively associated with landslide occurrence. A value less than 0 indicates a probability below the regional average and a negative association, whereas CF = 0 indicates approximate equality with the regional average. After the CF value of each factor class was obtained, the mean absolute CF strength of each factor was calculated and used to determine the factor weight:

$$S_j = \frac{\sum_i N_{ij}|CF_{ij}|}{\sum_i N_{ij}}$$
$$W_j = \frac{S_j}{\sum_j S_j} \tag{3}$$

where S_j represents the overall discriminatory power of factor j and W_j is its normalized weight. For each valid grid cell x, the corresponding CF value was retrieved from the class to which the cell belonged for each conditioning factor. The weighted values were then summed to obtain the initial landslide susceptibility index:

$$LSI(x) = \sum_j W_j \cdot CF_j(x) \tag{4}$$

The study area was divided by quantiles into five susceptibility classes: very low, low, moderate, high, and very high, producing the initial CF susceptibility map. To reduce local overestimation that might result from a single statistical method, the integrated SHALSTAB result was introduced as a constraint on the initial CF zonation. SHALSTAB is a representative physically based model that generally uses a DEM and integrates topographic conditions, geotechnical parameters, and stability criteria to identify spatial units with a clear tendency toward failure (Montgomery and Dietrich, 1994; Ye et al., 2025). Model stability was represented

https://doi.org/10.5194/egusphere-2026-4550
Preprint. Discussion started: 6 August 2026

by the critical saturation ratio, which describes the relative degree of saturation required for a slope to reach the threshold of instability, as follows:

$$m_c(x) = \frac{\gamma}{\gamma_w}\left(1 - \frac{\tan\theta}{\tan\varphi} + \frac{C}{\gamma z \cos^2\theta \tan\varphi}\right) \tag{5}$$

where m_c is the relative degree of saturation required for grid cell x to reach the critical failure state; γ is the natural unit weight of the soil or rock mass; γ_w is the unit weight of water; C is cohesion; z is the effective soil thickness; θ is the slope angle; and φ is the internal friction angle. When $m_c(x) \geq 1$, the cell does not reach the critical failure state even under full saturation and is therefore classified as stable; all other cells are classified as unstable. The SHALSTAB result was accordingly used as a physical constraint on the initial CF susceptibility zonation. SHALSTAB input data were constructed from the 30 m DEM, and all model outputs were resampled to a common spatial resolution. Topographic parameters were primarily derived from the DEM, whereas geotechnical parameters were assigned by lithological type; gravitational acceleration was treated as a constant. Drawing on published parameters and related information for landslides in the Qinling-Daba Mountains (Qiang et al., 2015; Song et al., 2022), initial values of cohesion C, internal friction angle φ, natural unit weight γ, and hydraulic conductivity K_s were specified for each rock type (Table 2). Using an inventory-based inverse-calibration approach, C and φ were calibrated through simulation trials against the historical landslides inventory. Because regional geotechnical parameters are difficult to determine and spatially uncertain, their values across the study area represent ranges rather than unique fixed values.

Table 2

Soil parameter values corresponding to different lithological classes in the SHALSTAB model.

Lithology type	C (kPa)	φ (°)	γ (kN/m³)	K_s (m/s)
Carbonate–clastic mixed rocks	50	31	24.5	5.0E-7
Granite / Monzogranite	95	38	26	1.0E-7
Marlstone	38	28	23.5	1.0E-7
Marble	75	36	26	5.0E-7
Weak interlayers in coal-bearing strata	18	20	21.5	1.0E-8
Granodiorite / Quartz diorite / Diorite	90	37	26.5	1.0E-7
Ophiolite / Ultramafic rocks	65	32	27	5.0E-8
Loess / Paleosol	18	24	17.5	5.0E-6
Loose alluvial–proluvial sand and gravel deposits	8	30	19.5	1.0E-4
Gabbro / Diabase / Mafic intrusive rocks	95	38	27.5	1.0E-7
Conglomerate / Sandy conglomerate	38	34	22.8	1.0E-5
Slate	30	25	23	1.0E-8
Limestone	70	35	25.5	1.0E-6
Rhyolite / Dacite / Acidic volcanic rocks	70	35	24.5	5.0E-7
Undifferentiated mixed lithology	35	28	23	5.0E-7
Silty clay / Clay / Sand–gravel mixed Quaternary deposits	16	22	18.8	2.0E-5
Dolomite	75	36	26	8.0E-7
Water bodies / Lakes / Rivers	—	—	—	—

The SHALSTAB result was converted into a stable-area mask and spatially overlaid on the CF result. Grid cells classified by CF as very high, high, moderate, or low susceptibility were downgraded by one class if they fell within a SHALSTAB stable area; very low susceptibility cells remained unchanged. This procedure produced the SHALSTAB-constrained initial CF susceptibility map.

3.3. Data processing and sample construction

The landslide conditioning factors are shown in Fig. 3. Topographic factors, including aspect, slope, and relief amplitude, were derived from the DEM. Based on the lithology, the study area were grouped into six classes: unconsolidated gravelly-soil deposits (A), massive hard metamorphic rocks (B), massive hard intrusive rocks (C), schistose-bedded weak metamorphic rocks (D), schistose-massive moderately hard metamorphic rocks (E), and medium- to thick-bedded intercalated hard and soft metamorphic rocks (F). Euclidean distances from each grid cell to the nearest fault, earthquake epicenter, river, and road were calculated in ArcGIS from the corresponding vector datasets. Soil was classified as luvisols, semi-luvisols, primosols, hydromorphic soils, anthroposols, or others. LULC was grouped into cultivated land, forest land, shrubland, grassland, water body, bare land, and impervious surface. All datasets were transformed to the WGS 84 spatial reference and a 30-m grid. Lower-resolution continuous data were resampled by bilinear interpolation, whereas categorical data were resampled using the nearest-neighbor method. After

https://doi.org/10.5194/egusphere-2026-4550
Preprint. Discussion started: 6 August 2026

harmonization of projection, resolution, and spatial extent, these factors served as the base data for susceptibility assessment and sample construction.

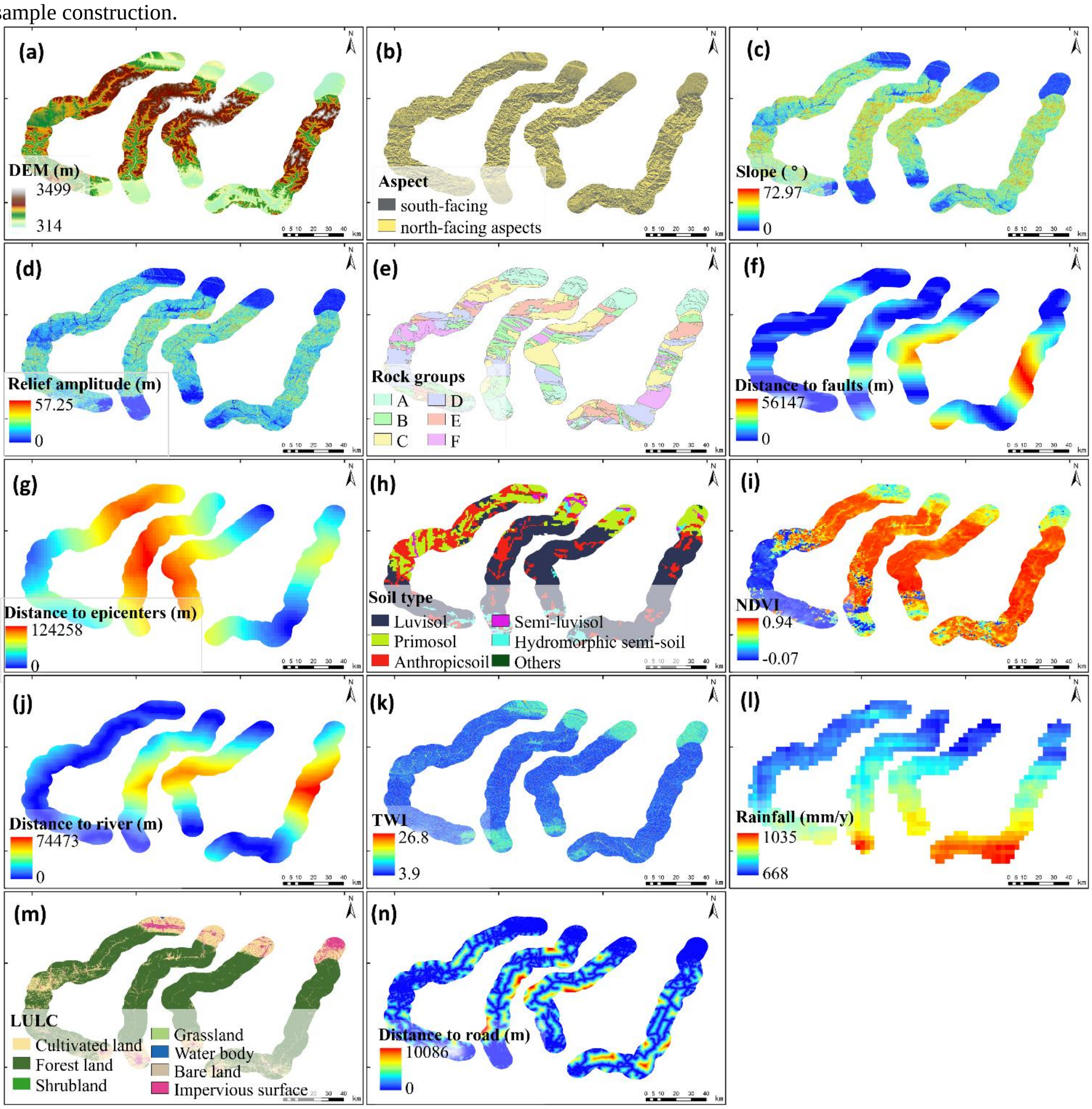

Fig. 3. Landslide conditioning factors. (a) DEM; (b) Aspect; (c) Slope; (d) Relief amplitude; (e) Rock groups; (f) Distance to faults; (g) Distance to earthquake epicenters; (h) Soil type; (i) NDVI; (j) Distance to rivers; (k) TWI; (l) Rainfall; (m) LULC; (n) Distance to roads.

Samples were constructed from the five-class CF susceptibility map after SHALSTAB constraint (Fig. 4). To ensure that all susceptibility classes were represented during model training, three landslides in the very low susceptibility zone were excluded. The grid cells containing the remaining 1,404 landslides were used as very high susceptibility samples. Additional samples were drawn from the very low, low, moderate, and high classes at rates of 17.15%, 23.88%, 26.54%, and 20.28%, respectively. The resulting five-class dataset contained 11,555 samples for subsequent model training, validation, and comparison. Random sampling was used to divide the dataset into training and test sets, with 70% of the samples used for training and 30% reserved for independent testing. The five susceptibility classes—very low, low, moderate, high, and very high—were assigned labels 1-5, respectively. Predictor variables were extracted from the conditioning-factor rasters, so each sample comprised a set of factor values and a susceptibility-class label.

https://doi.org/10.5194/egusphere-2026-4550
Preprint. Discussion started: 6 August 2026

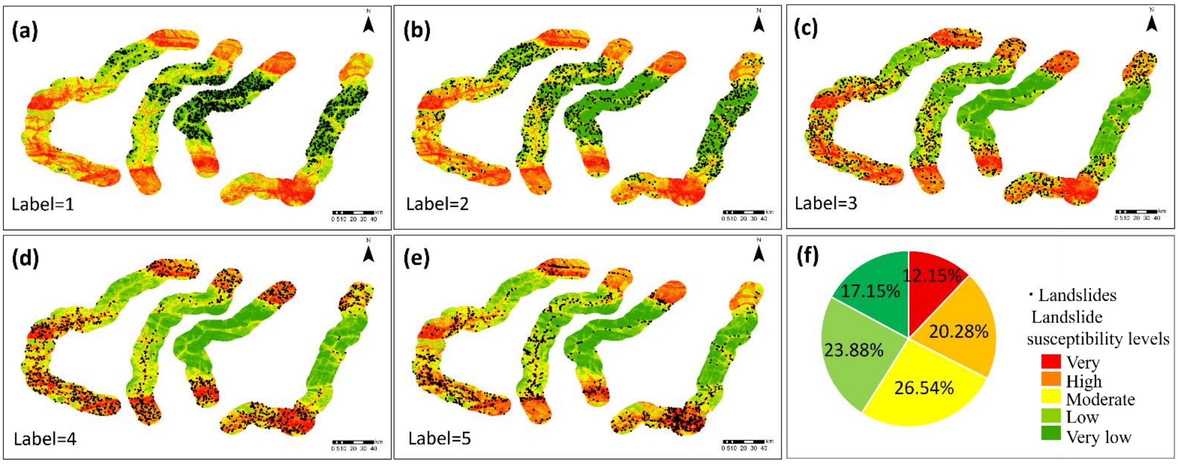

Fig. 4. Construction of the multi-class landslide susceptibility samples. (a) Very low susceptibility samples; (b) Low susceptibility samples; (c) Moderate susceptibility samples; (d) High susceptibility samples; (e) Very high susceptibility samples; (f) Sample composition.

3.4. Landslide susceptibility models

Six representative classifiers were selected for landslide susceptibility modeling: LR, SVM, RF, XGBoost, FT-Transformer, and TabPFN. Together, these methods cover linear statistical, kernel-based, ensemble-learning, and deep tabular-learning models and can characterize relationships between susceptibility and conditioning factors from complementary perspectives. LR served as the linear statistical baseline. An SVM with a radial basis function (RBF) kernel represented nonlinear decision boundaries (Cortes and Vapnik, 1995). RF reduced model variance through bootstrap sampling and random subspaces (Breiman, 2001). XGBoost used regularized gradient boosting to fit complex nonlinear relationships (Chen and Guestrin, 2016). FT-Transformer learned interactions among conditioning factors through feature attention (Gorishniy et al., 2021), whereas TabPFN used pretrained priors for rapid inference on small and medium-sized tabular datasets (Hollmann et al., 2025). Each model was trained for five-class classification and produced the probability that each sample belonged to each susceptibility class; the class with the highest predicted probability was assigned as the final result. To reduce the influence of random data partitioning, repeated random holdout validation was performed. In each repetition, the data were randomly divided into training and test sets at a 7:3 ratio. Experiments were repeated ten times using identical parameter settings, and the mean and variance of each metric were reported. The principal model parameters are listed in Table 3.

Table 3
Models and parameter settings for landslide susceptibility modeling.

Model	Type	Key parameters
LR	Linear classification model	max_iter = 3000; class_weight = balanced
SVM	Kernel-based classification model	kernel = rbf; C = 2.0; gamma = scale; probability = True; class_weight = balanced
RF	Ensemble learning model	n_estimators = 500; class_weight = balanced; n_jobs = -1
XGBoost	Gradient boosting decision tree model	n_estimators = 500; max_depth = 3; learning_rate = 0.03; subsample = 0.85; colsample_bytree = 0.85; objective = multi:softprob; num_class = 5; eval_metric = mlogloss; tree_method = hist;
FT-Transformer	Deep tabular Transformer model	epochs = 100; patience = 15; batch_size = 256; learning_rate = 8×10^{-4}; weight_decay = 2×10^{-4}; d_token = 48; nhead = 4; num_layers = 3; dropout = 0.12; optimizer: AdamW
TabPFN	Pretrained tabular foundation model	n_estimators = 4; maximum training samples per class = 512

3.5. Evaluation metrics

Model performance was evaluated using class-specific recall together with F1, accuracy, and AUC for the multi-class task, thereby preventing overall accuracy from obscuring poor recognition of an individual class (Hand and Till, 2001; Sokolova and Lapalme, 2009). The metrics were calculated as follows:

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \tag{6}$$

https://doi.org/10.5194/egusphere-2026-4550
Preprint. Discussion started: 6 August 2026

$$Precision_i = \frac{TP_i}{TP_i + FP_i}$$
$$F1_i = \frac{2 \times Precision_i \times Recall_i}{Precision_i + Recall_i} \quad (7)$$
$$F1 = \frac{1}{K}\sum_{i=1}^{K} F1_i$$

$$Accuracy = \frac{N_{correct}}{N_{total}} \quad (8)$$

$$AUC = \frac{1}{K}\sum_{i=1}^{K} AUC_i \quad (9)$$

264 where TP_i, TP_i, and FN_i denote the true positives, false positives, and false negatives for class i, respectively; K is the number of classes (K = 5); $N_{correct}$ is the number of correctly classified samples; N_{total} is the total number of samples; and AUC_i is the one-versus-rest AUC for class i.

267 Spatial mapping results were further evaluated using HOI, which measures the ability of a relatively small very high susceptibility zone to contain historical landslides (Zhang et al., 2026). HOI jointly considers the areal proportion of the very high susceptibility zone and the hit rate of historical landslides within that zone and is calculated as follows:

$$HOI = 1 - a_5 + h_5 \quad (10)$$

270 where a_5 is the areal proportion of the very high susceptibility zone and h_5 is the proportion of historical landslides falling within that zone. The metric therefore evaluates whether a model captures more historical landslides within a smaller very high susceptibility area. The original HOI was designed for five-class susceptibility maps and evaluates only the very high susceptibility zone. For a binary model, however, the positive-class area represents a landslide-prone zone and is not equivalent to the very high susceptibility class in a five-class map. In particular, when a binary probability map is further divided into five classes using natural breaks, the landslide-prone area originally defined by probabilities greater than 0.5 is subdivided into high and very high classes. Calculating HOI only for the resulting very high class would omit a substantial portion of the area correctly identified by the binary model as landslide-prone and would consequently underestimate its spatial recognition ability. To provide a consistent comparison between multi-class and binary maps, the evaluation region was therefore defined differently for the two settings. For multi-class results, HOI continued to use the very high susceptibility zone. For binary results, in addition to the very high class obtained by natural breaks, the landslide-prone zone defined by the binary threshold (p > 0.5) was evaluated using its areal proportion and historical-landslide hit rate. The modified HOI for a binary result can thus be expressed as:

$$HOI = 1 - a_{pos} + h_{pos} \quad (11)$$

282 where a_{pos} is the areal proportion classified by the binary model as landslide-prone and h_{pos} is the proportion of historical landslides falling within that zone. This modified form more accurately reflects a binary model's ability to identify landslide-prone areas and avoids underestimation caused by evaluating only the very high class produced by natural breaks.

4. Results and analysis

4.1. Model comparison

287 To systematically evaluate the suitability of different models for multi-class landslide susceptibility mapping, LR, FT-Transformer, XGBoost, SVM, TabPFN, and RF were compared. Experiments were conducted using both binary landslide and multi-class susceptibility datasets. The binary task evaluated discrimination between landslide and non-landslide samples, whereas the multi-class task focused on distinctions among susceptibility classes. Performance results are presented in Tables 4 and 5. Model performance was strongly affected by task complexity. F1 and accuracy were generally higher for binary than for multi-class classification. RF performed best in the binary task, with F1, accuracy, and AUC values of 83.15%, 82.49%, and 91.08%, respectively. RF also achieved the best overall multi-class performance, with corresponding values of 68.65%, 68.60%, and 92.53%. Although multi-class classification reduced F1 and accuracy relative to binary classification, AUC remained high, indicating that the models largely preserved the risk ordering among susceptibility classes. Overall, RF outperformed the deep-learning models FT-Transformer and TabPFN in both settings and provided the strongest comprehensive performance.

Table 4

Performance of landslide susceptibility models on the binary classification dataset (best values in bold).

Model	Type	Recall		Overall			
		Non-landslide	Landslide	F1	Accuracy	AUC	AUC variance

https://doi.org/10.5194/egusphere-2026-4550
Preprint. Discussion started: 6 August 2026

							(×10⁻⁵)
LR	Linear statistical model	72.01	74.84	73.79	73.43	82.74	22.56
FT-Transformer	Tabular deep learning model	75.00	82.45	79.48	78.72	87.47	29.62
XGBoost	Gradient-boosted decision-tree ensemble	77.98	85.22	82.22	81.60	89.95	18.52
SVM	Margin discriminant model	77.84	82.37	80.54	80.11	88.46	13.56
TabPFN	Transformer-based Bayesian deep-learning model	79.12	85.15	82.65	82.13	90.83	**9.43**
RF	**Bagging ensemble tree model**	78.55	**86.42**	**83.15**	**82.49**	**91.08**	19.08

Table 5
Performance of landslide susceptibility models on the multi-class dataset (best values in bold).

Model	Type	Recall					Overall			
		Very low	Low	Moderate	High	Very high	F1	Accuracy	AUC	AUC variance (×10⁻⁵)
LR	Linear statistical model	77.78	55.56	41.94	48.03	55.91	55.75	55.84	86.79	1.37
FT-Transformer	Tabular deep learning model	78.49	55.20	68.10	73.48	56.99	66.34	66.45	91.20	**0.67**
XGBoost	Gradient-boosted decision-tree ensemble	78.49	60.93	55.91	70.97	62.37	65.69	65.73	91.28	0.68
SVM	Margin discriminant model	76.34	59.86	63.44	69.89	59.14	65.76	65.73	91.28	1.19
TabPFN	Transformer-based Bayesian deep-learning model	**86.04**	**66.42**	62.07	69.10	60.93	68.25	68.51	92.30	1.18
RF	**Bagging ensemble tree model**	79.93	65.95	**63.08**	**71.33**	**62.72**	**68.65**	**68.60**	**92.53**	1.41

4.2. Evaluation of landslide susceptibility maps

To evaluate the historical-landslide coverage and spatial concentration of the susceptibility maps, the areal proportion of susceptibility class (a), the historical-landslide hit rate (h), and HOI were calculated. Binary results were summarized in two ways: Table 6 reports statistics after dividing binary probability maps into five classes by natural breaks, and Table 7 reports the landslide-prone zones defined by the binary threshold (p > 0.5). Multi-class results are given in Table 8. As shown in Tables 6 and 7, RF achieved the highest threshold-based binary HOI (1.68); its landslide-prone zone occupied 19.3% of the study area and contained 87.3% of the historical landslides. Its HOI based on the five-class natural-breaks result was 1.53, lower than the threshold-based HOI and confirming that evaluating a binary model solely by its very high susceptibility class underestimates model capability. Table 8 shows that RF also performed best in the multi-class setting. Its very high susceptibility zone occupied only 13.5% of the study area but contained 91.8% of the historical landslides, producing an HOI of 1.78. It is noteworthy that TabPFN delineated a comparatively conservative very high susceptibility zone: it occupied only 12.1% of the study area but contained 54.8% of the historical landslides. The binary and multi-class HOI results are compared in Fig. 5 and demonstrate that multi-class classification produces a more spatially concentrated high-susceptibility zone.

Table 6
Statistical results of landslide susceptibility maps based on binary classification using natural breaks (best values in bold).

Model	Very low		Low		Moderate		High		Very high		HOI
	a_1	h_1	a_2	h_2	a_3	h_3	a_4	h_4	a_5	h_5	
LR	43.6	12.9	26.2	34.4	15	26.7	9.7	17.8	5.6	8.2	1.03
FT-Transformer	52.7	9.8	18.6	18.6	12.1	16.9	9.8	25.5	6.8	29.2	1.22
XGBoost	43	0.9	21.1	7.4	16.1	20.9	12.1	33.1	7.8	37.6	1.30
SVM	36.9	17.6	30.8	23	18.2	27.7	9.4	16.7	4.8	14.9	1.10
TabPFN	45.07	4.8	26.46	21.1	16.14	30.2	8.74	24.8	3.61	19.2	1.16

https://doi.org/10.5194/egusphere-2026-4550
Preprint. Discussion started: 6 August 2026

RF	29.2	**0.1**	21.9	**0.4**	20.4	3.8	19.7	**34**	8.7	**61.7**	1.53

Table 7
Statistical results of landslide susceptibility maps based on binary classification using threshold (best values in bold).

Model	Landslides		Non-landslides		HOI
	a_{pos}	h_{pos}	a_{neg}	h_{neg}	
LR	11.6	20.1	88.4	79.9	1.09
FT-Transformer	15.6	51.7	84.4	48.3	1.36
XGBoost	15.4	61.3	84.6	38.7	1.46
SVM	8.6	22.1	91.4	77.9	1.14
TabPFN	6.6	28	93.4	72	1.21
RF	19.3	**87.3**	80.7	12.7	**1.68**

Table 8
Statistical results of landslide susceptibility maps based on multi-class classification (best values in bold).

Model	Very low		Low		Moderate		High		Very high		HOI
	a_1	h_1	a_2	h_2	a_3	h_3	a_4	h_4	a_5	h_5	
LR	20.1	0.9	22.2	7.5	19.5	16.2	21.7	21.5	16.4	53.7	1.37
FT-Transformer	18.7	0.5	19.6	4.3	24.4	14.7	25.5	21.5	11.8	58.9	1.47
XGBoost	19.2	0.6	22.0	5.7	20.0	10.7	24.5	18.3	14.4	64.7	1.50
SVM	18.1	0.7	22.2	4.7	22.1	10.0	24.0	17.9	13.6	66.7	1.53
TabPFN	18.4	0.8	23.9	6.0	22.1	15.6	23.5	**22.7**	12.1	54.8	1.43
RF	18.1	**0.3**	22.8	**1.3**	21.7	**1.9**	23.9	4.6	13.5	**91.8**	**1.78**

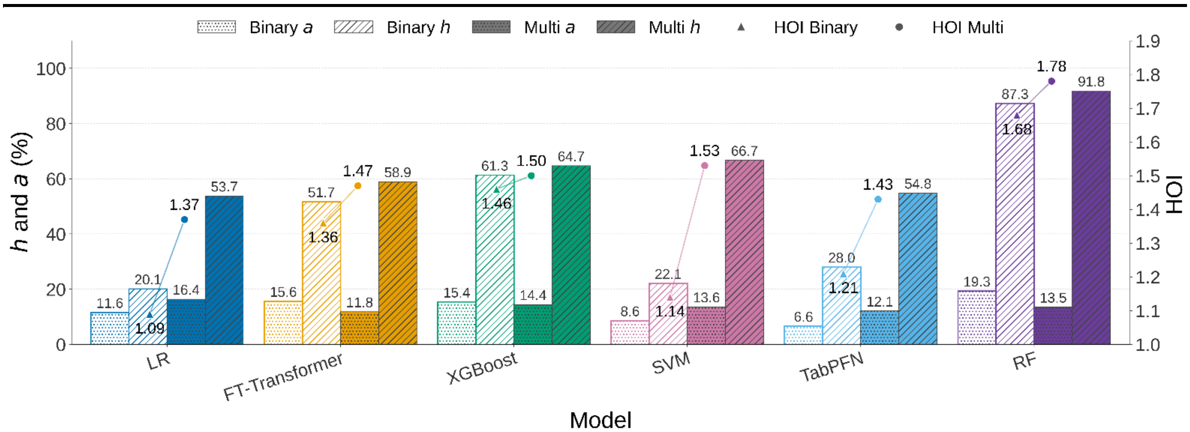

Fig. 5. Comparison of area ratio (a), hit rate (h), and HOI for binary and multi-class susceptibility maps.

The spatial susceptibility patterns produced by the six models under the binary and multi-class settings are shown in Fig. 6. Overall, very high susceptibility zones formed belts or patches concentrated mainly along valleys. In the binary results, most historical landslides occurred within the moderate, high, and very high susceptibility classes. Thus, binary classification identified broadly landslide-prone areas, but the resulting high-risk regions were relatively extensive and increased the area requiring attention. In the multi-class results, historical landslides were more concentrated within the very high susceptibility class, producing a more focused high-risk area. Enlarged views of the RF results show that binary classification generated a broad transition zone of moderate to high susceptibility around landslide locations, whereas multi-class classification assigned the landslide locations more directly to the very high susceptibility class and reduced overestimation in surrounding areas. This further demonstrates that multi-class classification improves the spatial concentration of the very high susceptibility zone.

https://doi.org/10.5194/egusphere-2026-4550
Preprint. Discussion started: 6 August 2026

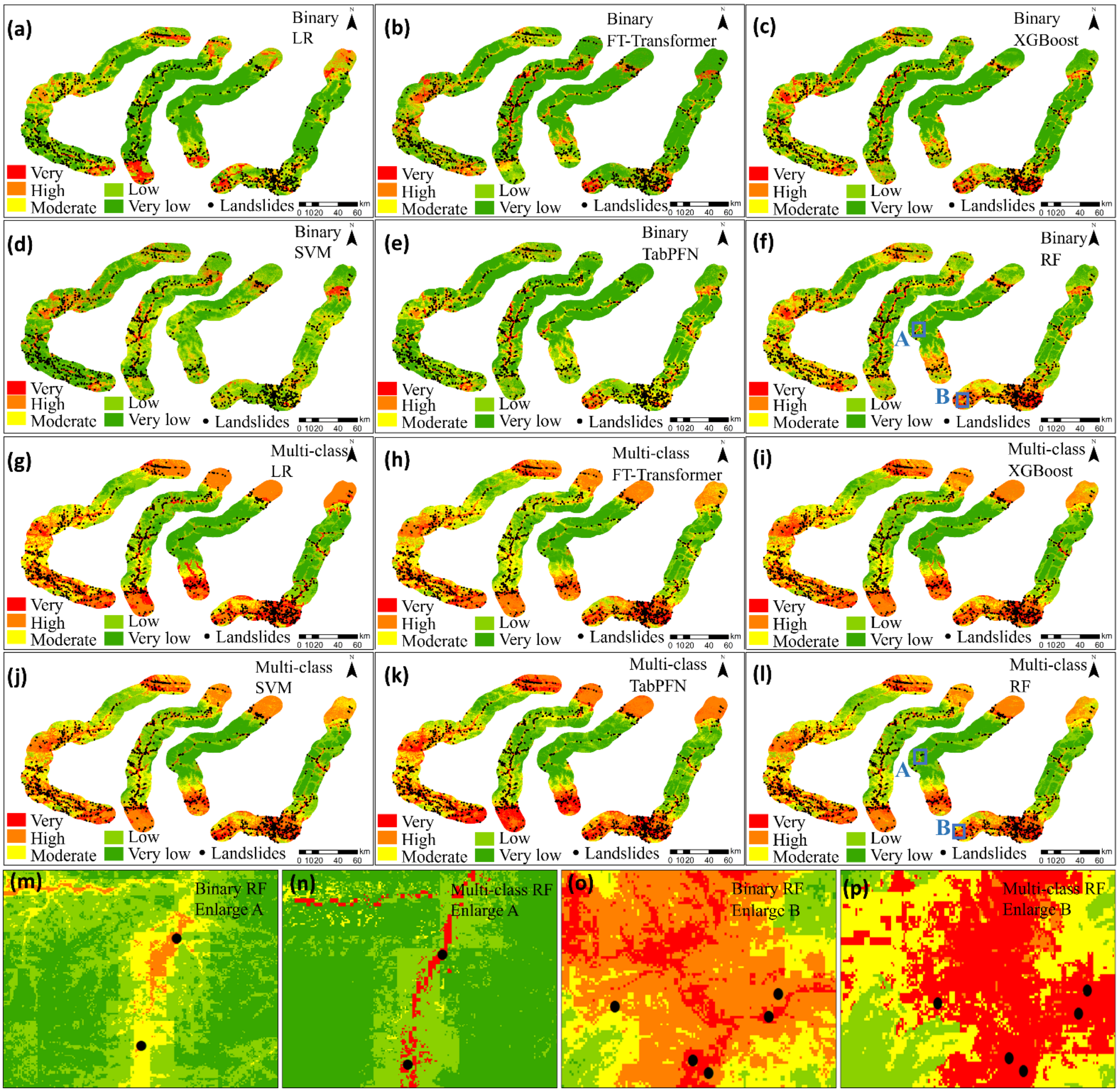

Fig. 6. Landslide susceptibility maps produced by six models under binary and multi-class classification settings. (a–f) Binary classification results of LR, FT-Transformer, XGBoost, SVM, TabPFN and RF; (g–l) multi-class classification results of LR, FT-Transformer, XGBoost, SVM, TabPFN and RF; (m–p) enlarged views of selected RF results.

5. Discussion

5.1. Effect of physics-constrained optimization

Initial susceptibility zonation before and after the SHALSTAB constraint is shown in Fig. 7. The initial CF map reflects the statistical relationships between historical landslides and conditioning factors. To capture as many landslides as possible, SHALSTAB classified most of the study area as unstable. After the SHALSTAB stability constraint was introduced, susceptibility classes within stable areas were downgraded. The proportion of the very high susceptibility class decreased from 16.25% to 12.15%, producing a more balanced spatial distribution in the initial susceptibility map.

https://doi.org/10.5194/egusphere-2026-4550
Preprint. Discussion started: 6 August 2026

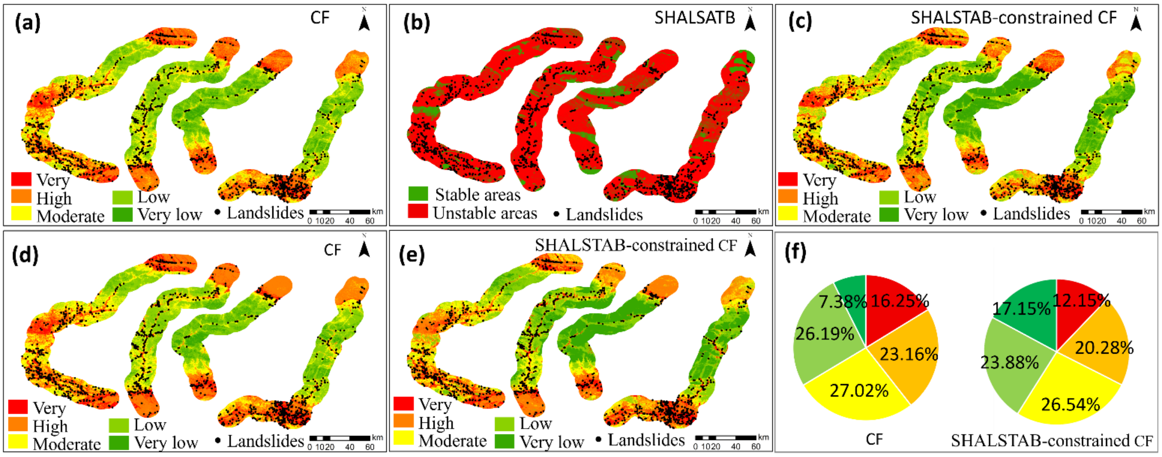

Fig. 7. Comparison of CF, SHALSTAB and SHALSTAB-constrained CF susceptibility maps. (a) CF-based initial susceptibility map; (b) SHALSTAB stability map; (c) SHALSTAB-constrained CF susceptibility map; (d) landslide distribution on the CF map; (e) landslide distribution on the SHALSTAB-constrained CF map; (f) area proportions of susceptibility classes before and after SHALSTAB constraint.

The effect of the SHALSTAB physical constraint on the initial CF zonation was tested using RF (Table 9). Compared with the result based on CF alone, the SHALSTAB-constrained CF map improved accuracy, AUC, and HOI. Accuracy increased from 65.05% to 68.60%, AUC from 91.30% to 92.53%, and HOI from 1.72 to 1.78. These results indicate that introducing the physical constraint improved the accuracy of the initial susceptibility zonation.

Table 9

Comparison of classification and spatial performance between CF and SHALSTAB-constrained CF susceptibility maps.

Model	Recall					Overall			
	Very low	Low	Moderate	High	Very high	F1	Accuracy	AUC	HOI
CF	68.28	**69.93**	**72.7**	**74.45**	58.79	68.44	65.05	91.30	1.72
SHALSTAB-constrained CF	**79.93**	65.95	63.08	71.33	**62.72**	**68.65**	**68.60**	**92.53**	1.78

5.2. Interpretation of conditioning factors

The ranking of mean absolute SHAP values for the very high susceptibility class is shown in Fig. 8. DEM was the most important factor, followed by LULC and distance to roads; all three were substantially more important than the remaining factors. This result indicates that geomorphic position, land-surface cover, and engineering disturbance jointly control the identification of very high susceptibility zones. The beeswarm plot shows the directions of effect for the main continuous factors. High DEM values were concentrated on the negative SHAP side, whereas low values extended toward positive contributions, indicating that low-elevation valleys and areas of concentrated human activity were more likely to be classified as very high susceptibility. Surface-cover types with relatively sparse vegetation or strong human disturbance, including cultivated land and grassland were more likely to form high-susceptibility zones, whereas forest had a protective effect on slope stability. Distance to roads exhibited a clear negative gradient: low values close to roads generally made positive contributions, whereas high values far from roads generally made negative contributions.

https://doi.org/10.5194/egusphere-2026-4550
Preprint. Discussion started: 6 August 2026

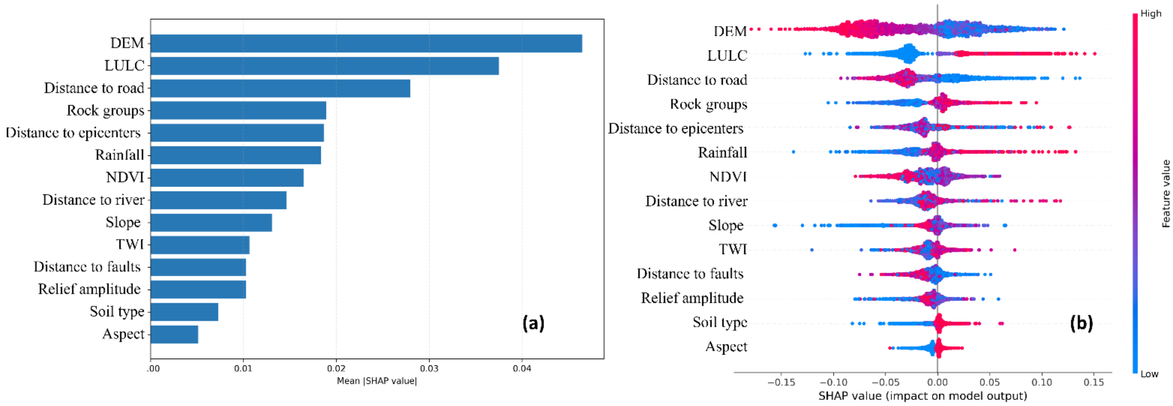

Fig. 8. SHAP-based interpretation of the multi-class for the very-high-susceptibility class. (a) Mean absolute SHAP values of conditioning factors; (b) SHAP beeswarm plot showing the contribution direction of each factor.

5. Conclusions

This study proposed a statistical-physical prior-guided framework for multi-class landslide susceptibility mapping and applied it to direct five-class mapping along the Ancient Qin-Shu Roads in the Qinling-Daba Mountains. The principal conclusions are as follows. (1) A five-class dataset representing very low, low, moderate, high, and very high landslide susceptibility was constructed from the SHALSTAB-constrained initial CF map. The dataset comprised 11,555 sample points, including 1,404 historical landslide records, and 14 conditioning-factor attributes were extracted for each sample for subsequent multi-class modeling. (2) Using identical samples, RF achieved the best overall performance in both binary and multi-class tasks and outperformed the deep-learning models FT-Transformer and TabPFN. In the binary task, RF attained F1, accuracy, and AUC values of 83.15%, 82.49%, and 91.08%, respectively; in the multi-class task, the corresponding values were 68.65%, 68.60%, and 92.53%. (3) Evaluation of the spatial susceptibility patterns showed that multi-class classification identified high-risk areas more compactly. The very high susceptibility zone in the multi-class RF result occupied only 13.5% of the study area but contained 91.8% of the historical landslides, yielding an HOI of 1.78. By contrast, the threshold-defined landslide-prone zone in the binary RF result occupied 19.3% of the area, contained 87.3% of the historical landslides, and had an HOI of 1.68. Multi-class classification therefore captured more historical landslides within a smaller very-high-susceptibility area, improving spatial concentration and facilitating the delineation of priority mitigation zones. (4) The SHALSTAB stability result produced a physically constrained initial CF susceptibility map. Relative to CF alone, the constraint increased accuracy from 65.05% to 68.60%, AUC from 91.30% to 92.53%, and HOI from 1.72 to 1.78, thereby improving the spatial concentration of landslides. (5) SHAP interpretation identified DEM, LULC, and distance to roads as the principal factors controlling the very high susceptibility class. Low-elevation valleys, cultivated land, and areas close to roads were more likely to be classified as very high susceptibility.

Several limitations should be acknowledged. First, SHALSTAB relies on regionally parameterized geotechnical parameters and provides relative stability estimates and therefore has limited ability to represent deep-seated landslides, rockslides, and preferential flow through fractures. It cannot fully capture the spatial heterogeneity of rock and soil parameters at the regional scale. Second, the initial CF map was constructed primarily from statistical relationships between historical landslides and conditioning-factor classes. It is consequently sensitive to inventory completeness, factor-classification schemes, and uneven sample distributions and may locally overestimate or underestimate susceptibility. Third, the landslide inventory is affected by survey accessibility and the completeness of historical records. Under-recording in high-elevation or inaccessible areas may influence both the statistical prior and SHAP interpretation. Fourth, this study focused on static susceptibility and did not fully consider temporal variations in rainfall, earthquake disturbance, and engineering activity. Future work can improve the framework in three directions: (1) incorporate measured geotechnical parameters, geological survey data, and uncertainty analysis to increase the reliability of the physical constraint; (2) use spatial cross-validation, cross-regional validation, or independent landslide-event validation to further test generalization across geomorphic regions and triggering conditions; and (3) introduce dynamic factors such as multi-temporal rainfall, LULC change, and road-construction evolution to support event-driven or spatiotemporally dynamic susceptibility assessment. Overall, the proposed statistical-physical prior-guided multi-class framework improves the representation and geomorphic plausibility of susceptibility gradients and increases the spatial concentration of very high susceptibility zones. It can support landslide-risk identification, priority mitigation zoning, and the protection of transportation heritage corridors in complex mountainous regions.

CRediT authorship contribution statement

Haixia Feng: Writing – original draft, Visualization, Methodology. **Qingwu Hu:** Project administration, Funding acquisition.

https://doi.org/10.5194/egusphere-2026-4550
Preprint. Discussion started: 6 August 2026

Yizhou Lan: Visualization, Methodology. **Daoyuan Zheng:** Visualization, Methodology. **Yuxiang Xu:** Visualization, Methodology. **Jiayuan Li**: Writing – review & editing. **Shunli Wang**: Investigation, Data curation. **Pengcheng Zhao**: Investigation, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Key R&D Program of China (2024YFB3908900).

Data availability

The historical landslide inventory used in this study was compiled from expert-verified records contained in national geological-hazard investigations and censuses, disaster reports provided by the Shaanxi Environmental Monitoring Center, and records supplied by relevant county and district authorities. These data are maintained by the contributing agencies and are subject to institutional regulations and access restrictions. The environmental data used in this study were obtained from the National 30 m Digital Elevation Model Dataset (https://www.resdc.cn), the Resource and Environment Science and Data Center (https://www.resdc.cn), the China Geological Cloud Platform (https://www.geocloud.cn), the National Earthquake Science Data Center (https://data.earthquake.cn), the National Aeronautics and Space Administration Earthdata (https://earthdata.nasa.gov), the Global Geographic Data Download Platform (https://www.webmap.cn), the Climate Hazards Group InfraRed Precipitation with Station data repository (https://www.chc.ucsb.edu/data/chirps), and the Global Land Cover Dataset (https://www.globallandcover.com), in accordance with the access and licensing conditions of the respective providers. The processed spatial layers, derived samples, and other data supporting the findings of this study are available from the corresponding author upon reasonable request. Access to the landslide inventory requires approval from the original data providers and must comply with their data-sharing policies.

References

Breiman, L., 2001. Random Forests. Machine Learning 45(1), 5-32. https://doi.org/10.1023/A:1010933404324.

Chang, Z., Catani, F., Huang, F., Liu, G., Meena, S.R., Huang, J., Zhou, C., 2023. Landslide susceptibility prediction using slope unit-based machine learning models considering the heterogeneity of conditioning factors. Journal of Rock Mechanics and Geotechnical Engineering 15(5), 1127-1143. https://doi.org/10.1016/j.jrmge.2022.07.009.

Chen, J., Li, Q., Huang, C., Hu, X., Shen, Z., 2026. Landslide susceptibility assessment in the Central Yunnan Plateau by assembling optimized statistical and machine learning models. Geoenvironmental Disasters 13(1), 28. https://doi.org/10.1186/s40677-026-00380-w.

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, New York, NY, USA, pp. 785-794. https://doi.org/10.1145/2939672.2939785.

Cifci, M.A., Hu, X., Öney, B., Misak, S., Moayedi, H., Dehrashid, H.A., 2026. Prediction of landslide susceptibility through ANN models optimized by evolutionary algorithms. Scientific Reports 16(1), 9471. https://doi.org/10.1038/s41598-026-39458-8.

Cortes, C., Vapnik, V., 1995. Support-Vector Networks. Machine Learning 20(3), 273-297. https://doi.org/10.1007/BF00994018.

Cui, H.-Z., Tong, B., Wang, T., Dou, J., Ji, J., 2025. A hybrid data-driven approach for rainfall-induced landslide susceptibility mapping: Physically-based probabilistic model with convolutional neural network. Journal of Rock Mechanics and Geotechnical Engineering 17(8), 4933-4951. https://doi.org/10.1016/j.jrmge.2024.08.005.

Edrich, A.-K., Yildiz, A., Roscher, R., Bast, A., Graf, F., Kowalski, J., 2024. A modular framework for FAIR shallow landslide susceptibility mapping based on machine learning. Natural Hazards 120(9), 8953-8982. https://doi.org/10.1007/s11069-024-06563-8.

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., Savage, W.Z., 2008. Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. Engineering Geology 102(3), 85-98. https://doi.org/10.1016/j.enggeo.2008.03.022.

Feng, H., Miao, Z., Hu, Q., 2022. Study on the Uncertainty of Machine Learning Model for Earthquake-Induced Landslide Susceptibility Assessment. Remote Sensing 14(13), 2968. https://doi.org/10.3390/rs14132968.

Feng, L., Xu, C., Tian, Y., Li, L., Sun, J., Huang, Y., Wang, P., Zhang, X., Li, T., Yang, W., Ma, S., Shao, X., Xu, J., Chen, J., 2024. Landslides of China's Qinling. Geoscience Data Journal 11(4), 725-741. https://doi.org/10.1002/gdj3.246.

Goetz, J.N., Brenning, A., Petschko, H., Leopold, P., 2015. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Computers & Geosciences 81(3), 1-11. https://doi.org/10.1016/j.cageo.2015.04.007.

Goetz, J.N., Guthrie, R.H., Brenning, A., 2011. Integrating physical and empirical landslide susceptibility models using generalized additive models. Geomorphology 129(3), 376-386. https://doi.org/10.1016/j.geomorph.2011.03.001.

Gorishniy, Y., Rubachev, I., Khrulkov, V., Babenko, A., 2021. Revisiting Deep Learning Models for Tabular Data, in: Advances in Neural Information Processing Systems 34. Curran Associates, Inc., Red Hook, NY, USA, pp. 18932-18943. https://doi.org/10.5555/3540261.3541708.
Han, Y., Semnani, S.J., 2025. Important considerations in machine learning-based landslide susceptibility assessment under future climate conditions. Acta Geotechnica 20(1), 475-500. https://doi.org/10.1007/s11440-024-02363-3.
Hand, D.J., Till, R.J., 2001. A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. Machine Learning 45(2), 171-186. https://doi.org/10.1023/A:1010920819831.
Hollmann, N., Müller, S., Purucker, L., Krishnakumar, A., Körfer, M., Hoo, S.B., Schirrmeister, R.T., Hutter, F., 2025. Accurate predictions on small data with a tabular foundation model. Nature 637(8045), 319-326. https://doi.org/10.1038/s41586-024-08328-6.
Huang, J., Wen, H., Zhou, X., Xiao, J., 2026. Is there difference in landslide susceptibility model based on explainable artificial intelligence from the perspective of slope units with different scales? Reliability Engineering & System Safety 266, 111701. https://doi.org/10.1016/j.ress.2025.111701.
Jiang, H., Zhou, H., Wu, J., Liu, M., Wu, Y., Guo, Y., 2026. Ensemble learning for landslide susceptibility mapping: a review of machine learning and hybrid approaches. Carbonates and Evaporites 41(1), 16. https://doi.org/10.1007/s13146-025-01221-x.
Karpatne, A., Atluri, G., Faghmous, J.H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., Kumar, V., 2017. Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data. IEEE Transactions on Knowledge and Data Engineering 29(10), 2318-2331. https://doi.org/10.1109/TKDE.2017.2720168.
Le, X.-H., Choi, C., Eu, S., Yeon, M., Lee, G., 2024. Quantitative evaluation of uncertainty and interpretability in machine learning-based landslide susceptibility mapping through feature selection and explainable AI. Frontiers in Environmental Science 12, 1424988. https://doi.org/10.3389/fenvs.2024.1424988.
Li, J., Wang, R., Shi, W., Yang, L., Wei, J., Liu, F., Xiong, K., 2025. Landslide Susceptibility Assessment in Ya'an Based on Coupling of GWR and TabNet. Remote Sensing 17(15), 2678. https://doi.org/10.3390/rs17152678.
Liu, S., Wang, L., Zhang, W., Sun, W., Fu, J., Xiao, T., Dai, Z., 2023. A physics-informed data-driven model for landslide susceptibility assessment in the Three Gorges Reservoir area. Geoscience Frontiers 14(5), 101621. https://doi.org/10.1016/j.gsf.2023.101621.
Liu, S., Wang, L., Zhang, W., Sun, W., Wang, Y., Liu, J., 2024. Physics-informed optimization for a data-driven approach in landslide susceptibility evaluation. Journal of Rock Mechanics and Geotechnical Engineering 16(8), 3192-3205. https://doi.org/10.1016/j.jrmge.2023.11.039.
Ma, T., Yi, X., Ci, H., Wang, R., Yang, H., Yan, Z., 2026. Landslide Susceptibility Evaluation Integrating Machine Learning and SBAS-InSAR-Derived Deformation Characteristics: A Case Study of Yining County, Xinjiang. Sensors 26(2), 707-734. https://doi.org/10.3390/s26020707.
Mejia-Manrique, S.A., Ramos-Scharrón, C.E., Hughes, K.S., Gonzalez-Cruz, J.E., Khanbilvardi, R.M., 2025. Dynamic landslide susceptibility for extreme rainfall events using an optimized convolutional neural network approach. Natural Hazards 121(13), 15383-15411. https://doi.org/10.1007/s11069-025-07396-9.
Merghadi, A., Yunus, A.P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D.T., Avtar, R., Abderrahmane, B., 2020. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. Earth-Science Reviews 207, 103225. https://doi.org/10.1016/j.earscirev.2020.103225.
Montgomery, D.R., Dietrich, W.E., 1994. A physically based model for the topographic control on shallow landsliding. Water Resources Research 30(4), 1153-1171. https://doi.org/10.1029/93WR02979.
Qiang, F., Zhao, F., Duan, Z., 2015. Development Characteristics and Spatial Distribution of Geological Hazards in the Qinba Mountains of Southern Shaanxi. Journal of Catastrophology 30(2), 193-198. https://doi.org/10.3969/j.issn.1000-811X.2015.02.037.
Qin, Y., Zhao, Z., Zhou, D., Chang, K., Mou, Q., Yang, Y., Hu, Y., 2024. Landslide Susceptibility Assessment in Yulong County Using Contribution Degree Clustering Method and Stacking Ensemble Coupled Model Based on Certainty Factor. Remote Sensing 16(19), 3582. https://doi.org/10.3390/rs16193582.
Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M., Guzzetti, F., 2018. A review of statistically-based landslide susceptibility models. Earth-Science Reviews 180, 60-91. https://doi.org/10.1016/j.earscirev.2018.03.001.
Sang, K., Sun, P., Li, R., Ke, C., Zhang, S., 2026. Landslide susceptibility mapping based on stratified non-landslides sampling from engineering geological petrofabric and optimized boosting models: a case of Maiji District, China. Geoenvironmental Disasters 13(1), 38. https://doi.org/10.1186/s40677-026-00389-1.
Sokolova, M., Lapalme, G., 2009. A Systematic Analysis of Performance Measures for Classification Tasks. Information Processing and Management 45(4), 427-437. https://doi.org/10.1016/j.ipm.2009.03.002.
Song, Y., Fan, W., Yu, N., Cao, Y., Jiang, C., Chai, X., Nan, Y., 2022. Rainfall-Induced Shallow Landslide Temporal Probability Modelling and Early Warning Research in Mountain Areas: A Case Study of the Qin-Ba Mountains, Western China. Remote Sensing 14(23), 5952. https://doi.org/10.3390/rs14235952.
Wei, Y., Qiu, H., Liu, Z., Huangfu, W., Zhu, Y., Liu, Y., Yang, D., Kamp, U., 2024. Refined and dynamic susceptibility assessment of landslides using InSAR and machine learning models. Geoscience Frontiers 15(6), 101890. https://doi.org/10.1016/j.gsf.2024.101890.

https://doi.org/10.5194/egusphere-2026-4550
Preprint. Discussion started: 6 August 2026

Yang, S., Tan, J., Luo, D., Wang, Y., Guo, X., Zhu, Q., Ma, C., Xiong, H., 2024. Sample size effects on landslide susceptibility models: A comparative study of heuristic, statistical, machine learning, deep learning and ensemble learning models with SHAP analysis. Computers & Geosciences 193, 105723. https://doi.org/10.1016/j.cageo.2024.105723.

Yang, W., Ge, Q., Tao, Z., Xu, D., Wang, Y., Hao, Z., 2026. Landslide susceptibility on the Qinghai-Tibet Plateau: Key driving factors identified through machine learning. Journal of Geographical Sciences 36(1), 199-218. https://doi.org/10.1007/s11442-026-2444-6.

Ye, C., Wu, H., Oguchi, T., Tang, Y., Pei, X., Wu, Y., 2025. Physically Based and Data-Driven Models for Landslide Susceptibility Assessment: Principles, Applications, and Challenges. Remote Sensing 17(13), 2280. https://doi.org/10.3390/rs17132280.

Zhang, F., Li, C., Wang, Y., Wang, Y., 2026. Statistical Knowledge-Guided Multi-Classification Machine Learning Modeling Scheme for Landslide Susceptibility Levels. Engineering Geology 360, 108458. https://doi.org/10.1016/j.enggeo.2025.108458.

Zhang, J., Ma, X., Zhang, J., Sun, D., Zhou, X., Mi, C., Wen, H., 2023. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. Journal of Environmental Management 332(15), 117357. https://doi.org/10.1016/j.jenvman.2023.117357.

Zheng, W., Fan, W., Cao, Y., Nan, Y., Jing, P., 2025. Landslide Hazard Assessment Under Record-Breaking Extreme Rainfall: Integration of SBAS-InSAR and Machine Learning Models. Remote Sensing 17(13), 2265. https://doi.org/10.3390/rs17132265.

Zhou, Y., Tang, Y., Feng, W., Feng, F., Liu, Z., Zhao, F., Wang, H., Hong, B., Wang, Z., 2026. Interpretable susceptibility mapping of loess landslides integrating soil parameters using machine learning and SHAP: a case study from the loess plateau in western Shanxi, China. Environmental Earth Sciences 85(7), 175. https://doi.org/10.1007/s12665-026-12840-5.