Global NEST Journal, **28** (8), 08709
Available online: 19/10/2026

Machine Learning-Based Environmental Risk Screening of Rainfall-Induced Landslides under Satellite Rainfall Uncertainty in Northern Vietnam

Doan Thi Noi[a,*], Nguyen Hoang Son[b], Dang Thu Thuy[a] and Nguyen Thanh Nga[a]

[a]Faculty of Civil and Environmental Engineering, University of Transport and Communications, 3 Cau Giay Street, Lang Ward, Hanoi 11528, Vietnam
[b]Faculty of Water Resources Engineering, Thuyloi University, 175 Tay Son, Kim Lien Ward, Hanoi 11520, Vietnam

*Correspondence e-mail: dtnoi@utc.edu.vn

Abstract

Rainfall-induced landslides threaten communities and transportation corridors in data-scarce mountainous regions, where the spatial representation of rainfall and the construction of non-landslide samples can strongly influence susceptibility estimates. This study develops a reproducible GIS-machine learning framework for Hoa Binh, northern Vietnam, using 182 catalogued landslide locations (from an original 184-record catalogue after excluding two boundary records lacking complete predictor coverage), six environmental variables, and CHIRPS-derived rainfall indicators. Five models-ANN, random forest, Extra Trees, LightGBM, and XGBoost-were evaluated using a stratified 70:30 split, five-fold cross-validation, and ten independent pseudo-absence realizations. LightGBM achieved the highest mean AUC (0.686 ± 0.050), closely followed by random forest (0.681 ± 0.039), Extra Trees (0.675 ± 0.037), XGBoost (0.674 ± 0.046), and ANN (0.656 ± 0.043). Model ranking varied among realizations, demonstrating non-negligible sampling uncertainty. Final susceptibility surfaces were therefore generated as pixel-wise mean probabilities across the ten realizations, accompanied by spatial uncertainty maps. The framework shows that robust regional screening depends not only on algorithm choice but also on transparent sampling, validation, and interpretation of satellite-rainfall limitations.

Keywords: susceptibility mapping; CHIRPS; pseudo-absence sampling; model robustness; LightGBM; Random Forest; GIS; transportation infrastructure.

OPEN ACCESS

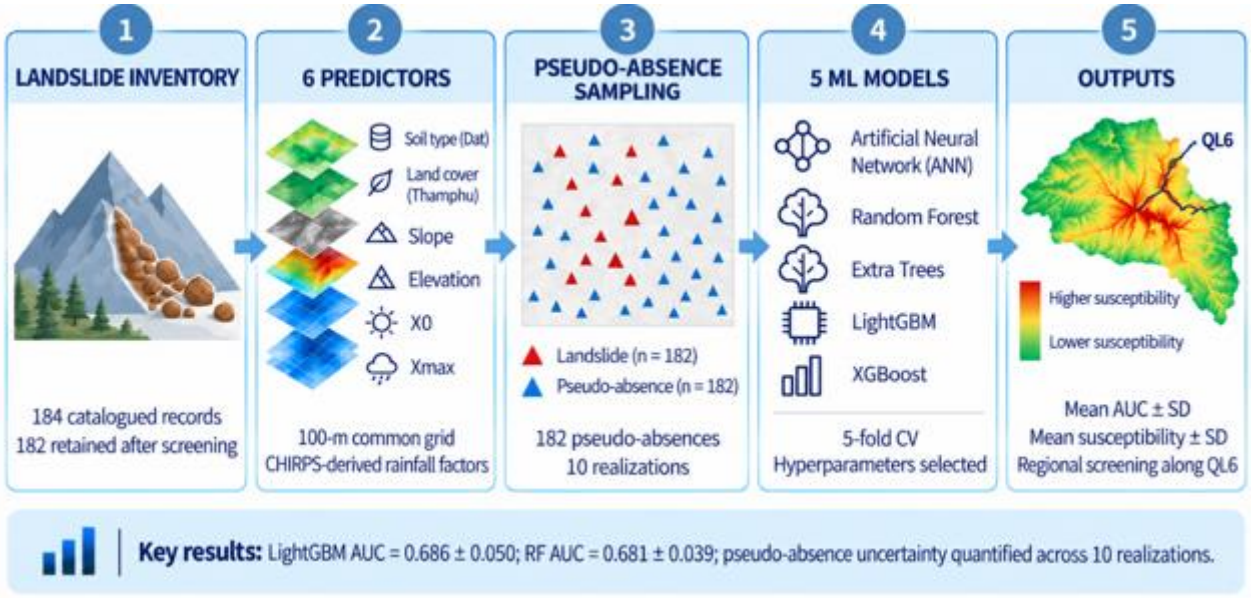
Received: 15/06/2026,

Accepted: 18/10/2026,

Available online: 19/10/2026

Copyright: © 2026 The Author(s). Published in Global NEST Journal. This article is licensed under the terms and conditions of the Creative Commons Attribution 4.0 International license (CC BY 4.0).

Graphical abstract



Graphical abstract. Landslide inventory, repeated pseudo-absence modelling, uncertainty-aware susceptibility mapping, and regional screening. AI-assisted graphical abstract prepared with ChatGPT (OpenAI); scientific content and final approval by the authors.

1. Introduction

Landslides are among the most damaging geohydrological hazards in mountainous regions, where steep terrain, dissected drainage networks, heterogeneous land cover, and intense monsoon rainfall interact to reduce slope stability. Their impacts extend beyond direct slope failure because road closures, damaged drainage systems, disrupted settlements, and loss of access can amplify social and economic consequences. These problems are particularly important in northern Vietnam, where major transport corridors traverse steep hillslopes and valley margins and where reliable ground observations remain spatially uneven. Regional-scale susceptibility mapping is therefore needed to identify areas where more detailed field investigation, monitoring, and engineering assessment should be prioritized.

Landslide susceptibility mapping integrates information on past landslide occurrence with terrain, hydrological, geological, land-surface, and climatic conditioning factors. Statistical, GIS-based, machine-learning, and deep-learning approaches have increasingly been used because they can represent nonlinear relationships among multiple predictors (Reichenbach *et al.*, 2018; Merghadi *et al.*, 2020; Ado *et al.*, 2022). Deep-learning methods have also been applied to susceptibility assessment, but their performance depends strongly on inventory size, input representation, and model configuration (Dao *et al.*, 2020). Recent work has further emphasized that high apparent predictive accuracy does not automatically imply robustness when sampling, preprocessing, and validation procedures differ among studies (Edrich *et al.*, 2024; Manopkawee *et al.*, 2026).

Rainfall is a major triggering-related factor for slope failure. Rainfall thresholds and event-based studies show that intensity, duration, antecedent wetness, and local hydrological response can be critical to landslide initiation (Guzzetti *et al.*, 2007). In data-scarce mountainous areas, satellite precipitation products provide spatially continuous information where gauge networks are insufficient. CHIRPS offers a long-term quasi-global rainfall record and has been widely used in hydrological and environmental applications (Funk *et al.*, 2015). Nevertheless, satellite rainfall estimates remain affected by retrieval uncertainty, gauge adjustment, spatial resolution, terrain effects, and the limited ability of temporally aggregated rainfall indicators to represent short-duration triggering events (Beck *et al.*, 2017; Dandridge *et al.*, 2019; Ji *et al.*, 2022; Trinh-Tuan *et al.*, 2019).

Recent Vietnamese studies demonstrate the value of machine learning for landslide susceptibility assessment but also illustrate how results depend on local data availability and modelling design. Long and De Smedt (2019) developed rainfall-induced susceptibility mapping in A Luoi under limited data conditions. Tran *et al.* (2024) applied RF, gradient boosting, and SVM to Van Yen District, Yen Bai Province, using 354 landslide locations and thirteen conditioning variables; RF achieved an AUC of 0.883. Das *et al.* (2024) also demonstrated the usefulness of machine learning for regional-scale susceptibility assessment in central Vietnam. These studies confirm the potential of data-driven mapping, while their relatively larger

inventories and richer predictor sets also indicate why model performance cannot be compared across regions on AUC alone.

A second methodological issue concerns the construction of non-landslide samples. Presence inventories are incomplete by nature, so locations treated as negative or pseudo-absence samples may include unrecorded or potentially unstable slopes. Gu *et al.* (2024) showed that different non-landslide sampling strategies can materially alter model performance and zoning outcomes, and explicitly evaluated buffer-controlled sampling among alternative strategies. Guo *et al.* (2024) likewise found that both positive- and negative-sampling design affected susceptibility accuracy. Liu *et al.* (2024) quantified substantial map uncertainty caused by sampling randomness and showed that buffer distance and repeated sampling influence susceptibility zoning. These findings support the use of repeated pseudo-absence realizations rather than relying on a single randomly generated negative sample set.

Reproducibility and data consistency have consequently become central concerns in recent susceptibility studies. Edrich *et al.* (2024) proposed a transparent, modular workflow intended to support FAIR and reproducible landslide mapping. More recently, Manopkawee *et al.* (2026) compared statistical, machine-learning, and deep-learning models under a unified data-consistent framework in northern Thailand and emphasized that robustness is influenced not only by algorithm selection but also by data structure, sampling, and parameter configuration. For rainfall-induced applications, Lee *et al.* (2025) integrated daily meteorological, topographic, and environmental information at 100-m resolution for short-term susceptibility and warning in South Korea, illustrating the additional information available when event-scale rainfall is explicitly represented.

Against this background, the present study addresses two linked uncertainties in a data-scarce mountainous setting: the limited physical representativeness of satellite-derived rainfall indicators and the sensitivity of machine-learning results to pseudo-absence selection. The novelty lies in three aspects. First, model performance is interpreted explicitly in the context of CHIRPS rainfall representativeness rather than treating rainfall as an error-free deterministic predictor. Second, pseudo-absence sampling uncertainty is quantified through ten independent realizations, and final susceptibility surfaces are produced as pixel-wise mean probabilities accompanied by spatial standard-deviation maps. Third, the resulting maps are interpreted as regional environmental-risk screening products for mountainous transport infrastructure, with particular relevance to the QL6 corridor. The objectives are to (i) construct a reproducible GIS-based susceptibility dataset from an audited landslide inventory and six traceable predictors; (ii) compare ANN, RF, Extra Trees, LightGBM, and XGBoost under a common tuning and validation framework; (iii) quantify sensitivity to pseudo-absence selection and interpret rainfall-input limitations; and (iv) provide spatial information for regional screening, land-use planning, and transportation-infrastructure protection.

2. Study Area

Hoa Binh Province is located in the mountainous region of northern Vietnam, approximately between latitudes 20°10′–21°20′ N and longitudes 104°20′–105°50′ E. The province covers an area of about 4,600 km² and is characterized by complex mountainous terrain, deeply dissected valleys, steep hillslopes, and dense drainage networks. Elevation ranges from low-lying river valleys to mountainous zones exceeding 1,200 m, creating strong spatial variability in terrain conditions and local hydrological responses. The location, topographic setting, transportation corridor, meteorological stations, and spatial distribution of the landslide inventory are shown in **Figure 1**.

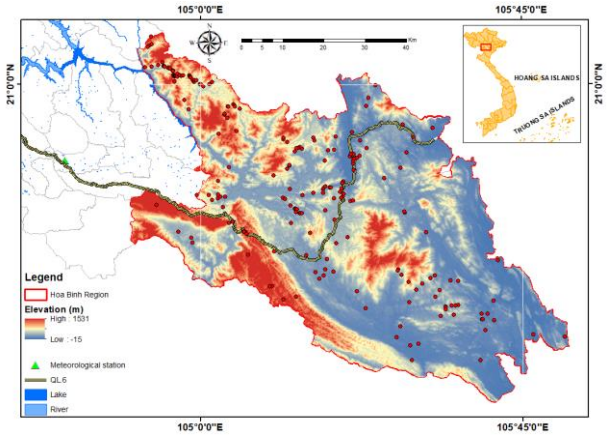


Figure 1. Hoa Binh study area and spatial distribution of the catalogue of 184 landslide locations. Two boundary records lacking complete predictor coverage were excluded from model development, leaving 182 locations for machine-learning analysis.

The climate of Hoa Binh Province is dominated by the tropical monsoon regime, with a rainy season from May to October and a dry season from November to April. Annual rainfall is generally high, and a large proportion of rainfall occurs during short-duration and high-intensity events in the rainy season. These rainfall conditions can increase soil moisture, reduce slope stability, and trigger landslides, particularly in steep terrain, road-cut slopes, and areas with disturbed land cover.

Hydrologically, the province belongs to the Red River basin and includes several important river systems and tributaries, including the Da River and its associated drainage network. The dense river network and strongly dissected terrain contribute to slope undercutting, rapid runoff concentration, and local instability along valley margins. Land cover in the province consists of forest, agricultural land, residential areas, and transportation infrastructure. Changes in land use, slope cutting, and road construction may further modify surface runoff and slope stability.

Hoa Binh Province is also crossed by important mountainous transportation corridors, including the QL6 corridor, which connects Hanoi with northwestern Vietnam. Many road sections are located along steep

hillslopes and valley margins, where rainfall-induced landslides can disrupt traffic, damage road infrastructure, and threaten nearby communities. These characteristics make Hoa Binh Province a suitable case study for evaluating machine learning-based landslide susceptibility mapping under satellite rainfall uncertainty.

3. Materials and Methods

3.1. *Landslide Inventory*

The source landslide catalogue contained 184 records with catalogue identifiers (TT 1–184). A data audit was performed before modelling to ensure consistency between the inventory and the environmental predictor rasters. Only catalogued records with traceable attribute information were considered for model development; supplementary placemarks lacking catalogue identifiers and associated descriptive fields were not included in the training inventory.

All catalogue points were reprojected to WGS 84 / UTM Zone 48N (EPSG:32648) and overlaid with the common predictor grid. Two boundary records (TT16 and TT63) had NoData values across the required predictor layers and were excluded rather than spatially relocated or imputed. Consequently, 182 landslide locations with complete predictor information were retained as positive samples for machine-learning analysis.

Pseudo-absence samples were generated from raster cells with complete predictor coverage. To reduce the probability of assigning a pseudo-absence immediately adjacent to a documented landslide, candidate cells were required to be at least 500 m from every retained landslide location. A 1:1 presence-to-pseudo-absence ratio was used, producing 182 landslide and 182 pseudo-absence samples in each modelling realization.

3.2. *Landslide Conditioning Factors*

Six predictor variables were used in the susceptibility modelling: soil type, land use/land cover, slope class, elevation, annualized mean rainfall (X0), and maximum monthly rainfall (Xmax). These predictors were selected because complete and traceable spatial data were available for the retained modelling samples. The six conditioning factors, their data sources, preprocessing procedures, and modelling roles are summarized in **Table 1**.

Table 1. Conditioning factors, data sources, preprocessing, and modelling roles used in landslide susceptibility assessment.

Factor group	Predictor	Data source	Preprocessing	Role in susceptibility modelling
Soil	Soil type (Dat)	Legacy project soil GIS layer (tho_nhuong48); original provider metadata unavailable.	Rasterized and aligned to the 100-m modelling grid	Represents spatial differences in soil/material and hydrological properties
Land surface	Land use/land cover (Thamphu)	2024 project land-cover raster (48Q_20240101–20241231.tif); original product/provider metadata unavailable.	Clipped to the study area and resampled to the 100-m modelling grid	Represents vegetation cover, land-surface condition, and human disturbance
Topography	Slope class (Dodoc)	Derived from DEM	Reclassified into five classes: 0–15°, 15–25°, 25–35°, 35–45°, and >45°	Represents terrain steepness and gravitational instability
Topography	Elevation	SRTM DEM (Farr *et al.*, 2007)	Reprojected and resampled to the 100-m modelling grid	Represents terrain setting and elevation-related environmental gradients
Rainfall	Annualized mean rainfall indicator (X0)	Long-term mean rainfall raster from the project dataset; source period unavailable in metadata.	Annualized as mean rainfall × 12 and aligned to the 100-m modelling grid	Represents the broad spatial rainfall background
Rainfall	Maximum monthly rainfall (Xmax)	CHIRPS monthly rainfall, 1981–2024	Maximum monthly rainfall derived from the monthly dataset, then reprojected/resampled for spatial alignment.	Represents spatial variability in high monthly rainfall

Table note: All predictor layers were projected to WGS 84 / UTM Zone 48N (EPSG:32648) and aligned to a common 100-m computational grid. Resampling of CHIRPS-derived rainfall layers was performed for spatial alignment only and does not increase the native spatial information content of the satellite product.

Categorical and ordered rasters were aligned using nearest-neighbour resampling, while continuous variables were resampled to the same grid for pixel-wise extraction. The 100-m grid is a computational alignment grid; resampling the CHIRPS-derived rainfall layers does not increase the effective spatial information content of the original satellite precipitation product.

The rainfall representation has two complementary components. X0 describes the broad long-term rainfall background, whereas Xmax represents the maximum monthly rainfall in the available 1981–2024 CHIRPS dataset. These variables characterize spatial rainfall conditions but do not reproduce the short-duration, event-

specific triggering rainfall associated with individual landslide dates.

The X0 layer was available as a project-derived long-term rainfall climatology. Because its original source period was not preserved in the archived metadata, its temporal representativeness cannot be independently verified from the project records; X0 is therefore used only as a broad spatial rainfall-background indicator rather than as an independently validated event-scale rainfall product. The 2024 land-cover raster represents current land-surface conditions, whereas the landslide inventory comprises events accumulated over a multi-year period and exact event dates were not consistently available for all records. This temporal mismatch may introduce additional uncertainty in locations affected by recent land-use change. Original provider metadata for the legacy soil and land-cover layers were also not preserved. Although the exact processed layers and preprocessing steps used in this study remain traceable within the project archive, these provenance gaps limit independent reconstruction from the original source products.

Figure 2 illustrates the spatial distribution of four representative predictor layers used in the susceptibility analysis, namely Xmax, land use/land cover, soil type, and elevation.

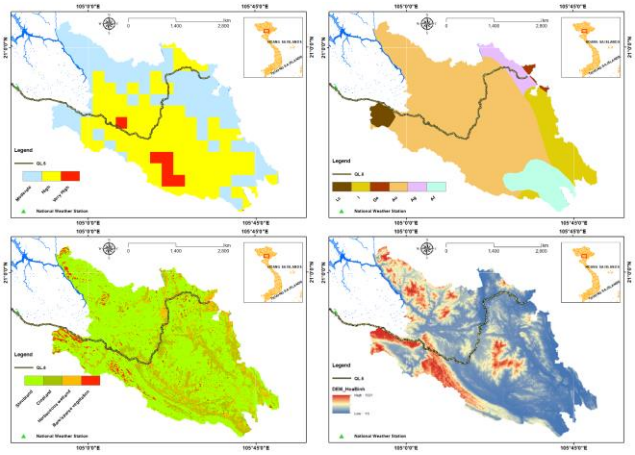


Figure 2. Selected spatial predictors used in the susceptibility analysis: (a) maximum monthly rainfall (Xmax; derived from CHIRPS 1981–2024), (b) land use/land cover (2024 project classification), (c) soil type (legacy project layer), and (d) elevation (m; SRTM). Inventory points represent the 182 retained landslide locations with complete predictor coverage. Slope class and X0 are described in **Table 1**.

3.3. *Data Preprocessing and Sampling Strategy*

All spatial layers were projected to WGS 84 / UTM Zone 48N and aligned to the 100-m reference grid. Cells containing NoData in any required predictor were excluded from candidate sampling and final map prediction. Soil type (Dat) and land-cover class (Thamphu) were treated as nominal categorical predictors and were one-hot encoded. Slope class (Dodoc) was retained as an ordered predictor, whereas elevation, X0, and Xmax were treated as continuous variables. For the ANN only, the numeric variables were standardized using parameters estimated from the training data.

For the primary modelling realization, a fixed random seed of 42 was used to generate 182 pseudo-absence samples from eligible cells located at least 500 m from the retained landslide points. The resulting balanced dataset contained 364 samples. A stratified 70:30 train–test split with random_state = 42 was applied, yielding 127 landslide and 127 pseudo-absence samples for training and 55 landslide and 55 pseudo-absence samples for independent testing. The stratified random split preserves class balance but does not impose spatial separation between training and test samples; its performance estimates should therefore be interpreted as internal discrimination under the sampled data rather than as a strict test of spatial transferability.

To quantify sensitivity to pseudo-absence selection, the full evaluation was repeated using ten independently generated pseudo-absence realizations with seeds 11, 22, 33, 42, 55, 66, 77, 88, 99, and 123. The landslide inventory, 1:1 class ratio, 500-m exclusion distance, train–test split rule, preprocessing rules, and fixed model hyperparameters were kept unchanged across realizations. This design isolates the uncertainty associated with pseudo-absence selection.

Table 2. Model hyperparameters and validation protocol used in the susceptibility analysis.

Model	Preprocessing	Selected hyperparameters	Tuning / validation
RF	One-hot: Dat, Thamphu; ordinal numeric: Dodoc; continuous: Elevation, X0, Xmax	n_estimators=400; min_samples_leaf=6; max_features=1.0; max_depth=5	RandomizedSearchCV; 5-fold stratified CV; scoring=ROC-AUC; training subset only
ExtraTrees	One-hot: Dat, Thamphu; ordinal numeric: Dodoc; continuous: Elevation, X0, Xmax	n_estimators=400; min_samples_leaf=6; max_features=1.0; max_depth=5	RandomizedSearchCV; 5-fold stratified CV; scoring=ROC-AUC; training subset only
LightGBM	One-hot: Dat, Thamphu; ordinal numeric: Dodoc; continuous: Elevation, X0, Xmax	subsample=0.85; reg_lambda=10.0; num_leaves=7; n_estimators=100; min_child_samples=5; max_depth=3; learning_rate=0.01; colsample_bytree=0.85	RandomizedSearchCV; 5-fold stratified CV; scoring=ROC-AUC; training subset only
XGBoost	One-hot: Dat, Thamphu; ordinal numeric: Dodoc; continuous: Elevation, X0, Xmax	subsample=1.0; reg_lambda=5.0; n_estimators=200; min_child_weight=1;	RandomizedSearchCV; 5-fold stratified CV; scoring=ROC-AUC; training subset only

		max_depth=4; learning_rate=0.03; colsample_bytree=1.0	
ANN	One-hot: Dat, Thamphu; ordinal numeric: Dodoc; continuous: Elevation, X0, Xmax; StandardScaler for numeric variables	learning_rate_init=0.01; hidden_layer_sizes=[32]; batch_size=16; alpha=0.01; activation=tanh	RandomizedSearchCV; 5-fold stratified CV; scoring=ROC-AUC; training subset only

3.4. *Machine Learning Models and Hyperparameter Tuning*

Five algorithms were evaluated: an artificial neural network implemented as a multilayer perceptron (ANN) (Rumelhart *et al.*, 1986), random forest (RF) (Breiman, 2001), Extra Trees (ERT) (Geurts *et al.*, 2006), Light Gradient Boosting Machine (LightGBM) (Ke *et al.*, 2017), and Extreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016). The four tree-based algorithms were selected because of their ability to model nonlinear relationships and interactions in tabular environmental data, whereas the ANN provided a neural-network comparator. A recurrent LSTM architecture was not included because the predictor matrix is static spatial/tabular data and does not contain a genuine temporal sequence.

Hyperparameter optimization was performed once on the training subset of the seed-42 modelling realization. RandomizedSearchCV with five-fold StratifiedKFold cross-validation was used, with ROC-AUC as the model-selection criterion. The independent 30% test subset was not used during tuning. After the optimal settings were selected, the hyperparameters were frozen and applied unchanged to all ten pseudo-absence realizations. This separation between tuning and final testing was used to improve reproducibility and reduce optimistic performance estimation. The selected model hyperparameters and the common tuning and validation protocol are summarized in **Table 2**.

The complete modelling workflow, from data quality control and repeated pseudo-absence sampling to model validation and uncertainty-aware susceptibility mapping, is illustrated in **Figure 3**.

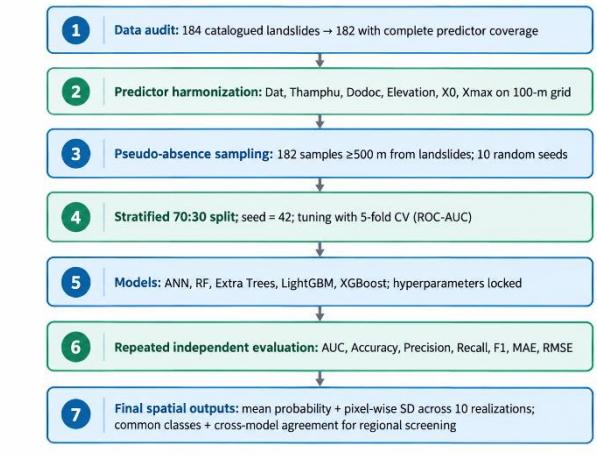


Figure 3. Workflow for data quality control, repeated pseudo-absence sampling, model tuning and validation, and uncertainty-aware susceptibility mapping.

3.5. *Model Evaluation, Robustness Analysis, and Susceptibility Mapping*

Model performance was assessed on the independent 30% testing subset using the area under the receiver operating characteristic curve (AUC), accuracy, precision, recall, F1-score, mean absolute error (MAE), and root mean square error (RMSE). Classification metrics were calculated at a probability threshold of 0.5, while AUC was used as the main threshold-independent discrimination measure. Confusion matrices were also retained for each realization.

Rather than reporting a single train–test realization, final performance was summarized across the ten pseudo-absence realizations as the mean ± standard deviation and range. Mean ROC curves were constructed from the per-realization test predictions. This repeated-sampling design was used to distinguish model performance from the variability introduced by pseudo-absence selection.

After independent validation was completed and hyperparameters were fixed, each model was refitted on the complete 182-presence + 182-pseudo-absence dataset for each of the ten pseudo-absence realizations and applied to all valid cells of the common 100-m grid. The final continuous susceptibility map for each algorithm was calculated as the pixel-wise mean predicted probability across the ten realizations. A corresponding pixel-wise standard-deviation map was calculated to represent spatial uncertainty associated with pseudo-absence selection.

For direct cross-model comparison, continuous probabilities were classified using one common set of fixed probability thresholds: very low (<0.20), low (0.20–<0.40), moderate (0.40–<0.60), high (0.60–<0.80), and very high (≥0.80). Common thresholds were preferred to model-specific natural-break classes because model-specific class boundaries would confound spatial comparison among algorithms. Continuous probability maps remain the primary modelling output.

4. Results and Discussion

4.1. *Model Performance and Robustness to Pseudo-absence Selection*

Table 3 summarizes model performance across the ten pseudo-absence realizations. The mean AUC values were 0.686 ± 0.050 for LightGBM, 0.681 ± 0.039 for RF, 0.675 ± 0.037 for Extra Trees, 0.674 ± 0.046 for XGBoost, and 0.656 ± 0.043 for ANN. The AUC ranges overlapped substantially, indicating that no single algorithm can be regarded as unequivocally superior under the present data conditions.

The model ranking also varied among pseudo-absence realizations. LightGBM achieved the highest AUC in four of

the ten realizations, RF and XGBoost each ranked first twice, and Extra Trees and ANN each ranked first once. This variation confirms that the selection of pseudo-absence locations is a non-negligible source of uncertainty and that performance conclusions based on a single random realization can be unstable.

At the default probability threshold of 0.5, RF achieved the highest mean accuracy (0.636 ± 0.034), XGBoost achieved the highest recall (0.600 ± 0.056) and F1-score (0.617 ± 0.046), and Extra Trees produced the highest mean precision (0.674 ± 0.061). The lowest mean MAE was obtained by XGBoost (0.429 ± 0.020), whereas RF produced the lowest mean RMSE (0.477 ± 0.014). These complementary results further indicate that the models emphasize different aspects of discrimination and classification performance. The recall of ANN was also the most variable across realizations (SD = 0.099), although Extra Trees showed a similarly high recall SD of 0.096. The elevated ANN variability likely reflects sensitivity to weight initialization and gradient-based optimization under changing pseudo-absence configurations, reinforcing the value of repeated sampling.

Table 3. Final performance of the five models across ten pseudo-absence realizations.

Model	AUC mean±SD	AUC range	Accuracy mean±SD	Precision mean±SD	Recall mean±SD	F1 mean±SD	MAE mean±SD	RMSE mean±SD	Best AUC count
LightGBM	0.686 ± 0.050	0.615–0.786	0.625 ± 0.040	0.673 ± 0.093	0.524 ± 0.070	0.581 ± 0.034	0.476 ± 0.007	0.481 ± 0.007	4
RF	0.681 ± 0.039	0.636–0.742	0.636 ± 0.034	0.660 ± 0.059	0.580 ± 0.039	0.615 ± 0.024	0.436 ± 0.016	0.477 ± 0.014	2
ExtraTrees	0.675 ± 0.037	0.632–0.735	0.617 ± 0.040	0.674 ± 0.061	0.467 ± 0.096	0.545 ± 0.068	0.443 ± 0.013	0.479 ± 0.013	1
XGBoost	0.674 ± 0.046	0.618–0.775	0.628 ± 0.050	0.641 ± 0.065	0.600 ± 0.056	0.617 ± 0.046	0.429 ± 0.020	0.482 ± 0.019	2
ANN	0.656 ± 0.043	0.557–0.714	0.606 ± 0.046	0.615 ± 0.054	0.589 ± 0.099	0.596 ± 0.059	0.442 ± 0.024	0.490 ± 0.015	1

Figure 4 illustrates model robustness across the ten pseudo-absence realizations, showing both the seed-to-seed variation in test AUC and the corresponding mean ROC performance.

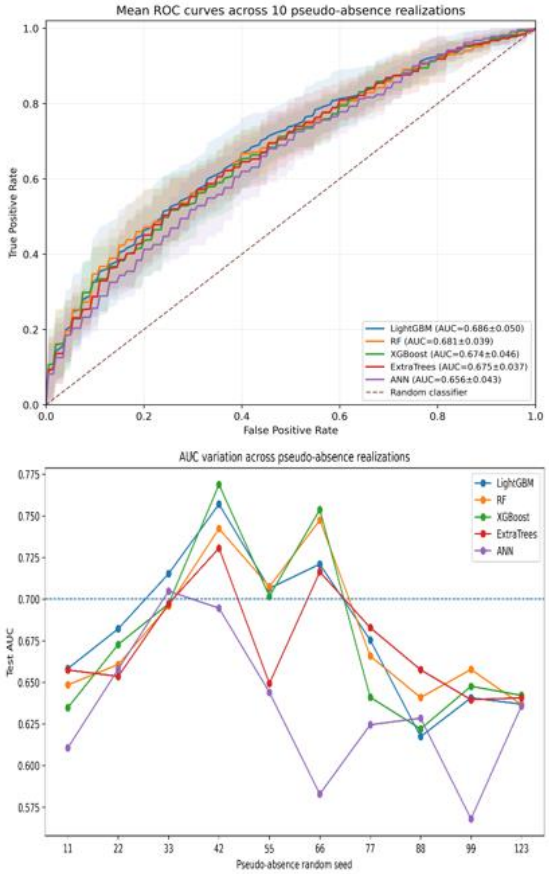


Figure 4. Robustness of model performance across ten pseudo-absence realizations: (a) mean ROC curves of ANN, RF, Extra Trees, LightGBM, and XGBoost across the ten realizations; (b) test AUC variation across random pseudo-absence seeds.

The resulting AUC values are moderate rather than high. This should not be interpreted solely as a limitation of the algorithms. The modelling problem is constrained by the relatively small inventory, uncertainty in pseudo-absence definition, the limited six-predictor set, and the use of climatological/monthly satellite-rainfall indicators rather than event-specific triggering rainfall. The repeated-sampling results therefore support a screening interpretation rather than deterministic landslide prediction.

The AUC values are lower than those reported in several recent regional studies, but direct comparison should be made cautiously because inventory size, predictor richness, sampling design, and validation protocols differ substantially. For example, Tran *et al.* (2024) reported an AUC of 0.883 for RF in Van Yen District using 354 landslide locations and thirteen conditioning factors, whereas Manopkawee *et al.* (2026) reported AUC values above 0.90 for several models using 1,752 landslides and a unified ten-factor dataset in northern Thailand. In contrast, the present analysis is based on 182 retained landslide locations, six traceable predictors, and rainfall indicators that do not contain event-specific triggering histories. The moderate mean AUC values should therefore be interpreted as performance under constrained observational information rather than as evidence that one algorithm is intrinsically inadequate. Importantly, the repeated pseudo-absence analysis shows that model ranking itself changes among realizations, reinforcing the need to report robustness rather than a single favorable AUC value.

4.2. *Spatial Distribution of Landslide Susceptibility*

For each model, the final spatial prediction was obtained by averaging ten full-coverage probability maps generated from the ten pseudo-absence realizations. This ensemble-by-sampling procedure reduces dependence on a single pseudo-absence realization and provides a corresponding pixel-wise uncertainty estimate.

The continuous probability maps showed substantial but not complete spatial agreement among algorithms. The strongest pairwise correlation occurred between RF and XGBoost (r = 0.961), followed by RF and LightGBM (r = 0.896) and LightGBM and XGBoost (r = 0.887). The overlap of areas classified as high or very high was lower than the correlation of continuous probabilities; the highest Jaccard similarity for high/very-high areas was 0.583 between RF and XGBoost. These results indicate that the principal spatial gradients are broadly shared, while the delineation of the highest-susceptibility zones remains model-dependent.

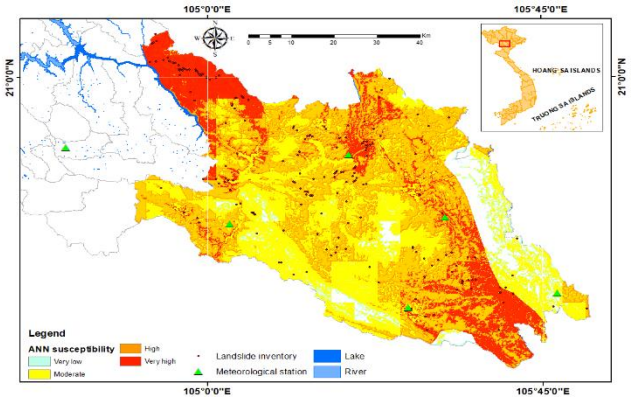


Figure 5. Mean rainfall-induced landslide susceptibility map generated by ANN across ten pseudo-absence realizations using common fixed probability classes.

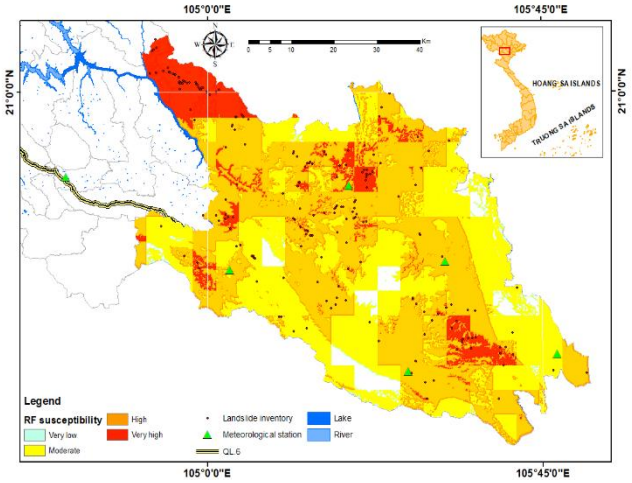


Figure 6. Mean rainfall-induced landslide susceptibility map generated by RF across ten pseudo-absence realizations using common fixed probability classes.

Using the common fixed probability thresholds, the proportion of mapped area classified as high or very high was 17.27% for ANN, 10.02% for RF, 3.98% for LightGBM, 15.09% for XGBoost, and 10.91% for Extra Trees. LightGBM produced a notably narrower probability range (0.381–0.668), causing 90.64% of the valid grid to fall within the common moderate class. This reflects differences in probability calibration and score dispersion among algorithms and should not be interpreted as direct evidence that LightGBM identifies little spatial variation. For this reason, the continuous probability surfaces and cross-model agreement are more informative than class-area percentages alone. **Figures 5–9** show the mean landslide susceptibility maps generated by ANN, RF, LightGBM, XGBoost, and Extra Trees, respectively, using the common fixed probability classes.

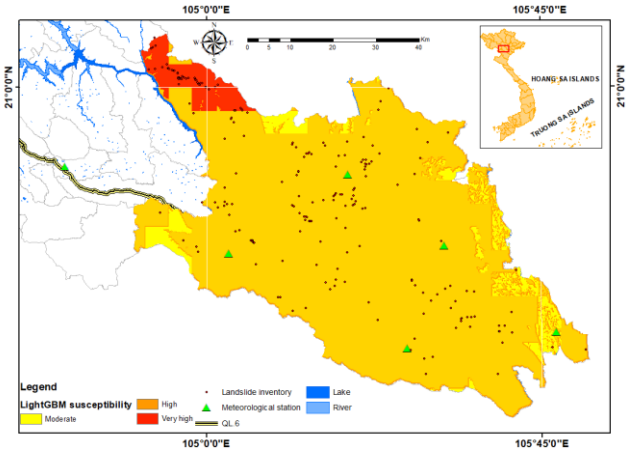


Figure 7. Mean rainfall-induced landslide susceptibility map generated by LightGBM across ten pseudo-absence realizations using common fixed probability classes.

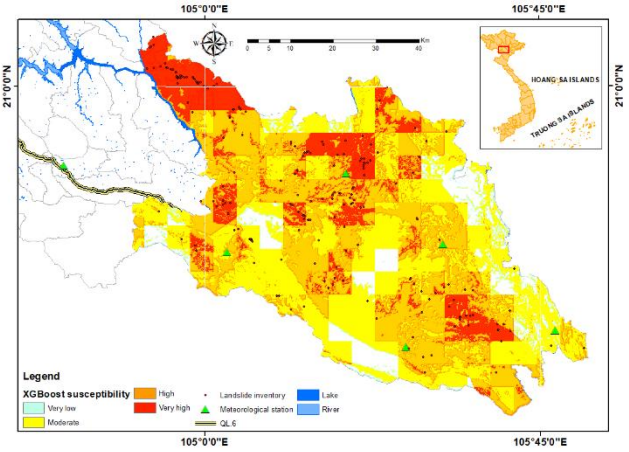


Figure 8. Mean rainfall-induced landslide susceptibility map generated by XGBoost across ten pseudo-absence realizations using common fixed probability classes.

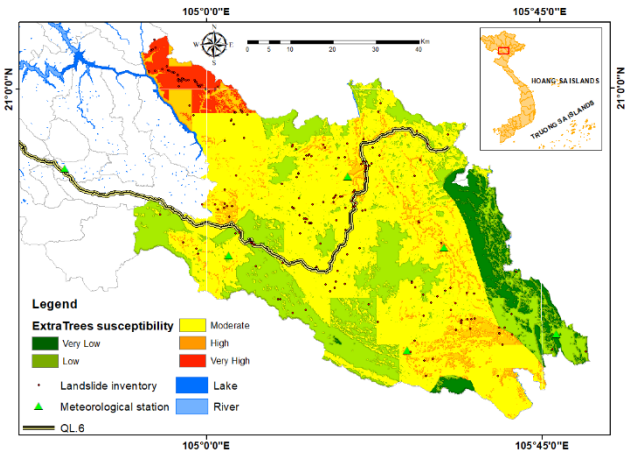


Figure 9. Mean rainfall-induced landslide susceptibility map generated by Extra Trees across ten pseudo-absence realizations using common fixed probability classes.

For practical screening, locations identified as high or very high by multiple tree-based models can be treated as higher-confidence priority areas for field verification. The resulting maps can be intersected with the QL6 transportation corridor and other exposed assets to identify candidate sections for detailed investigation; however, the maps do not quantify the probability of failure of an individual road segment.

4.3. *Influence of Conditioning Factors on Landslide Susceptibility*

Relative predictor importance was calculated for RF, Extra Trees, LightGBM, and XGBoost and aggregated back to the six original predictor variables. Averaged across the four tree-based models, soil type (Dat) had the highest relative importance (0.341), followed by maximum monthly rainfall Xmax (0.273), the long-term rainfall indicator X0 (0.225), elevation (0.095), slope class (0.033), and land use/land cover (0.033). Thus, the two rainfall indicators together accounted for approximately one-half of the mean normalized importance across the tree-based models.

The model-specific importance patterns were not identical. RF and LightGBM assigned comparatively large importance to the rainfall variables, whereas Extra Trees and XGBoost assigned greater weight to soil type. This model dependence is important because feature-importance values are algorithm-specific diagnostics rather than direct measures of physical causality. The low average importance of slope and land cover in the present framework should therefore be interpreted in the context of their discretized representation, interactions with other predictors, and the limited inventory rather than as evidence that they are physically unimportant. Figure 10 compares the relative importance of the six conditioning factors across the four tree-based ensemble models.

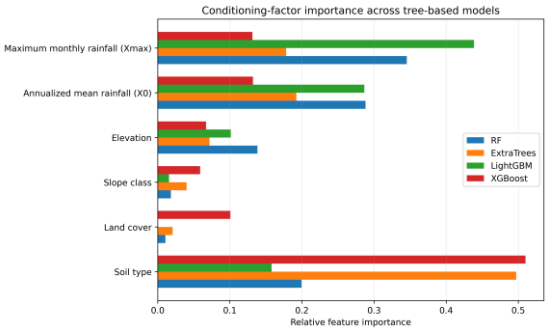


Figure 10. Relative importance of the six conditioning factors across the four tree-based ensemble models.

4.4. *Rainfall-Input and Sampling Uncertainty*

Uncertainty in the modelling framework can be separated into two components. The first is rainfall-input uncertainty. The CHIRPS-derived predictors describe long-term and monthly rainfall conditions, not the event-specific rainfall associated with each landslide occurrence. Xmax was derived as the maximum monthly rainfall from the available 1981–2024 CHIRPS dataset, while X0 was generated from a long-term mean rainfall raster multiplied by 12. The rainfall rasters were subsequently reprojected and resampled to the common 100-m computational grid. This resampling was performed for spatial alignment only and does not create independent rainfall information at 100-m resolution.

Because exact landslide occurrence dates and corresponding triggering-rainfall histories were not consistently available for the inventory, the present rainfall variables represent broad spatial forcing rather than event-scale triggers. Short-duration rainfall intensity, multi-day antecedent rainfall, local orographic enhancement, and retrieval errors in satellite precipitation are therefore not explicitly resolved. The study consequently evaluates the implications of rainfall-input limitations rather than performing a full probabilistic decomposition of CHIRPS measurement error.

The observed sensitivity to pseudo-absence selection is consistent with recent evidence that non-landslide sampling can substantially alter both predictive metrics and susceptibility zoning (Gu *et al.*, 2024; Guo *et al.*, 2024; Liu *et al.*, 2024). The present study extends this concern by propagating ten independently generated pseudo-absence realizations into both performance statistics and final spatial predictions. At the same time, rainfall-related uncertainty remains qualitatively different from sampling uncertainty. Lee *et al.* (2025) demonstrated the value of daily meteorological information for rainfall-induced susceptibility and short-term warning, whereas the present CHIRPS-derived indicators describe broader rainfall climatology and monthly extremes. Therefore, the study does not claim to predict event timing; it evaluates regional susceptibility under incomplete information and makes the remaining uncertainty explicit.

The second component—pseudo-absence sampling uncertainty—was quantified directly. Across the ten realizations, AUC standard deviations were 0.050 for LightGBM, 0.039 for RF, 0.037 for Extra Trees, 0.046 for XGBoost, and 0.043 for ANN. At the map level, the mean pixel-wise standard deviation of predicted susceptibility was 0.035 for LightGBM, 0.057 for Extra Trees, 0.073 for RF, 0.094 for XGBoost, and 0.104 for ANN. LightGBM therefore produced the most spatially stable probabilities with respect to pseudo-absence selection, although its prediction range was also more compressed.

These results demonstrate that uncertainty arises not only from the rainfall input but also from the way non-landslide conditions are represented. Accordingly, the mean susceptibility map should be interpreted together with the spatial standard-deviation map and, where possible, with cross-model agreement. Areas with consistently high mean susceptibility and low sampling uncertainty provide stronger screening evidence than areas where high probabilities are accompanied by large inter-realization variability.

A separate reproducibility limitation concerns source provenance. The original provider metadata for the legacy soil layer and the 2024 land-cover layer, as well as the source period of the X0 rainfall raster, were not retained in the available project metadata. The archived processed

layers and preprocessing workflow are traceable and sufficient to reproduce the reported calculations from the project dataset, but the missing upstream metadata limits independent verification of the original source products and their temporal representativeness.

A further limitation concerns spatial validation. The stratified random 70:30 train–test split does not explicitly account for spatial dependence among nearby landslide and pseudo-absence samples. Consequently, the reported discrimination metrics may be somewhat optimistic with respect to transferability to spatially independent areas. The present AUC values are therefore interpreted as internal regional-screening performance under the available sample configuration, not as evidence of independent spatial generalization. Future work should evaluate spatial block cross-validation or leave-one-area-out validation to provide a stricter assessment of spatial predictive transferability.

4.5. *Implications for Environmental Management and Disaster Risk Reduction*

The susceptibility maps are intended for regional-scale environmental-risk screening rather than deterministic prediction of individual landslides. Their principal value is to prioritize locations for field verification, slope inspection, drainage assessment, monitoring, and more detailed geotechnical analysis. For transportation management, the continuous susceptibility and cross-model agreement layers can be overlaid with the QL6 corridor to identify road-adjacent slopes where additional investigation may be warranted.

The moderate discrimination performance and the quantified sensitivity to pseudo-absence selection also provide an important caution for decision-making. A single model or a single random training sample should not be used as the sole basis for engineering intervention. Greater confidence is associated with locations where several tree-based models indicate elevated susceptibility and where the sampling-uncertainty map is comparatively low.

Site-specific engineering design requires information beyond the present regional model, including field observations, local drainage conditions, geotechnical parameters, failure geometry, and event-scale rainfall. The present framework is therefore best used as an initial prioritization layer that directs limited field and monitoring resources toward areas of greater potential concern.

5. Conclusions

This study developed a reproducible machine-learning framework for regional landslide susceptibility screening in Hoa Binh, northern Vietnam. Following data quality control, 182 catalogued landslide locations with complete predictor coverage were retained and paired with an equal number of pseudo-absence samples located at least 500 m from known landslides. Five models—ANN, RF, Extra Trees, LightGBM, and XGBoost—were evaluated using a stratified 70:30 train–test design, five-fold cross-validation for hyperparameter selection, and ten independent pseudo-absence realizations for robustness assessment.

Across the ten realizations, LightGBM achieved the highest mean AUC (0.686 ± 0.050), closely followed by RF (0.681 ± 0.039), Extra Trees (0.675 ± 0.037), XGBoost (0.674 ± 0.046), and ANN (0.656 ± 0.043). However, the best-performing algorithm changed among realizations (LightGBM in 4/10, RF and XGBoost in 2/10 each, and Extra Trees and ANN in 1/10 each), demonstrating that pseudo-absence selection materially affects apparent model ranking. The analysis therefore does not identify a single universally superior model; instead, it emphasizes robust performance summaries and model agreement.

The final susceptibility surfaces were generated as pixel-wise mean probabilities across ten pseudo-absence realizations, accompanied by pixel-wise standard-deviation maps. LightGBM showed the lowest mean spatial sampling uncertainty, whereas ANN and XGBoost were more sensitive to pseudo-absence selection. Feature-importance analysis of the four tree-based models indicated that soil type and the two rainfall indicators were the dominant predictors on average, although importance patterns differed among algorithms.

The results also clarify the role of satellite rainfall uncertainty. The CHIRPS-derived variables capture broad long-term and monthly rainfall patterns but cannot fully represent short-duration or antecedent rainfall conditions associated with individual landslide events. Consequently, the resulting susceptibility maps should be used as regional screening products and interpreted together with sampling uncertainty, model agreement, and field information. Future work should prioritize event-dated landslide inventories and higher-temporal-resolution rainfall information to improve the physical linkage between rainfall forcing and slope failure.

Availability of data and materials

CHIRPS precipitation data are publicly available from the Climate Hazards Center (https://www.chc.ucsb.edu/data/chirps), and SRTM elevation data are available through NASA Earthdata (https://earthdata.nasa.gov/). The legacy project GIS layers (soil type and land cover), processed datasets, model inputs, and derived outputs supporting the findings of this study are available from the corresponding author upon reasonable request, subject to any applicable data-use restrictions. Provenance limitations for the legacy project layers and X0 rainfall raster are described in Section 3.2.

Competing interests

The authors declare that they have no competing interests.

References

Ado, M., Amitab, K., Maji, A.K., Jasińska, E., Gono, R., Leonowicz, Z., Jasiński, M. (2022). Landslide susceptibility mapping using machine learning: A literature survey. Remote Sensing, 14, 3029. https://doi.org/10.3390/rs14133029

Beck, H.E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A.I.J.M., Weedon, G.P., Brocca, L., Pappenberger, F., Huffman, G.J., Wood, E.F. (2017). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. Hydrology and Earth System Sciences, 21, 6201–6217. https://doi.org/10.5194/hess-21-6201-2017

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

Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

Dandridge, C., Lakshmi, V., Bolten, J., Srinivasan, R. (2019). Evaluation of satellite-based rainfall estimates in the Lower Mekong River Basin, Southeast Asia. Remote Sensing, 11, 2709. https://doi.org/10.3390/rs11222709

Dao, D.V., Jaafari, A., Bayat, M., Mafi-Gholami, D., Qi, C., Moayedi, H., Phong, T.V., Ly, H.-B., Le, T.-T., Trinh, P.T., Luu, C., Quoc, N.K., Thanh, B.N., Pham, B.T. (2020). A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. CATENA, 188, 104451. https://doi.org/10.1016/j.catena.2019.104451

Das, R., Tien, P.V., Wegmann, K.W., Chakraborty, M. (2024). Machine learning-based assessment of regional-scale variation of landslide susceptibility in central Vietnam. PLOS ONE, 19(10), e0308494. https://doi.org/10.1371/journal.pone.0308494

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, 8953–8982. https://doi.org/10.1007/s11069-024-06563-8

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D. (2007). The Shuttle Radar Topography Mission. Reviews of Geophysics, 45(2), RG2004. https://doi.org/10.1029/2005RG000183

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. Scientific Data, 2, 150066. https://doi.org/10.1038/sdata.2015.66

Geurts, P., Ernst, D., Wehenkel, L. (2006). Extremely randomized trees. Machine Learning, 63, 3–42. https://doi.org/10.1007/s10994-006-6226-1

Gu, T., Duan, P., Wang, M., Li, J., Zhang, Y. (2024). Effects of non-landslide sampling strategies on machine learning models in landslide susceptibility mapping. Scientific Reports, 14, 7201. https://doi.org/10.1038/s41598-024-57964-5

Guo, Z., Tian, B., Zhu, Y., He, J., Zhang, T. (2024). How do the landslide and non-landslide sampling strategies impact landslide susceptibility assessment? A catchment-scale case study from China. Journal of Rock Mechanics and Geotechnical Engineering, 16(3), 877–894. https://doi.org/10.1016/j.jrmge.2023.07.026

Guzzetti, F., Peruccacci, S., Rossi, M., Stark, C.P. (2007). Rainfall thresholds for the initiation of landslides in central and southern Europe. Meteorology and Atmospheric Physics, 98, 239–267. https://doi.org/10.1007/s00703-007-0262-7

Ji, H., Peng, D., Gu, Y., Liang, Y., Luo, X. (2022). Evaluation of multiple satellite precipitation products and their potential utilities in the Yarlung Zangbo River Basin. Scientific Reports, 12, 13334. https://doi.org/10.1038/s41598-022-17551-y

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3146–3154.

Lee, S., Roh, M., Jo, H.-W., Joon, K., Lee, W.-K. (2025). Machine learning-based rainfall-induced landslide susceptibility model and short-term early warning assessment in South Korea. Landslides, 22(8), 2809–2827. https://doi.org/10.1007/s10346-025-02513-y

Liu, L.-L., Zhao, S.-L., Yang, C., Zhang, W. (2024). Quantifying uncertainty in landslide susceptibility mapping due to sampling randomness. International Journal of Disaster Risk Reduction, 114, 104966. https://doi.org/10.1016/j.ijdrr.2024.104966

Long, N.T., De Smedt, F. (2019). Analysis and mapping of rainfall-induced landslide susceptibility in A Luoi District, Thua Thien Hue Province, Vietnam. Water, 11, 51. https://doi.org/10.3390/w11010051

Manopkawee, P., Mankhemthong, N., Tongsae, M., Panna, N. (2026). Evaluating model robustness in landslide susceptibility mapping using a unified data-consistent framework in northern Thailand. Scientific Reports, 16, 22568. https://doi.org/10.1038/s41598-026-50703-y

Merghadi, A., Yunus, A.P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D.T., Avtar, R., Boumezbeur, A. (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

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

Rumelhart, D.E., Hinton, G.E., Williams, R.J. (1986). Learning representations by back-propagating errors. Nature, 323, 533–536. https://doi.org/10.1038/323533a0

Tran, V.A., Khuc, T.D., Truong, X.Q., Nguyen, A.B., Phi, T.T. (2024). Application of potential machine learning models in landslide susceptibility assessment: A case study of Van Yen District, Yen Bai Province, Vietnam. Quaternary Science Advances, 14, 100181. https://doi.org/10.1016/j.qsa.2024.100181

Trinh-Tuan, L., Matsumoto, J., Ngo-Duc, T., Nodzu, M.I., Inoue, T. (2019). Evaluation of satellite precipitation products over Central Vietnam. Progress in Earth and Planetary Science, 6, 54. https://doi.org/10.1186/s40645-019-0297-7.