Comparing Five Machine Learning Approaches for Local-Scale Landslide Susceptibility Mapping in the Huoi Reng Watershed, Central Vietnam

Dao, M. N.,[1*] Phan, T. M. H.[2] and Nguyen, Q. P.[3]
[1]Northwest Geological Division, Vietnam Geological Department, Hanoi, Vietnam; Hanoi University of Mining and Geology, Hanoi, Vietnam
E-mail: NCS229108014@humg.edu.vn* ORCID ID: 0009-0004-3889-7068*
[2]Vietnam Academy of Science and Technology, Hanoi, Vietnam; Hanoi University of Mining and Geology, Hanoi, Vietnam, E-mail: phanthimaihoa@humg.edu.vn, ORCID ID: 0009-0002-7129-2170
[3]Hanoi University of Mining and Geology, Hanoi, Vietnam
E-mail: nguyenquocphi@humg.edu.vn, ORCID ID: 0000-0001-9944-8216
Corresponding Author

DOI: https://doi.org/10.52939/ijg.v22i8.5139

Abstract
Landslide susceptibility mapping supports disaster-risk reduction in steep tropical catchments where rainfall, deeply weathered materials, lithology and land-cover disturbance interact. This study compares artificial neural network (ANN), J48 decision tree, Random Forest (RF), Bagging with Random Forest base learner (Bagging-RF) and Support Vector Machine (SVM) models for the Huoi Reng watershed, Central Vietnam. The inventory contains 319 mapped landslide locations. Calibration used pre-2023 landslides, while 39 landslides mapped in 2024 and six field-checked landslides from 2025 were kept for a limited occurrence-only plausibility assessment. Presence samples were extracted from landslide cells on a common 30 m grid, and non-landslide samples were randomly selected outside mapped landslides and a three-cell exclusion buffer. Model comparison was revised from a cell-level random split to a five-fold spatial block validation design. Bagging-RF obtained the highest mean ROC-AUC (0.807 ± 0.063), followed closely by RF (0.805 ± 0.063); the difference between these two ensemble models was not statistically significant. Bagging-RF and RF also produced the lowest Brier scores (0.184 ± 0.027 and 0.187 ± 0.028), indicating smaller fold-level probability error than ANN, J48 and SVM. The 45 later landslide points were used only as an occurrence-only class-concentration check, not as a full validation set. RF captured 61.85% of the later landslide signal within 9.89% of the area mapped as high or very high susceptibility. The results indicate that tree-based ensembles provide the most stable local screening products, but the maps should be used for field-priority zoning rather than quantitative risk prediction until independent stable-site sampling and event-based rainfall validation are available.

Keywords: Artificial Neural Network, Bagging-RF, Landslide Susceptibility, Random Forest, Spatial Block Validation

1. Introduction

Central Vietnam is frequently affected by tropical storms, prolonged rainfall and short, steep catchments that favor shallow landslides, debris movement and flash-flood processes. Local-scale susceptibility maps are therefore needed to guide field inspection, road maintenance, forest management and land-use screening in mountainous districts [1] and [2].

Landslide susceptibility mapping estimates the relative spatial tendency of slope failure under a defined set of terrain, geological, hydro-climatic, environmental and land-cover conditions. Machine learning is widely used because nonlinear classifiers can combine heterogeneous predictors without imposing a single linear response form [3] and [4]. However, high reported accuracy in landslide susceptibility studies can be strongly affected by landslide inventory quality, absence sampling, cell-level pseudo-replication, spatial autocorrelation and validation design [5] and [6].

Previous studies in comparable tropical and mountainous settings have demonstrated the usefulness of decision trees, random forests, support vector machines, neural networks and ensemble learning for susceptibility mapping [7] and [8], including recent applications in Vietnam [9] and [10]. Nevertheless, three unresolved problems remain important for local-scale applications. First, many studies still rely on random cell-level splits, which can overestimate predictive performance when neighboring raster cells from the same landslide or geomorphic unit are divided between training and testing sets. Second, absence-sample selection and temporally separated landslide checks are often insufficiently described. Third, probability-based maps are frequently interpreted as operational predictions even when rainfall is represented only by long-term climatological layers.

This study addresses these issues by revising the Huoi Reng analysis around a spatially blocked validation framework, a transparent distinction between calibration data and later occurrence-only checks, and a cautious interpretation of the resulting maps as local screening products. The study does not claim to introduce a new classifier. Its scientific contribution is instead methodological and geomorphic: it shows how model ranking, map plausibility and management recommendations change when spatial dependence, occurrence-only validation and practical map use are treated explicitly.

The five selected algorithms were chosen to represent complementary model families available in Weka: J48 as an interpretable tree benchmark, RF as a standard tree ensemble, Bagging-RF as an outer bagging ensemble with RF base learner, ANN as a nonlinear connectionist classifier, and SVM as a margin-based classifier. This selection allows comparison between transparent, ensemble, neural-network and kernel-based learning strategies while avoiding an excessively large algorithm set for a single local watershed. The objectives of the study are: (i) to construct a reproducible GIS-based landslide susceptibility database for the Huoi Reng watershed using a common 30 m grid; (ii) to compare five machine learning classifiers under a five-fold spatial block validation design; and (iii) to evaluate the geomorphic plausibility and planning relevance of the susceptibility maps using spatial block metrics and a 45-point occurrence-only class-concentration check.

2. Materials and Methods

2.1 Study Area

The Huoi Reng watershed is located in the western part of Le Thuy District, Quang Binh Province, Central Vietnam, close to the Laos border. The mapped watershed extends approximately from 16°58′05″ to 17°07′15″N and from 106°23′40″ to 106°33′40″E, covering about 144 km². The area contains rugged hills, dissected valleys, forested slopes, road-cut corridors and sparsely distributed settlements. The audited GIS layers available for the revised manuscript show an elevation range of about 88.8-1,253.1 m and local slope gradients exceeding 60°, indicating strongly dissected terrain. Mean elevation, mean slope and formation-area proportions were not available from the archived modeling table and are therefore not used as inferential evidence in this revision.

The watershed was selected because it combines steep terrain, strongly weathered geological materials, structural discontinuities, intense rainfall exposure and field accessibility for 2024-2025 checking. Long-term rainfall was represented by mean annual rainfall for 2010-2023, ranging from approximately 2,383 mm to 2,745 mm in the GIS layer. Historical landslide evidence from provincial-scale landslide investigation reports and the 2024-2025 field checks indicates that slope failures are concentrated along dissected slopes, road corridors and weathered lithological units [11] and [12]. Quantified economic-loss data specific to this small watershed were not available in the source materials; therefore, the practical contribution of the study is framed as local field-priority screening for road-slope inspection, drainage maintenance and land-use screening rather than as economic risk assessment. The spatial distribution of the landslide inventory over the watershed terrain is shown in Figure 1.

2.2 Data Sources and Scale Harmonization

The database integrates landslide inventory data, topographic derivatives, geological maps, engineering-geological and hydrogeological layers, weathering-crust information, remote-sensing indices, land-use/land-cover data and rainfall. Raster processing, area calculation and sample extraction were performed in WGS 84/ UTM Zone 48N (EPSG:32648) using a common 30 m grid. Map layouts may be displayed using a WGS 84 geographic graticule, but the analytical grid was kept consistent across all predictors. The data sources are detailed in Table 1.

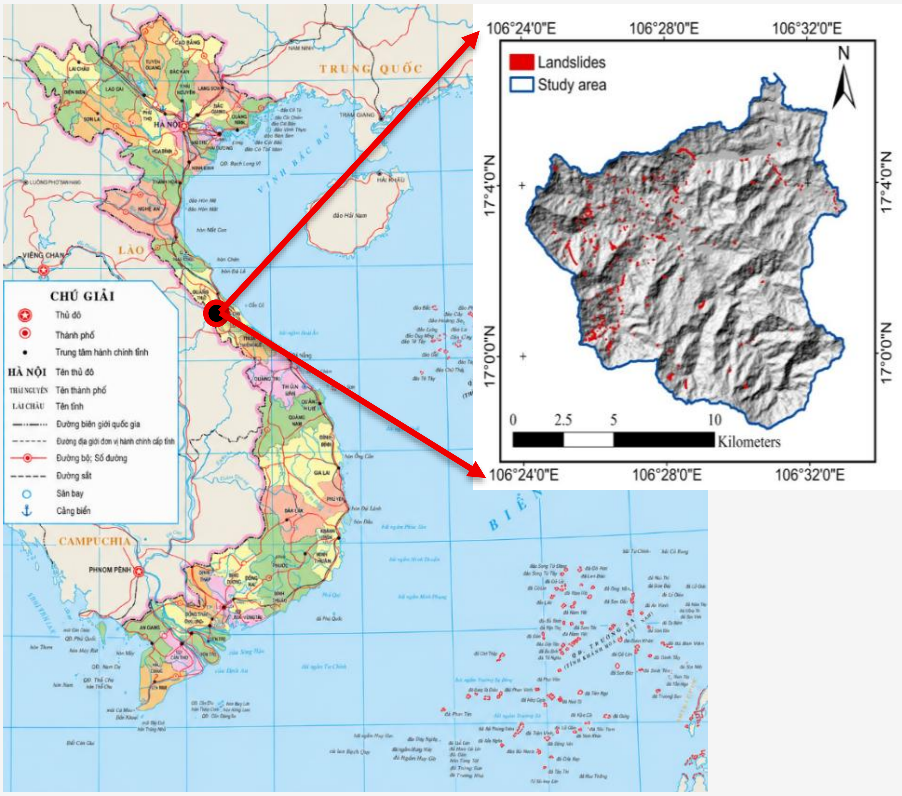

Figure 1: Location of the Huoi Reng watershed and spatial distribution of the landslide inventory

Table 1: Data sources, conditioning factors and processing steps used in the landslide susceptibility database

Group	Factor(s)	Source	Processing and scale harmonization
Landslide inventory	Presence locations	Field surveys, Google Earth interpretation and 2024-2025 checking	Pre-2023 landslides used for calibration; 2024-2025 points held out for occurrence-only plausibility assessment.
Topography	Elevation, slope, aspect, Terrain Ruggedness Index (TRI), Topographic Wetness Index (TWI), Mass Balance Index (MBI)	DEM interpolated from 1:10,000 contour data using Topo to Raster (ANUDEM)	Hydrologically conditioned and resampled to the 30 m UTM grid.
Geology	Lithology, engineering geology, hydrogeology, weathering crust, fault distance	DGMV geological, engineering-geological, hydrogeological and weathering maps; field checking	Vector layers reclassified and rasterized to the 30 m grid; fault distance derived from mapped faults.
Remote sensing	NDVI, NDMI, BSI	Sentinel-2 Level-2A surface reflectance scene acquired on 25 March 2025	Cloud-free dry-season scene; indices computed from bands B8, B4, B2 and B11 and resampled to the 30 m grid.
Rainfall	Mean annual rainfall 2010-2023	Satellite precipitation data supported by station-based interpolation	Used as long-term hydro-climatic conditioning layer, not as event-trigger rainfall.
Land cover	LULC 2020	Land-use map, remote-sensing interpretation and field checking	Reclassified into common land-cover categories and converted to the 30 m grid.

The 30 m grid was adopted to harmonize data sources with different original scales and resolutions. The DEM was interpolated from 1:10,000 contour data using the hydrologically correct Topo to Raster (ANUDEM) method, and terrain derivatives were resampled to the common grid. Geological, engineering-geological, hydrogeological and weathering-crust vector layers were digitized or reclassified from source maps and converted to the same raster grid. Because the geological layers have coarser original mapping scales than the DEM, the final products should be interpreted as catchment-scale screening outputs, not as site-specific geotechnical design maps.

Normalized Difference Vegetation Index (NDVI), Normalized Difference Moisture Index (NDMI) and Bare Soil Index (BSI) were retained because vegetation density, canopy or soil moisture and exposed ground conditions are relevant to slope-material strength, infiltration, root reinforcement, erosion and landslide detectability. These indices are temporally variable and therefore represent the selected image-period surface condition rather than an invariant terrain property [13] and [14]. The three indices were computed from Sentinel-2 Level-2A surface reflectance imagery acquired on 25 March 2025. Because Level-2A data are atmospherically corrected by the Sen2Cor processing chain, no additional atmospheric correction was applied. The scene was checked against the Scene Classification Layer, and cloud, cloud-shadow, cirrus and no-data pixels were excluded before index calculation. Bands B8, B4, B2 and B11 were used for NIR, red, blue and SWIR1 reflectance, and the resulting indices were resampled to the common 30 m analytical grid. The scene was selected because it was a cloud-free dry-season acquisition covering the whole watershed with stable illumination and minimal atmospheric disturbance.

2.3 Landslide Inventory and Sample Design

The inventory contains 319 mapped landslide locations. Pre-2023 landslides were used for model calibration and sample extraction. Later landslides mapped on 28 April 2024 and six field-checked landslides from April 2025 were excluded from model calibration and used only for a limited occurrence-only temporal plausibility assessment.

Presence samples were generated by intersecting mapped pre-2023 landslide footprints with the common 30 m grid. Raster cells whose centroids fell within mapped landslide scarps or footprints were retained as landslide-presence cells and duplicate cells were removed. Non-landslide samples were selected outside mapped landslides and outside a three-cell exclusion buffer. The 90 m exclusion distance was chosen as three times the analytical grid size to remove the immediate neighborhood of mapped failures and reduce contamination by marginal landslide cells, mapping-boundary uncertainty and small unmapped failures. This rule is a pragmatic GIS sampling control rather than a geomorphic distance threshold; because a formal buffer-sensitivity analysis was not available for this revision, the effect of alternative buffers is acknowledged as a limitation and a priority for future work.

The use of raster cells increases the sample size but can introduce pseudo-replication because neighboring cells within the same landslide or slope unit may share similar terrain and predictor values. For this reason, the main model comparison was revised from a random cell-level split to a five-fold spatial block validation design. This design better separates training and testing samples spatially and provides a stricter estimate of model transferability within the watershed [15] and [16]. Table 2 summarizes the inventory subsets and their validation roles.

Table 2: Landslide inventory subsets and validation roles

Subset	Time period	Number	Role in study
Calibration inventory	Pre-2023 landslides	274 landslides; 3,714 balanced raster samples in the original table	Used to build the training database and spatial block folds.
Later mapped points	28 April 2024	39	Held out; occurrence-only temporal plausibility assessment.
New field-checked points	April 2025	6	Held out; limited field plausibility assessment.
Total inventory	2010-2025	319	Complete landslide inventory used in the study.

2.4 Conditioning Factors and Index Equations

Sixteen conditioning factors were retained to represent terrain, geological, hydrological, environmental and land-cover controls of slope failure under a scale-consistent GIS framework. Continuous predictors include rainfall, elevation, slope, TRI, TWI, MBI, fault distance, NDVI, NDMI and BSI. Categorical predictors include aspect, lithology, engineering geology, hydrogeology, weathering crust and LULC. The categorical variables were used as nominal classes for modeling and were not interpreted as ordinal numeric variables. The predictor set was restricted to variables that were available across the whole watershed, reproducible at the 30 m grid, and geomorphically relevant to the local landslide mechanism.

Rainfall represents long-term spatial hydro-climatic exposure rather than the trigger of individual slope failures. Lithology, engineering geology and weathering crust describe material strength, weathering degree, regolith development and expected geotechnical behavior. Hydrogeology and TWI represent groundwater-bearing conditions and potential moisture accumulation. Slope, elevation, aspect, TRI and MBI describe gravitational stress, relief position, radiation-related moisture differences, terrain dissection and morphodynamic balance. Fault distance represents proximity to fractured and structurally weakened zones. LULC, NDVI, NDMI and BSI capture vegetation cover, land-surface disturbance and moisture-related surface condition [7] [8] and [14]. Curvature, stream power index, sediment transport index, road distance and drainage density were not added because complete and scale-consistent source layers were not available for the whole watershed or because their information partly overlapped with slope, TRI, TWI and fault-distance patterns at the 30 m analytical scale. This conservative selection avoids increasing dimensionality with unstable or inconsistently mapped predictors and is treated as a data-availability constraint rather than proof that the omitted variables are unimportant.

The engineering-geological, hydrogeological and weathering-crust layers were derived from DGMV source maps and classified by grouping mapped units according to lithological composition, rock fabric, weathering state and expected engineering-geological behavior. The grouping follows the engineering-geological and geological-hazard classification approaches applied by [11] and [12].

In the general scheme, rock units are grouped into seven engineering-geological rock-material categories: (1) aluminosilicate-rich clastic sedimentary and volcaniclastic rocks, (2) quartz-rich clastic sedimentary rocks, (3) carbonate sedimentary rocks, (4) ultramafic and mafic igneous rocks, (5) acidic to intermediate igneous rocks, (6) aluminosilicate-rich metamorphic rocks, and (7) quartz-rich metamorphic rocks. In the Huoi Reng analytical layer, the engineering-geological classes represented in the map are Geo-engineering types 2, 3, 5, 6 and 7; the class numbers are therefore retained as source-map labels rather than re-numbered ordinal ranks.

The hydrogeological classes were classified from the same source-map framework according to groundwater-bearing capacity. Non-aquifer 1 and Non-aquifer 2 indicate units with limited regional groundwater-bearing capacity, whereas Aquifer 2 indicates a groundwater-bearing unit. The weathering-crust layer was aligned with the lithological classification used in the Huoi Reng map so that each weathering-crust code can be interpreted through its corresponding parent lithology. Weathering crust 3 corresponds to carbonate weathering crust developed on light gray limestone and dark gray limestone; Weathering crust 4 corresponds to granitic weathering crust developed on biotite-hornblende granite; Weathering crust 5 corresponds to shale-derived weathering crust developed on shale-dominated units, including shale, siliceous shale and banded shale; and Weathering crust 6 corresponds to weathering crust developed on metamorphosed shale. These categories describe lithology-associated near-surface weathering and regolith conditions, but they should not be read as direct field measurements of weathering thickness, shear strength or permeability.

All three factors were used as nominal categorical predictors in the machine-learning models. Consequently, labels such as Geo-engineering type 2, Non-aquifer 2 and Weathering crust 3 identify source-map classes linked to mapped lithological or hydrogeological units rather than ordinal numerical ranks. They do not imply a monotonic increase or decrease in rock strength, groundwater storage, weathering thickness or landslide susceptibility. Their relevance is interpreted together with lithology, slope, fault proximity, moisture-related indices and field-based geomorphic understanding. The corresponding class interpretations used in the study are summarized in Table 3.

Table 3: Classification of engineering-geological, hydrogeological and weathering-crust classes used as nominal predictors

Predictor layer	Map label in the Huoi Reng layer	Corresponding lithological or hydrogeological interpretation
Engineering geology	Geo-engineering type 2	Quartz-rich clastic sedimentary rocks.
	Geo-engineering type 3	Carbonate sedimentary rocks.
	Geo-engineering type 5	Acidic to intermediate igneous rocks.
	Geo-engineering type 6	Aluminosilicate-rich metamorphic rocks.
	Geo-engineering type 7	Quartz-rich metamorphic rocks.
Hydrogeology	Non-aquifer 1	Non-aquifer or very weakly water-bearing unit according to the hydrogeological source map.
	Non-aquifer 2	Non-aquifer or weakly water-bearing unit according to the hydrogeological source map.
	Aquifer 2	Groundwater-bearing unit according to the hydrogeological source map.
Weathering crust	Weathering crust 3	Light gray limestone and dark gray limestone; carbonate sedimentary rocks.
	Weathering crust 4	Biotite-hornblende granite; granitic intrusive rocks.
	Weathering crust 5	Shale-dominated units, including shale, siliceous shale and banded shale.
	Weathering crust 6	Metamorphosed shale or foliated metasedimentary shale units.

Hydrogeological and weathering-crust codes are inherited from source-map legends. In this revision, the weathering-crust classes are aligned with the parent-lithology classification used in the lithology map: Weathering crust 3 with limestone, Weathering crust 4 with biotite-hornblende granite, Weathering crust 5 with shale-dominated units, and Weathering crust 6 with metamorphosed shale. These classes remain nominal categories rather than ordinal ranks or measured weathering-thickness classes. The spectral and terrain indices were computed using Equations 1 to 5:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

Equation 1

$$NDMI = \frac{NIR - SWIR1}{NIR + SWIR1}$$

Equation 2

$$BSI = \frac{(SWIR1 + RED) - (NIR + BLUE)}{(SWIR1 + RED) + (NIR + BLUE)}$$

Equation 3

$$TWI = \ln \frac{A_s}{\tan \beta}$$

Equation 4

$$TRI = \sqrt{\left(z_i + z_c\right)^2}$$

Equation 5

Where *NIR*, *RED*, *BLUE* and *SWIR1* are the Sentinel-2 surface reflectance bands B8, B4, B2 and B11, respectively; A_s is upslope contributing area per unit contour width; β is local slope angle; z_i is the elevation of each neighboring cell; and z_c is the elevation of the central cell. MBI was generated as a morphodynamic terrain indicator in the GIS workflow and used as a continuous screening predictor. The locally relevant expectation for each factor is summarized in Table 4, and the factor maps are presented in Figures 2 to 4.

2.5 Model Training and Spatial Block Validation

Five Weka classifiers were implemented: ANN, J48, RF, Bagging-RF and SVM [17] [18] [19] and [20]. Bagging-RF denotes a Bagging meta-classifier with an RF base learner; the model is deliberately not labeled as an optimized variant because no formal hyperparameter optimization procedure was applied. The study was intentionally limited to these five model families because they represent the original analytical scope and cover transparent tree learning, ensemble learning, neural-network learning and margin-based learning. A separate generalized linear or logistic baseline was not added in the revised version because it would require a different dummy-coding strategy for nominal geological and land-cover units and would shift the manuscript from a focused comparison of the five selected methods to a broader benchmarking study. The absence of such a baseline is acknowledged as a limitation and the study does not claim universal superiority of any model family. The main validation was revised to five-fold spatial block validation. In this design, samples were grouped into spatial folds before modeling, and each fold was tested once while the remaining folds were used for training. Mean performance and standard deviation were calculated from the five folds. This reduces spatial leakage relative to a random cell-level split and provides a more defensible algorithm comparison for susceptibility mapping [5] and [16].

Table 4: Conditioning factors and locally relevant landslide expectations

No.	Factor	Locally relevant expectation in the Huoi Reng watershed
1	Mean annual rainfall 2010-2023	Higher long-term rainfall exposure increases preconditioning by infiltration and pore-pressure build-up, but it is not an event trigger.
2	Lithology	Weak, weathered or fractured lithological units are expected to show higher susceptibility.
3	Fault distance	Areas closer to mapped faults may contain fractured and weakened rock masses.
4	Weathering crust	Lithology-aligned weathering-crust classes reflect near-surface regolith behavior developed on limestone, biotite-hornblende granite, shale-dominated units and metamorphosed shale rather than a simple ordinal weathering-intensity scale.
5	Engineering geology	Units with poorer geotechnical behavior are expected to be more susceptible.
6	Hydrogeology	Water-bearing or seepage-prone units may promote saturation and instability.
7	TRI	Rugged and dissected terrain is associated with steep gradients and active erosion.
8	MBI	Morphodynamic imbalance may indicate zones of erosion, transport or deposition relevant to failure.
9	Elevation	Relief position influences drainage, rainfall exposure and slope development.
10	Aspect	Aspect affects radiation, vegetation, soil moisture and weathering conditions.
11	Slope	Steeper slopes increase gravitational shear stress.
12	TWI	Higher wetness potential can indicate moisture accumulation and reduced shear strength.
13	LULC 2020	Forest, bare soil, cultivation and disturbed land have different root reinforcement and runoff behavior.
14	BSI	Higher exposed soil values may indicate disturbance, erosion or sparse vegetation.
15	NDVI	Lower vegetation density may reduce root reinforcement and surface protection.
16	NDMI	Surface moisture differences may indicate wetter terrain or disturbed vegetation conditions.

ISSN: 1686-6576 (Printed) | ISSN 2673-0014 (Online) |

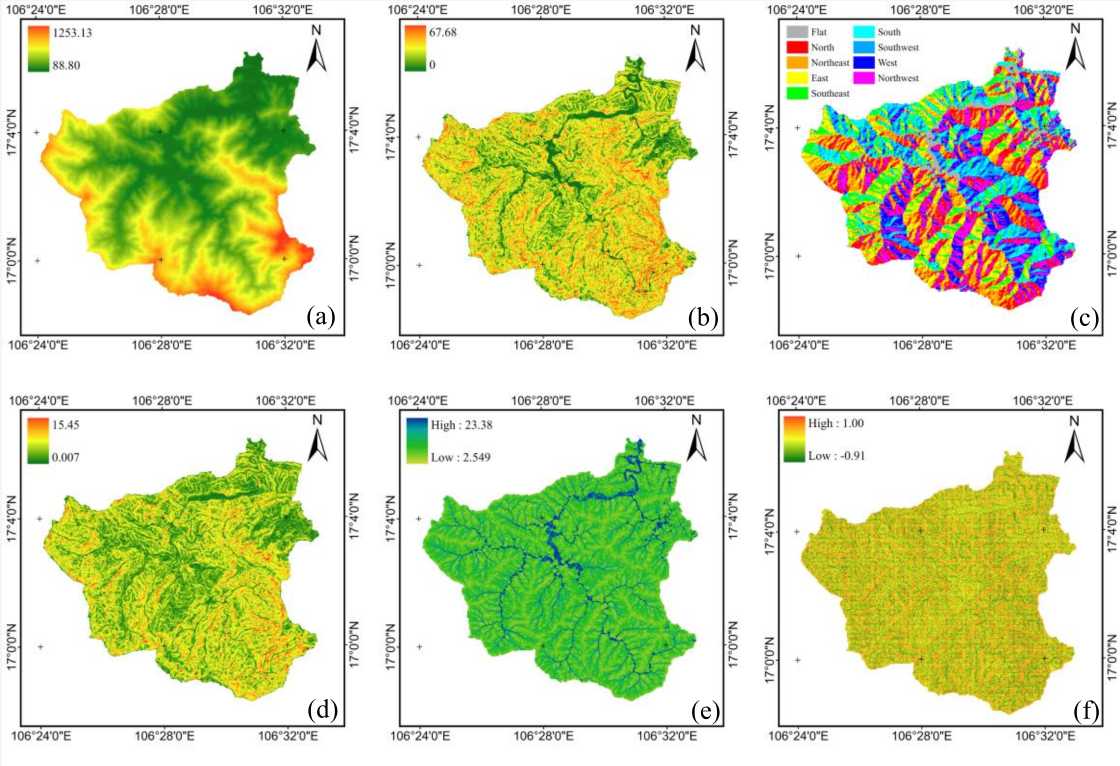

Figure 2: Topographic conditioning-factor maps:
(a) elevation, (b) slope, (c) aspect, (d) TRI, (e) TWI, and (f) MBI

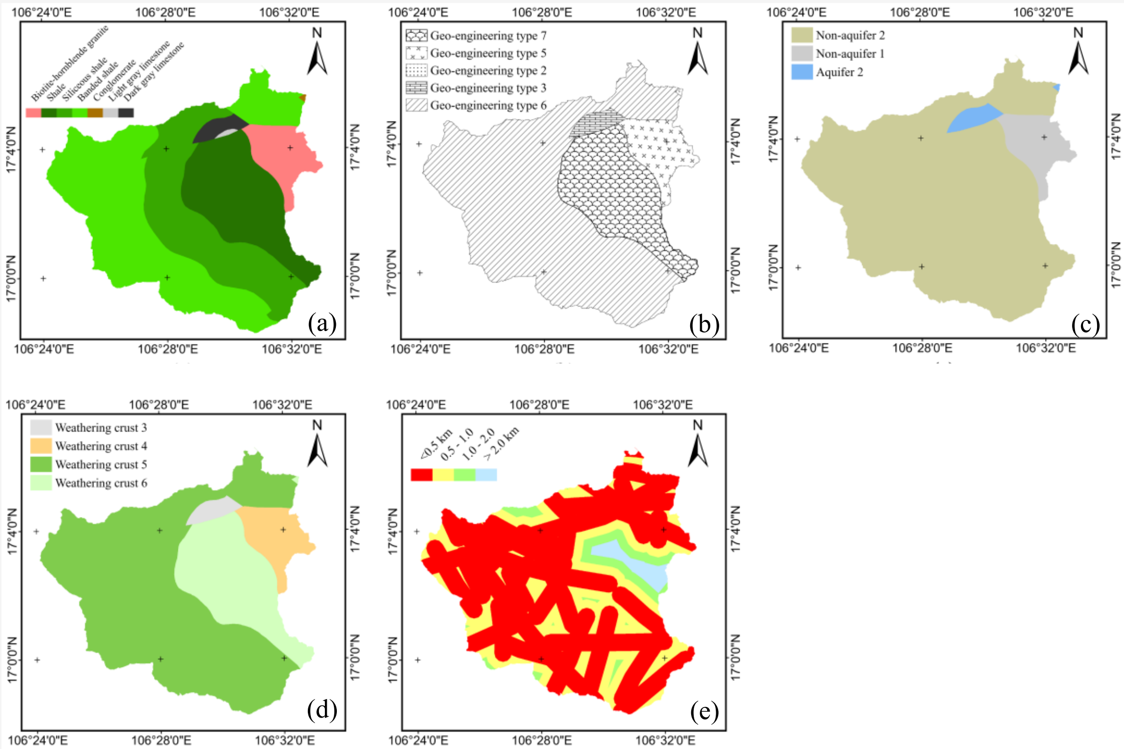

Figure 3: Geological conditioning-factor maps: (a) lithology, (b) engineering geology, (c) hydrogeology, (d) weathering crust, and (e) fault distance

ISSN: 1686-6576 (Printed) | ISSN 2673-0014 (Online) | © Geoinformatics International

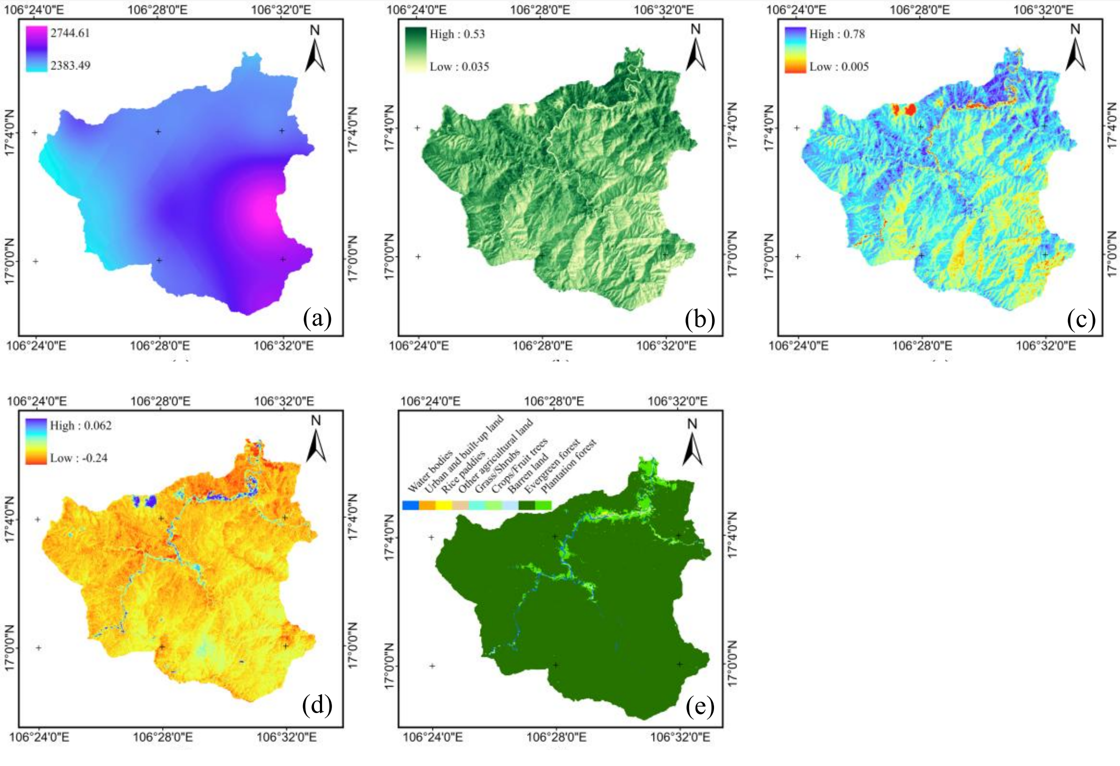

Figure 4: Environmental and land-cover conditioning-factor maps: (a) rainfall, (b) NDVI, (c) NDMI, (d) BSI, and (e) LULC

Classifier comparison used accuracy, Cohen's Kappa, RMSE, ROC-AUC, Brier score, landslide-class recall and landslide-class F1-score. ROC-AUC was used as the main ranking criterion because it evaluates discrimination across thresholds. Recall and F1-score were also reported because false negatives are more critical than false positives in preliminary landslide screening. RMSE and Brier score were used as complementary probability-error metrics derived from the supplied test folds; lower values indicate smaller disagreement between predicted landslide probability and the observed binary class. Brier score was computed for each fold as the mean squared difference between predicted probability and observed class label, and the fold-level mean and standard deviation were then reported. Paired fold-level tests were used to evaluate whether ROC-AUC differences between models were statistically meaningful, but the small number of folds means that p-values should be interpreted cautiously. The classifiers and their main parameters are listed in Table 5. Hyperparameters follow widely used Weka default configurations rather than exhaustive grid-search tuning. RF was retained at 100 trees to represent a reproducible Weka baseline; the ANN hidden layer follows the Weka automatic setting; and the SVM polynomial kernel corresponds to the exported SMO configuration. The comparison therefore represents a default-configuration baseline across model families; the absence of systematic hyperparameter search, RBF-kernel benchmarking and architecture optimization is acknowledged as a limitation in Section 6.

2.6 Predictor Diagnostics

Predictor diagnostics were revised to avoid treating nominal class codes as continuous variables. Variance Inflation Factor (VIF) and tolerance are reported only for continuous variables as a first-order linear redundancy screen [21] and [22]. Categorical predictors are modeled as nominal classes and should be evaluated using class-level association measures such as Cramer's V if the full attribute table is available. Information Gain, Gain Ratio, OneR and correlation-based evaluator outputs are interpreted as univariate diagnostics, not as automatic variable deletion rules [23] and [24]. Zero Information Gain for TRI, MBI or hydrogeology means that these variables have limited individual separability under the selected evaluator; it does not prove that they are causally irrelevant or useless through interactions. Permutation importance and SHAP-based diagnostics were not produced because the revised analysis was constrained to the retained Weka outputs and archived fold summaries. The discussion therefore avoids causal overstatement from ranking results and treats multivariate importance analysis as future work.

3. Results

3.1 Spatial Block Model Performance

Table 6 shows that spatially blocked validation produced lower and more conservative performance than the earlier cell-level random split. Bagging-RF had the highest mean ROC-AUC (0.807 ± 0.063), followed very closely by RF (0.805 ± 0.063). The same two ensemble models also had the lowest RMSE and Brier score values, with Bagging-RF showing the smallest probability error (RMSE = 0.428 ± 0.031; Brier score = 0.184 ± 0.027) and RF giving a nearly equivalent result (RMSE = 0.431 ± 0.033; Brier score = 0.187 ± 0.028). ANN and J48 produced intermediate probability errors, whereas SVM had the highest Brier score (0.288 ± 0.055) and RMSE (0.535 ± 0.051). The SVM model had relatively high landslide recall (0.790 ± 0.088), but its lower ROC-AUC and larger probability-error metrics indicate a tendency toward more liberal landslide classification rather than better overall discrimination or probability reliability.

Fold-level paired testing indicated no significant ROC-AUC difference between Bagging-RF and RF (mean difference = 0.001; p = 0.311). Bagging-RF and RF had higher ROC-AUC than ANN, J48 and SVM in the fold-level paired comparisons (Tables 7 and 8), but these tests are interpreted cautiously because only five folds were available. The result supports treating Bagging-RF and RF as statistically comparable tree-ensemble alternatives rather than claiming strong superiority of one over the other. Figure 5 summarizes discrimination performance and probability-error behavior, separating ROC-AUC variability (Figure 5(a)) from RMSE and Brier score variability (Figure 5(b)).

Table 5: Weka classifiers and main parameters used for model training

Model	Weka classifier	Main parameters
ANN	weka.classifiers.functions.MultilayerPerceptron	Learning rate = 0.3; momentum = 0.2; epochs = 500; one hidden layer with nine nodes, following the Weka automatic setting a = (number of attributes + number of classes)/2 = (16 + 2)/2.
J48	weka.classifiers.trees.J48	C = 0.25; M = 2; pruned tree; used as transparent benchmark.
RF	weka.classifiers.trees.RandomForest	100 trees; bag size = 100%; K = 0; min leaf = 1.0; seed = 1.
Bagging-RF	weka.classifiers.meta.Bagging with RF base learner	10 outer bagging iterations with an RF base learner (100 trees per iteration); no additional hyperparameter optimization applied.
SVM	weka.classifiers.functions.SMO	C = 1.0; tolerance = 0.001; polynomial kernel (exported Weka SMO configuration); model scores were used only for comparative validation metrics.

Table 6: Five-fold spatial block validation performance of the machine learning models

Model	Accuracy (%) mean±SD	ROC-AUC mean±SD	Kappa mean±SD	Recall landslide mean±SD	F1 landslide mean±SD	RMSE mean±SD	Brier score mean±SD	Rank
Bagging-RF	72.21 ± 5.35	0.807 ± 0.063	0.443 ± 0.108	0.684 ± 0.114	0.706 ± 0.070	0.428 ± 0.031	0.184 ± 0.027	1
RF	71.62 ± 5.29	0.805 ± 0.063	0.431 ± 0.107	0.659 ± 0.108	0.694 ± 0.069	0.431 ± 0.033	0.187 ± 0.028	2
ANN	67.53 ± 5.13	0.745 ± 0.065	0.351 ± 0.103	0.685 ± 0.143	0.672 ± 0.065	0.487 ± 0.044	0.239 ± 0.042	3
SVM	71.19 ± 5.47	0.713 ± 0.055	0.424 ± 0.110	0.790 ± 0.088	0.730 ± 0.055	0.535 ± 0.051	0.288 ± 0.055	4
J48	68.85 ± 4.68	0.712 ± 0.042	0.377 ± 0.094	0.696 ± 0.091	0.688 ± 0.052	0.504 ± 0.036	0.255 ± 0.035	5

ISSN: 1686-6576 (Printed) | ISSN 2673-0014 (Online) |

Table 7: Fold-level paired comparisons based on ROC-AUC

Comparison	Metric	Mean difference	p-value	Interpretation
Bagging-RF vs RF	ROC-AUC	0.001	0.311	Not significant
Bagging-RF vs ANN	ROC-AUC	0.062	0.001	Significant by paired t-test; small fold count
Bagging-RF vs J48	ROC-AUC	0.094	0.001	Significant by paired t-test; small fold count
Bagging-RF vs SVM	ROC-AUC	0.094	<0.001	Significant by paired t-test; small fold count

Table 8: Fold-level validation results from the five-fold spatial block validation

Model	Fold	Instances	Accuracy (%)	ROC Area	Kappa	Recall Yes	F1 Yes	RMSE
ANN	1	744	63.71	0.676	0.275	0.582	0.619	0.516
	2	715	74.69	0.839	0.497	0.897	0.776	0.416
	3	763	61.99	0.695	0.242	0.668	0.628	0.532
	4	716	70.39	0.770	0.406	0.534	0.641	0.485
	5	776	66.88	0.744	0.336	0.746	0.696	0.486
Bagging-RF	1	744	67.20	0.738	0.345	0.609	0.652	0.459
	2	715	80.28	0.882	0.606	0.840	0.807	0.384
	3	763	67.50	0.754	0.347	0.619	0.647	0.458
	4	716	72.35	0.857	0.445	0.582	0.675	0.414
	5	776	73.71	0.802	0.474	0.769	0.748	0.427
J48	1	744	63.44	0.664	0.268	0.662	0.647	0.533
	2	715	74.55	0.752	0.493	0.854	0.767	0.450
	3	763	66.84	0.685	0.336	0.649	0.653	0.526
	4	716	72.91	0.761	0.458	0.689	0.716	0.485
	5	776	66.50	0.700	0.331	0.627	0.655	0.527
RF	1	744	66.53	0.735	0.332	0.582	0.638	0.463
	2	715	79.30	0.881	0.586	0.814	0.794	0.384
	3	763	66.71	0.752	0.330	0.591	0.631	0.461
	4	716	72.35	0.853	0.445	0.576	0.673	0.422
	5	776	73.20	0.805	0.464	0.731	0.735	0.427
SVM	1	744	66.67	0.666	0.332	0.737	0.691	0.577
	2	715	75.94	0.763	0.522	0.926	0.790	0.490
	3	763	64.61	0.648	0.295	0.700	0.656	0.595
	4	716	76.95	0.770	0.539	0.766	0.767	0.480
	5	776	71.78	0.716	0.434	0.822	0.747	0.531

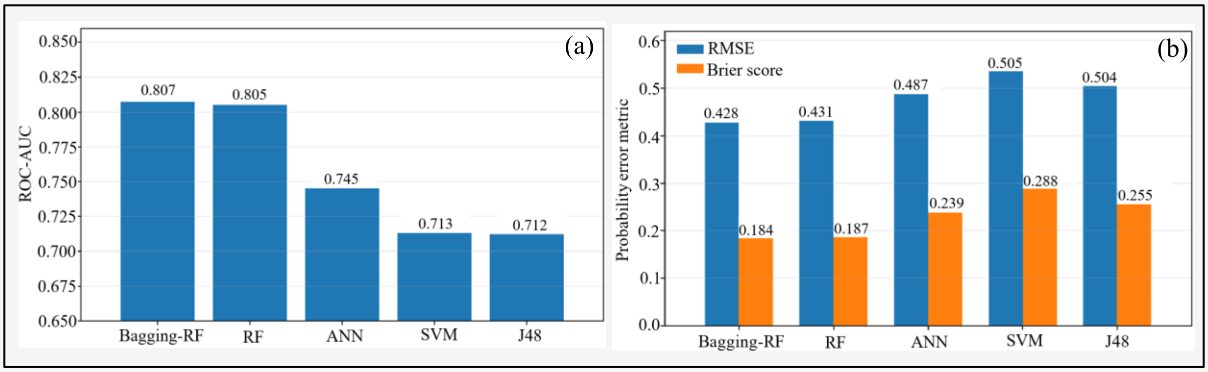

Figure 5: Spatial-block model performance summary:
(a) ROC-AUC and (b) probability-error metrics based on RMSE and Brier score

3.2 Classification Errors and Threshold Interpretation

Confusion matrices are reported at the default 0.5 probability threshold to allow direct comparison with the Weka outputs. These matrices should be interpreted together with ROC-AUC, RMSE and Brier score because a single threshold may advantage models with better probability separation or different score distributions. In landslide susceptibility screening, false negatives are generally more costly than false positives because they indicate landslide locations that would not be prioritized for field inspection. However, excessive false positives can also reduce practical usefulness by expanding the area requiring field checks beyond local capacity. Accordingly, the final operational threshold should depend on management purpose. A conservative threshold may be appropriate for disaster-preparedness screening where missing unstable slopes is unacceptable. A more restrictive threshold may be appropriate for limited field resources, where the objective is to identify the highest-priority slopes for inspection. The present manuscript therefore reports default-threshold matrices but avoids presenting them as final operational thresholds. Figure 6 summarizes the fold-aggregated confusion matrices used to support this threshold-sensitive interpretation.

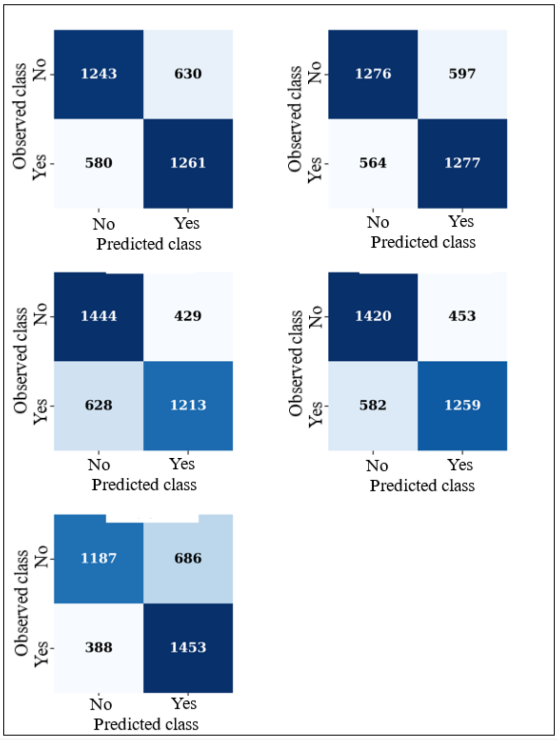

Figure 6: Fold-aggregated confusion matrices at the default 0.5 probability threshold: (a) ANN, (b) J48, (c) RF, (d) Bagging-RF, and (e) SVM

3.3 Predictor Ranking and Redundancy

Continuous-variable VIF results did not indicate severe linear redundancy among the continuous predictors used in the modeling table. The revised VIF values range from 1.000 to 1.194, which is well below commonly used concern thresholds [21]. VIF is not reported for nominal class-coded variables because numeric codes for lithology, aspect, LULC, engineering geology, hydrogeology and weathering crust do not represent meaningful quantitative distances.

The univariate ranking results indicate that mean annual rainfall, lithology, elevation, engineering geology, weathering crust, fault distance and vegetation/moisture indices provide measurable individual separability between landslide and non-landslide samples. Different ranking evaluators can produce different orderings, so the results are treated as complementary diagnostics [25] and [26]. Information Gain-based rankings have also been combined with classifiers such as SVM in previous landslide studies [27]. These rankings are interpreted as associations within the sample design rather than causal evidence, because spatial correlation among predictors and between predictors and landslide samples may inflate univariate importance values. Table 9 reports the revised predictor-ranking statistics and VIF diagnostics, and Figure 7 visualizes the ranking. VIF values were calculated only for continuous predictors; categorical predictors were excluded from VIF computation to avoid interpreting nominal class codes as ordered numerical variables.

3.4 Susceptibility Maps and Class Distributions

The predicted susceptibility maps should be interpreted as relative local screening products. High and very high susceptibility zones occur where unfavorable terrain, weathered materials, structural proximity and land-surface conditions coincide. Map comparison should therefore focus not only on visual coherence but also on class-area distribution, agreement between models and the ability to concentrate later landslide occurrences in smaller high-priority areas. Area distribution differs among models because each classifier transforms predictor space into probability or class scores differently. Tree ensembles tend to produce spatially coherent probability gradients by averaging many decision trees. J48 produces sharper rule-based boundaries. ANN and SVM may produce broader or more compressed high-score regions depending on parameter settings, calibration and threshold behavior. These differences mean that class-area proportions reflect both model structure and the probability-to-class reclassification method.

ISSN: 1686-6576 (Printed) | ISSN 2673-0014 (Online) |

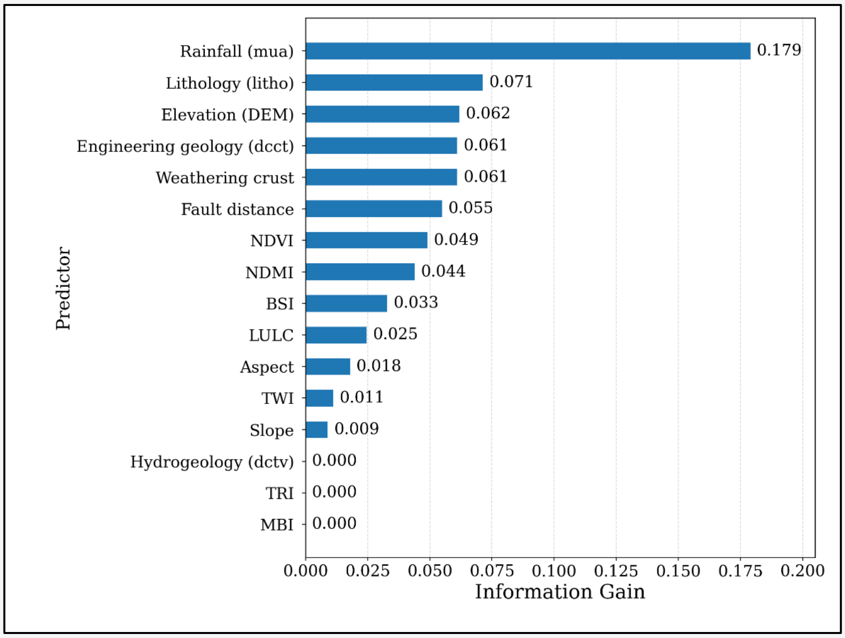

Figure 7: Predictor ranking based on the revised diagnostic procedure

Table 10: Occurrence-only class-concentration metrics for the 45 later landslide points

Model	High + very high area (%)	Later landslide signal in high + very high (%)	Concentration ratio	Highest FR class	Highest FR
RF	9.89	61.85	6.26	4	7.218
J48	10.66	60.33	5.66	5	6.461
Bagging-RF	12.39	58.05	4.69	5	5.122
ANN	15.88	57.92	3.65	5	4.994
SVM	0.59	18.63	31.52	4	32.044

The continuous susceptibility outputs were reclassified into five relative classes (very low, low, medium, high and very high) using the same GIS reclassification workflow for all five models. The revised text treats the class boundaries as cartographic and management-screening thresholds rather than absolute probability thresholds. If the final production maps are regenerated, the exact GIS rule used for classing (for example, Natural Breaks/Jenks, equal interval or quantile) should be reported together with the class break values to make the map classification fully reproducible. Figure 8 presents the spatial patterns of the five model outputs using a common susceptibility legend and the same later landslide points, enabling visual comparison of map coherence and occurrence-only plausibility across models.

3.5 Occurrence-Only Class-Concentration Assessment Using 45 Later Points

The 39 landslides mapped in 2024 and six field-checked landslides from 2025 were used only as an occurrence-only class-concentration assessment. They cannot support a complete confusion matrix, accuracy or Kappa because independently sampled stable locations were not collected. The assessment therefore examines whether later landslide occurrences are concentrated in the high and very high susceptibility classes more strongly than expected from the mapped area share. Table 10 summarizes the class-concentration metrics derived from the 45-point comparison workbook. The comparison is based on the areal extent of the high and very high susceptibility classes, the percentage of later landslide signal falling within those classes, the concentration ratio, and the highest frequency-ratio class. RF concentrated 61.85% of the later landslide signal within 9.89% of the watershed mapped as high or very high susceptibility.

ISSN: 1686-6576 (Printed) | ISSN 2673-0014 (Online) | © Geoinformatics International

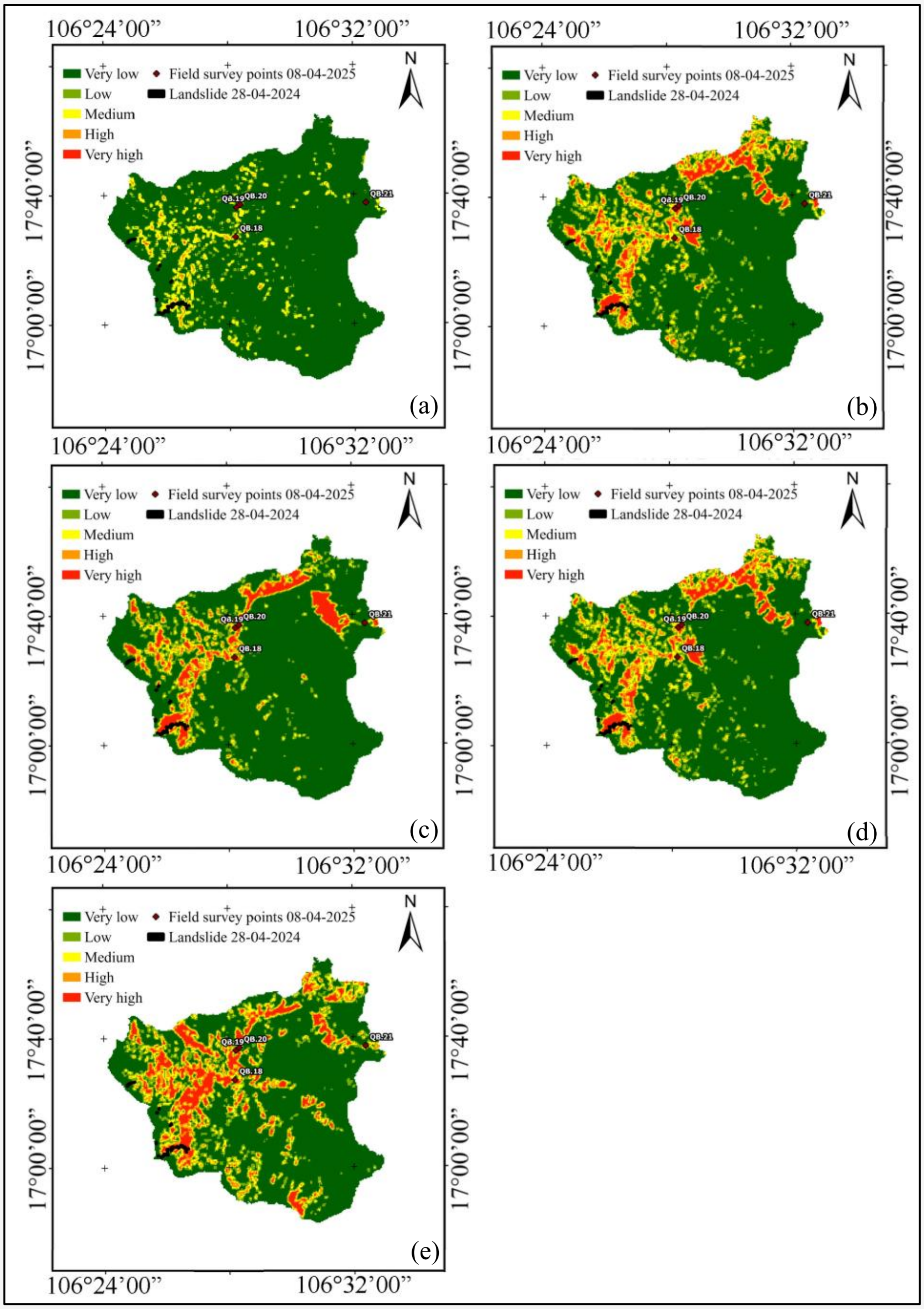

Figure 8: Predicted landslide susceptibility maps from:
(a) ANN, (b) J48, (c) RF, (d) Bagging-RF, and (e) SVM

J48 and Bagging-RF also concentrated more than 58% of the later landslide signal within about 10-12% of the area. SVM produced a very small high and very high area share (0.59%) and captured only 18.63% of the later landslide signal, which makes it less useful for broad screening despite a high frequency ratio in a very small class. Table 10 is retained to provide the exact numerical comparison, while Figure 9 is retained only as a visual summary of the same occurrence-only pattern; either item may be moved to supplementary material if required by the editor. Figure 10 provides a compact summary of the discrimination-probability error trade-off. Bagging-RF and RF occupy the most favorable region because they combine high ROC-AUC with low Brier score, whereas SVM shows relatively high probability error despite high landslide recall at the default threshold.

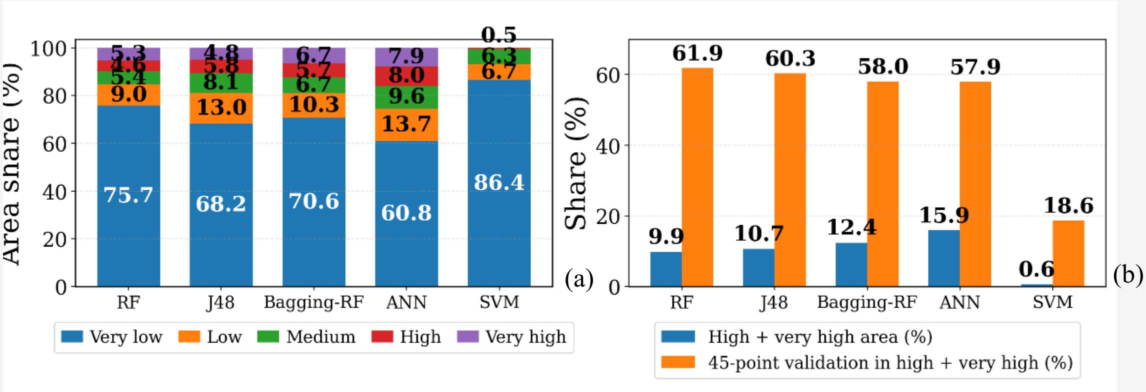

Figure 9: (a) Susceptibility-class area distribution and (b) occurrence-only class-concentration assessment for the 45 later landslide points

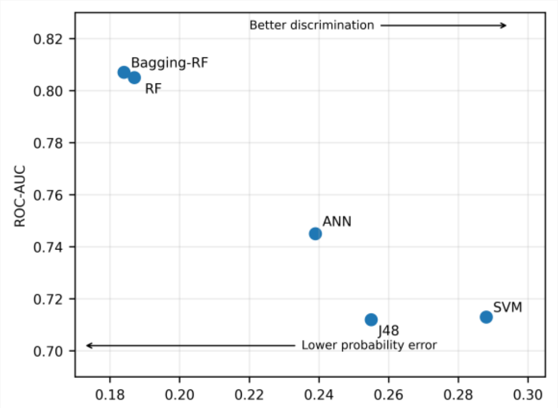

Figure 10: Model-level trade-off between discrimination and probability error under five-fold spatial block validation

4. Discussion

4.1 Effect of Spatial Validation on Model Interpretation

The spatial block results show that the earlier random cell-level split should not be treated as a definitive estimate of spatial predictive capability. Random splitting can place neighboring cells from the same landslide or slope unit in both training and testing subsets, which artificially reduces the apparent generalization gap. The revised analysis therefore uses the five-fold spatial block results as the main performance evidence and treats the random split, if mentioned, only as an internal diagnostic. The moderate ROC-AUC values under spatial blocking are useful for local screening but not sufficient for operational forecasting or quantitative risk estimation, and they are within the range reported by GIS-based comparative landslide susceptibility studies in other mountainous regions [28] and [29]. This is expected in a small mountainous catchment where landslide samples, absence samples, predictor scales and rainfall representation all introduce uncertainty. The main interpretation is therefore not that one model is universally superior, but that tree-based ensemble models provide the most stable local screening behavior under the available data.

4.2 Model Behavior in the Huoi Reng Watershed

Bagging-RF and RF performed similarly because both rely on bootstrap aggregation and random feature selection. The additional outer bagging layer in Bagging-RF provides only marginal improvement over the standard RF base learner and may be structurally redundant. This explains why the two models have nearly identical ROC-AUC under spatial block validation and why strong claims of RF superiority should be avoided without additional independent testing. Similar sensitivity of model ranking to algorithm family and study design has been reported in other comparative susceptibility analyses [30].

J48 is less discriminative than the ensemble models but remains useful because it provides interpretable decision rules. This interpretability-performance trade-off is important for local authorities who may need transparent screening logic for field inspection. ANN and SVM underperformed in ROC-AUC, RMSE or Brier score likely because the available sample size, mixed continuous-categorical predictors and limited parameter tuning did not fully support stable nonlinear optimization. The SVM configuration followed the exported Weka SMO setting with a polynomial kernel and logistic probability calibration. This choice was retained to keep the comparison consistent with the original five-model workflow, not because it is assumed to be optimal. SVM also showed threshold-sensitive behavior, which can increase false positives or create overly compressed high-susceptibility zones depending on the probability classification. More complex deep-learning architectures have been proposed for susceptibility mapping [31], but their additional data and tuning requirements were not justified for this local watershed.

4.3 Rainfall, Remote-Sensing Timing and Static Susceptibility

Mean annual rainfall for 2010-2023 is useful as a long-term hydro-climatic conditioning layer because it captures persistent spatial differences in rainfall exposure across the watershed. This use of a long-term rainfall layer as a conditioning factor is consistent with previous susceptibility studies in Vietnam [4]. However, it cannot represent the high-intensity short-duration rainfall events that trigger individual landslides [2]. The susceptibility maps should therefore be interpreted as static predisposing-condition maps, not as time-dependent warning products.

NDVI, NDMI and BSI also depend on image acquisition season, cloud conditions and land-surface state, even though vegetation and land-surface indices have shown clear explicative value for landslides in comparable tropical and mountainous contexts [32]. The 25 March 2025 Sentinel-2 scene was selected because it is a cloud-free dry-season acquisition that provides stable illumination and minimal atmospheric disturbance across the whole watershed. Because the scene postdates most of the calibration inventory, the indices describe the recent land-surface condition rather than the pre-failure condition of individual landslides, and a different image period could alter vegetation, moisture and bare-soil patterns and therefore change model outputs. This temporal-representativeness issue is treated explicitly as a limitation in Section 6.

4.4 Practical Implications for Local Management

For the Huoi Reng watershed, the maps are most appropriate for tiered field-priority zoning. High and very high susceptibility zones should be prioritized for road-slope inspection, drainage maintenance, forest-road planning and detailed geotechnical checks. Moderate zones should be treated as conditional buffer areas where new slope cuts, deforestation or drainage alteration may increase instability. Low and very low zones should not be interpreted as permanently safe because extreme rainfall or local engineering disturbance can still trigger failures. Table 11 translates the susceptibility classes into practical screening actions. The proposed actions are indicative and should be combined with field verification and local management capacity before operational use.

Because field resources are limited, the RF and Bagging-RF maps can be used to allocate inspection effort in stages. First-priority checks should focus on high and very high zones intersecting roads, settlements, drainage lines and previously mapped landslide clusters. Second-priority checks should cover moderate zones adjacent to high-susceptibility belts. This approach reduces inspection area while retaining a high proportion of observed later landslide occurrences. The maps should be integrated with local knowledge from district disaster-management staff, forest managers and road-maintenance units before operational use. Stakeholder consultation is necessary to define acceptable false-negative and false-positive costs, inspection capacity and map classes that can be translated into local planning actions. Where detailed inventories or modeling capacity are unavailable, knowledge-driven zonation approaches remain common [33], but the data-driven maps produced here provide a stronger evidence base for staged inspection in this watershed.

Table 11: Practical interpretation of susceptibility classes for local field-priority screening

Susceptibility class	Interpretation for local management	Recommended use
Very low and low	Areas with comparatively lower modeled susceptibility under the selected predictor set. These zones are not permanently safe under extreme rainfall or local disturbance.	Routine monitoring; avoid using these classes as proof of safety for new slope cuts or infrastructure works.
Medium	Transition zones where susceptibility may increase if drainage, vegetation cover or slope geometry is modified.	Use as buffer zones; require site checking before road widening, deforestation or new drainage alteration.
High	Priority zones where unfavorable terrain, material conditions, structural proximity and surface disturbance coincide.	Prioritize road-slope inspection, drainage maintenance, erosion control and detailed geotechnical checks.
Very high	Critical screening zones with the strongest modeled susceptibility and/or concentration of later landslide evidence.	Avoid new exposure where possible; require urgent field verification and engineering assessment before intervention.

5. Conclusion

This revised study compares five machine learning approaches for local-scale landslide susceptibility mapping in the Huoi Reng watershed using a stricter five-fold spatial block validation framework. The revised design reduces the optimistic interpretation associated with random cell-level splitting and provides a more defensible estimate of local spatial transferability within the watershed. Tree-based ensemble models, especially Bagging-RF and RF, provided the most stable discrimination and the smallest probability errors under spatially separated folds, although their ROC-AUC difference was not statistically significant based on five folds. The 45 later landslide points were used only as an occurrence-only class-concentration check. RF concentrated 61.85% of the later landslide signal within 9.89% of the watershed mapped as high or very high susceptibility, indicating useful screening potential but not independent validation because stable sites were not sampled. The maps should be used as moderate-confidence, field-priority screening products for road-slope inspection, drainage-maintenance planning and identification of areas requiring local verification. Their main operational value lies in reducing the area requiring immediate field inspection while retaining a large proportion of later observed landslide occurrences. They should not be interpreted as quantitative risk maps or operational forecasts until stable-site validation, probability calibration, systematic tuning, event-based rainfall analysis and more complete geological-class documentation are available.

6. Limitations and Future Work

The revised interpretation is constrained by several unresolved data and modeling limitations. First, spatial autocorrelation remains important even after five-fold spatial block validation because fold design, block size and raster-cell sampling can influence performance. Second, the 45 later landslide points are occurrence-only data and cannot establish specificity, accuracy or Kappa without independently sampled stable sites. Third, rainfall and spectral indices are static or period-based representations that cannot capture event-scale triggering processes, and the 25 March 2025 Sentinel-2 scene postdates most of the calibration inventory. Fourth, mean terrain statistics, formation-area proportions and watershed-specific economic-loss data were not available from the archived revision materials; consequently, the study uses range-based geomorphic descriptors and frames management value as field-priority screening rather than economic risk assessment. Fifth, the three-cell (90 m) absence-sample exclusion buffer was selected as a pragmatic GIS control but was not tested through a buffer-sensitivity experiment. Sixth, DGMV geological, hydrogeological and weathering-crust classes were used as nominal map units.

The weathering-crust categories were clarified as lithology-associated classes corresponding to limestone or carbonate rocks, granite, shale and metamorphosed shale, but independent field measurements of weathering thickness, weathering grade, shear-strength parameters and permeability were not available for every class. Seventh, hyperparameters follow default Weka configurations without systematic grid search, ANN architecture optimization or SVM kernel benchmarking. Eighth, Brier score provides fold-level probability-error information, but full probability calibration curves and reliability diagrams could not be produced because retained per-instance probability outputs were not available. Ninth, SHAP and permutation-importance diagnostics were not produced, so predictor ranking remains a univariate screening exercise rather than a full multivariate interpretation of model behavior. Finally, the susceptibility-class maps depend on the probability-to-class reclassification workflow; exact class breaks should be reported whenever the maps are regenerated for publication.

Future work should implement spatial blocking before attribute extraction, test several block sizes and absence-buffer distances, maintain landslide-object separation where polygons are available, collect probability-sampled stable validation sites, retain per-sample probability outputs for reliability diagrams and calibration plots, run systematic hyperparameter tuning under a common validation design, compare polynomial and RBF SVM kernels, test a baseline model under the same nominal-variable encoding scheme, add permutation or SHAP-based model interpretation, compile the full DGMV legend for geological class definitions, compute audited mean terrain statistics, and separate landslide inventories by triggering event when event rainfall data become available. These steps are required before the maps can be used for quantitative risk assessment or operational early warning.

References

[1] National Steering Committee for Natural Disaster Prevention and Control, (2020). *Damages Caused by Natural Disasters from the Beginning of 2020 and in October 2020.* Natural Disaster Prevention and Control. [Online]. Available: https://phongchongthientai.mard.gov.vn/. [Accessed: Apr. 28, 2026].

[2] Luu, L. N., Scussolini, P., Kew, S., Philip, S., Vautard, R., van Oldenborgh, G. J., Otto, F. E. L., Hauser, M., Singh, R. and Ha, L. M., (2021). Attribution of Typhoon-Induced Torrential Precipitation in Central Vietnam, October 2020. *Climatic Change*, Vol. 169. https://doi.org/10.1007/s10584-021-03261-3.

[3] Pradhan, B., (2013). A Comparative Study on the Predictive Ability of the Decision Tree, Support Vector Machine and Neuro-Fuzzy Models in Landslide Susceptibility Mapping Using GIS. *Computers & Geosciences,* Vol. 51; 350-365. https://doi.org/10.1016/j.cageo.2012.08.023.

[4] Bui, D. T., Pradhan, B., Lofman, O., Revhaug, I. and Nguyen, D. T., (2012). Landslide Susceptibility Mapping at Hoa Binh Province (Vietnam) Using an Adaptive Neuro-Fuzzy Inference System and GIS. *Computers & Geosciences,* Vol. 45; 199-211. https://doi.org/10.1016/j.cageo.2011.10.031.

[5] Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schroeder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F. and Dormann, C. F., (2017). Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure. *Ecography,* Vol. 40; 913-929. https://doi.org/10.1111/ecog.02881.

[6] Steger, S., Brenning, A., Bell, R., Petschko, H. and Glade, T., (2016). Exploring Discrepancies between Quantitative Validation Results and the Geomorphic Plausibility of Statistical Landslide Susceptibility Maps. *Geomorphology,* Vol. 262; 8-23. https://doi.org/10.1016/j.geomorph.2016.03.015.

[7] Pawluszek, K. and Borkowski, A., (2017). Impact of DEM-Derived Factors and Analytical Hierarchy Process on Landslide Susceptibility Mapping in the Region of Roznow Lake, Poland. *Natural Hazards,* Vol. 86; 919-952. https://doi.org/10.1007/s11069-016-2725-y.

[8] Roy, J., Saha, S., Arabameri, A., Blaschke, T. and Pham, B. T., (2019). A Novel Ensemble Approach for Landslide Susceptibility Mapping in Darjeeling and Kalimpong Districts, West Bengal, India. *Remote Sensing,* Vol. 11. https://doi.org/10.3390/rs11232866.

[9] Nguyen, D. D., Le, H. V. and Pham, B. T., (2023). Landslide Susceptibility Modeling and Mapping at Dien Bien Province, Vietnam Using Bagging Based MLP Neural Network. IOP Conference Series: Materials Science and Engineering, Vol. 1289. https://doi.org/10.1088/1757-899X/1289/1/012020.

[10] Tran, T. T., Bui, V. C., Nguyen, M. D. and Pham, V. D., (2024). Integrating Remote Sensing, GIS and Machine Learning Approaches in Evaluation of Landslide Susceptibility in Mountainous Region of Nghe An Province, Vietnam. *IOP Conference Series: Earth and Environmental Science,* Vol. 1345. https://doi.org/10.1088/1755-1315/1345/1/012008.

[11] Trinh, X. H., (2017). *Report on Investigation and Mapping of Landslide Inventory at a Scale of 1:50,000 in the Mountainous Areas of Quang Binh Province.* Institute of Geological Sciences and Mineral Resources, Ministry of Natural Resources and Environment, Hanoi, Vietnam. (in Vietnamese).

[12] Tran, T. V., (2002). *Assessment of Geological Hazards in the Coastal Provinces from Quang Binh to Phu Yen: Current Status, Causes, Forecasting and Proposed Mitigation Measures.* Institute of Geological Sciences and Mineral Resources, Hanoi, Vietnam. (in Vietnamese).

[13] Nguyen, C. T., Chidthaisong, A., Phan, K. D. and Lian, Z. H., (2021). A Modified Bare Soil Index to Identify Bare Land Features during Agricultural Fallow Period in Southeast Asia Using Landsat 8. *Land,* Vol. 10. https://doi.org/10.3390/land10030231.

[14] Ariza, A., Davila, N. A., Kemper, H., Fritz, A., Zhao, L., Wang, W. and Meyer, H., (2021). Landslide Detection in Central America Using the Differential Bare Soil Index. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* Vol. XLIII-B3-2021; 679-684. https://doi.org/10.5194/isprs-archives-XLIII-B3-2021-679-2021.

[15] More, N., Patil, S., Trimbakkar, V., Shrivastav, S., Shah, P., Pasi, A., and Nikam, V. (2026). Evaluating Quantum Machine Learning for GIS-Based Land Suitability Analysis: A Case Study of Reforestation . *International Journal of Geoinformatics*, Vol. 22(4); 134–147. https://doi.org/10.52939/ijg.v22i4.4943.

[16] Brenning, A., (2012). Spatial Cross-Validation and Bootstrap for the Assessment of Prediction Rules in Remote Sensing: The R Package Sperrorest. *IEEE International Geoscience and Remote Sensing Symposium,* 5372-5375. https://doi.org/10.1109/IGARSS.2012.6352393.

[17] Quinlan, J. R., (1996). Improved Use of Continuous Attributes in C4.5. *Journal of Artificial Intelligence Research,* Vol. 4; 77-90. https://doi.org/10.1613/jair.279.

[18] Breiman, L., (2001). Random Forests. *Machine Learning,* Vol. 45; 5-32. https://doi.org/10.1023/A:1010933404324.

[19] Cortes, C. and Vapnik, V., (1995). Support-Vector Networks. *Machine Learning,* Vol. 20; 273-297. https://doi.org/10.1007/BF00994018.

[20] Kadavi, P. R., Lee, C. W. and Lee, S., (2018). Application of Ensemble-Based Machine Learning Models to Landslide Susceptibility Mapping. *Remote Sensing,* Vol. 10. https://doi.org/10.3390/rs10081252.

[21] O'Brien, R. M., (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity,* Vol. 41; 673-690. https://doi.org/10.1007/s11135-006-9018-6.

[22] Shrestha, N., (2020). Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics,* Vol. 8(2); 39-42. https://doi.org/10.12691/ajams-8-2-1.

[23] Ahmad, A. and Dey, L., (2005). A Feature Selection Technique for Classificatory Analysis. *Pattern Recognition Letters,* Vol. 26; 43-56. https://doi.org/10.1016/j.patrec.2004.08.015.

[24] Al Janabi, K. B. S. and Kadhim, R., (2018). Data Reduction Techniques: A Comparative Study for Attribute Selection Methods. *International Journal of Advanced Computer Science and Technology,* Vol. 8(1); 1-13. https://www.ripublication.com/ijacst18/ijacstv8n1_01.pdf.

[25] Thammaboribal, P., Triapthti, N., and Lipiloet, S. (2025). Using of Analytical Hierarchy Process (AHP) in Disaster Management: A Review of Flooding and Landslide Susceptibility Mapping. *International Journal of Geoinformatics*, Vol. 21(4); 177–196. https://doi. org/10.52939/ijg.v21i4.4091..

[26] Sawane, R., Ahmad, M. A., Ansari, F., Kamal, M. and Yadav, A. K., (2022). Comparison of Various Machine Learning Techniques Based on Variable Selection under Imbalanced Data. *International Journal of Engineering Development and Research,* Vol. 10(4); 60-70.

[27] Yu, L., Cao, Y., Zhou, C., Zhang, Y. and Li, W., (2019). Landslide Susceptibility Mapping Combining Information Gain Ratio and Support Vector Machines: A Case Study from Wushan Segment in the Three Gorges Reservoir Area, China. *Applied Sciences,* Vol. 9. https://doi.org/10.3390/app9224756.

[28] Wang, G., Chen, X., Chen, W. and Li, L., (2020). Spatial Prediction of Landslide Susceptibility Based on GIS and Discriminant Functions. *ISPRS International Journal of Geo-Information,* Vol. 9. https://doi.org/10.3390/ijgi9030144.

[29] Zhao, P., Masoumi, Z., Kalantari, M., Rahmati, O. and Pham, B. T., (2022). A GIS-Based Landslide Susceptibility Mapping and Variable Importance Analysis Using Artificial Intelligent Training-Based Methods. *Remote Sensing,* Vol. 14. https://doi.org/10.3390/rs14010211.

[30] Yuan, X., Liu, C., Nie, R., Zhang, X. and Zhang, H., (2022). A Comparative Analysis of Certainty Factor-Based Machine Learning Methods for Collapse and Landslide Susceptibility Mapping in Wenchuan County, China. *Remote Sensing,* Vol. 14. https://doi.org/10.3390/rs14143259.

[31] Ghasemian, B., Aghdam, I., Rahmati, O., Shahabi, H., Pham, T. A. and Tiefenbacher, J., (2022). A Robust Deep-Learning Model for Landslide Susceptibility Mapping: A Case Study of Kurdistan Province, Iran. *Sensors,* Vol. 22. https://doi.org/10.3390/s22041573.

[32] Cobos Mora, S. L., Rodriguez Galiano, V. and Lima, A., (2023). Analysis of Landslide Explicative Factors and Susceptibility Mapping in an Andean Context: The Case of Azuay Province, Ecuador. *Heliyon,* Vol. 9. https://doi.org/10.1016/j.heliyon.2023.e20170.

[33] Liu, X., Shao, S. and Shao, S., (2024). Landslide Susceptibility Zonation Using the Analytical Hierarchy Process in the Great Xian Region, China. *Scientific Reports,* Vol. 14. https://doi.org/10.1038/s41598-024-53630-y.