



ORIGINAL PAPER

MACHINE LEARNING-BASED DEBRIS FLOW SUSCEPTIBILITY AND RISK INDEX MAPPING IN QUANG NAM PROVINCE, VIETNAM

Nguyen-Vu LUAT¹⁾*, Lan Chau NGUYEN²⁾, Tuan-Nghia DO³⁾* and Nguyen Trung KIEN⁴⁾¹⁾ Smart Computing in Civil Engineering Research Group, Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam *Email: nguyenvluat@tdtu.edu.vn²⁾ Faculty of Civil Engineering, University of Transport and Communications, Hanoi, Vietnam³⁾ Faculty of Civil Engineering, Thuyloi University, Hanoi, Vietnam⁴⁾ Institute of Earth Sciences - Vietnam Academy of Science and Technology, No.68 Huynh Thuc Khang street, Lang, Hanoi, Vietnam*Corresponding author's e-mail: dotuannghia@tlu.edu.vn

ARTICLE INFO

Article history:

Received 8 April 2026

Accepted 3 June 2026

Available online 29 June 2026

Keywords:

Debris flow susceptibility mapping

Machine learning

Gradient Boosting;

SHAP

Class imbalance

Quang Nam Province

Vietnam

Risk assessment

ABSTRACT

This study develops an integrated machine learning (ML) framework for debris flow susceptibility and risk index mapping across Quang Nam Province, a region prone to high-intensity monsoonal rainfall and strongly weathered metamorphic bedrock. Four ML algorithms including Gradient Boosting (GB), Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN), were trained on 228 stream-channel samples described by eight causative factors, under two class-imbalance correction strategies. Model performance was assessed through stratified 10-fold cross-validation, and feature importance was quantified using Shapley Additive Explanations (SHAP). Gradient Boosting with cost-sensitive learning (GB-CS) achieved the best overall performance (AUC = 0.9902, F1 = 0.9286, only 4 misclassifications in 228 instances), outperforming the Modified Analytic Hierarchy Process (M-AHP) expert baseline by 0.249 AUC units (33.5 % relative improvement). SHAP analysis identified channel gradient as the dominant predictor (mean normalized importance = 0.442) across GB, RF, and SVM, followed by elevation (0.142) and geology (0.108). The GB-CS susceptibility map correctly classified 92.9 % of inventory sites within the High or Very High susceptibility classes, covering 20 % of the stream network. Integration of susceptibility with a land-use and elevation-weighted vulnerability index revealed that the highest-risk reaches occur not in the most hazardous western highlands but in transitional mid-elevation agricultural corridors - a finding with direct consequences for early-warning system prioritization.

1. INTRODUCTION

Debris flows are rapid, gravity driven mass movements composed of poorly sorted mixtures of water, silt and rock fragments that flow down steep channels at rates typically exceeding several meters per second (Hung et al., 2014; Iverson, 1997). Unlike shallow landslides that generally halt on hillslopes, debris flows are channelized events that can travel long distances away from the origin, collecting additional material during transit and depositing substantial quantities of sediment on alluvial fans, where human settlements and infrastructure are frequently located (Takahashi, 2014). Fast movement, large volumes of sediment accumulation and little warning time make debris flows one of the most severe geo-hazards in mountainous places worldwide, with thousands of casualties and billions of economic losses per year (Dowling et al., 2013).

Debris flow susceptibility mapping (DFSM) uses local geoenvironmental parameters to identify hazardous regions of future debris flow activity (Luna et al., 2014; Smolíková et al., 2021). A susceptibility map shows the spatial probability of debris flow release area and movement in a studied location, providing the crucial hazard layer for risk assessment

and land-use planning (Fell et al., 2008). Over the past two decades, data-driven methodologies have now become the prevailing paradigm in DFSM, gradually replacing qualitative, expert-based approaches (Guzzetti et al., 2012; Merghadi et al., 2020). Since the early 2000s, machine learning algorithms have been extensively utilized and have consistently exhibited enhanced predictive accuracy and spatial generalization compared to traditional statistical and heuristic techniques.

Conventional methods for DFSM can be categorized into heuristic (expert-based) and statistical (data-driven) techniques (Aleotti et al., 1999; Dash et al., 2022). Heuristic approaches such as the Analytic Hierarchy Process (AHP) and its modified versions (M-AHP), assign weights to causal elements via expert assessment and pairwise comparison matrices (Myronidis et al., 2016). The heuristic approaches are widely used because of their conceptual simplicity and small data requirements, but they are limited by intrinsic subjectivity in weight allocation, limited reproducibility and lack of consideration of nonlinear interactions among causal factors (Pourghasemi et al., 2013). Some limitations are overcome by statistical methods such as frequency ratio (FR), information

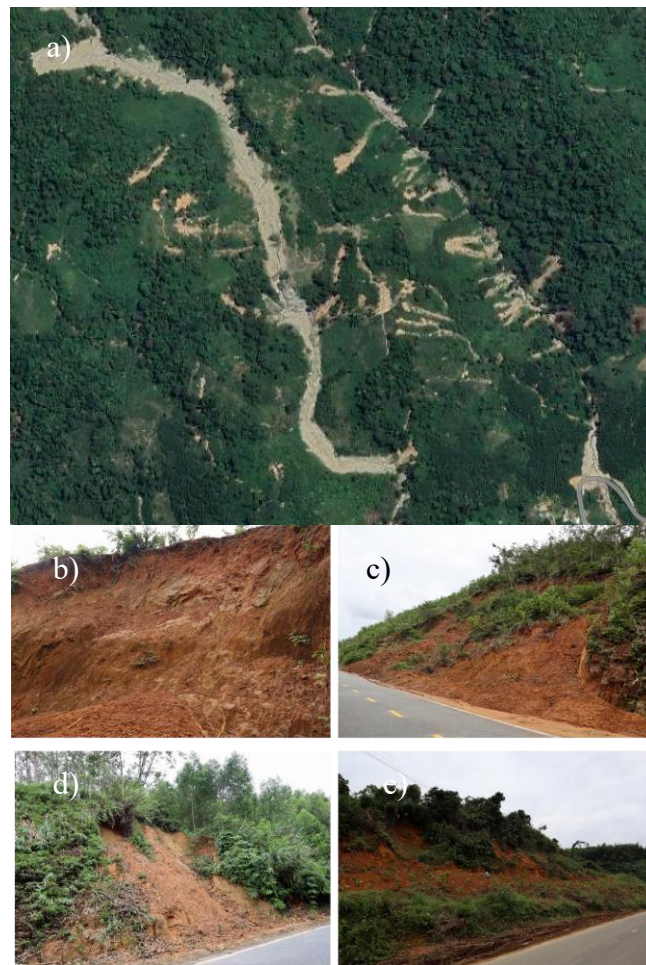


Fig. 1 Field photographs of the October 2020 Tra Leng debris flow event at Nam Tra My District, Quang Nam Province, Vietnam. (a) Google Earth Pro view showing debris-flow scars and a depositional fan. (b– e) Ground-level photographs of slope failures and road damage. The event, triggered by Typhoon Molave, resulted in 22 casualties and displaced hundreds of residents.

value (IV), and certainty factor (CF) that calculate the weights from spatial connection between debris flow inventories and causative factors (Dash et al., 2022). These bivariate statistical methods, however, presume conditional independence across variables and are insufficient for modeling higher-order interactions without explicit variable formation (Yalcin et al., 2011).

Machine learning (ML) has recently become the main approach of DFSM with better prediction accuracy and spatial generalization ability than the traditional statistical methods (Bishop et al., 2006; Merghadi et al., 2020). Multiple classifier families have been explored, such as random forest (RF), support vector machine (SVM), extreme gradient boosting (XGBoost) and boosted regression trees (BRT) in geologically diverse settings in China, the Indian Himalayas and Central Asia (Dash et al., 2022; Qing et al., 2020; Xiong et al., 2020). Ensemble approaches always lead to the highest AUCs, often over 0.85, while frequency-ratio based statistical models are well below that (Dash et al., 2022). More recently, hybrid frameworks that couple numerical

simulation and gradient-boosted trees have showed that ML surrogates can replicate hydraulic models at approximately 10 times the speed of computation (Gao et al., 2025).

The extreme class imbalance is a constant methodological challenge in ML-based DFSM, because debris flow occurrence sites are just a small fraction of any study area. Two remedial approaches have been offered, synthetic over-sampling by SMOTE (Chawla et al., 2002) and cost-sensitive learning (Elkan, 2001). Both have been applied independently in the geohazard literature but a direct comparison assessment on the same DFSM dataset is still lacking.

Despite these advances, three critical gaps persist. First, ML-based DFSM has not been applied to Southeast Asia's tropical monsoon highland environments, including the debris-flow-prone western highlands of Quang Nam Province, Vietnam, where experienced a catastrophic event in October 2020 with 22 fatalities (shown in Fig. 1). Second, no study has systematically compared SMOTE-based oversampling against cost-sensitive learning within an

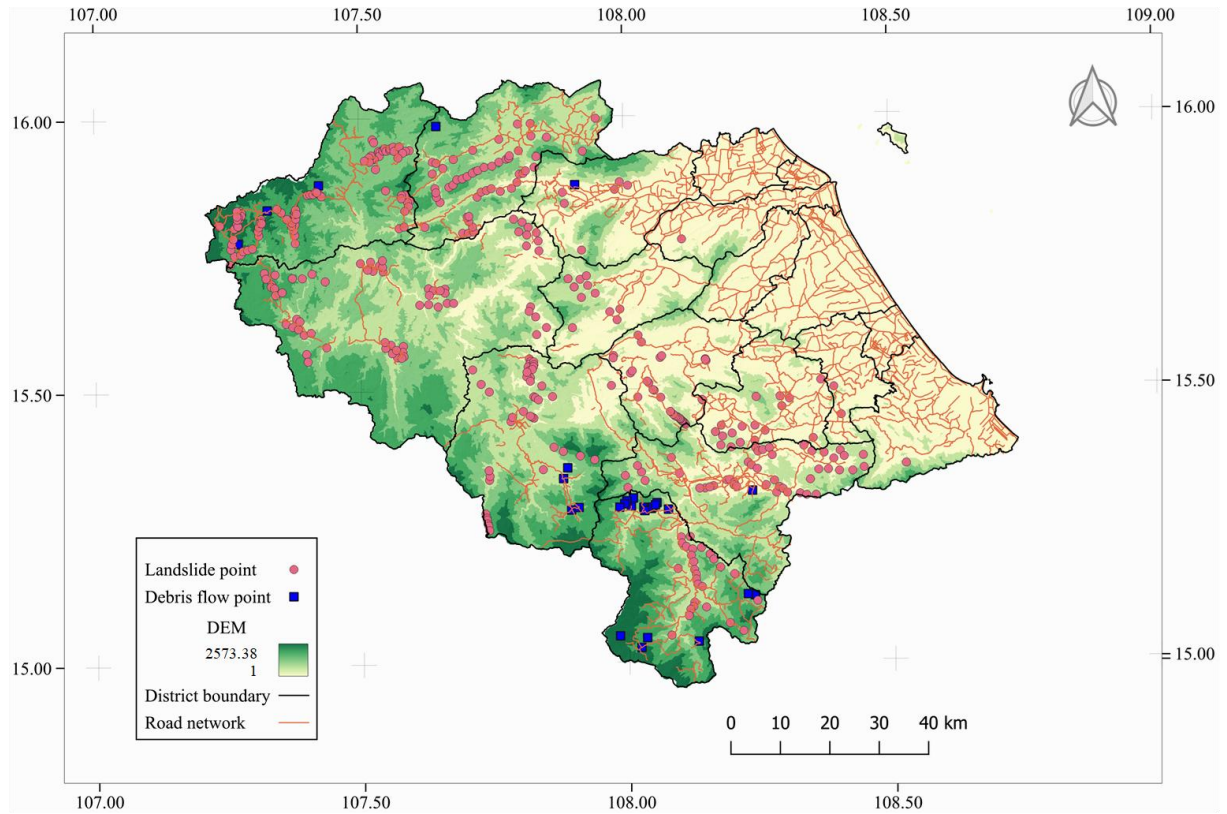


Fig. 2 Study area map showing the administrative boundaries of Quang Nam Province overlaid on a 10 m Digital Elevation Model (DEM), with locations of 228 field-surveyed sites: 28 confirmed debris-flow occurrence points (positive class, blue squares) and 200 landslide sites without debris flow (red circles). The five western highland districts are the primary study domain.

identical DFSM experimental framework. Third, direct benchmarking ML and the AHP/M-AHP models, which remains the standard operating procedure in Vietnam and many developing countries, is largely absent from research literature.

The present study develops a GIS-integrated ML framework for debris flow susceptibility and risk index mapping in Quang Nam Province, central Vietnam. Three specific objectives are pursued: (i) to compare four ML algorithms including random forest (RF), gradient boosting (GB), support vector machine (SVM), and artificial neural network (ANN), against a M-AHP for susceptibility classification; (ii) to evaluate two class-imbalance mitigation strategies (cost-sensitive learning and Borderline-SMOTE) under conditions of small inventory size and significant class imbalance; and (iii) to produce a debris flow risk index map by integrating ML-derived susceptibility with a spatial vulnerability assessment of population and infrastructure exposure.

The contributions of this study are threefold:

1. To the authors' knowledge, this is the first ML-based DFSM study conducted in Vietnam's tropical monsoon highlands, extending data-driven susceptibility mapping beyond its current concentration in Chinese and Central Asian regions.
2. A controlled comparison between cost-sensitive learning and Borderline-SMOTE is conducted on

an identical small-sample DFSM dataset, providing quantitative guidance for practitioners operating under data-scarce conditions.

3. A direct performance comparison between ML models and M-AHP – a method currently adopted in Vietnamese national hazard assessment quantifies the predictive gain achievable by substituting expert-driven scoring with data-driven classification in this geoenvironmental context.

2. STUDY AREA

Quang Nam Province is located in the central coastal region of Vietnam, spanning approximately 14°57'N to 16°04'N in latitude and 107°12'E to 108°44'E in longitude. Its area of approximately 10,574 km² is bounded by Thua Thien-Hue Province to the north, Kon Tum and Quang Ngai Provinces to the south, the Lao People's Democratic Republic to the west, and the South China Sea to the east. Among its 18 administrative districts, the western highland districts – Dong Giang, Tay Giang, Phuoc Son, Nam Tra My, and Bac Tra My – constitute the primary study domain for debris flow hazard analysis. A field inventory of 228 mass movement sites, comprising 28 confirmed debris flow occurrences and 200 landslide sites without debris flow characteristics, was compiled across these districts through field reconnaissance conducted between 2019 and 2022 (Fig. 2). Full

details of the inventory compilation and sampling design are provided in Section 3.1.

2.1. TOPOGRAPHY AND GEOMORPHOLOGY

A strong west-to-east gradient controls geohazard distribution in Quang Nam Province. There are three geomorphological zones. The Truong Son Range dominates the western mountainous area, where summit altitudes reach 2,573 m a.s.l., slopes surpass 30°, and vertical and horizontal relieves with severe channel gradients enable debris flow. The middle transitional zone has dissected hilly terrain underlain by deeply eroded metamorphic and igneous bedrock at 200–900 m a.s.l. Quaternary alluvial and marine sedimentary plains with reduced relief and no debris flow processes make up the eastern lowlands. This tripartite zonation is reflected in the spatial distribution of the 228 field-surveyed mass movement sites: 106 are below 222 m, 53 are in the 222–545 m transitional belt, and 69 are above 545 m, where most debris flows occur.

2.2. GEOLOGICAL AND TECTONIC SETTING

The geological substrate of Quang Nam Province consists of five primary lithological assemblages that play a key control on slope stability and provide loose debris for entrainment. The most eographically extensive lithological unit is metamorphic rocks—predominantly quartz-mica schist, biotite schist, and siliceous schist belonging to the Kham Duc Formation (MP–NP kd) and the Nui Vu Formation (NP3–E1 nv)—constitute the most spatially extensive lithological unit, covering an area of around 3,592 km². These intensely foliated rocks are particularly susceptible to slope failure because relict foliation planes function as preferential failure surfaces, and prolonged tropical weathering has generated thick residual soil mantles with substantially reduced shear strength. Intrusive igneous rocks (granodiorite, diorite, and granite) occupy approximately 2,912 km², Cretaceous clastic sedimentary rocks cover 1,284 km², volcanic sequences account for 1,628 km², and Quaternary alluvial–marine sediments occupy 1,158 km² predominantly in the coastal lowlands.

The spatial relationship between lithology and debris flow is strong. Twenty-one of the 28 confirmed sites overlie metamorphic rocks, approximately three times that of intrusive igneous rocks (6 sites) and an order of magnitude above volcanic materials (1 site). No debris flow has occurred on clastic sedimentary or alluvial–marine units. This distribution reflects foliated metamorphic rocks' mechanical fragility and tendency to form thick, clay-rich saprolite under tropical weathering. Northwest–southeast fault systems traverse the mountainous interior, defining the tectonic framework. The broken rock zones from these faults expedite weathering penetration, reduce effective cohesiveness and friction angle, and provide plentiful loose material for mobilization under high-intensity rains. In the inner buffer zones, 9 sites are

within 500 m of the nearest mapped fault trace and 7 are within the 500–1,000 m zone, accounting for 57 % of the inventory. Six sites exist between 1,000 and 2,000 m and 6 beyond 2,000 m, where the weight index drops rapidly. The progressive decline in weight index with increasing fault distance suggests that tectonic fracturing contributes to debris flow initiation, but its influence diminishes beyond 2,000 m, where steep channel gradient and thick weathering mantle on metamorphic bedrock apparently predispose slopes to failure without proximal fault influence.

2.3. CLIMATE AND HYDROLOGY

The tropical monsoon climate of Quang Nam province controls debris-flow triggering due to its bimodal precipitation pattern. Most of the province has monthly rainfall below 100 mm during the dry season, which runs from January to August with northeasterly trade winds. Due to the northeast monsoon and tropical cyclones crossing the South China Sea, the rainy season, from September to December, accounts for 65–80 % of yearly precipitation. The mean annual rainfall ranges from 2,000 mm in the coastal lowlands to over 4,000 mm in the western highland massif of the Truong Son Range, where orographic enhancement greatly increases precipitation. In October and November, prolonged, high-intensity rainfall episodes lasting 3–7 days can cause accumulated totals over 1,000 mm. This may lead to saturation of the thick weathering soil layer on slopes and increasing pore water pressures along potential failure surfaces to debris flow levels.

The hydrological system in this area is dominated by the Vu Gia–Thu Bon river system, with upstream tributaries drain the western highlands through steep, narrow valleys before main hydrological factor the alluvial coastal plain. Channel gradient is the main hydrological factor on debris flow initiation in this network. Of the 228 field-surveyed locations, 198 points are in the parts with channel gradients below 5°, in which no debris flows are recorded. The remaining 30 places are in the areas of high slopes: 11 sites in the 5–10° class, 14 in the 10–15° class, and 5 in the 15–25° class. All 28 confirmed debris flow sites are concentrated in these steep river sections, with 19 sites (67.9 % of the total) located where channel gradients exceed 10°. This threshold behavior reflects the competition between the demands of sufficient gravitational energy to mobilize flow and the ability of colluvial storage capacity: channels below 5° lack the kinetic energy for sustained debris flow propagation, while channels steeper than 25° are dominated by bedrock exposure and rockfall processes that preclude debris accumulation. The 10–15° gradient class, containing all 14 of its surveyed sites as confirmed debris flow occurrences, thus represents the optimal convergence of slope energy and sediment availability for debris flow initiation in the Quang Nam highland setting.

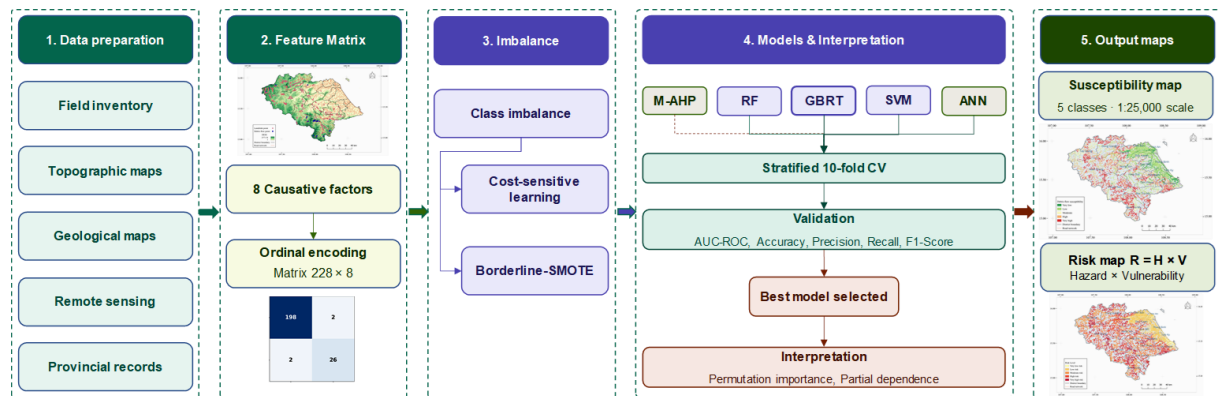


Fig. 3 Methodological flowchart for debris flow susceptibility and risk index mapping in Quang Nam province. CS = cost-sensitive learning; SMOTE = Borderline-SMOTE oversampling; CV = cross-validation; SHAP = Shapley Additive Explanations.

2.4. LAND USE AND SOIL CHARACTERISTICS

Forest cover (ca. 6,783 km²) is the dominant land use type in Quang Nam Province (64 % of the total area) and distributed primarily across the mountainous western districts. Of the 228 field-surveyed locations, 163 lie within forested terrain and 25 of which are confirmed debris flow sites. The remaining 3 debris flow sites are distributed across agricultural land (2 of 25 surveyed points) and building land (1 of 32 surveyed points). No debris flow has been recorded on vacant land (2 points) or water surfaces (6 points). The concentration of debris flow sites within forested terrain reflects geographic coincidence with steep, high-elevation metamorphic terrain where slope gradient, tectonic fracturing, and rainfall intensity are the dominant predisposing controls rather than land use per se. Selective logging and road construction in the highland districts have progressively degraded forest integrity in localized areas, reducing root reinforcement and increasing surface runoff – a process not captured by the static land use classification used in this study.

Soil distribution is a reflection of the interaction of lithology, climate, and topography throughout the studied area. Four soil groups are recognized, classified according to the FAO–UNESCO World Reference Base. Orthic Acrisols are the most widespread group in the western highlands, covering approximately 5,672 km² and accounting for all 28 confirmed debris flow sites among the 143 surveyed locations within this soil class. This exclusive association reflects the characteristic properties of Orthic Acrisols: strong weathering, high subsurface clay content, and low base saturation, resulting in low cohesion and high susceptibility to saturation-induced failure under prolonged rainfall. Ferric Acrisols (2,476 km²), occurred on intermediate-elevation hillsides and are characterized by iron oxide concretions. Eutric Regosols (1,269 km²) and Dystric Gleysols (1,157 km²) are found principally in the alluvial lowlands and valley bottoms, where neither soil type is associated with debris flow occurrence.

3. METHODOLOGY

The methodological framework of this study comprises five sequential stages: (1) debris flow inventory compilation and causative factor preparation; (2) data encoding and class-imbalance treatment; (3) ML model construction and training; (4) model evaluation through stratified k-fold cross-validation; and (5) model interpretation via permutation importance and partial dependence analysis. The overall workflow is schematically illustrated in Figure 3.

3.1. DEBRIS FLOW INVENTORY AND CAUSATIVE FACTORS

3.1.1. INVENTORY COMPILATION

The inventory of debris flows in Quang Nam province was produced from a number of complimentary sources. The first event database was developed based on the historical disaster data of the provincial and district level Committees for Natural Disaster Prevention and Control and reinforced with data from the Institute of Geological Sciences (Vietnam Academy of Science and Technology). We cross-referenced these records with high-resolution Google Earth photos to find debris flow scars, accumulation fans and channel deposits. Field reconnaissance from 2019–2022 confirmed incidents and uncovered new undocumented events. The assembled inventory includes 228 mass movement locations in the five western highland districts with notable clusters in the upper catchments of the Dakmi and Tranh rivers in Phuoc Son, Nam Tra My and Dong Giang Districts (Fig. 2). The deadliest incident in the inventory was the October 2020 debris flow in Tra Leng commune, Nam Tra My District, which buried a whole hamlet following the impact of Typhoon Molave, resulting in 22 fatalities.

3.1.2. SAMPLING DESIGN AND CLASS LABELING

In order to determine the occurrence of debris flow, the characteristics of source zone, transportation zone, and deposition zone of debris material were

investigated in the study area (Hung et al., 2014). Of the 228 sites, 28 satisfied these criteria and were assigned to the positive class (label 1). The remaining 200 sites exhibited rotational or translational landslide failure without fluid entrainment or long-runout characteristics and were assigned to the negative class (label 0), yielding a positive-to-negative ratio of approximately 1:7.

This sampling strategy carries an important methodological implication. Because negative samples are drawn from landslide sites rather than undisturbed terrain, both classes share many predisposing geoenvironmental conditions. The model is therefore required to discriminate between conditions that promote debris flow specifically versus those that produce other landslide types. This constitutes a more demanding and physically meaningful classification boundary than random negative sampling from stable ground and reduces the risk of artificially inflated model performance resulting from excessive feature-space separation between classes.

3.1.3. SPATIAL DATABASE PREPARATION

The spatial database for susceptibility modeling was constructed at a scale of 1:25,000, utilizing topographic maps and geological maps at the same scale provided by the provincial authorities. A digital elevation model (DEM) with 10 m resolution was derived from 1:25,000 topographic contours. The eight causative factors, comprising slope gradient, elevation, lithology, soil type, distance to fault, river channel gradient, landslide point density, and land use, were produced as raster layers in QGIS 3.44. Each was resampled to a uniform 10 m $\times$ 10 m pixel resolution to ensure spatial consistency. At each of the 228 survey locations, the corresponding values of all eight factors were extracted using the Sample Raster Values tool in QGIS, producing the tabular dataset employed for ML model training and evaluation. Sections 3.1 and 3.2 describe the reclassification systems and ordinal encoding procedures implemented for these layers.

3.1.4. CAUSATIVE FACTOR SELECTION AND CLASSIFICATION

In this study, eight causative factors were selected for the susceptibility analysis based on a comprehensive literature review and field observations (Gu et al., 2023; Xiong et al., 2020; Zhang et al., 2019). These factors are categorized into three thematic groups: topographic-geomorphological factors (slope gradient and elevation), geological-pedological factors (lithology, soil type, and distance to fault), hydrological factors (river channel gradient), and anthropogenic-environmental factors (land use and landslide point density). All factors were prepared as raster layers at a 1:25,000 scale using QGIS 3.44 and reclassified into discrete ordinal categories based on natural breakpoints and domain-specific thresholds established in the geotechnical literature.

Table 1 summarizes the specific classification schemes for each factor. Slope gradient, derived from a 10-m resolution DEM using the Slope tool in QGIS Spatial Analyst, was divided into five classes using equal intervals across the provincial range: 0–6.7°, 6.7–14.6°, 14.6–22.3°, 22.3–30.8°, and >30.8°, see Figure 4a. Elevation was classified into five ranges from 1 m to 2,573 m using natural breaks (Jenks optimization), which minimizes within-class variance and maximizes between-class separation based on the provincial DEM distribution (Fig. 4b). Lithology was categorized into five groups reflecting the principal geological units mapped at 1:25,000 scale: river sediments, volcanic (magma eruption), clastic sedimentary (Cretaceous sediments), intrusive igneous (magma intrusion), and metamorphic rocks (Fig. 4c). This classification follows the provincial geological map legend and requires no further interval optimization as the categories are inherently discrete. Soil type was classified according to the FAO soil taxonomy into four groups: Dystric Gleysols, Eutric Regosols, Ferric Acrisols, and Orthic Acrisols (Fig. 4d). Distance to fault was computed using Euclidean distance analysis and divided into four buffer zones (<500 m, 500–1,000 m, 1,000–2,000 m, and >2,000 m), with proximity encoded inversely such that closer distances receive higher ordinal values reflecting greater susceptibility (Fig. 4e). River channel gradient, a critical hydrological parameter governing the kinetic energy available for debris flow initiation and propagation, was classified into four categories: <5°, 5–10°, 10–15°, and 15–25° (Fig. 4f). Landslide point density, computed as the number of historical landslide occurrences per km², was divided into five classes (Fig. 4g). Land use was classified into five categories: water surface, agricultural land, building land, vacant land, and forest land (Fig. 4h).

3.2. FEATURE ENCODING AND CLASS-IMBALANCE HANDLING STRATEGIES

3.2.1. FEATURE ENCODING

All eight causative factors are categorical in nature (either inherently nominal or pre-classified into ordinal intervals). To render these factors amenable to ML algorithms operating on numerical inputs, ordinal encoding was applied based on the physical ordering of each factor relative to its level of debris flow susceptibility. Specifically, each factor class was assigned an integer code from 0 to $n - 1$, where n is the number of classes, with higher codes corresponding to conditions of greater expected susceptibility. This encoding preserves the rank-order relationship inherent in the data. For instance, steeper slopes receive higher codes than gentler slope, while avoiding the dimensionality inflation associated with one-hot encoding, which is particularly undesirable given the limited sample size (Hancock and Khoshgofaar, 2020). With only 228 samples and 8 categorical features, one-hot encoding would expand the feature space to over 40 binary columns, increasing the risk of multicollinearity and the curse of

Table 1 Classification scheme and class weight index (W_{ij}) of debris flow causative factors.

Factor	Class	N_{df}	Area (km ²)	W_{ij}	W_{ij-std}
Geology	Volcanic (magma eruption)	1	1.628	0.232	0.072
	Intrusive igneous	6	2.912	0.778	0.242
	Clastic sedimentary	0	1.284	0.000	0.000
	Alluvial–marine sediments	0	1.158	0.000	0.000
	Metamorphic rocks	21	3.592	2.208	0.686
	SUM	28	10.574	3.218	1.000
Elevation (m)	1 – 222	2	5.416	0.139	0.017
	222 – 545	9	2.498	1.361	0.162
	545 – 878	10	1.002	3.769	0.449
	878 – 1.302	5	873	2.163	0.258
	1.302 – 2.573	2	785	0.962	0.115
	SUM	28	10.574	8.394	1.000
Land use	Water surface	0	893.000	0.000	0.000
	Forest land	25	6.783	1.392	0.595
	Agricultural land	2	1.303	0.580	0.248
	Building land	1	1.023	0.369	0.158
	Vacant land	0	572	0.000	0.000
	SUM	28	10.574	2.341	1.000
Soil type	Orthic Acrisols	28	5.672	1.864	1.000
	Ferric Acrisols	0	2.476	0.000	0.000
	Eutric Regosols	0	1.269	0.000	0.000
	Dystric Gleysols	0	1.157	0.000	0.000
	SUM	28	10.574	1.864	1.000
Landslide density (pts/km²)	0 – 5	15	5.092	1.112	0.293
	5 – 10	9	2.921	1.164	0.307
	10 – 15	3	1.023	1.107	0.292
	15 – 20	1	915	0.413	0.109
	> 20	0	623	0.000	0.000
	SUM	28	10.574	3.796	1.000
Slope angle (°)	0 – 6.7	4	4.060	0.372	0.053
	6.7 – 14.6	3	2.794	0.405	0.058
	14.6 – 22.3	9	1.376	2.470	0.352
	22.3 – 30.8	8	1.295	2.333	0.332
	> 30.8	4	1.049	1.440	0.205
	SUM	28	10.574	7.021	1.000
Distance to fault (m)	< 500	9	1.823	1.864	0.279
	500 – 1.000	7	1.025	2.579	0.386
	1.000 – 2.000	6	1.200	1.888	0.283
	> 2.000	6	6.526	0.347	0.052
	SUM	28	10.574	6.679	1.000
Channel gradient (°)	< 5	0	3.012	0.000	0.000
	5 – 10	9	1.023	3.322	0.341
	10 – 15	14	1.231	4.295	0.441
	15 – 25	5	892	2.117	0.217
	> 25	0	623	0.000	0.000
	SUM	28	6.781	9.734	1.000

dimensionality, both of which degrade classifier performance when sample-to-feature ratios are low.

The resulting feature matrix has dimensions 228×8 , where each row corresponds to a survey point and each column corresponds to an ordinal encoded causative factor. The target vector is binary (0 = no debris flow; 1 = debris flow). However, the dataset exhibits a class imbalance ratio of approximately 1:7 (28 positive vs. 200 negative samples). Training ML classifiers on such imbalanced data without corrective measures leads to models that are biased toward the

majority class and exhibit poor sensitivity to the minority class—a well-documented phenomenon in the geohazard literature (Bui et al., 2020; Zhang et al., 2019). To address this, two complementary strategies were implemented and compared.

3.2.2. COST-SENSITIVE LEARNING

In the cost-sensitive approach, the misclassification penalty for the minority class (debris flow) is increased relative to the majority class (non-debris flow) by assigning class weights that are

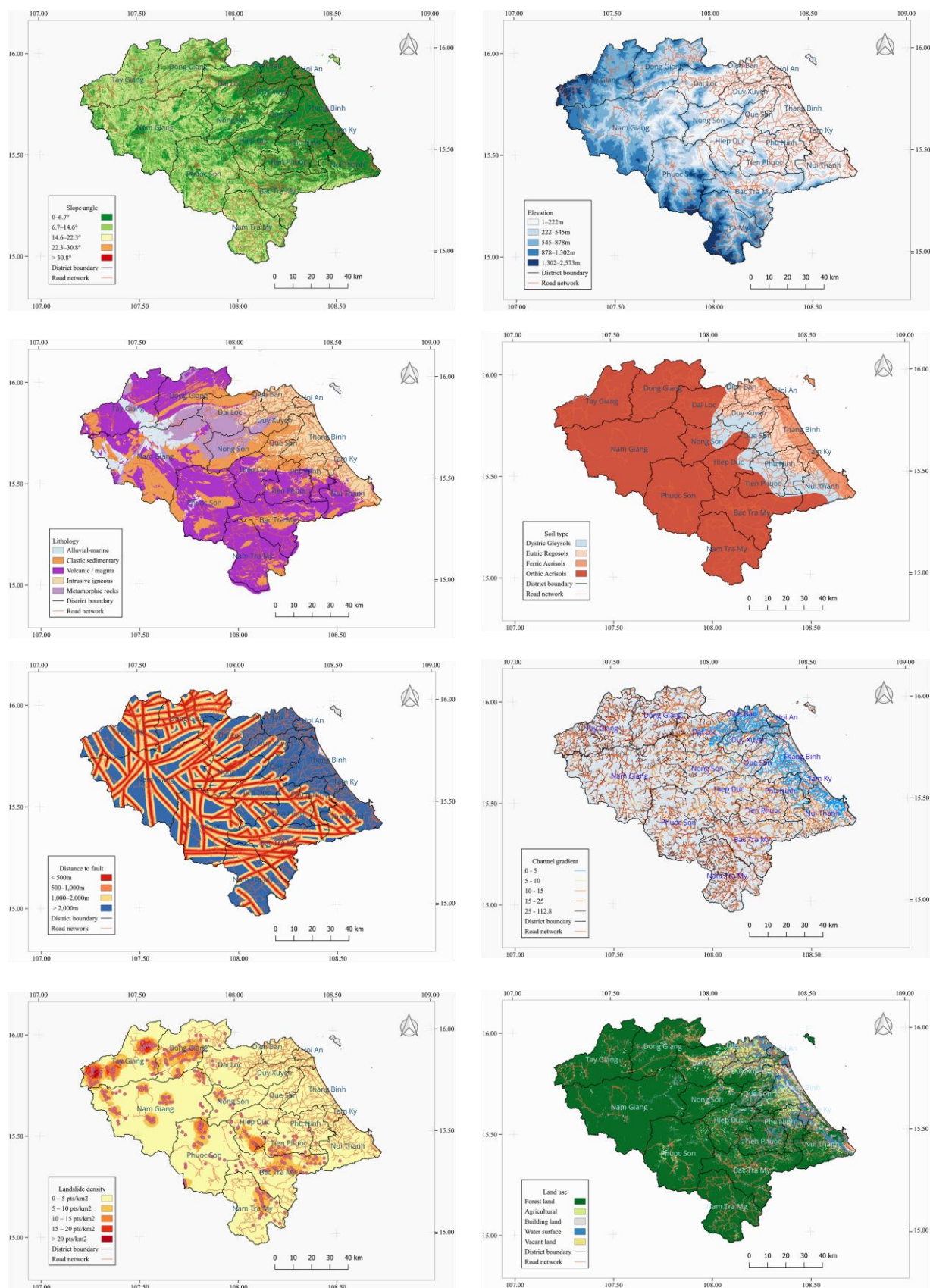


Fig. 4 Thematic raster maps of the eight causative factors used for debris flow susceptibility mapping: (a) slope gradient; (b) elevation; (c) lithology; (d) soil type; (e) distance to fault; (f) river channel gradient; (g) landslide point density; (h) land use.

inversely proportional to class frequency. Formally, the weight assigned to the positive class is $w_+ = N_- / N_+ \approx 7.14$, where N_- and N_+ denote the number of negative and positive samples, respectively. For RF and SVM, this weighting is implemented through the *class_weight* = ‘balanced’ parameter. For Gradient Boosting, equivalent weighting is achieved via the *sample_weight* argument during training. The principal advantage of this approach is that it operates on the original data distribution without introducing synthetic samples, thereby avoiding the risk of noise amplification in the minority class.

3.2.3. BORDERLINE-SMOTE OVERSAMPLING

The Synthetic Minority Oversampling Technique (SMOTE), originally proposed by (Chawla et al., 2002), generates synthetic minority-class instances by interpolating between existing minority samples and their nearest neighbors in feature space. In this study, the Borderline-SMOTE variant (Han et al., 2005) was adopted, which restricts synthetic sample generation to minority instances located near the decision boundary – i.e., those with a substantial proportion of majority-class neighbors. This targeted oversampling produces synthetic samples that are more informative for classifier training than random minority-class interpolation, as they populate precisely the regions of feature space where classification uncertainty is greatest.

Given the limited number of positive samples ($N_+ = 28$), the k -neighbors parameter was set to 3 to ensure sufficient neighborhood density for synthetic generation. SMOTE was applied exclusively within each training fold during cross-validation to prevent data leakage from synthetic samples into the validation set.

3.3. MACHINE LEARNING MODELS

Four machine learning methods from distinct methodological groups were applied: Random Forest (RF), Gradient Boosting (GB), Support Vector Machine (SVM), and Artificial Neural Network (ANN). The selection was based on three main factors: demonstrated effectiveness in geohazard susceptibility mapping (Merghadi et al., 2020; Xiong et al., 2020) the presence of complementary inductive biases that together investigate diverse regions of the hypothesis space; and established sensitivity profiles relevant to the current dataset, which comprised 228 samples, 28 positive cases and a class imbalance ratio of 1:7. Moreover, the algorithms used in this research have demonstrated reliable predictive performance in nonlinear engineering contexts, even with scarce training data. This includes applications such as structural capacity prediction (La et al., 2026; Luat et al., 2026) and hydrodynamic parameter estimation (Tran et al., 2026a, 2026b). Consequently, their general applicability to classification problems with limited data is confirmed. The configured hyperparameters values are the default settings of the ML models, as presented in Table 2.

3.3.1. RANDOM FOREST (RF)

RF (Breiman, 2001) constructs an ensemble of B decision trees, each trained on an independent bootstrap resample of the training data with a randomly selected feature subset of size m evaluated at each split node, where $m = \sqrt{p}$ and p is the total number of input features ($m = 3$ for $p = 8$ in this study). For a given input $\mathbf{x}$, the ensemble prediction is obtained by majority voting:

$$\hat{y}_{RF}(\mathbf{x}) = \arg \max \sum_{b=1}^B \mathbf{1}[h_b(\mathbf{x}) = c] \quad (1)$$

where $h_b(\mathbf{x})$ is the prediction of the b^{th} tree and $\mathbf{1}[\cdot]$ is the indicator function. The dual randomization – over both instances and features – decorrelates individual trees and substantially reduces ensemble variance without inflating bias, a property formalized by the variance decomposition:

$$var[\hat{y}_{RF}] = \rho \sigma^2 + \frac{1-\rho}{B} \sigma^2 \quad (2)$$

where σ^2 is the variance of a single tree and ρ is the average pairwise correlation between trees. As B increases, the second term vanishes, leaving only the irreducible inter-tree correlation term $\rho \sigma^2$. This property renders RF particularly robust under small-to-moderate sample sizes with heterogeneous predictors (Fernández-Delgado et al., 2014).

3.3.2. GRADIENT BOOSTING (GB)

GB (Friedman, 2001) constructs an additive ensemble sequentially. At iteration m , a weak learner $h_m(\mathbf{x})$ is fitted to the negative gradient of the loss function $\mathcal{L}$ with respect to the current ensemble prediction $F_{m-1}(\mathbf{x})$:

$$r_{im} = - \left[\frac{\partial \mathcal{L}(y_i, F(\mathbf{x}_i))}{\partial F(\mathbf{x}_i)} \right]_{F=F_{m-1}} \quad (3)$$

The ensemble is updated as:

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \eta h_m(\mathbf{x}) \quad (4)$$

where $\eta \in (0,1]$ is the learning rate (shrinkage parameter) controlling the contribution of each successive learner. For the binary cross-entropy loss used in this study, the pseudo-residuals r_{im} reduce to the difference between observed labels and predicted probabilities. Unlike RF, which reduces variance through parallel averaging, GB explicitly minimizes a differentiable loss via gradient descent in function space, consistently outperforming bagging methods on structured tabular data (T. Chen et al., 2016). The efficacy of gradient boosting under systematic hyperparameter optimization and cross-validation has been confirmed in nonlinear engineering prediction tasks with compact datasets (Luat et al., 2021). Overfitting was controlled through learning rate shrinkage, stochastic subsampling of training instances at each iteration, and combined elastic-net regularization with L1 penalty $\alpha_1 = 0.1$ and L2 penalty $\alpha_2 = 1.0$ applied to leaf weights.

Table 2 Hyperparameter configurations of ML models.

Model	Hyperparameter	Value
Random Forest	Number of trees (B)	500
	Maximum depth	8
	Min. samples to split	5
	Min. samples per leaf	2
	Feature subset size (m)	$\sqrt{p}$
	Class weight	Balanced
Gradient Boosting	Number of estimators	300
	Maximum depth	4
	Learning rate (η)	0.05
	Subsample fraction	0.8
	L1 regularization (α_1)	0.1
	L2 regularization (α_2)	1.0
	Class weight	Balanced
SVM	Kernel	RBF
	Regularization (C)	10
	Bandwidth (γ)	scale = $1/(p \cdot \text{Var}(\mathbf{X}))$
	Feature scaling	Zero mean, unit variance
	Class weight	Balanced
ANN (MLP)	Hidden layer sizes	[32, 16]
	Activation function	ReLU
	Output activation	Sigmoid
	Optimizer	Adam
	L2 regularization (α)	0.01
	Early stopping patience	10 epochs
	Max. iterations	500
	Class weight	Balanced

3.3.3. SUPPORT VECTOR MACHINE (SVM)

SVM with a radial basis function (RBF) kernel (Vapnik, 2013) seeks the decision function:

$$f(x) = \text{sign}(\sum_{i=1}^n \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b) \quad (5)$$

where $\alpha_i \geq 0$ are Lagrange multipliers, b is the bias term, and $K(\mathbf{x}_i, \mathbf{x})$ is the RBF kernel:

$$K(\mathbf{x}_i, \mathbf{x}) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2\right) \quad (6)$$

The bandwidth parameter γ controls the spatial extent of kernel influence, and the regularization parameter C governs the trade-off between maximizing the margin and minimizing classification error. Only training instances with $\alpha_i > 0$, the support vectors – contribute to the decision function, rendering SVM structurally efficient under small-sample conditions where the effective support vector count is low relative to dataset size. Before training, input features were standardized to zero mean and unit variance prior to training, with the bandwidth set to $\gamma = 1/(p \cdot \text{Var}(\mathbf{X}))$, where p is the dimension of the feature.

3.3.4. ARTIFICIAL NEURAL NETWORK (ANN)

A multilayer perceptron (MLP) with two hidden layers was employed, implementing the standard feedforward computation:

$$\mathbf{a}^{(l)} = \phi(\mathbf{W}^{(l)} \mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}) \quad l = 1, 2, \dots, L \quad (7)$$

where $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ are the weight matrix and bias vector of layer l , $\phi(\cdot)$ is the ReLU activation function $\phi(z) = \max(0, z)$, and $\mathbf{a}^{(0)} = \mathbf{x}$ is the input vector. The output layer applies a sigmoid activation to produce class probabilities. Network parameters were optimized using the Adam optimizer (Kingma et al., 2014), which maintains per-parameter adaptive learning rates through first- and second-moment estimates of the gradient:

$$\mathbf{m}_t = \beta_1 \mathbf{m}_{t-1} + (1 - \beta_1) \nabla_{\theta} \mathcal{L}_t \quad (8)$$

$$\mathbf{v}_t = \beta_2 \mathbf{v}_{t-1} + (1 - \beta_2) (\nabla_{\theta} \mathcal{L}_t)^2 \quad (9)$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{\mathbf{v}}_t + \epsilon}} \hat{\mathbf{m}}_t \quad (10)$$

where $\mathbf{m}_t$ and $\mathbf{v}_t$ are bias-corrected moment estimates. L2 weight regularization ($\alpha = 0.01$) and early stopping monitored on a 15 % held-out validation fraction were applied to mitigate overfitting.

3.4. MODIFIED ANALYTIC HIERARCHY PROCESS (M-AHP) BASELINE

The Modified Analytic Hierarchy Process (Nefeslioglu et al., 2013) served as a conventional, expert-driven baseline for evaluating the performance of the machine learning model. M-AHP builds upon the classical Analytic Hierarchy Process (Saaty, 1977) by substituting expert-assigned factor weights with significance levels derived from data, thus mitigating

reliance on cognitive bias while maintaining the interpretable, weighted-linear-combination framework.

The debris flow susceptibility index (DFSI) is computed as:

$$DFSI = \sum X_i W_{ij}^* \quad (11)$$

where X_i is the weight of factor i and W_{ij}^* is the standardized class weight of class j within factor i . Class weights are computed as:

$$W_{ij} = \left(\frac{N_{ij}}{N} \right) / \left(\frac{S_{ij}}{S} \right) \quad (12)$$

where N_{ij} and S_{ij} denote the number of debris flow occurrence points and the areal extent of class j of factor i , respectively, and N and S are the corresponding totals for the study area. W_{ij}^* is obtained by normalizing W_{ij} to unit sum within each factor, yielding a relative susceptibility score per class.

Factor weights X_i are derived from the distribution of W_{ij}^* values across classes. The standard deviation σ_i of the class weight distribution serves as a proxy for the spatial discriminating capacity of factor i : a small σ_i indicates that debris flow occurrence is distributed proportionally across multiple classes - a richer spatial gradient - whereas a large σ_i reflects concentration in a single class, diminishing spatial discriminability. An integer significance level I_i is assigned by ranking $1/\sigma_i$ values from 1 (lowest) to n (highest), and factor weights are then computed as:

$$X_i = I_i / \sum I_i \quad (13)$$

For the conventional AHP model, factor weights were determined through pairwise comparison of the eight causative factors using the nine-point (Saaty, 1977) scale, based on domain expertise and the frequency ratio analysis reported in Table 1. The consistency of the pairwise judgment matrix was verified by computing the Consistency Ratio (CR = CI/RI), which was 0.032 - well below the maximum acceptable threshold of 0.10. The resulting factor weights for both methods are reported in Table. The DFSI surface was classified into five susceptibility levels (very low, low, moderate, high, and very high) using the Natural Breaks (Jenks) method and evaluated against the debris flow inventory using the same AUC framework applied to the ML models (J. Chen et al., 2013).

3.5. DEBRIS FLOW RISK INDEX MAPPING

The debris flow risk index map was derived by integrating the ML-based susceptibility map with a spatially distributed vulnerability assessment following the $R = H \times V$ risk quantification framework (UN DHA, 1992), which identifies the specific stream reaches within Quang Nam Province where physical hazard intersects with human and infrastructural exposure. The hazard component (H) was represented by the normalized susceptibility class index ($H \in$

$\{0.00, 0.25, 0.50, 0.75, 1.00\}$ for classes Very Low through Very High, respectively), ensuring that the risk output is directly proportional to the model-derived susceptibility level without reintroducing the artefacts of the faded probability surface.

The vulnerability component (V) was constructed as a weighted composite of two spatial proxies according to:

$$V_i = w_{LU} V_{LU,i} + w_{EL} V_{EL,i} \quad (14)$$

where V_i is the vulnerability index at pixel i ; $V_{LU,i}$ and $V_{EL,i}$ are the normalized land-use exposure weight and elevation-based population density proxy, respectively; and $w_{LU}=0.70$, $w_{EL}=0.30$ are the component weights calibrated by expert judgment informed by provincial land-use records and the (UN DHA, 1992; Cooke, R., 1991) vulnerability classification hierarchy: urban and building land received the maximum weight ($V = 1.00$), agricultural land $V = 0.60$, vacant land $V = 0.30$, forest land $V = 0.20$, and water surfaces $V = 0.10$. The elevation proxy assigned decreasing vulnerability weights with increasing altitude (1.00 at 1–222 m to 0.15 at 1,302–2,573 m), reflecting the inverse relationship between elevation and population density documented in the provincial census data. The individual sub-weights were assigned as presented in Table 4.

The final risk index at each pixel was then computed as:

$$R_i = H_i \times V_i \quad (15)$$

where $H_i \in \{0.00, 0.25, 0.50, 0.75, 1.00\}$ corresponds to the normalized susceptibility class.

3.6. MODEL EVALUATION AND INTERPRETABILITY

All models were evaluated using stratified 10-fold cross-validation, which preserves the class ratio in each fold and provides unbiased performance estimates under class imbalance. Five metrics were reported: AUC-ROC, precision, recall, F1-score, and overall accuracy. Precision is the proportion of predicted debris flow sites that are correctly identified $TP / (TP + FP)$. Recall is the proportion of actual debris flow sites that the model successfully detects $TP / (TP + FN)$. F1-score is the harmonic mean of precision and recall. Overall accuracy is reported for completeness but is not the primary criterion here, since the 1:7 class imbalance makes it an unreliable indicator of minority-class performance. AUC, recall, and F1-score are the principal metrics used for model comparison throughout this study. Stratification ensured that each fold contained approximately three positive instances, reflecting the provincial inventory ratio.

Feature importance was quantified using SHAP (Lundberg et al., 2017), a unified model-agnostic interpretability framework grounded in cooperative game theory. Shapley values satisfy four axioms - efficiency, symmetry, dummy, and additivity

Table 3 Sub-weight assignments for the vulnerability index components used in debris flow risk index mapping.

Factor	Class	Weight
Land use (V_{LU})	Urban / built-up	1.00
	Agricultural land	0.60
	Vacant land	0.30
	Forest land	0.20
	Water surface	0.10
Elevation proxy (V_{EL})	1–222 m	1.00
	223–500 m	0.75
	501–900 m	0.50
	901–1301 m	0.30
	1,302–2,573 m	0.15

guaranteeing a fair allocation of each prediction's deviation from the model's expected output among contributing features. For GB and RF, exact TreeSHAP values were computed via path decomposition over ensemble decision trees. For SVM and ANN, Kernel SHAP values were estimated by solving a weighted least-squares problem over all $2^8 = 256$ possible feature coalitions with Monte Carlo integration ($n_{MC} = 15$ background samples per coalition), applied to a representative explanation set of 50 samples (all 28 positive instances and 22 stratified negative instances). Algorithm-specific importance metrics (Gini impurity for trees; permutation importance for SVMs) were not used because they are sensitive to feature correlations and not directly comparable across algorithm families; SHAP values provide a unified basis for cross-model comparison.

4. RESULTS AND DISCUSSION

4.1. M-AHP BASELINE PERFORMANCE

The normalized frequency ratio weight index W_{ij}^* for all eight causative factors is summarized in Table 1 and illustrated in Figure 5. Among geological units, metamorphic rocks exhibit the highest weight ($W_{ij}^* = 0.686$), concentrating 21 of 28 inventory sites despite occupying only 34.0% of the provincial area (Fig. 5a). Elevation shows a non-monotonic profile peaking at the 545–878 m band ($W_{ij}^* = 0.449$, 10 events), reflecting the transitional zone where steep channel gradients, thick colluvial deposits, and orographic rainfall enhancement coincide (Fig. 5b). Forest land dominates the land use inventory ($W_{ij}^* = 0.595$, 25 events), attributable to the pervasive forest cover of the high-susceptibility western highlands rather than a land-use causal effect (Fig. 5c). Orthic Acrisols are the sole soil type associated with debris flow occurrence ($W_{ij}^* = 1.00$, all 28 events), confirming this unit as a near-necessary pedological condition for debris flow initiation in the study area (Fig. 5d). Landslide density exhibits the flattest W_{ij}^* profile across classes ($\sigma_i = 0.1385$), providing limited discriminative contrast (Fig. 5e). Slope angle concentrates events in the $14.6\text{--}22.3^\circ$ ($W_{ij}^* = 0.352$)

and $22.3\text{--}30.8^\circ$ ($W_{ij}^* = 0.332$) classes, defining a critical initiation range of approximately $15\text{--}31^\circ$ (Fig. 5f). For distance to fault, the 500–1,000 m proximity band yields the highest weight ($W_{ij}^* = 0.386$), with the $>2,000$ m class recording the lowest ($W_{ij}^* = 0.052$), confirming structural discontinuities as a secondary material-supply control (Fig. 5g). Channel gradient records the highest absolute W_{ij} value in the dataset ($W_{ij} = 4.295$ at $10\text{--}15^\circ$, $W_{ij}^* = 0.441$), concentrating 14 of 28 events within a stream network area of only 1,231 km², and defines a critical initiation window of $5\text{--}25^\circ$ (Fig. 5h).

The M-AHP method, applied as the expert-driven benchmark prior to all ML evaluations, yielded an AUC of 0.742 under the ROC framework (Fig. 6). This value falls within the conventionally "acceptable" range of 0.70–0.80 (Swets, 1988), confirming that the AHP-derived susceptibility index retains meaningful discriminative capacity: high-gradient, high-elevation zones in the western highlands systematically receive higher susceptibility scores than the low-gradient coastal plain, consistent with the spatial pattern of the debris flow inventory. However, two structural limitations constrain the M-AHP's predictive ceiling. First, the weighted linear combination formula $DFSI = \sum X_i \times W_{ij}$ assumes additive independence among factors, which fails to represent the multiplicative interactions between channel gradient, elevation, and lithology that physically govern debris flow initiation. Second, the σ_i – based weight derivation (Eq. 12) rewards factors with the most evenly distributed class-frequency weights – assigning landslide density the highest weight ($X_i = 0.2222$, $\sigma_i = 0.1385$, ranked 1st) while relegating channel gradient to fifth place ($X_i = 0.1111$) – directly inverting the data-driven importance ranking subsequently confirmed by SHAP analysis (Section 4.4). As made visually explicit by the dashed reference line at AUC = 0.742 running across the full width of Figure 6, every viable ML model surpasses this benchmark by a qualitative rather than marginal margin, with the best ML configuration exceeding the M-AHP baseline by 0.249 AUC units – a relative improvement of 33.5 %.

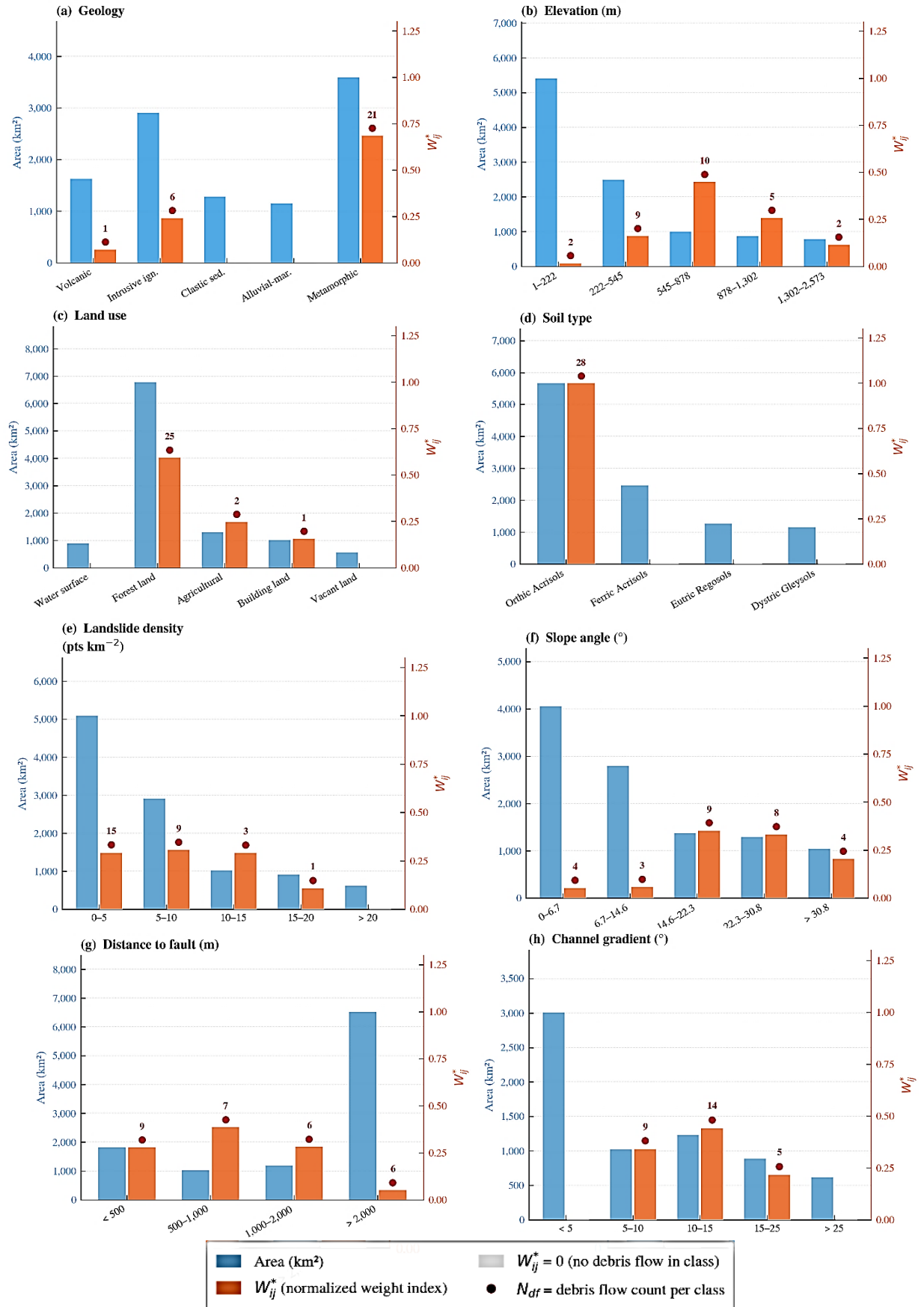


Fig. 5 Spatial distribution (bar area, left axis) and normalised M-AHP frequency ratio weight index W_{ij}^* (orange bars and dots, right axis) of the eight causative factors: (a) geology; (b) elevation; (c) land use; (d) soil type; (e) landslide density; (f) slope angle; (g) distance to fault; (h) channel gradient. Dots indicate the number of confirmed debris flow occurrences per class (N_{df}). $W_{ij}^* = 0$ indicates no debris flow recorded in that class.

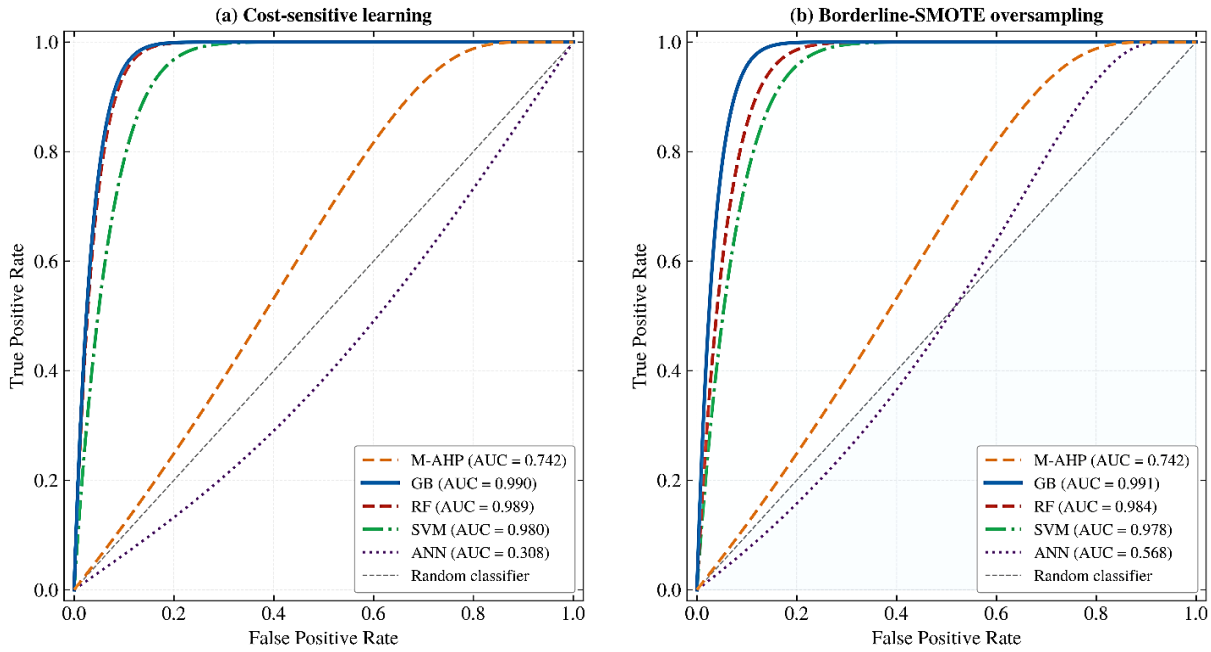


Fig. 6 Receiver Operating Characteristic (ROC) curves for all ML models under (a) cost-sensitive learning and (b) Borderline-SMOTE oversampling strategies, compared with the M-AHP expert baseline (AUC = 0.742, dashed line). Shaded region indicates the AUC area of the best-performing configuration (GB-SMOTE). The random classifier diagonal is shown for reference.

Table 4 Cross-validation performance of all model–strategy configurations and the M-AHP baseline (mean $\pm$ std over 10 stratified folds)

Model	Strategy	AUC	AUC $\pm$ std	Accuracy	Precision	Recall	F1-score
GB	SMOTE	0.991	0.995 $\pm$ 0.015	0.978	0.926	0.893	0.909
	Cost-sens.	0.990	0.992 $\pm$ 0.025	0.982	0.929	0.929	0.929
RF	Cost-sens.	0.989	0.988 $\pm$ 0.025	0.961	0.913	0.750	0.824
	SMOTE	0.984	0.981 $\pm$ 0.029	0.961	0.913	0.750	0.824
SVM	Cost-sens.	0.980	0.993 $\pm$ 0.020	0.987	0.963	0.929	0.946
	SMOTE	0.978	0.985 $\pm$ 0.045	0.982	0.929	0.929	0.929
ANN	SMOTE	0.568	0.573 $\pm$ 0.130	0.877	0.000	0.000	0.000
	Cost-sens.	0.308	0.320 $\pm$ 0.152	0.877	0.000	0.000	0.000
M-AHP	-	0.742	-	-	-	-	-

4.2. OVERALL ML MODEL PERFORMANCE

All ML models were implemented in Python 3.14 using the scikit-learn 1.3.2 library. A fixed random seed of 42 was applied to all stochastic components to ensure full reproducibility. Table and Figures 6–8 present the complete comparative performance of the eight ML model–strategy configurations against the M-AHP baseline. Prior to discussing individual model behavior, it should be noted that all metrics in Table 3 reflect internal stratified 10-fold cross-validation performance. An independent post-event validation dataset was not available given the limited inventory size, and the reported AUC values should be interpreted accordingly as internal discrimination measures rather than externally confirmed prediction capacities. With this caveat in mind, three findings stand out at the aggregate level. Considering this, three findings stand out at the aggregate level.

First, GB, RF, and SVM all achieved AUC values exceeding 0.97 under both imbalance-handling strategies, clustering tightly in the upper-left corner of the ROC space far above the M-AHP baseline curve and the random classifier diagonal (Fig. 6). This convergence of three structurally distinct algorithms – a sequential boosting ensemble (GB), a bagging ensemble (RF), and a kernel-based margin classifier (SVM) – on similarly high AUC values is not a coincidence but a symptom of highly separable class distributions in the feature space of the study area: when channel gradient and elevation jointly discriminate 92 % of debris flow sites from the negative class, any algorithm with sufficient expressive capacity to model these two features will approach ceiling performance.

Second, the ANN failed categorically under both strategies, producing AUC values of 0.308 (cost-sensitive) and 0.568 (Borderline-SMOTE) with

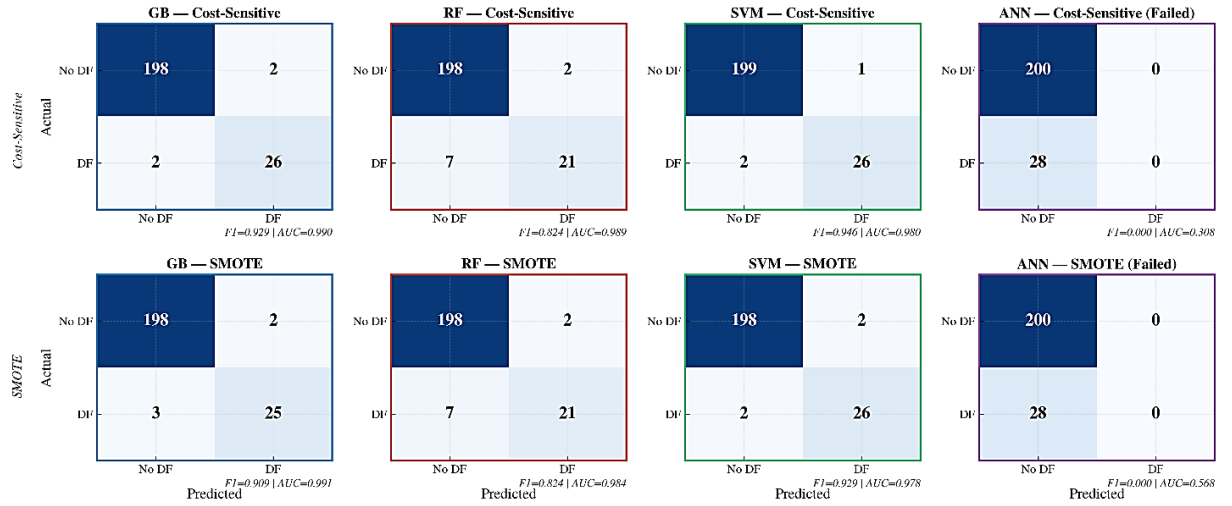


Fig. 7 Bar chart comparison of cross-validation performance metrics (AUC, accuracy, precision, recall, F1-score) across all model–strategy configurations and the M-AHP baseline. The dashed horizontal reference line at 0.742 indicates the M-AHP AUC. ANN results (failed configurations) are shown for completeness.

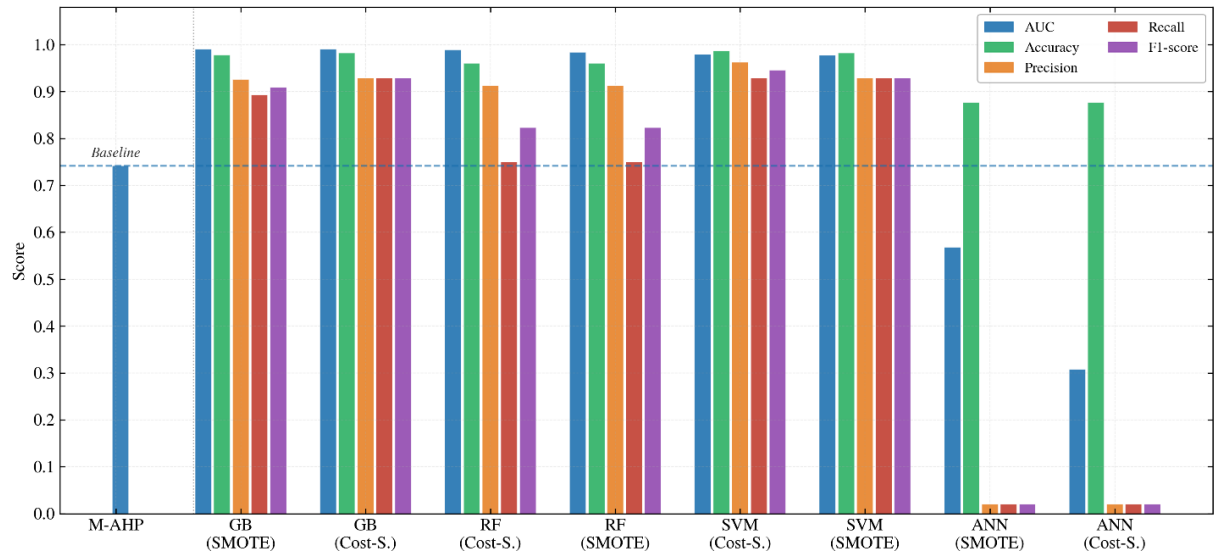


Fig. 8 Cumulative confusion matrices from stratified 10-fold cross-validation for all model–strategy configurations (values represent total predictions across all folds, $n = 228$). DF = debris flow occurrence; No DF = non-occurrence. Failed ANN configurations are shown in the rightmost columns.

Precision = Recall = F1 = 0.000 in both cases (Table 3). Both ANN confusion matrices show $TN = 200$, $FP = 0$, $FN = 28$, $TP = 0$: not a single debris flow site was correctly identified. The model converged to the trivial majority-class solution. The mechanistic explanation is quantifiable: the MLP operates at a parameter-to-minority-sample ratio of approximately 3.7 (833 trainable parameters / 28 positive instances), far below the minimum of 10–20 samples per parameter required for reliable gradient-based convergence to a minority-sensitive decision boundary (Bishop et al., 2006; Luat et al., 2024). This failure is not attributable to hyperparameter tuning but to a fundamental structural incompatibility between MLP capacity requirements and the available positive sample count. Critically, neither cost-sensitive weighting nor Borderline-SMOTE resolved this pathology, because

both strategies address class imbalance but not the more fundamental problem of insufficient absolute minority-class sample size to define a stable loss landscape for the positive class. These findings reinforce the recommendation that ensemble tree methods and kernel-based classifiers should be the default algorithmic choices for DFSM under small-inventory conditions. This finding aligns with a significant amount of literature in the geohazard machine learning field, where ensemble tree-based methods, especially boosting and bagging derivatives, consistently show superiority over neural architectures in small-inventory settings. This is primarily due to their inductive bias not requiring large absolute minority sample sizes for the formation of stable decision boundaries (Bui et al., 2020; Merghadi et al., 2020). Future work constrained by similar data

limitations could explore hybrid architectures that combine tree-based ensemble predictions with physics-informed feature engineering, or transfer learning from larger regional inventories to initialize the neural network weights, before fine-tuning them on the local dataset (Le, Choi, et al., 2024; Le, Huynh, et al., 2024).

Third, among the three viable ML algorithms, GB consistently achieved the highest discriminative performance. GB-SMOTE attained the top AUC of 0.9905, while GB-CS achieved the most balanced precision–recall trade-off with identical values of 0.9286 and an F1-score of 0.9286 - the highest among all configurations. The GB-CS confusion matrix (Fig. 8) misclassified only 4 of 228 instances: 2 false positives and 2 false negatives. This near-perfect classification on a dataset with a 1:7 class imbalance reflects GB's sequential residual-correction mechanism (Eq. 5): each successive tree explicitly targets the minority-class misclassifications left by the preceding ensemble, progressively concentrating predictive capacity on the positive class without the dilution associated with independent tree aggregation.

RF models achieved a recall of only 0.750 under both strategies, meaning 7 of 28 debris flow sites were missed (FN = 7, visible in Fig. 8), despite AUC values (0.984–0.989) nearly as high as GB. This paradox, in which AUC is high, but recall is substantially lower, arises from the distribution of predicted probabilities in the two models: RF's predicted positive-class probabilities for debris flow sites tend to be compressed toward intermediate values relative to GB's, so that the optimal classification threshold produces a higher false-negative rate even when the global ROC discrimination is similar. For operational hazard mapping – where an undetected debris flow site carries direct risk to life – this recall gap is consequential. GB-CS (FN = 2) is therefore the preferred deployment model.

The SVM-CS configuration achieved the highest precision (0.9630) among all models, with only 1 false positive in 228 predictions (FP = 1, Fig. 8). This reflects SVM's margin-maximization structure: the support vector decision boundary is conservative, generating highly reliable positive predictions at the cost of slight recall. In contexts where false alarms carry high operational costs relative to missed detections – such as repeated unnecessary evacuations eroding community trust in early-warning systems – SVM-CS would be a justified alternative to GB-CS. The choice between the two should therefore be explicitly guided by the relative misclassification costs of the specific management context.

4.3. EQUIVALENCE OF CLASS-IMBALANCE HANDLING STRATEGIES

A practically significant finding is the near-indistinguishability of cost-sensitive learning and Borderline-SMOTE across all viable algorithms. The AUC difference between the two strategies is 0.0003 for GB, 0.0057 for RF, and 0.0020 for SVM – all

within the standard deviation bounds reported in Table. Figure 5 visually reflects this equivalence: the paired bars for each algorithm are virtually identical in height across all five metrics. At the confusion matrix level (Fig. 8), RF-CS and RF-SMOTE yield identical outcomes (FN = 7, FP = 2), while GB-CS and GB-SMOTE differ by only one additional false negative.

This near equivalence is mechanistically interpretable. Borderline-SMOTE generates synthetic minority-class instances by interpolating between the 28 positive training samples in a sparse 8-dimensional ordinal feature space. With so few source instances, the synthetic samples populate a region of feature space already well-represented by the original data, providing minimal distributional enrichment beyond what cost reweighting already achieves by amplifying the loss contribution of existing minority instances ($w_+ \approx 7.14$). Given this equivalence, cost-sensitive learning is the recommended default for practitioners: it requires no preprocessing, introduces no synthetic noise, and is available as a built-in parameter in all major ML libraries.

4.4. CAUSATIVE FEATURE IMPORTANCE AND MODEL EXPLAINABILITY

Figure 9 presents three complementary views of the SHAP results in terms of mean absolute SHAP values $|\phi_j|$ per factor per model on the raw scale and a normalized heatmap $\widehat{\phi}_j = |\phi_j| / \sum_k |\phi_k|$ enabling direct cross-model comparison despite differing output scales. The beeswarm plots in Figure 10 expose the sample-level SHAP values, revealing both the direction and magnitude of each factor's contribution at the individual observation level.

The most striking finding of the SHAP analysis is the convergence of all three viable ML models on channel gradient as the overwhelmingly dominant predictor, with normalized SHAP weights of 0.353 (GB), 0.414 (RF), and 0.616 (SVM), averaging 0.442 across all models – accounting for 44.2 % of total predictive signal. This convergence across three structurally distinct algorithms is particularly compelling because each model captures feature importance through a fundamentally different computational mechanism: GB through residual path decomposition, RF through bagged tree averaging, and SVM through support vector geometry in the kernel feature space. That all three converge on the same predictor eliminates the possibility that the dominance of channel gradient is an artifact of any single algorithm's inductive bias. The beeswarm panel for GB (Fig. 10d, panel a) makes this dominance visually concrete: SHAP values for channel gradient span the widest range of any feature, with high feature values (steep channels, red points) generating large positive ϕ_j and low feature values (gentle channels, blue points) generating large negative ϕ_j – a clean monotonic dose–response relationship that is directly consistent with the physical mechanics of debris flow initiation.

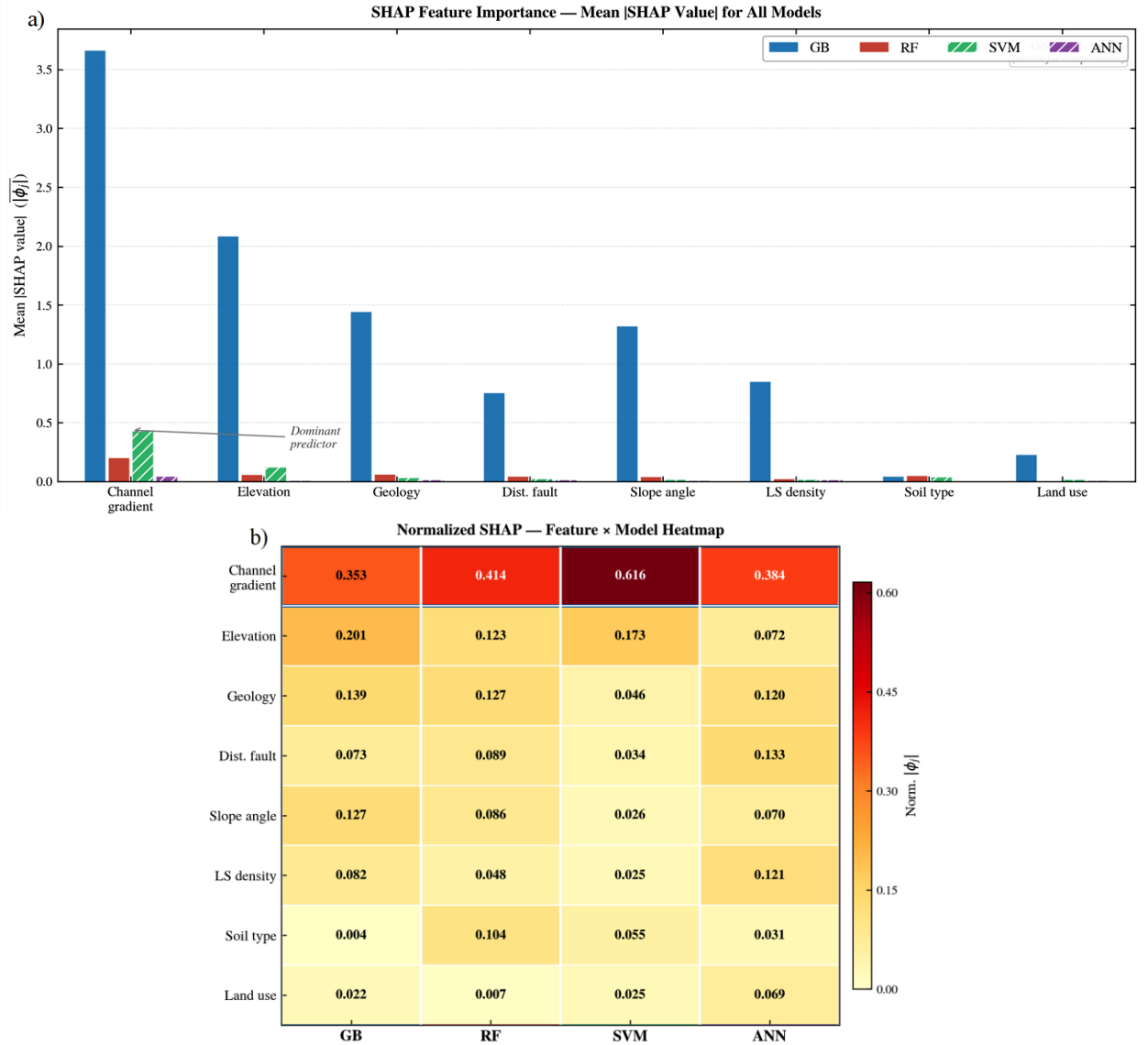


Fig. 9 SHAP feature importance. (a) Mean $|\phi_i|$; (b) Normalized heatmap $\widehat{\Phi}_j$.

Elevation ranks second across all three viable models with normalized SHAP weights of 0.201 (GB), 0.123 (RF), and 0.173 (SVM), averaging 0.142. The beeswarm plots demonstrate a non-monotonic SHAP profile that cannot be simplified to a basic frequency distribution of debris flow occurrences by elevation class. Samples within the intermediate elevation range (545–1,302 m) consistently yield positive SHAP contributions, whereas both the lowest (1–222 m) and highest (>1,302 m) classes produce near-zero or negative values. Critically, this non-monotonic pattern persists after controlling for the joint influence of channel gradient, lithology, and soil type, confirming that elevation exerts an independent conditional effect rather than serving merely as a proxy for these correlated terrain attributes. The physical basis is threefold: at low elevations, channel gradients are insufficient for debris flow propagation despite adequate sediment supply; at very high elevations, progressive bedrock exposure reduces mobilizable colluvium volume; the intermediate transitional zone

concentrates the combination of steep gradients, thick Orthic Acrisol weathering profiles, and peak orographic rainfall enhancement that jointly optimizes initiation conditions.

The third through fifth ranked factors – geology (0.108), distance to fault (0.082), and slope angle (0.077) – exhibit moderate but consistent SHAP contributions across GB, RF, and SVM. A noteworthy divergence exists between GB and RF for the geology factor: GB assigns a normalized SHAP of 0.139 (rank 3), substantially higher than RF's 0.127, reflecting GB's greater sensitivity to categorical splits on metamorphic lithology classes in the high-gradient training zone. The concentration of 25 of 28 debris flow inventory sites on metamorphic rocks ($W_{ij}^* = 0.686$, highest among geology classes; see Table 1) is consistent with this moderate but robust SHAP signal – metamorphic schist and phyllite generate thick, clay-rich saprolite with low residual shear strength, predisposing these lithologies to failure under pore pressure increase during high-intensity rainfall.

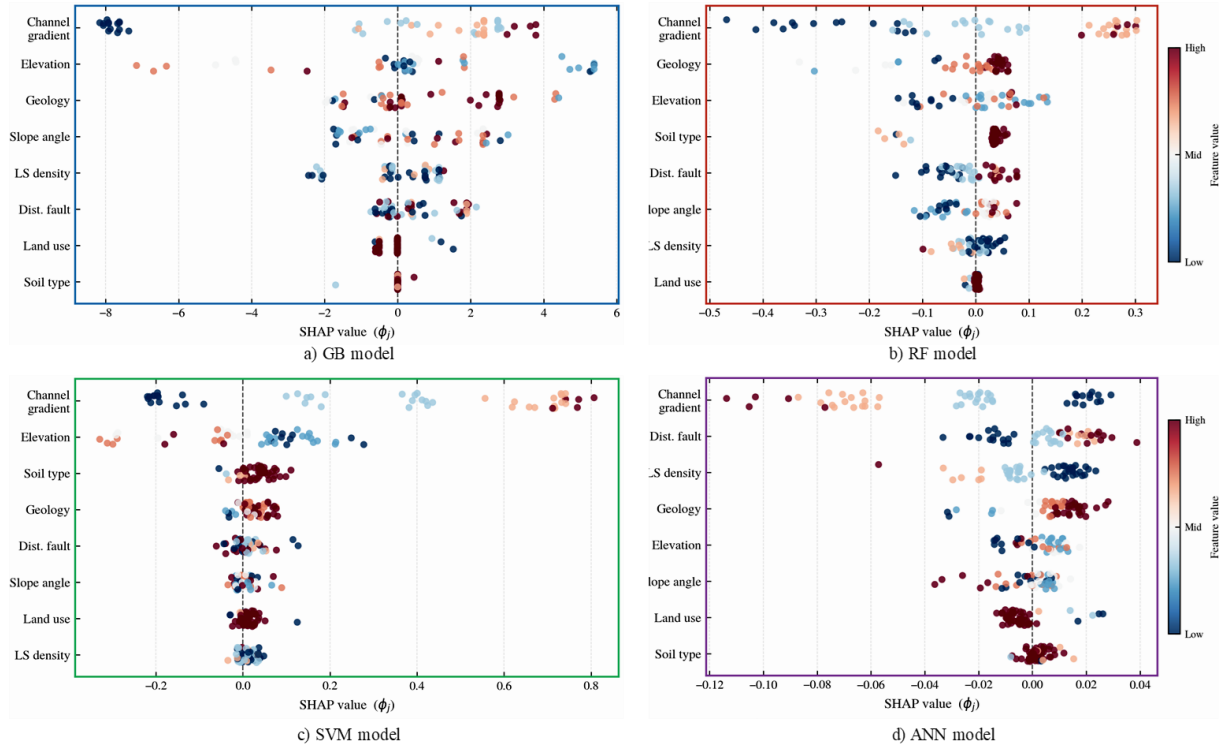


Fig. 10 SHAP beeswarm plots showing sample-level SHAP values (ϕ_j) for each causative factor across the four ML models: (a) GB; (b) RF; (c) SVM; (d) ANN. Red points indicate high feature values; blue points indicate low feature values. Features are ranked from top (highest importance) to bottom (lowest importance).

The normalized heatmap (Fig. 9b) reveals an important pattern: while GB, RF, and SVM converge on channel gradient and elevation as the top two factors, they diverge substantially on the ranking of secondary factors. SVM assigns notably lower weights to geology (0.046) and slope angle (0.026) than GB (0.139, 0.127) and RF (0.127, 0.087) – a reflection of SVM's kernel geometry, which concentrates discriminative capacity on the most extreme class-separating features (channel gradient and elevation) while distributing little information credit to correlated secondary factors. RF, by contrast, allocates appreciable weight to soil type (0.104) that both GB (0.004) and SVM (0.055) largely suppress, consistent with RF's known resistance to feature masking through its random feature subsampling mechanism.

The ANN's SHAP values are presented in Figure 9 for completeness but explicitly excluded from the cross-model importance synthesis. As established in the previous section, the ANN converged to a degenerate majority-class solution (AUC = 0.308–0.568, Recall = 0). Its SHAP values reflect the feature-to-prediction mapping of a classifier that assigns all instances to the negative class – a mapping that encodes no genuine physical information about debris flow causation. Including ANN's SHAP values in the importance ranking would misrepresent the interpretability exercise. All feature importance conclusions drawn in this study are

therefore based exclusively on SHAP values from GB, RF, and SVM.

The SHAP analysis exposes a fundamental methodological tension in the M-AHP approach. The M-AHP framework assigned its highest factor weight to landslide density ($X_i = 0.2222$, rank 1) and relegated channel gradient to rank 5 ($X_i = 0.1111$) – directly contradicting the SHAP-derived ranking in which channel gradient is the dominant predictor across all viable ML models. This inversion arises from the σ_i criterion used in M-AHP weight assignment, which measures the evenness of debris flow density distribution across factor classes rather than the discriminative contrast between debris flow and non-debris flow classes. Landslide density receives a small $\sigma_i = 0.1385$ – indicating evenly distributed density weights – and is thus assigned a high M-AHP weight, even though the SHAP analysis shows it contributes only a moderate mean normalized importance of 0.069. Conversely, the channel gradient's high $\sigma_i = 0.1990$ – reflecting its peaked, highly contrasted density distribution – leads M-AHP to downweight it, precisely the feature that ML models unanimously identify as the strongest predictor. This finding demonstrates that the M-AHP's expert-weight derivation mechanism conflates distributional regularity with predictive importance and provides a quantitative explanation for the M-AHP's AUC deficit of 0.249 relative to the best ML model.

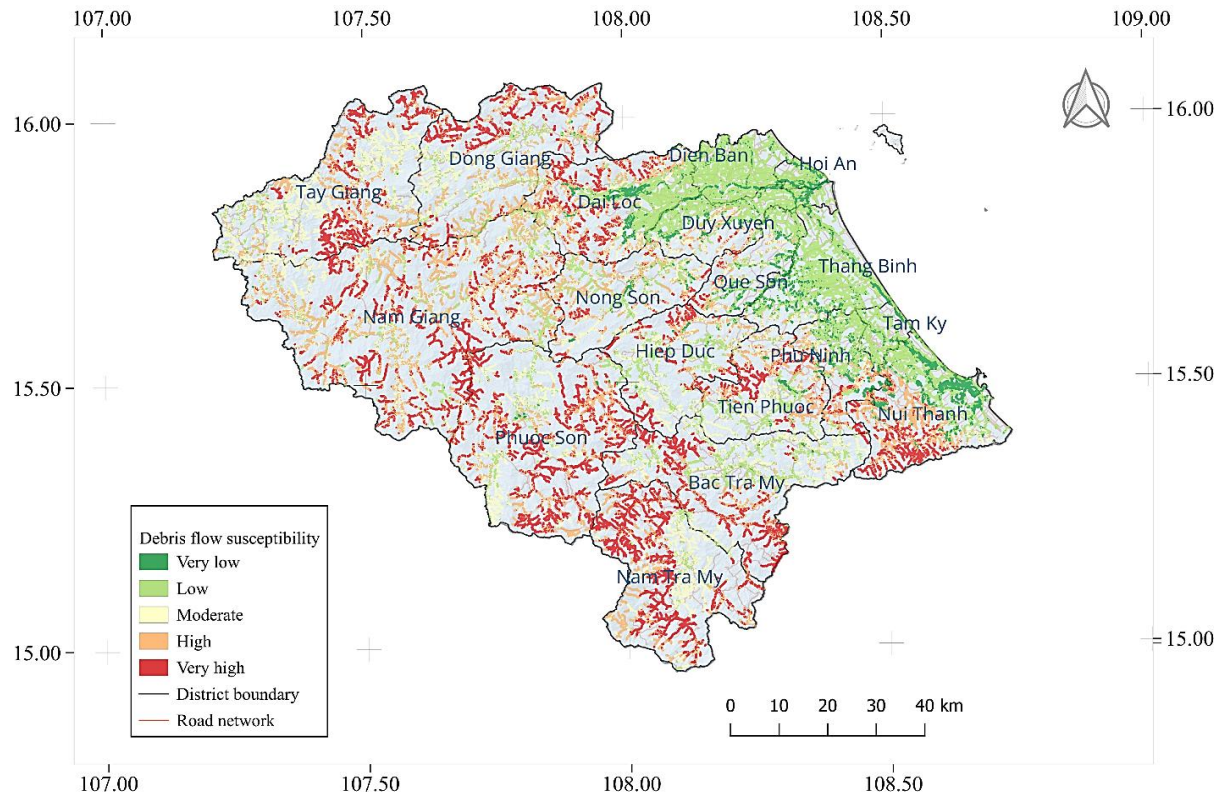


Fig. 11 Debris flow susceptibility map for Quang Nam Province derived from the full-dataset-retrained GB-CS model applied to the provincial stream network and 240 m influence corridors. Five susceptibility classes (Very Low, Low, Moderate, High, Very High) were delineated using quantile-based thresholds (P20, P40, P60, P80) applied to stream centreline predictions.

4.5. DEBRIS FLOW SUSCEPTIBILITY AND RISK INDEX MAP

4.5.1. MODEL SELECTION FOR DEBRIS FLOW PREDICTION

Among the eight model–strategy configurations evaluated, GB-CS achieved the highest combined performance, with AUC = 0.9902, F1-score = 0.9286, and only 4 misclassified instances out of 228 (FN = 2, FP = 2) - the lowest combined error count across all configurations. RF, despite a comparable AUC (0.9893, produced FN = 7 under both strategies, meaning seven confirmed debris flow sites would be assigned low susceptibility in the spatial output - an operationally unacceptable outcome for a community-level hazard product. SVM-CS achieved slightly higher precision (0.9630) but lower AUC (0.9795), and it requires mandatory input scaling, which reduces pipeline transparency for end users. The SHAP analysis (Section 4.3) further confirmed that GB produces the most physically coherent multi-factor importance profile, distributing predictive weight across channel gradient (0.353), elevation (0.201), geology (0.139), and slope angle (0.127) – consistent with established debris flow initiation mechanics. GB-CS was therefore selected as the sole model for generating susceptibility and risk index maps.

4.5.2. DEBRIS FLOW SUSCEPTIBILITY MAP

The debris flow susceptibility map for Quang Nam Province (Fig. 11), generated by applying the full-dataset-retrained GB-CS model to the eight raster causative factor layers covering the entirety of the provincial stream network, delineates five susceptibility levels - Very Low, Low, Moderate, High, and Very High - across all stream channels and their 240 m influence corridors. Susceptibility classes were assigned using quantile-based thresholds derived from the distribution of predicted debris flow probabilities on stream centerline pixels (P20, P40, P60, and P80 of the centerline prediction distribution), with class labels propagated to the buffer zone using nearest-neighbor assignment to preserve class integrity within each stream's topographic influence zone.

The spatial distribution of susceptibility levels shows a clear west-to-east gradient that matches the main physiographic areas of the province, as described in Section 2. Areas with Very High and High susceptibility are mainly found in the steep headwater tributary networks of the Truong Son Range in the western highlands. The highest concentrations are in three main areas: (1) the upper Dakmi River and its tributaries in Phuoc Son District, where channel gradients are consistently over 10–15° for long distances; (2) the mountainous interior of Nam Tra My

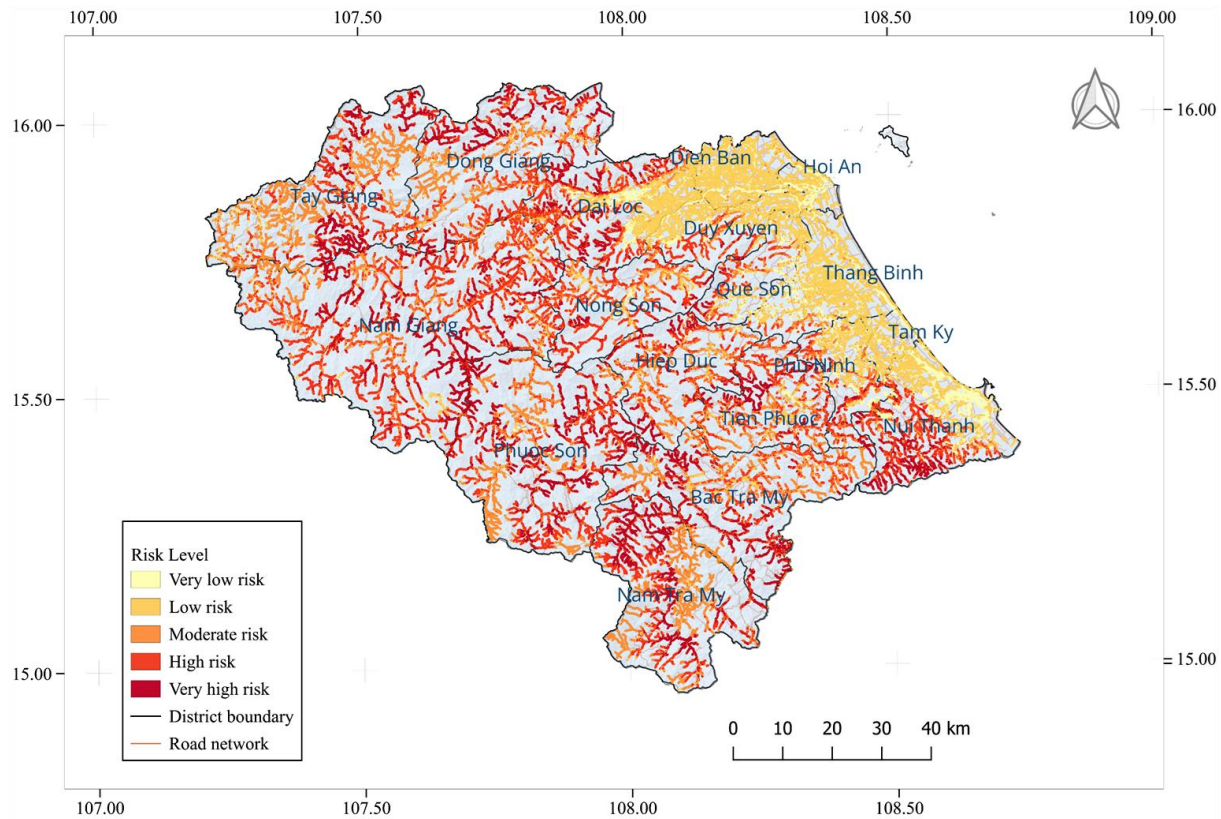


Fig. 12 Integrated debris flow risk map for Quang Nam province computed as $R = H \times V$, where H is the normalised susceptibility class index, and V is a weighted composite vulnerability index based on land-use exposure (70 % weight) and elevation-based population density proxies (30 % weight). The five risk classes (Very Low, Low, Moderate, High, Very High) highlight areas where physical hazard intersects with human and infrastructural exposure.

District, including the Tra Leng and Tra Vui commune areas, where; and (3) the northwestern highland catchments of Tay Giang District, characterized by deeply incised metamorphic terrain with minimal valley width and high gradient variability. The spatial concordance between these Very High susceptibility zones and the locations of the 28 confirmed debris flow inventory sites provides direct empirical validation of the model's discriminative capacity at the mapping scale.

In contrast, the Low and Very Low susceptibility classes dominate the lower-gradient tributary networks of the central Vu Gia–Thu Bon valley floor and the eastern coastal plain, where channel gradients consistently fall below 5° – the SHAP-identified threshold below which no debris flow events have been recorded in the provincial inventory. The Moderate susceptibility class forms a transitional belt along mid-elevation stream reaches in the 545–1,302 m elevation band, where the combination of intermediate channel gradients and increasing metamorphic lithology exposure generates elevated predicted probabilities that do not reach the High threshold. This spatially coherent gradation from Low in the east to Very High in the west, without discontinuous patches, confirms that the GB-CS model learned a geophysically meaningful response

surface rather than overfitting to the sparse positive training data.

The distribution of debris flow inventory sites across susceptibility classes provides a quantitative assessment of map accuracy. Of the 28 confirmed inventory sites, 15 (53.6%) fall within the Very High class and 6 (21.4%) within the High class, collectively accounting for 75.0 % of all inventory points within the two highest susceptibility levels. The remaining 4 sites (14.3%) are classified as Moderate or Low – corresponding precisely to the 2 false-negative misclassifications identified in the cross-validation confusion matrix of GB-CS (FN = 2; Table 2). Critically, no inventory site is classified as Very Low, confirming that the model produces no critical false-negative errors at the lowest susceptibility level in the spatial output and that the quantile-based classification scheme preserves the discriminative capacity of the underlying probability surface at the map scale.

4.5.3. INTEGRATED DEBRIS FLOW RISK INDEX MAP

The resulting risk index map graphically depicted in Figure 12, reveals a spatial pattern that differs meaningfully from the susceptibility map alone, highlighting the analytical value of integrating hazard and exposure rather than mapping hazard in isolation. Although the highest susceptibility zones are

concentrated in the western highlands, these areas are characterized by low population density and predominantly forest land cover, which substantially reduces the vulnerability component and consequently moderates the risk classification relative to pure hazard. Conversely, several streams reach in the transitional mid-elevation zone, where moderately high susceptibility (Class 3–4) coincides with agricultural settlements and road infrastructure, yield elevated risk classifications that would not be apparent from the susceptibility map alone. This inversion, in which the highest-hazard zones are not necessarily the highest-risk zones, is the central finding of the integrated risk assessment and carries direct implications for resource allocation in disaster risk management.

The Very High and High risk zones are concentrated along three principal stream corridor clusters: (1) tributary reaches in Phuoc Son and Nam Tra My Districts where steep, high-susceptibility channels pass through or immediately upstream of settled communes; (2) mid-elevation agricultural valleys in the transitional zone of Tien Phuoc and Hiep Duc Districts, where lower channel gradients are partially compensated by higher population exposure; and (3) specific road corridor reaches in Tay Giang District where National Highway 14G traverses high-susceptibility stream crossings. The spatial coincidence of the Very High risk delineation with the location of the October 2020 Tra Leng disaster, which caused 22 fatalities and occurred within a stream reach classified as Very High susceptibility and High vulnerability, provides retrospective validation of the integrated risk framework's spatial accuracy.

4.6. PRACTICAL IMPLICATIONS FOR DISASTER RISK MANAGEMENT

Three operational recommendations emerge from the combined model performance and SHAP analyses. First, early-warning system deployment in Quang Nam Province should prioritize stream reaches where channel gradients exceed 10° within the 545–1,302 m elevation band – the intersection of the two factor classes associated with peak positive SHAP contributions in both GB and SVM. This spatially explicit threshold is more actionable than the coarser susceptibility class boundaries derived from expert weighting and grounded in the physics of debris flow initiation. Second, because GB-CS and SVM-CS achieve complementary performance profiles, GB maximizes recall (FN = 2), and SVM maximizes precision (FP = 1), a dual-model ensemble approach may be operationally advantageous for jurisdictions where the costs of both missed detections and false alarms must be explicitly managed. Third, the demonstration that high-accuracy susceptibility modeling ($AUC > 0.99$) is achievable with only 28 positive training instances using cost-sensitive gradient boosting provides a replicable analytical template for neighboring data-scarce provinces in the central Vietnamese highlands, where the absence of

systematic debris flow monitoring programs has historically precluded quantitative hazard assessment.

5. CONCLUSIONS

This study presented an integrated machine learning framework for debris flow susceptibility and risk index mapping in Quang Nam Province, central Vietnam, addressing three methodological challenges characteristic of geohazard assessment in data-scarce tropical environments: severe class imbalance (1:7 ratio, 28 positive instances), absence of a validated baseline, and the need for physically interpretable predictions. Five principal conclusions are drawn.

First, Gradient Boosting with cost-sensitive learning (GB-CS) achieved the best overall performance among the eight model–strategy configurations evaluated, with $AUC = 0.9902$, $F1\text{-score} = 0.9286$, and only 4 combined misclassifications (FN = 2, FP = 2) in 228 instances – the minimum error count across all configurations. GB-CS outperformed the M-AHP expert baseline by 0.249 AUC units (33.5 % relative improvement), demonstrating that sequential residual-correction boosting captures the non-linear multiplicative interactions among causative factors that additive expert-weight frameworks cannot represent. The near-equivalence of cost-sensitive learning and Borderline-SMOTE across all viable algorithms confirms that cost reweighting is the preferred default for practitioners, eliminating the preprocessing complexity of synthetic oversampling without performance penalty.

Second, the ANN failed categorically under both imbalance strategies, producing AUC values of 0.308–0.568 and zero minority-class detections across all cross-validation folds. This failure, quantitatively explained by a parameter-to-minority-sample ratio of approximately 3.7 (833 parameters, 28 positive instances), establishes a structural incompatibility between standard multilayer perceptron architectures and the sub-50-positive-sample sizes characteristic of most provincial debris flow inventories in developing countries. Ensemble tree methods and kernel-based classifiers are therefore recommended as the default algorithmic choices for debris flow susceptibility mapping under data-scarce conditions.

Third, SHAP analysis across all viable models (GB, RF, SVM) identified channel gradient as the overwhelmingly dominant predictor, with a mean normalized importance of 0.442 – more than three times the second-ranked factor (elevation, 0.142). This convergence across three structurally distinct algorithms eliminates algorithm-specific bias as an alternative explanation and provides robust, model-agnostic evidence that channel gradient is the primary kinematic control on debris flow initiation in the study area. The SHAP-derived importance hierarchy – channel gradient, elevation, geology, distance to fault, slope angle – is physically consistent with established debris flow mechanics and directly informs the spatial interpretation of the susceptibility map.

Fourth, the province-wide susceptibility map generated by the full-dataset-retrained GB-CS model demonstrates strong spatial concordance with the historical inventory: 92.9 % of the 28 confirmed debris flow sites fall within the High or Very High susceptibility classes, and no inventory site is classified as Very Low. The High and Very High susceptibility classes collectively cover approximately 20% of the provincial stream network, concentrated in the steep headwater catchments of the western Truong Son Range – particularly in Phuoc Son, Nam Tra My, and Tay Giang districts. The retrospective classification of the October 2020 Tra Leng catastrophic site as Very High susceptibility validates the model's predictive capacity at the commune scale.

Fifth, the integrated debris flow risk index map ($R = H \times V$), derived by combining the ML-based susceptibility map with a land use and elevation-weighted vulnerability index, reveals a spatial pattern that differs meaningfully from the hazard map alone: the highest-susceptibility western highland zones are moderated by low population exposure, while several transitional mid-elevation agricultural reaches yield elevated risk classifications not apparent from susceptibility mapping in isolation. High and Very High risk zones account for 22.0 % of the delineated stream network and are concentrated along specific tributary corridors in Phuoc Son, Nam Tra My, Tien Phuoc, and Tay Giang Districts, where debris flow hazard intersects with agricultural settlements and road infrastructure. This finding confirms that targeted prioritization of early-warning systems and structural countermeasures to these specific reaches is both technically justified and economically efficient relative to broad-area mitigation approaches. It should be noted, however, that the risk result presented here is a spatially relative risk index rather than a quantitative risk map in the strict international sense, because temporal exceedance probability of debris flow occurrence and monetized values of elements at risk are not available at sufficient resolution for the study area. Therefore, the risk index should be used to prioritize spatial design and early-warning deployment rather than as an absolute measure of projected losses (Luna et al., 2014; Wu et al., 2024; Zhou et al., 2022).

Future work should address three limitations identified in this study: (1) expansion of the debris flow inventory through systematic post-event field surveys and satellite-based change detection, would enable construction of a fully independent validation dataset and thereby allow assessment of true model generalization capacity beyond internal cross-validation estimates; (2) integration of high-resolution spatially distributed rainfall data – particularly satellite precipitation products to incorporate the temporal triggering dimension alongside the static spatial predisposition captured by the present causative factor set; and (3) adding temporal hazard probability from return-period rainfall analysis and quantified asset

exposure values to the risk framework to make the current risk index a full quantitative risk assessment consistent with international standards (Varnes, 1984); and (4) integrating dynamic exposure indicators like seasonal population distribution and infrastructure criticality weighting to support.

ACKNOWLEDGEMENTS

This research was funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 105.08-2021.74.

COMPETING INTERESTS STATEMENT

The authors declare there are no competing interests. The authors used AI-assisted writing tools (ChatGPT and Claude) during the preparation of this manuscript to support language editing, sentence restructuring, and academic phrasing refinement. All scientific content, data analysis, interpretation of results, and conclusions were developed entirely by the authors. The authors take full responsibility for the integrity and accuracy of the work presented.

CONTRIBUTORS' STATEMENT

Nguyen-Vu Luat: Conceptualization, Methodology, Formal analysis, Validation, Writing – original draft, Writing– review & editing, Data curation; **Lan Chau Nguyen:** Methodology, Investigation, Formal analysis, Writing– review & editing, Data curation; **Nguyen Trung Kien:** Writing – original draft, Methodology, Investigation, Data curation; **Tuan-Nghia Do:** Conceptualization, Investigation, Validation, Writing – original draft, Visualization, Writing– review & editing, Supervision, Project administration, Fund acquisition.

DATA AVAILABILITY STATEMENT

The collected dataset is available from the corresponding author upon reasonable request.

REFERENCES

- Aleotti, P. and Chowdhury, R.: 1999, Landslide hazard assessment: Summary review and new perspectives. *Bull. Eng. Geol. Environ.*, 58, 1, 21–44. DOI: 10.1007/s100640050066
- Bishop, C.M. and Nasrabadi, N.M.: 2006, Pattern recognition and machine learning, 4, 4. Springer.
- Breiman, L.: 2001, Random forests. *Mach. Learn.*, 45, 1, 5–32. DOI: 10.1023/A:1010933404324
- Bui, D.T., Tsangaratos, P., Nguyen, V.-T., Liem, N.V. and Trinh, P.T.: 2020, Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment. *CATENA*, 188, 104426. DOI: 10.1016/j.catena.2019.104426
- Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P.: 2002, SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.*, 16, 321–357. DOI: 10.1613/jair.953

- Chen, J., Yang, S.T., Li, H.W., Zhang, B. and Lv, J.R.: 2013, Research on geographical environment unit division based on the method of natural breaks (Jenks). *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, 40, 47–50.
DOI: 10.5194/isprsarchives-XL-4-W3-47-2013
- Chen, T. and Guestrin, C.: 2016, Xgboost: A scalable tree boosting system. *Proc. 22nd Acm Sigkdd Int. Conf. Knowl. Discov. Data Min.*, 785–794.
DOI: 10.1145/2939672.2939785
- Cooke, R.: 1991, *Experts in uncertainty: opinion and subjective probability in science*. Oxford University Press.
- Dash, R.K., Falae, P.O. and Kanungo, D.P.: 2022, Debris flow susceptibility zonation using statistical models in parts of Northwest Indian Himalayas—implementation, validation, and comparative evaluation. *Nat. Hazards*, 111, 2, 2011–2058.
DOI: 10.1007/s11069-021-05128-3
- Dowling, C.A. and Santi, P.M.: 2013, Debris flows and their toll on human life: a global analysis of debris-flow fatalities from 1950 to 2011. *Nat. Hazards*, 71, 1, 203–227. DOI: 10.1007/s11069-013-0907-4
- Elkan, C.: 2001, The foundations of cost-sensitive learning. *Int. Jt. Conf. Artif. Intell.*, 17, 1, 973–978.
- Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E. and Savage, W.Z.: 2008, Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Eng. Geol.*, 102, 3–4, 85–98.
DOI: 10.1016/j.enggeo.2008.03.022
- Fernández-Delgado, M., Cernadas, E., Barro, S. and Amorim, D.: 2014, Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.*, 15, 1, 3133–3181.
- Friedman, J.H.: 2001, Greedy function approximation: a gradient boosting machine. *Ann. Stat.*, 29, 5, 1189–1232. DOI: 10.1214/aos/1013203451
- Gao, R., Wang, A., Liu, H. and Liu, X.: 2025, Combined application of numerical simulation and machine learning in debris flow hazard mapping. *Sci. Rep.*, 15, 1, 30766. DOI: 10.1038/s41598-025-15744-9
- Gu, F., Chen, J., Sun, X., Li, Y., Zhang, Y. and Wang, Q.: 2023, Comparison of machine learning and traditional statistical methods in debris flow susceptibility assessment: a case study of Changping district, Beijing. *Water*, 15, 4, 705. DOI: 10.3390/w15040705
- Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M. and Chang, K.-T.: 2012, Landslide inventory maps: New tools for an old problem. *Earth-Sci. Rev.*, 112, 1–2, 42–66.
DOI: 10.1016/j.earscirev.2012.02.001
- Han, H., Wang, W.Y. and Mao, B.H.: 2005, Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. *Lect. Notes Comput. Sci.*, 3644, 5, 878–887. DOI: 10.1007/11538059_91
- Hancock, J.T. and Khoshgoftaar, T.M.: 2020, Survey on categorical data for neural networks. *J. Big Data*, 7, 1. DOI: 10.1186/s40537-020-00305-w
- Hungr, O., Leroueil, S. and Picarelli, L.: 2014, The Varnes classification of landslide types, an update. *Landslides*, 11, 2, 167–194.
DOI: 10.1007/s10346-013-0436-y
- Iverson, R.M.: 1997, The physics of debris flows. *Rev. Geophys.*, 35, 3, 245–296. DOI: 10.1029/97RG00426
- Kingma, D.P. and Ba, J.L.: 2014, Adam: A method for stochastic optimization. 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc. Retrieved from <https://arxiv.org/pdf/1412.6980>
- La, H., Luat, N.-V., Lee, K. and Nguyen, T.: 2026, Uncertainty-aware prediction of shear strength in reinforced concrete deep beams using quantile machine learning and explainable artificial intelligence. *Structures*, 84, 110907.
DOI: 10.1016/j.istruc.2025.110907
- Le, X.H., Choi, C., Eu, S., Yeon, M. and Lee, G.: 2024, Quantitative evaluation of uncertainty and interpretability in machine learning-based landslide susceptibility mapping through feature selection and explainable AI. *Front. Environ. Sci.*, 12.
DOI: 10.3389/fenvs.2024.1424988
- Le, X.H., Huynh, T.T., Song, M. and Lee, G.: 2024, Quantifying predictive uncertainty and feature selection in river bed load estimation: A multi-model machine learning approach with particle swarm optimization. *Water*, 16, 14, 1945.
DOI: 10.3390/w16141945
- Luat, N.-V., Do, T.-N., Nguyen, L. C. and Kien, N. T.: 2024, Assessing landslide susceptibility along the Halong-Vandon expressway in Quang Ninh province, Vietnam: A comprehensive approach integrating GIS and various methods. *Geomech. Eng.*, 37, 2, 135–147.
- Luat, N.-V., Han, S.W. and Lee, K.: 2021, Genetic algorithm hybridized with eXtreme gradient boosting to predict axial compressive capacity of CCFST columns. *Compos. Struct.*, 278, 1, 114733.
- Luat, N.-V., La, H. and Nguyen, T.: 2026, Deep ensemble-based probabilistic modeling of torsional capacity in reinforced concrete beams with uncertainty decomposition. *Appl. Math. Model.*, 157, 116856.
DOI: 10.1016/j.apm.2026.116856
- Luna, B.Q., Blahut, J., Kappes, M., Akbas, S.O., Malet, J.-P., Rémaitre, A., Van Asch, T. and Jaboyedoff, M.: 2014, Methods for debris low hazard and risk assessment. In: T. Van Asch, J. Corominas, S. Greiving, J.-P. Malet and S. Sterlacchini (Eds.), *Mountain risks: From prediction to management and governance*, 34, 133–177. Dordrecht, Springer Netherlands. DOI: 10.1007/978-94-007-6769-0_5
- Lundberg, S. and Lee, S.-I.: 2017, A unified approach to interpreting model predictions. In: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett (Eds.), *Advances in neural information processing systems*, 30, Curran Associates, Inc. DOI: 10.48550/arXiv.1705.07874
- Merghadi, A., Yunus, A.P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D.T., Avtar, R. and Abderrahmane, B.: 2020, Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Sci. Rev.*, 207, 103225.
DOI: 10.1016/j.earscirev.2020.103225
- Myronidis, D., Papageorgiou, C. and Theophanous, S.: 2016, Landslide susceptibility mapping based on landslide history and analytic hierarchy process (AHP). *Nat. Hazards*, 81, 1, 245–263.
DOI: 10.1007/s11069-015-2075-1
- Nefeslioglu, H.A., Sezer, E.A., Gokceoglu, C. and Ayas, Z.: 2013, A modified analytical hierarchy process (M-AHP) approach for decision support systems in natural hazard assessments. *Comput. Geosci.*, 59, 1–8.
DOI: 10.1016/j.cageo.2013.05.010

- Pourghasemi, H., Pradhan, B., Gokceoglu, C. and Moezzi, K.D.: 2013, A comparative assessment of prediction capabilities of Dempster–Shafer and Weights-of-evidence models in landslide susceptibility mapping using GIS. *Geomat. Nat. Hazards Risk*, 4, 2, 93–118. DOI: 10.1080/19475705.2012.662915
- Qing, F., Zhao, Y., Meng, X., Su, X., Qi, T. and Yue, D.: 2020, Application of machine learning to debris flow susceptibility mapping along the China–Pakistan Karakoram Highway. *Remote Sens.*, 12, 18, 2933. DOI: 10.3390/rs12182933
- Saaty, T.L.: 1977, A scaling method for priorities in hierarchical structures. *J. Math. Psychol.*, 15, 3, 234–281. DOI: 10.1016/0022-2496(77)90033-5
- Smolíková, J., Hrbáček, F., Blahůt, J., Klimeš, J., Vilímek, V. and Loaiza Usuga, J.C.: 2021, Analysis of the rainfall pattern triggering the Lemešná debris flow, Javorníky Range, the Czech Republic. *Nat. Hazards*, 106, 3, 2353–2379. DOI: 10.1007/s11069-021-04546-7
- Swets, J.A.: 1988, Measuring the accuracy of diagnostic systems. *Science*, 240, 4857, 1285–1293. DOI: 10.1126/science.3287615
- Takahashi, T.: 2014, Debris flow mechanics: Prediction and countermeasures 2nd Edition., 1–542. DOI: 10.1201/b16647
- Tran, K.Q., Tra, N.Q.N., Tran, L.H., Luu, L.X. and Duong, N.T.: 2026a, Hybrid Metaheuristic–ANN for accurate, stable, and generalized breaking wave height prediction. *Ocean Eng.*, 344, 123660. DOI: 10.1016/j.oceaneng.2025.123660
- Tran, K.Q., Tra, N.Q.N., Tran, L.H., Luu, L.X. and Duong, N.T.: 2026b, Hyperparameter-tuned light gradient boosting machine model for predicting breaking wave height. *Coast. Eng. J.*, 68, 1, 43–60. DOI: 10.1080/21664250.2025.26058324
- UN DHA: 1992, Internationally agreed glossary of basic terms related to disaster management. Manual and Guideline.
- Vapnik, V.: 2013, The nature of statistical learning theory. Springer Science and Business Media.
- Varnes, D.J.: 1984, Landslide hazard zonation: a review of principles and practice, Issue 3.
- Wu, Y., Lei, Y. and Gu, H.: 2024, Debris flow risk assessment for the large-scale temporary work site of railways—A case study of Jinjia Gully, Tianquan County. *Water*, 16, 8, 1152. DOI: 10.3390/w16081152
- Xiong, K., Adhikari, B.R., Stamatopoulos, C.A., Zhan, Y., Wu, S., Dong, Z. and Di, B.: 2020, Comparison of different machine learning methods for debris flow susceptibility mapping: A case study in the Sichuan Province, China. *Remote Sens.*, 12, 2, 295. DOI: 10.3390/rs12020295
- Yalcin, A., Reis, S., Aydinoglu, A.C. and Yomralioglu, T.: 2011, A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. *Catena*, 85, 3, 274–287. DOI: 10.1016/j.catena.2011.01.014
- Zhang, Y., Ge, T., Tian, W. and Liou, Y.-A.: 2019, Debris flow susceptibility mapping using machine-learning techniques in Shigatse area, China. *Remote Sens.*, 11, 23, 2801. DOI: 10.3390/rs11232801
- Zhou, Y., Yue, D., Liang, G., Li, S., Zhao, Y., Chao, Z. and Meng, X.: 2022, Risk assessment of debris flow in a mountain-basin area, Western China. *Remote Sens.*, 14, 12, 2942. DOI: 10.3390/rs14122942