Dynamic Landslide Susceptibility Assessment Using Machine Learning Models

Yuya Tanaka [1], Shintaro Goto [2]

[1] Graduate School of Geo-Environmental Science, Rissho University - 253h10001@rissho-univ.jp
[2] Faculty of Geo-Environmental Science, Rissho University - got@ris.ac.jp

Keywords: Landslide susceptibility, Dynamic triggering factors, Spatial cross-validation, Class imbalance, Explainable AI.

Abstract

Landslide susceptibility assessments have traditionally used static rainfall statistics that do not reflect the actual meteorological conditions when slopes fail. This study develops a machine learning framework that aligns high-resolution radar rainfall (XRAIN, 250 m / 10 min) and modeled soil moisture (XSWI) with documented landslide occurrence times as dynamic triggering factors. Applied to the Heavy Rain Event of July 2018 in Hiroshima Prefecture, the framework combines watershed-based spatial cross-validation, systematic comparison of four class imbalance strategies (no treatment, sample weighting, random under-sampling, and SMOTE) across eight algorithms (XGBoost, LightGBM, CatBoost, HGBoost, Random Forest, Balanced Random Forest, Easy Ensemble Classifier, and Logistic Regression), and spatially explicit SHAP interpretation. Two key findings emerged. First, soil moisture state — not instantaneous rainfall intensity — was the dominant triggering predictor: XSWI variables ranked 2nd and 3rd in importance after slope angle, operating as independent axes (r = 0.074). The no-treatment condition consistently outperformed all resampling strategies across algorithms. Second, spatial SHAP mapping revealed that predisposing factors produce time-invariant contribution patterns governed by terrain, while dynamic triggers produce event-specific patterns reflecting rainfall distribution; their spatial overlap identifies the highest-risk locations. Time-series susceptibility maps confirmed that the framework captures within-event risk evolution as rainfall progresses — a capability unattainable with static approaches. These results indicate that incorporating occurrence-time-aligned soil moisture dynamics substantially improves both the predictive and explanatory capacity of landslide susceptibility assessment.

1. Introduction

Rainfall-induced landslides cause severe human and economic losses worldwide; between 2004 and 2016, 55,997 fatalities were recorded across 4,862 fatal events (Froude and Petley, 2018). Increases in the frequency and intensity of extreme rainfall have been observed in association with climate change (IPCC, 2023), and a continued rise in landslide risk is anticipated. Against this background, advancing methods for spatially assessing landslide risk has become an urgent priority.

Landslide occurrence is generally understood to involve the interaction between predisposing factors, such as topography, geology, and soil properties, and triggering factors, such as rainfall. Based on this understanding, landslide susceptibility assessment (LSA) evaluates the spatial likelihood of landslide occurrence from predisposing and triggering factors and is widely employed as a foundation for disaster risk management. LSA methods are broadly categorized into knowledge-driven approaches based on expert judgment and data-driven approaches that extract statistical relationships from observational data (Reichenbach et al., 2018). In recent years, machine learning (ML) algorithms have achieved high predictive accuracy among data-driven approaches and have been increasingly applied to regional-scale spatial assessments (Lee and Lee, 2024; Wang et al., 2024). We have also undertaken data-driven LSA; however, linear models did not adequately capture the nonlinear interactions between predisposing and triggering factors, necessitating the adoption of ML approaches (Tanaka and Goto, 2024).

The hydrometeorological conditions that trigger landslides, however, act across multiple temporal scales, ranging from short-term rainfall intensity to medium-term antecedent rainfall and soil moisture dynamics that reflect cumulative infiltration. Most data-driven LSA studies have adopted aggregated precipitation values at low temporal resolution, such as annual or monthly totals, as triggering variables (Zhang et al., 2023; Lee and Lee, 2024). These aggregated values inadequately represent the physical reality that landslide occurrence is governed by rainfall over timescales of hours to days. Although the incorporation of soil moisture data has been reported to improve predictive performance (Palazzolo et al., 2025), studies that integrate both the intensity characteristics of rainfall events and the soil moisture state before and during events as dynamic triggering factors remain scarce (Halter et al., 2025).

Several methodological challenges have also been identified in ML-based LSA. First, the spatial autocorrelation inherent in geospatial data can lead to overestimation of generalization performance when cross-validation is conducted using random partitioning (Roberts et al., 2017). Second, in regional-scale spatial assessments, landslide locations constitute only a small fraction of the study area, resulting in pronounced class imbalance. This class imbalance substantially affects model prediction accuracy (Lombardo and Mai, 2018); however, few studies have systematically compared multiple sampling strategies to address this issue. Third, ML models inherently suffer from the black-box problem, in which prediction rationale is difficult to interpret. Although explainable AI (XAI) has recently attracted attention for quantifying feature contributions, few attempts have been made to visualize the spatial heterogeneity of these contributions on maps and to verify the geographic validity of prediction rationale.

Given these challenges, an ML framework for dynamic LSA targeting a single rainfall event is proposed. The framework addresses the following three methodological issues. First, instead of the annual or monthly precipitation aggregates predominantly used in previous studies, high-spatiotemporal-resolution rainfall data and soil moisture data linked to the rainfall event are introduced as dynamic triggering factors. This approach enables the model to reflect hydrometeorological conditions immediately preceding landslide occurrence. Second,

Category	Variable [unit]	Abbreviation	Resolution/Scale	Source
Topographic factors	Elevation [m]	Elevation	10 x 10 m	GSI [1]
	Slope angle [degree]	SlopeAng	10 x 10 m	GSI
	Slope length [m]	SlopeLeng	10 x 10 m	GSI
	Aspect type (*)	Aspect	10 x 10 m	GSI
	Plan curvature [m^{-1}]	PlanCurv	10 x 10 m	GSI
	Profile curvature [m^{-1}]	ProfCurv	10 x 10 m	GSI
	Topographic Position Index [m]	TPI	10 x 10 m	GSI
	Landform type (*)	Land	1:200,000	NLNI [2]
Geological factors	Distance to fault network [km]	DistFault	10 x 10 m	GSJ [3]
	Surface geology type (*)	Geo	1:200,000	NLNI
Land cover factors	Normalized Difference Vegetation Index	NDVI	10 x 10 m	Sentinel-2 [4]
	Normalized Difference Built-up Index	NDBI	10 x 10 m	Sentinel-2
	Vegetation type (*)	Veg	1:25,000	BCJ [5]
	Soil type (*)	Soil	1:200,000	NLNI
Hydrological factors	Flow accumulation [number of cells]	FlowAcc	10 x 10 m	GSI
	Topographic Wetness Index	TWI	10 x 10 m	GSI
	Distance to stream network [km]	DistStream	10 x 10 m	NLNI
Anthropogenic factors	Distance to road network [km]	DistRoad	10 x 10 m	OSM [6]
	Population density [persons km^{-2}]	PopDen	250 x 250 m	Census [7]
Triggering factors	XRAIN_now [mm h^{-1}]	XRAIN_now	250 x 250 m	MLIT [8]
	XSWI_now [mm]	XSWI_now	250 x 250 m	MLIT
	XSWI_delta_θh [mm] (**)	XSWI_delta_2h	250 x 250 m	MLIT

(*) Categorical variables. (**) Parameters (e.g., θ) were selected from multiple candidates and optimized.
[1] GSI: Geospatial Information Authority of Japan (https://service.gsi.go.jp/kiban/)
[2] NLNI: National Land Numerical Information (https://nlftp.mlit.go.jp/ksj/)
[3] GSJ: Geological Survey of Japan, AIST (https://gbank.gsj.jp/seamless/)
[4] Sentinel-2: Sentinel-2 satellite imagery (https://registry.opendata.aws/sentinel-2)
[5] BCJ: Biodiversity Center of Japan (https://www.biodic.go.jp/)
[6] OSM: OpenStreetMap (https://www.openstreetmap.org/)
[7] Census: 2020 Population Census (https://www.stat.go.jp/data/kokusei/2020/index.html)
[8] MLIT: Ministry of Land, Infrastructure, Transport and Tourism (https://search.diasjp.net/en/dataset/MLIT_XRAIN)

Table 1. Predisposing and triggering variables used for landslide susceptibility assessment.

watershed-based spatial cross-validation that accounts for spatial autocorrelation in geospatial data is adopted, and the effects of multiple sampling strategies for class imbalance on model performance are systematically compared to ensure the reliability of model evaluation. Third, spatial visualization of feature contributions based on XAI is employed to interpret prediction rationale at individual locations on maps, enabling verification of whether model predictions are geographically and physically valid.

2. Study Area and Materials

2.1 Study Area and Target Event

Figure 1 shows the geographic location, elevation distribution, river network, and landslide polygons from the Heavy Rain Event of July 2018 in Hiroshima Prefecture, the study area. The study area is situated between 34° 10′ N-34° 40′ N and 132° 20′ E-133° 30′ E. Elevations range from 0 to 1,300 m, and steep slopes with gradients of 30° or more are widespread. Geologically, the Late Cretaceous Hiroshima Granite Group underlies approximately 70% of the prefectural area. Its long-term weathering product, decomposed granite soil, exhibits low cohesion and poor erosion resistance, with a marked reduction in strength upon saturation (Miura and Yamanouchi, 1975). Owing

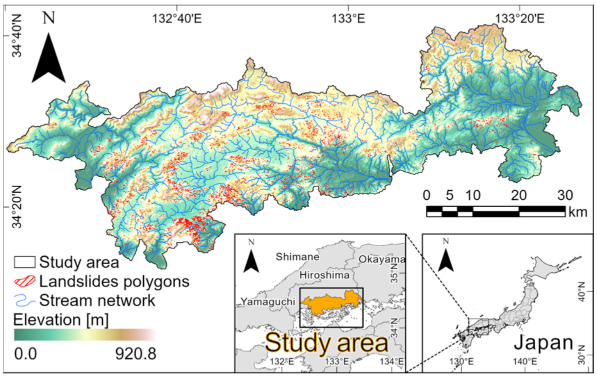

Figure 1. Location of the study area in Hiroshima, Japan, showing elevation, the stream network, and landslide polygons triggered by the July 2018 heavy rainfall.

to these topographic and geological predisposing factors, Hiroshima Prefecture has the highest number of designated sediment disaster warning zones in Japan, and exceptionally high landslide counts were recorded during the heavy rainfall events of 2014 and 2018 (Cabinet Office, Government of Japan, 2018).

This study targets the Heavy Rain Event of July 2018, which occurred from June 28 to July 8, 2018. The stagnation of a stationary Baiu front combined with the influx of warm, moist air from a typhoon produced a rainfall pattern in which sustained antecedent rainfall accumulated soil moisture before being followed by short-duration intense rainfall. Cumulative precipitation exceeded 400 mm over extensive portions of the study area (Shimpo et al., 2019).

2.2 Landslide Inventory

The landslide inventory used in this study was compiled through stereoscopic interpretation of aerial photographs taken after the disaster by the Geospatial Information Authority of Japan (GSI) (Hiroshima University July 2018 Heavy Rain Disaster Investigation Team (Geography Group), 2019). The landslide polygons are classified into source areas (mean area: approximately 490 m²) and depositional areas (mean area: approximately 2,105 m²). Because LSA generally targets the initiation point of failure, source areas were adopted as the landslide polygons for analysis. The study area was divided into 100-m grid cells; cells that spatially overlapped, even partially, with a source area were labeled as positive, and all remaining cells were labeled as negative.

Occurrence times were obtained from records estimated by the Sediment Control Division of Hiroshima Prefecture on the basis of reports from police, fire departments, and other emergency services (Hiroshima Prefecture Sediment Control Division, 2018). In this study, rainfall and soil moisture data corresponding to the occurrence time are extracted as dynamic triggering factors. Assigning a different triggering-factor timestamp to each individual landslide location, however, would mix meteorological conditions from different times across space, complicating spatial comparison within a single event. Therefore, among the 26 documented occurrence records, 18 records (approximately 70%) occurring before 14:00 UTC on July 6, 2018 were defined as an early cluster, and their median time of 11:50 UTC on July 6, 2018, was adopted as the reference occurrence time t_0 applied uniformly to all samples.

2.3 Predisposing Factors

According to a review of LSA, the explanatory variables employed are broadly categorized into topography, geology, land cover, hydrology, and other factors, with the majority being topography-related (Reichenbach et al., 2018). From these categories, we selected 19 predisposing variables related to slope stability and hydrological processes (topography: 8; geology: 2; land cover: 4; hydrology: 3; anthropogenic: 2; Table 1). Topographic variables were derived from a Digital Elevation Model (DEM) with a spatial resolution of 10 m. The Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Built-up Index (NDBI) among the land cover variables were calculated from a Sentinel-2 image acquired on October 21, 2018, with a cloud cover of less than 1% and complete spatial coverage of the study area. Esri's pan-sharpening procedure was applied to integrate bands with different spatial resolutions.

2.4 Triggering Factors

To capture the dynamic state of rainfall intensity and soil moisture at the occurrence time, we selected triggering-factor data based on the eXtended RAdar Information Network (XRAIN), a high-resolution precipitation observation system operated by the Ministry of Land, Infrastructure, Transport and

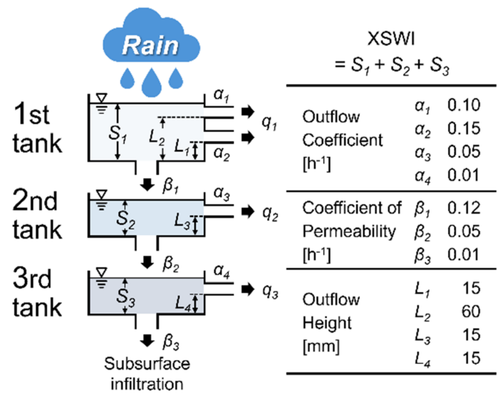

Figure 2. Schematic diagram of the three-layer tank model used to calculate XSWI.

Tourism (MLIT), and the XRAIN Soil Water Index (XSWI). XSWI is defined as the sum of storage heights across a three-layer tank model (Okada et al., 2001) driven by XRAIN rainfall input and represents a continuous indicator of moisture conditions from shallow to deep soil layers (Figure 2). Standard parameter values for granitic terrain, which is widely distributed in the study area, were adopted for the tank model (Ishihara and Kobatake, 1979). XRAIN and XSWI were obtained as time-series rasters at 10-minute intervals; missing values were filled through bicubic resampling and linear interpolation using valid observations within ±60 minutes. Using the reference occurrence time t_0, 56 candidate triggering variables were generated from time windows spanning 1–6 hours before t_0. After removing redundancy through correlation clustering and evaluating statistical significance via permutation testing, three variables were ultimately retained: XRAIN_now, XSWI_now, and XSWI_delta_2h (Table 1).

3. Methodology

This section presents the workflow of the LSA framework, from the introduction of dynamic triggering factors through model construction, evaluation, and interpretation (Figure 3). First, a model construction and evaluation scheme is designed by combining data preprocessing, multiple class imbalance strategies, watershed-based spatial cross-validation, and eight ML algorithms (Table 2). Second, model interpretation is performed through spatial visualization of feature contributions using SHapley Additive exPlanations (SHAP), an XAI method. All spatial analyses were performed using ArcGIS Pro 3.2.0 and SAGA GIS 9.3.1.

3.1 Data Preprocessing

Samples containing missing values (2.7% of the total) were excluded, yielding 224,378 samples (10,518 positive, 213,860 negative, positive rate: 4.7%) for analysis. Outliers were retained because they may reflect extreme topographic or rainfall conditions rather than observational errors. One-hot encoding was applied to categorical variables, except for CatBoost, which handles categorical variables internally. Standardization was applied only to Logistic Regression (LR), which is sensitive to variable scale; no scaling was performed for the other algorithms.

3.2 Watershed-Based Spatial Cross-Validation

Because geospatial data exhibit spatial autocorrelation, conventional random cross-validation tends to overestimate generalization performance for unseen areas (Roberts et al., 2017). In spatial block cross-validation, which addresses this problem, the definition of spatial blocks is critical. Geometric partitioning using arbitrary grids or hexagons, however, lacks a physical connection to landslide occurrence processes. In this study, we adopted watersheds as spatial blocks, as they represent the fundamental hydrological units governing infiltration-runoff processes and mass movement triggered by rainfall (Figure 4). Watersheds were delineated using Arc Hydro Tools (Maidment, 2002) by sequentially applying sink filling, followed by flow direction, flow accumulation, and stream raster generation based on the D8 algorithm, to a 10-m DEM. A drainage area threshold of 20 km² yielded 53 watersheds. Watersheds whose positive sample count or total sample count fell below one-third of the expected value across all watersheds were merged with the adjacent watershed having the shortest minimum point-to-point distance, resulting in 35 spatial blocks. These blocks were partitioned into five folds using Stratified Group K-Fold.

For model training and tuning, we adopted nested cross-validation with the watershed-based spatial cross-validation described above as the outer loop. In each outer fold, hyperparameter search was conducted using the Tree-structured Parzen Estimator (TPE) sampler of Optuna (Akiba et al., 2019), with the area under the precision-recall curve (PR-AUC) in the inner cross-validation as the objective function over 10 trials. The search space was configured per algorithm: tree complexity was controlled by max_leaf_nodes for decision tree-based algorithms and by max_depth or num_leaves for gradient boosting algorithms. The total number of trials was approximately 1,600 (5 outer folds × 10 inner trials × 8 algorithms × 4 strategies). Platt scaling was applied to the out-of-fold predicted probabilities

Model (Abbreviation)	Type
Balanced Random Forest (BRF)	Imbalanced
Categorical Boosting (CatBoost)	Boosting
Easy Ensemble Classifier (EEC)	Imbalanced
Histogram-Based Gradient Boosting (HGBoost)	Boosting
Light Gradient Boosting Machine (LightGBM)	Boosting
Logistic Regression (LR)	Linear
Random Forest (RF)	Bagging
eXtreme Gradient Boosting (XGBoost)	Boosting

Table 2. Machine learning algorithms and their types.

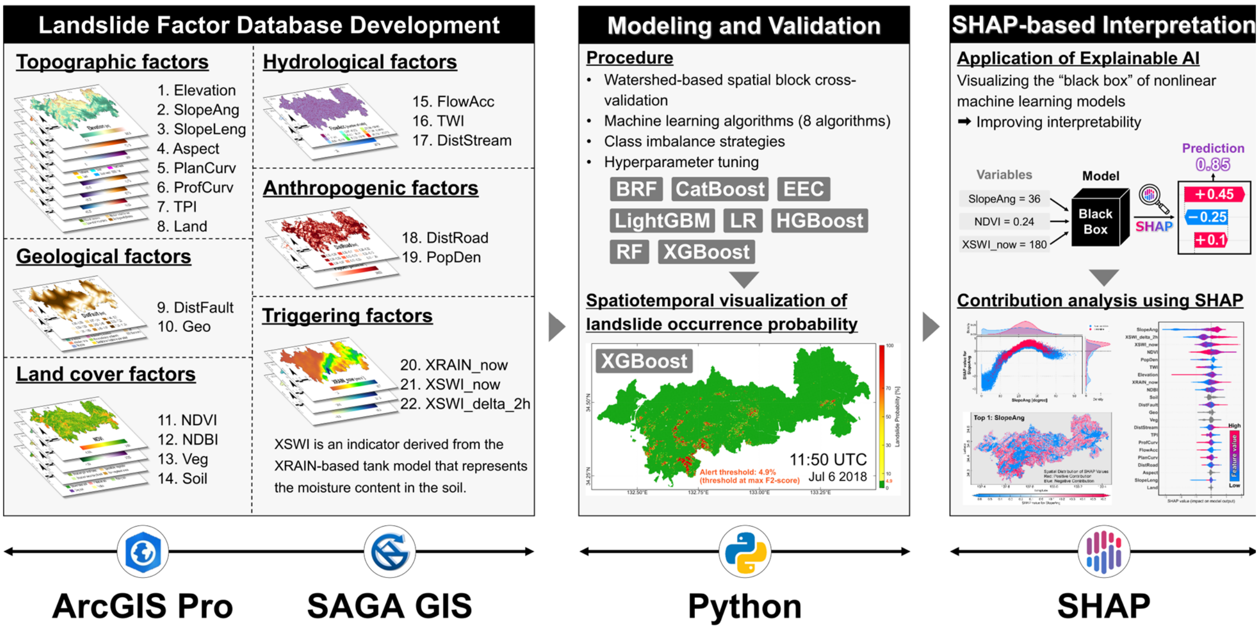

Figure 3. Workflow of the machine learning framework for landslide susceptibility assessment.

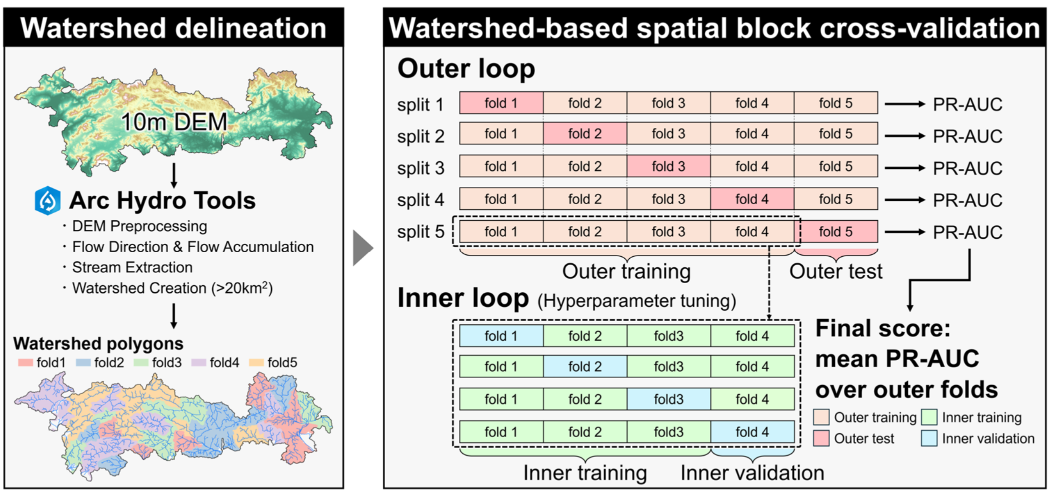

Figure 4. Watershed delineation and watershed-based spatial block cross-validation.

obtained from the outer-loop training data for probability calibration, after which the optimal classification threshold was determined based on the calibrated predicted probabilities.

3.3 Class Imbalance Strategies

Because the dataset exhibits pronounced class imbalance with a positive rate of 4.7%, the effect of imbalance treatment on predictive performance must be evaluated. We systematically compared the following four strategies (Figure 5): (i) no imbalance treatment as a baseline (NONE); (ii) sample weighting (SW), which assigns a weight equal to the negative-to-positive ratio to positive samples; (iii) random under-sampling (US), which reduces the number of negative samples to match the number of positive samples; and (iv) Synthetic Minority Over-sampling Technique (SMOTE), proposed by Chawla et al. (2002), which synthetically augments positive samples. SMOTE generates synthetic samples by linear interpolation between minority-class samples based on k-nearest neighbors and is defined by Equation (1):

$$S_{new}(m) = S_i(m) + \lambda\{S_z(m) - S_i(m)\} \tag{1}$$

where $S_i(m)$ is the original sample value, $S_z(m)$ is the neighboring sample value, λ is a random number from 0 to 1, and $S_{new}(m)$ is the generated sample value. Resampling strategies were not applied to Balanced Random Forest (BRF) and Easy Ensemble Classifier (EEC), as these algorithms incorporate internal resampling mechanisms.

3.4 Model Performance Evaluation

To evaluate the discrimination performance of the models, we adopted PR-AUC and the area under the receiver operating characteristic curve (ROC-AUC) as threshold-independent metrics. Because ROC-AUC tends to underestimate performance differences in highly imbalanced datasets (Davis and Goadrich, 2006), PR-AUC was designated as the primary metric and ROC-AUC as the secondary metric. For assessing the reliability of probabilistic outputs, the Brier Score, which indicates the overall error of predicted probabilities, and the Expected Calibration Error (ECE), which evaluates the correspondence between predicted probabilities and observed positive rates, were adopted. The Brier Score and ECE are defined by Equations (2) and (3), respectively:

$$Brier\ Score = \frac{1}{N}\sum_{i=1}^{N}(p_i - y_i)^2 \tag{2}$$

$$ECE = \sum_{m=1}^{M}\frac{|B_m|}{N}|acc(B_m) - conf(B_m)| \tag{3}$$

where p_i is the predicted probability, $y_i \in \{0,1\}$ is the true label, M is the number of bins (equal-frequency binning with M = 10 was adopted), B_m is the set of samples in bin m, $acc(B_m)$ is the observed positive rate within the bin, and $conf(B_m)$ is the mean predicted probability within the bin. Each metric was computed as the mean of fold-level scores, and 95% confidence intervals were estimated using a block bootstrap method with folds as the resampling unit (n = 1,000 iterations).

3.5 Model Interpretation Using SHAP

To enhance the interpretability of the trained models, SHapley Additive exPlanations (SHAP), proposed by Lundberg and Lee

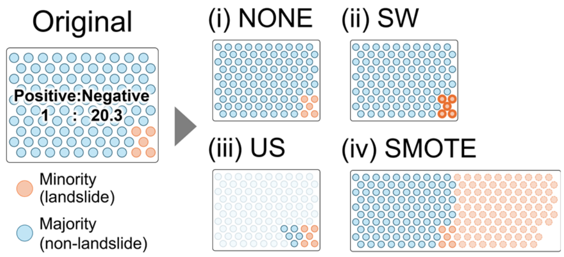

Figure 5. Schematic illustration of four class imbalance handling strategies: NONE, SW, US, and SMOTE.

(2017), was applied. SHAP is a method rooted in cooperative game theory that quantitatively evaluates the contribution of each feature to the model output by computing the marginal contribution of each feature across all possible feature combinations. The SHAP value ϕ_i for feature i is defined as follows:

$$\phi_i = \sum_{S \subseteq N \setminus i}\frac{|S|!\,(|N| - |S| - 1)!}{|N|!}[f(S \cup \{i\}) - f(S)] \tag{4}$$

where N is the set of all features, $|N|$ is the total number of features, S is a subset of N excluding the target feature i ($S \subseteq N \setminus i$), $|S|$ denotes the number of features in S, and $f(S \cup \{i\}) - f(S)$ represents the marginal contribution of feature i when added to the subset S. Defining ϕ_0 as the base value (the expected value of the model output over the training dataset), the final prediction $\hat{f}(X)$ is expressed as the sum of individual feature contributions as follows:

$$\hat{f}(X) = \phi_0 + \sum_{i=1}^{|N|}\phi_i \tag{5}$$

SHAP was selected because it is the only additive feature attribution method that simultaneously satisfies three desirable theoretical properties—local accuracy, missingness, and consistency—and is model-agnostic, enabling a consistent and fair comparison of feature importance across different model architectures. In this study, SHAP values were computed for each sample point and mapped spatially to visualize how feature contributions vary across the study area. This spatial mapping enables verification of whether model predictions are geographically and physically plausible, addressing the limitation of conventional XAI studies that typically present only global or aggregate feature importance without spatial context.

4. Results and Discussion

4.1 Model Performance

Figure 6 compares PR-AUC, ROC-AUC, Brier Score, and ECE across algorithms and strategies. Under the NONE condition, XGBoost achieved the highest PR-AUC (0.323), followed by LightGBM (0.323), CatBoost (0.318), and HGBoost (0.314). Gradient boosting algorithms occupied the top ranks; however, the 95% confidence intervals based on 5-fold cross-validation were wide (XGBoost: 0.240–0.404), and the performance differences among algorithms were small relative to the width of the confidence intervals. ROC-AUC differences among algorithms were considerably smaller than those of PR-AUC (NONE condition: 0.806–0.859), indicating that ROC-AUC cannot sufficiently discriminate performance differences under pronounced class imbalance. This result supports the decision to adopt PR-AUC as the primary metric.

Regarding the effects of sampling strategies, SW maintained PR-AUC at levels comparable to the NONE condition (LightGBM-SW: 0.313; XGBoost-SW: 0.307) while reducing ECE by 0.007–0.012 points relative to NONE, contributing to improved probability calibration. In contrast, US and SMOTE tended to decrease PR-AUC below the NONE condition, with particularly pronounced decreases observed for RF (0.175) and LightGBM (0.284) under SMOTE. Under US, the reduction of majority-class information may have destabilized decision boundary learning; under SMOTE, synthetic samples generated in unoccupied regions of the feature space may have introduced additional noise. The Brier Score remained within a narrow range of 0.038–0.043 across all conditions, with no substantial differences among algorithms or strategies. For the best-performing model (XGBoost, NONE condition), the performance at the threshold maximizing the F2-score (4.9%) was as follows: Recall = 0.607, F2-score = 0.435, and Balanced Accuracy = 0.745.

4.2 Landslide Susceptibility Mapping

Figure 7 presents the susceptibility maps for each algorithm under the sampling condition that yielded the highest PR-AUC (NONE for all algorithms). Predicted probabilities were classified into five levels using Jenks natural breaks: Very Low, Low, Moderate, High, and Very High.

The gradient boosting algorithms (XGBoost, LightGBM, CatBoost, and HGBoost) confined the Very High class to 1.3–2.0% of the total area and exhibited a distinct spatial pattern in which high-risk zones concentrated along steep slopes adjacent to stream channels. Boundaries with ridgelines and gentle slopes were well defined, and the Very Low class accounted for 91.7–92.6%, producing a clear spatial contrast between high-risk and low-risk zones. Landslide polygons were concentrated in the Very High class, suggesting that these models possess strong spatial discrimination capability.

RF and BRF yielded a Very High class proportion (1.0–1.7%) comparable to that of the gradient boosting algorithms; however, the Low class was broader (6.2–8.3%), and high-risk zones tended to be more diffuse. EEC produced a Very Low class of only 64.3%, with the Low (20.2%) and Moderate (9.2%) classes being substantially larger than those of other algorithms, resulting in high-risk zones dispersed across the entire study area. This spatial pattern corresponds to the low PR-AUC of EEC (0.224). LR produced the smallest Very High class among all algorithms (0.4%), with most landslide polygons falling within the High to Moderate classes, reflecting insufficient spatial discrimination. The NONE condition yielded the most distinct spatial discrimination pattern in the susceptibility maps across all algorithms, consistent with the finding that NONE also achieved the highest PR-AUC.

4.3 Variable Importance and Spatial Distribution of SHAP Values

Figure 8 presents the beeswarm plot for XGBoost under the NONE condition. Slope angle (SlopeAng) exhibited the highest contribution (mean |SHAP| = 0.724), followed by XSWI_delta_2h (0.516) and XSWI_now (0.333) in the top three. XRAIN_now ranked eighth (0.228); nevertheless, all three variables introduced as dynamic triggering factors appeared within the top eight of 22 variables, indicating that hydrometeorological conditions at the occurrence time played an important role in model discrimination. The SHAP values of all 22 variables differed significantly between landslide and non-landslide cells (Mann–Whitney U test, $p < 0.001$ for all variables).

Figure 9 presents the dependence plots for the top 9 variables. XSWI_delta_2h exhibited a sign change in SHAP values from negative to positive at approximately 36 mm, with a mean SHAP value of +0.490 in the upper quartile (43–66 mm), indicating a pronounced positive contribution. XSWI_now showed a sign change at approximately 191 mm, with a positive contribution of +0.389 in the upper quartile (191–222 mm). The correlation between SHAP values of XSWI_now and XSWI_delta_2h was

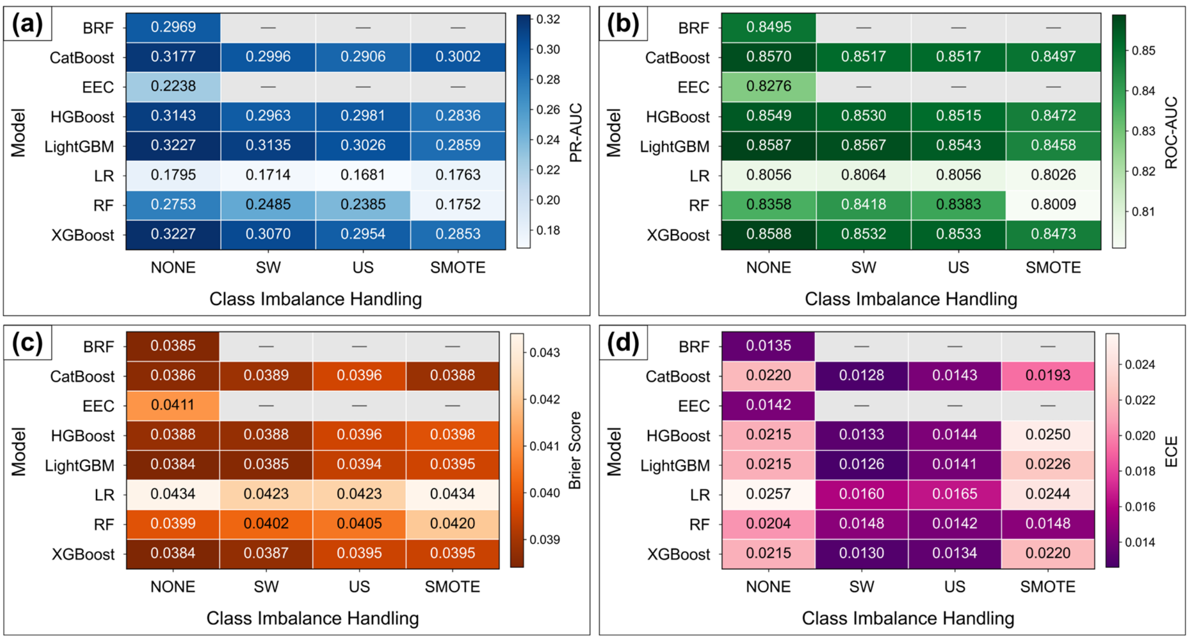

Figure 6. Comparison of model performance across eight algorithms and four class imbalance handling strategies: (a) PR-AUC, (b) ROC-AUC, (c) Brier Score, and (d) ECE.

weak (r = +0.074), suggesting that the absolute soil moisture level and the short-term rate of change function as independent triggering axes. XRAIN_now exhibited a positive sign change at approximately 11 mm h⁻¹, and positive contributions increased in the high-intensity range above 36 mm h⁻¹; however, in the intermediate range (17–36 mm h⁻¹), the mean SHAP value was near zero, indicating that rainfall intensity alone has limited discriminative power between landslide and non-landslide conditions. This finding is consistent with the physical understanding that cumulative soil moisture state is more dominant in landslide discrimination than instantaneous rainfall intensity.

Among predisposing variables, slope angle (SlopeAng) exhibited a positive sign change at approximately 17°, with the maximum positive contribution observed between 19° and 26°, followed by a decline above 26°. This nonlinear response corresponds to a mechanism in which risk increases on steeper slopes, whereas on the steepest slopes, the thickness of collapsible soil mantle decreases (Xie et al., 2025). NDVI exhibited SHAP values dispersed in both positive and negative directions: elevated risk was observed in low-NDVI areas (reflecting the susceptibility of bare land and developed areas), and a mixture of positive and negative values appeared in moderate-NDVI areas (attributable to the coexistence of steep forested slopes and lowland vegetation). Population density (PopDen) showed a sign change

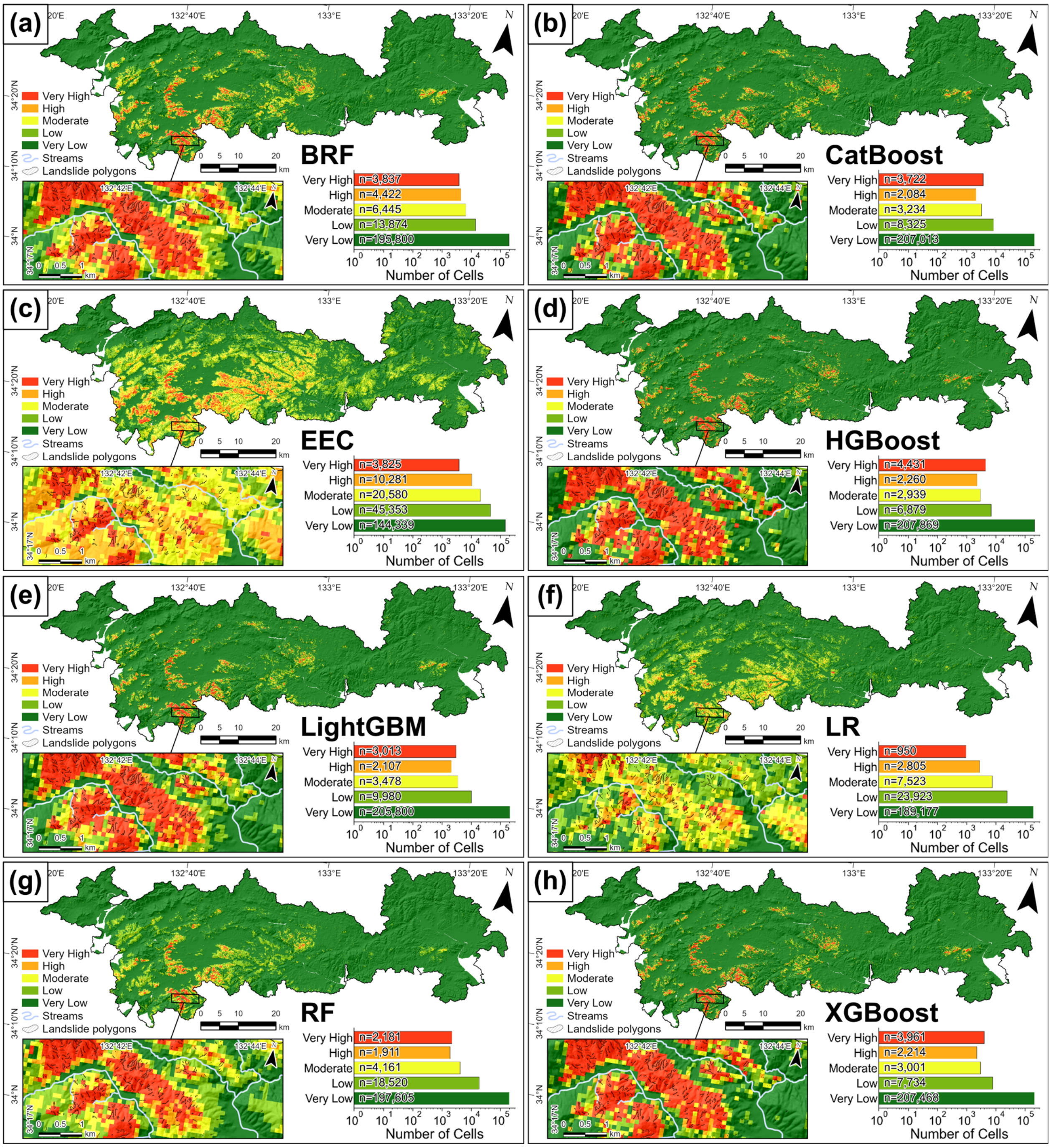

Figure 7. Landslide susceptibility maps for the eight models under the NONE condition: (a) BRF, (b) CatBoost, (c) EEC, (d) HGBoost, (e) LightGBM, (f) LR, (g) RF, and (h) XGBoost. Susceptibility was classified into five levels (Very Low to Very High) using Jenks natural breaks.

from positive to negative at approximately 5 persons km⁻², and the Topographic Wetness Index (TWI) displayed a strong negative correlation with SHAP values (r = −0.925). Both variables exhibit high values in plains and lowlands, and they are considered to function as proxy variables for topographic position, reflecting the spatial separation from landslide-prone zones. The response patterns of the remaining variables are presented in Figure 9(g)–(i).

Figure 10 presents the Spatial SHAP maps for the top 9 variables. Positive SHAP values (red) of slope angle (SlopeAng) were distributed in bands along steep slopes adjacent to stream channels, exhibiting a spatial pattern governed by topography. In contrast, positive SHAP values of XSWI_delta_2h were concentrated in the western to central portions of the study area, whereas negative contributions (blue) predominated in the eastern portion. This spatial pattern is consistent with the meteorological progression in which the rainfall zone of the Heavy Rain Event of July 2018 developed earlier in the western part, indicating that the model correctly captured the rapid increase in soil moisture as a spatially resolved factor promoting landslide occurrence. XSWI_now also showed stronger positive

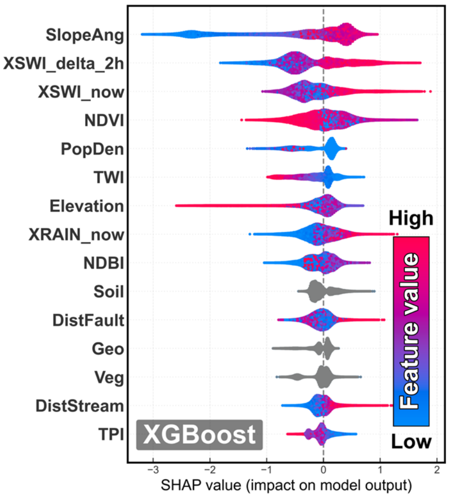

Figure 8. Beeswarm plot of SHAP values for the top 15 variables in the XGBoost model.

contributions in the western portion, although the spatial contrast was somewhat more gradual compared with XSWI_delta_2h. Positive SHAP values of XRAIN_now corresponded to the areas of intense rainfall at time t_0 ; however, the widespread distribution of the rainfall zone resulted in more limited spatial discrimination than that provided by the XSWI-based variables. Negative SHAP values of TWI, PopDen, and Elevation were concentrated in coastal lowlands and urban areas, consistent with the spatial pattern of non-landslide zones. These results indicate that the SHAP contributions of static predisposing factors exhibit temporally stable spatial patterns controlled mainly by terrain, whereas those of dynamic triggering factors show event-specific patterns reflecting the rainfall and soil-moisture conditions of the event. The model's decision structure, in which locations where both patterns spatially overlap are identified as the highest risk, is thus visualized.

4.4 Temporal Variation of Dynamic Landslide Susceptibility

To verify the practical applicability of the dynamic LSA framework, time-series susceptibility maps were generated at 30-minute intervals by replacing only the triggering variables with data from each respective time step while keeping predisposing variables fixed, using the XGBoost model trained under the NONE condition at t_0 (11:50 UTC) (Figure 11). The alert threshold was set at 4.9%, which maximizes the F2-score.

At 10:30 UTC, intense rainfall was concentrated in the western portion of the study area and XSWI showed an increasing trend in that area; however, cells exceeding the threshold were limited to steep slopes in the west. By 11:00 UTC, the rainfall zone had expanded eastward and XSWI accumulation had progressed, resulting in an expansion of high-risk zones from the western to central portions. At 11:30 UTC (immediately before t_0), intense rainfall covered a broad area in the central portion, and the concurrent rise in XSWI led to a marked increase in high-risk zones. At 12:00 UTC, the rainfall peak had shifted eastward, yet XSWI in the western and central portions remained at elevated levels, and high-risk zones were distributed extensively. This temporal progression demonstrates that both XRAIN_now (instantaneous rainfall intensity distribution) and XSWI_now/XSWI_delta_2h (cumulative effects of antecedent rainfall) are dynamically reflected in model output. At 10:30

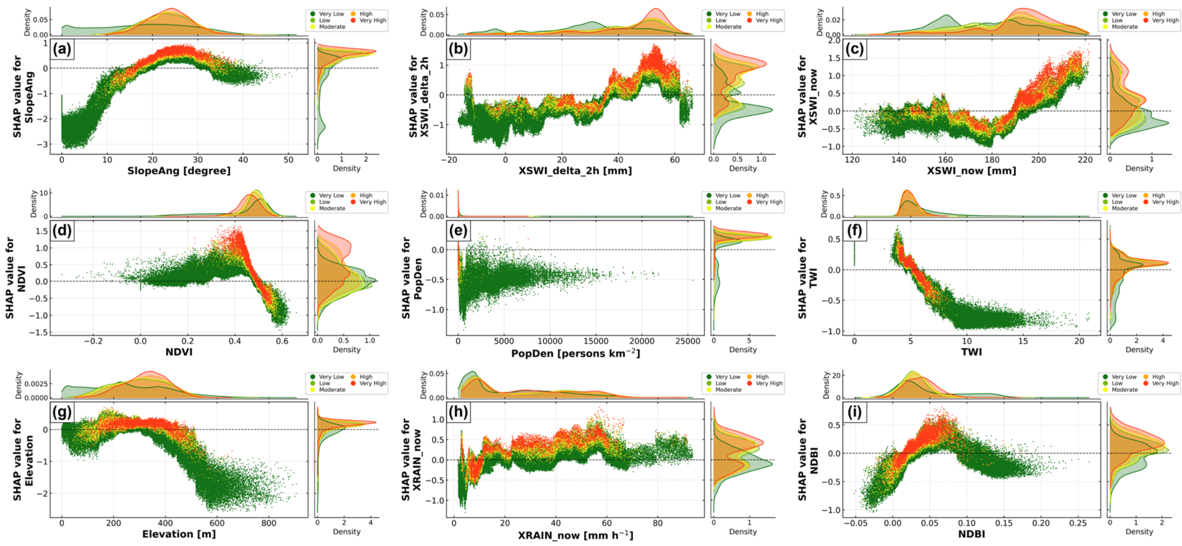

Figure 9. Dependence plots showing the relationships between feature values and SHAP values for the top 9 variables: (a) SlopeAng, (b) XSWI_delta_2h, (c) XSWI_now, (d) NDVI, (e) PopDen, (f) TWI, (g) Elevation, (h) XRAIN_now, and (i) NDBI.

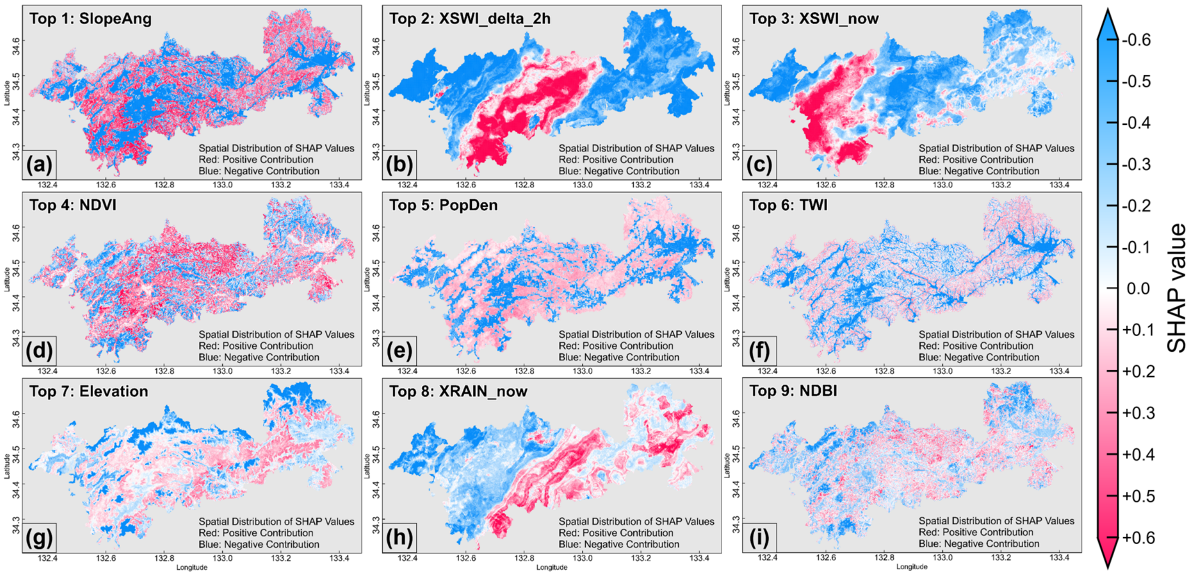

Figure 10. Spatial distribution of SHAP values for the top 9 variables: (a) SlopeAng, (b) XSWI_delta_2h, (c) XSWI_now, (d) NDVI, (e) PopDen, (f) TWI, (g) Elevation, (h) XRAIN_now, and (i) NDBI.

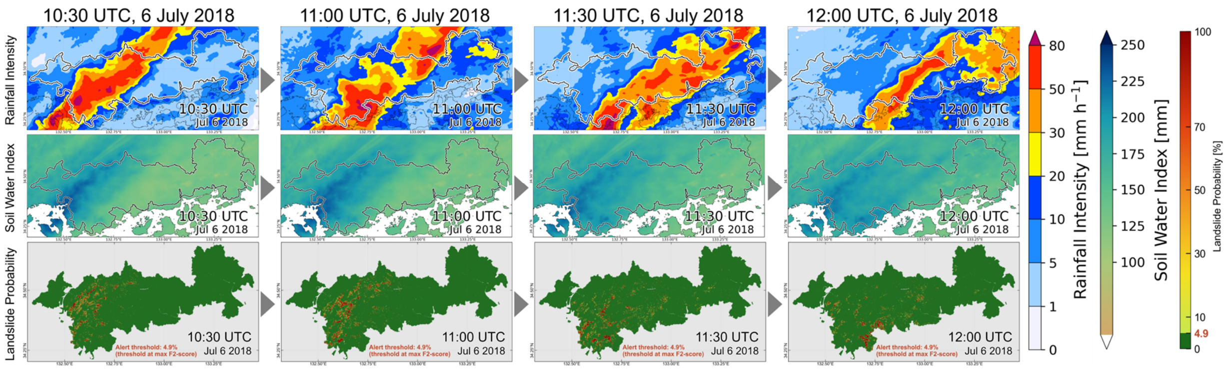

Figure 11. Spatiotemporal snapshots of rainfall intensity (upper row), soil water index (middle row), and predicted landslide probability (lower row) during the Heavy Rain Event of July 2018 (supplementary video: https://youtu.be/4silN5bNZuo).

UTC, when XSWI accumulation was still insufficient despite high rainfall intensity, high-risk zones remained limited; extensive high-risk zones emerged only at 11:30–12:00 UTC, after XSWI accumulation had progressed.

These results demonstrate that the proposed framework can represent intra-event spatiotemporal risk variations that cannot be captured by conventional LSA using static annual or monthly precipitation as triggering factors. The design, in which predisposing variables are fixed and only triggering variables are updated, enables susceptibility maps to be generated for any point in time without model retraining. Because XRAIN and XSWI are delivered in real time, the framework has the potential to be extended to dynamic susceptibility assessment during an ongoing rainfall event. The model, however, was trained on a single event (the Heavy Rain Event of July 2018), and its generalization performance under different rainfall patterns and seasonal conditions requires further investigation.

5. Conclusions

In this study, we proposed an ML framework for dynamic LSA targeting a single rainfall event by introducing high-spatiotemporal-resolution rainfall and soil moisture data as dynamic triggering factors, integrated with watershed-based spatial cross-validation and spatial interpretation using SHAP. Application to the Heavy Rain Event of July 2018 revealed that dynamic triggering factors (XSWI_delta_2h and XSWI_now) were the leading predictors after slope angle and that cumulative soil moisture state was more dominant in landslide discrimination than instantaneous rainfall intensity. Regarding class imbalance treatment, resampling strategies (US and SMOTE) instead degraded performance, and the no-treatment condition (NONE) yielded the best results across all algorithms. Spatial SHAP visualized the model's decision structure, in which locations where the time-invariant spatial patterns of predisposing factors overlap with the event-specific spatial patterns of dynamic triggering factors are identified as high risk, providing a framework for verifying the geographic validity of prediction rationale. A limitation of this study is that validation was conducted for a single event in a single region, and generalization performance under different rainfall patterns and in other regions remains unexamined. Future work should focus on verifying generalization performance through application to multiple events and on extending the framework toward dynamic susceptibility assessment by integrating real-time data streams.

Acknowledgements

We express our sincere sympathies to all those affected by the Heavy Rain Event of July 2018. This study used XRAIN data provided by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT).

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A Next-generation Hyperparameter Optimization Framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)*, ACM, New York, pp. 2623–2631. doi.org/10.1145/3292500.3330701

Cabinet Office, Government of Japan, 2018. *2018 White Paper on Disaster Management: The Threat of Meteorological Disasters — Focusing on the Northern Kyushu Heavy Rain Disaster*. https://www.bousai.go.jp/kaigirep/hakusho/pdf/H30_hakusho.pdf (21 March 2026).

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. J. Artif. Intell. Res., 16, 321–357. doi.org/10.1613/jair.953

Davis, J., Goadrich, M., 2006. The Relationship between Precision-Recall and ROC Curves. *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*, ACM, Pittsburgh, pp. 233–240. doi.org/10.1145/1143844.1143874

Froude, M.J., Petley, D.N., 2018. Global Fatal Landslide Occurrence from 2004 to 2016. Nat. Hazards Earth Syst. Sci., 18(8), 2161–2181. doi.org/10.5194/nhess-18-2161-2018

Halter, T., Lehmann, P., Bast, A., Aaron, J., Stähli, M., 2025. In Situ Soil Moisture Data Improve Precipitation-Based Shallow Landslide Early Warning Through Innovative Machine Learning Methods. Landslides, 22, 3599–3614. doi.org/10.1007/s10346-025-02599-4

Hiroshima Prefecture Sediment Control Division, 2018. *Verification of the Disaster-Causing Factors in the Tenchigawa River (Saka Town). Material 2 of the 1st Sediment Control Subcommittee, Review Committee on Future Flood and Sediment Disaster Countermeasures Based on the July 2018 Heavy Rain Disaster*. https://www.pref.hiroshima.lg.jp/uploaded/attachment/330356.pdf (21 March 2026).

Hiroshima University July 2018 Heavy Rain Disaster Investigation Team (Geography Group), 2019. *Detailed Distribution Map of Slope Failures in Hiroshima Prefecture Due to the July 2018 Heavy Rain (Second Report)*. The Association of Japanese Geographers. https://www.ajg.or.jp/disaster/files/201807_report008.pdf (21 March 2026).

IPCC, 2023. *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, Geneva. doi.org/10.59327/IPCC/AR6-9789291691647

Ishihara, Y., Kobatake, S., 1979. Runoff Model for Flood Forecasting. Bull. Disaster Prev. Res. Inst., 29(1), 27–43. http://hdl.handle.net/2433/124881

Lee, S.-M., Lee, S.-J., 2024. Landslide Susceptibility Assessment of South Korea Using Stacking Ensemble Machine Learning. Geoenviron. Disasters, 11(1), 7. doi.org/10.1186/s40677-024-00271-y

Lombardo, L., Mai, P.M., 2018. Presenting Logistic Regression-Based Landslide Susceptibility Results. Eng. Geol., 244, 14–24. doi.org/10.1016/j.enggeo.2018.07.019

Lundberg, S.M., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems (NeurIPS 2017)*, 30, 4768–4777.

Maidment, D.R., 2002. *Arc Hydro: GIS for Water Resources*. ESRI Press, Redlands, CA.

Miura, N., Yamanouchi, T., 1975. Effect of Water on the Behavior of a Quartz-Rich Sand Under High Stresses. Soils Found., 15(4), 23–34. doi.org/10.3208/sandf1972.15.4_23

Okada, K., Makihara, Y., Shimpo, A., Nagata, K., Kunitsugu, M., Saito, K., 2001. Soil Water Index. Tenki, 48(5), 349–356. [in Japanese]

Palazzolo, N., Cancelliere, A., Zofei, R.D., Peres, D.J., 2025. Use of Delayed ERA5-Land Soil Moisture Products for Improving Landslide Early Warning. Nat. Hazards Earth Syst. Sci., 25, 4907–4919. doi.org/10.5194/nhess-25-4907-2025

Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M., Guzzetti, F., 2018. A Review of Statistically-Based Landslide Susceptibility Models. Earth-Sci. Rev., 180, 60–91. doi.org/10.1016/j.earscirev.2018.03.001

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., et al., 2017. Cross-validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure. Ecography, 40(8), 913–929. doi.org/10.1111/ecog.02881

Shimpo, A., Takemura, K., Wakamatsu, S., Togawa, H., Mochizuki, Y., Takekawa, M., et al., 2019. Primary Factors behind the Heavy Rain Event of July 2018 and the Subsequent Heat Wave in Japan. SOLA, 15(A), 13–18. doi.org/10.2151/sola.15A-003

Tanaka, Y., Goto, S., 2024. Early Detection of Disaster Areas Using a Landslide Occurrence Probability Model Considering Predisposing Factors and Triggers Combined with Deep Learning. J. Jpn. Soc. Photogramm. Remote Sens., 63(3), 54–68. doi.org/10.4287/jsprs.63.54 [in Japanese]

Wang, N., Zhang, H., Dahal, A., Cheng, W., Zhao, M., Lombardo, L., 2024. On the Use of Explainable AI for Susceptibility Modeling: Examining the Spatial Pattern of SHAP Values. Geosci. Front., 15(4), 101800. doi.org/10.1016/j.gsf.2024.101800

Xie, C., Xu, C., Huang, Y., Liu, J., Jin, J., Xu, X., et al., 2025. Detailed Inventory and Initial Analysis of Landslides Triggered by Extreme Rainfall in the Northern Huaiji County, Guangdong Province, China, from June 6 to 9, 2020. Geoenviron. Disasters, 12(1), 7. doi.org/10.1186/s40677-025-00311-1

Zhang, J., Ma, X., Zhang, J., Sun, D., Zhou, X., Mi, C., Wen, H., 2023. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. J. Environ. Manage., 332, 117357. doi.org/10.1016/j.jenvman.2023.117357

Appendix

A supplementary video showing the temporal evolution of dynamic landslide susceptibility during the Heavy Rain Event of July 2018 is available at:
https://youtu.be/4silN5bNZuo