Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Explainable Machine Learning for Multi-Temporal Landslide Detection along the Mechi– Tamor Corridor, Eastern Nepal"

Sudeep Bista[1]

sudeepbista09@gmail.com

Council for Technical Education and Vocational Training

Sarthak Regmi[2]

Kathmandu University

Sudin Adhikari[2]

Kathmandu University

Sarjun Khatri[1]

Council for Technical Education and Vocational Training

Sujan Sapkota[2]

Kathmandu University

Research Article

Keywords: landslide detection, SAR–optical fusion, multi-temporal analysis, Random Forest, XGBoost, explainable machine learning (SHAP)

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11282655/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

This study explores a multi-temporal, multi-sensor framework for landslide detection along the Mechi Highway–Tamor River corridor in Ilam, Panchthar, Taplejung and Dhankuta districts of eastern Nepal. The framework is evaluated on two contrasting rainfall-triggered events, a late-monsoon withdrawal event in early October 2025 and a peak-monsoon event on 20 July 2026, which differ in rainfall intensity, duration, structure and antecedent wetness. An event-based inventory of 116 landslide scars was compiled through visual interpretation of Sentinel-2 imagery and cross-checked against a regional baseline of 255 landslide records. This inventory was combined with Sentinel-1 SAR backscatter change, optical spectral indices, terrain derivatives, antecedent rainfall and lithology to create a feature table of 410 units with 820 observations (unit × event). Random Forest and XGBoost classifiers were trained and evaluated under a naive random split and two spatially independent validation schemes (leave-one-municipality-out and spatial block cross-validation); the spatial schemes yielded F1 scores of 0.47–0.56 and ROC-AUC of 0.80–0.84, substantially lower than the optimistic random-split estimate (F1 = 0.77–0.80, ROC-AUC = 0.95), demonstrating the necessity of spatial validation for this problem. SHAP analysis showed that SAR backscatter change was the most important predictor overall, but that optical spectral indices, especially the Normalized Burn Ratio (NBR) difference, carried more signal wherever cloud-free images were available. SAR's value in this setting lies in cloud penetration rather than in superior predictive power. Combining predicted susceptibility with proximity to schools, health facilities and road bridges identified 35 critical corridor segments.

1. Introduction

Landslides are among the most pervasive natural hazards in the Nepal Himalaya, where steep terrain, monsoon-concentrated rainfall, active tectonics and expanding road infrastructure combine to produce frequent, often infrastructure-disrupting slope failures, while the roads themselves provide the access corridors that settlements and services depend on (Amatya et al., 2019; McAdoo et al., 2018). At the same time, the roads provide the access corridors that settlements and services depend on. In eastern Nepal, the Mechi Highway and Tamor Corridor, linking Jhapa, Ilam, Panchthar and Taplejung to the Terai and major market centres, have been repeatedly blocked by monsoon and post-monsoon landslides, with recent events in 2025–2026 causing multi-day closures and substantial reconstruction costs (The Kathmandu Post, 2025).

Two complementary remote sensing approaches dominate landslide detection: optical change detection, which is intuitive and well validated but frequently obstructed by persistent monsoon-season cloud cover, and Synthetic Aperture Radar (SAR) backscatter change detection, which penetrates cloud cover but is more difficult to interpret physically and is sensitive to viewing geometry, soil moisture and vegetation state (Lindsay et al., 2025). Machine learning classifiers, particularly Random Forest (Breiman, 2001) and XGBoost (Chen & Guestrin, 2016), are now standard tools for fusing such heterogeneous predictors into a single susceptibility or detection model, but their opacity has motivated growing use of post-hoc explainability methods, most notably SHapley Additive exPlanations (SHAP;

(Lundberg & Lee, 2017), to verify that model behaviour is physically plausible rather than an artefact of spurious correlation.

We focus on the Mechi–Tamor road corridor in eastern Nepal, spanning Ilam, Panchthar, Taplejung and Jhapa districts along the Mechi and Tamor river valleys, where recent highway expansion and slope cutting have increased landslide exposure (Amatya et al., 2019; McAdoo et al., 2018). Two landslide episodes are analysed: (i) a late-monsoon/withdrawal event in early October 2025, characterised by prolonged, moderate rainfall and higher antecedent soil moisture following an extended monsoon season, and (ii) a peak-monsoon event in July 2026, associated with intense, short-duration rainfall typical of the core monsoon (Department of Hydrology and Meteorology (DHM, 2025; ACONF, 2026). These events differ not only in timing relative to the climatological monsoon cycle but also in rainfall intensity–duration structure and antecedent wetness, providing contrasting conditions under which to evaluate SAR, optical and rainfall signatures.

This study addresses four gaps in the existing literature. First, most Himalayan road-corridor susceptibility studies are static, single-epoch assessments based on one landslide inventory or monsoon season (Reichenbach et al., 2018); we instead analyse two contrasting events, one at the monsoon's peak and one at its withdrawal, within the same corridor, enabling direct comparison of SAR, optical and rainfall signatures under differing antecedent conditions. Second, whereas many studies fuse SAR and optical data at the pixel level, we aggregate these predictors with terrain, rainfall and lithological information at the scale of municipality and road-segment units, keeping the model interpretable and directly actionable for infrastructure planning and maintenance prioritization (McAdoo et al., 2018).

Third, most landslide susceptibility models are validated using random train/test splits that ignore spatial autocorrelation and yield systematically optimistic performance (Brenning, 2005; Roberts et al., 2017). We evaluate performance under spatial cross-validation (leave-one-municipality-out and spatial block cross-validation), providing a more realistic estimate of transferability and reducing variance inflation from spatial leakage. Fourth, although SHAP is increasingly used to rank feature importance in landslide and geoscience models (Jiang, et al., 2024; Zhou, et al, 2022), few studies exploit it to diagnose methodological artefacts, that is, the pitfalls of careless interpretability practice that can produce misleading conclusions (Quiroz, 2026) we use SHAP not only to quantify predictor contributions but also to identify and correct an imputation-dilution artefact, whereby cloud-obscured optical observations, imputed with neutral values, have their apparent importance diluted, and to translate model outputs into an infrastructure-exposure monitoring priority product.

2. Study Area

The study corridor spans the Mechi Highway and Tamor River across four districts of eastern Nepal — Ilam, Panchthar, Taplejung, and Dhankuta — covering a total area of approximately 1,126 km². The corridor was delineated using a 2km buffer applied on each side of the highway and river centerlines

(sourced via OpenStreetMap through Overpass Turbo), yielding a total corridor width of approximately 4km.

Within the corridor, the Mechi Highway extends approximately 275.5km and the Tamor River approximately 182.7km. Corridor coverage varies by district, with Panchthar contributing the largest share (approximately 107.0km of corridor length, 145 analytical units), followed by Taplejung (94.5km, 114 units), Ilam (70.1km, 103 units), and Dhankuta (41.3km, 48 units). Elevation within the corridor ranges from 114m to 5,395m (mean: 1,432m), reflecting the transition from lower-elevation river valleys and foothill terrain to high mountain ridgelines characteristic of the Himalayan Lesser and Middle hill zones. This elevational range, combined with steep terrain gradients, positions the corridor as a setting of significant landslide susceptibility, particularly where road cuts and river undercutting intersect with unstable slope material.

3. Data and Methods

The methodological framework adopted in this study integrates multi-source remote sensing, terrain, meteorological, and geological data within a supervised machine-learning pipeline to characterize landslide occurrence along the Mechi–Tamor corridor. The methodological workflow (Fig. 2) comprises six stages: (i) compilation of a multi-source landslide inventory; (ii) extraction of terrain, SAR, optical, rainfall, and lithological features; (iii) supervised classification using Random Forest and XGBoost; (iv) spatial cross-validation; (v) model interpretation via SHAP (SHapley Additive exPlanations); and (vi) translation of model output, combined with infrastructure exposure, into a corridor monitoring-priority map. Each stage is detailed in the following subsections.

3.1 Landslide Inventory Compilation and Optical Detection

The landslide inventory was compiled from three sources: (1) the ICIMOD Regional Database System baseline (Landslide Inventory of the Koshi River Basin, digitised from December 2024 Sentinel-2 imagery; n = 255 within the corridor; ICIMOD, 2024); (2) five event-specific records from official news and disaster reports for the two anchor events; and (3) 116 scars digitised through visual interpretation of pre- and post-event Sentinel-2 imagery in Google Earth Engine (95 for October 2025, 21 for July 2026). Pre- and post-event composites were built using least-cloud-percentage mosaicking with s2cloudless-based cloud and shadow masking. Scars were identified through comparison of true-colour and false-colour (NIR-Red-Green) composites.

To validate the digitised inventory, mean pre–post Bare Soil Index (BSI) difference and terrain slope were extracted per polygon. Of the 116 scars, 94 (81%) showed a clear bare-soil increase consistent with genuine disturbance; 13 (11%), on steep terrain adjacent to river channels, showed weaker BSI change and were retained but flagged as likely fluvial/riverbank failures; and 9 (8%) coincided spatially with OpenStreetMap (OpenStreetMap contributors, 2026) building footprints and were flagged for manual re-verification. For the October 2025 event, this was further corroborated by full-corridor pre–post

differencing of NDVI, NBR and BSI (8 pre-event, 38 post-event images): digitised scars showed mean differences of − 0.33 (NDVI), − 0.36 (NBR) and + 0.24 (BSI), against a near-zero corridor background (+ 0.01, − 0.01, + 0.01 respectively), independently corroborating the SAR-based signal (Section 3.4).

Digitised scars were cross-checked against the ICIMOD 2024 baseline to distinguish new failures from reactivations; only 3 of 95 October 2025 scars intersected the baseline, indicating that most mapped disturbance represents new slope failure. A limitation concerns the July 2026 event: post-event imagery availability (20 July–8 September 2026, within active monsoon season) suffered persistent cloud cover, most severe in Taplejung district, consistent with its higher average elevation and associated orographic cloud cover; the gap in neighbouring Panchthar was confined to the narrower pre-event window and resolved with a wider post-event compositing period. This precluded full-corridor optical differencing for the July 2026 event and motivated the multi-sensor design of this study.

3.2 Terrain Data and Derivatives

Terrain data were derived from the Copernicus GLO-30 Digital Elevation Model (30 m; European Space Agency, 2022), downloaded from the Copernicus Data Space Ecosystem and clipped to the corridor. Slope, aspect, general curvature (following Zevenbergen and Thorne, 1987), the Topographic Wetness Index (TWI; Beven and Kirkby, 1979) and the Topographic Position Index (TPI; Weiss, 2001) were computed at native resolution:

$$TWI = \ln (a / \tan\beta) \quad (1)$$

Where a is the specific catchment area and β is local slope angle (radians); and

$$TPI = Z_0 - \text{mean}(Z) \quad (2)$$

Where Z_0 is the elevation of the focal cell and mean (Z) is the mean elevation within a defined neighbourhood, used to characterise landform position (ridge, mid-slope, and valley).

3.3 Analytical Unit Delineation

Rather than pixel- or slope-unit classification, the corridor was divided into 410 fixed-length (~ 1 km) segmented buffer units along the Mechi Highway and Tamor River centrelines, following the general mapping-unit taxonomy (Reichenbach et al., 2018) and corridor-buffer precedents established for Himalayan highway landslide studies (Amatya et al., 2019). Because this study is linear and infrastructure-centred, and explicitly distinguishes road-associated from river-associated landslide occurrence, fixed-length segmented buffer units were adopted as the primary analytical unit. Each unit carries district, municipality (GaPa/NaPa) and corridor-type (road/river) attributes, providing an interpretable, directly actionable spatial unit for infrastructure-oriented reporting.

3.4 SAR Change Analysis

Sentinel-1 GRD imagery (Interferometric Wide swath, VV and VH polarisation) was used for SAR change detection. Ascending and descending orbit coverage was compared per event window; the descending

pass offered consistently greater image availability and was used throughout to avoid backscatter variability attributable to differing viewing geometry. Pre- and post-event median composites were differenced (POST − PRE) for each polarisation.

Visual inspection indicated that SAR response varied systematically with proximity to active river channels. Scars were therefore stratified using the JRC Global Surface Water dataset (Pekel et al., 2016; seasonality ≥ 8 months), buffered by 50 m to define a river-edge zone. For October 2025, hillslope scars (n = 81 of 95) showed mean VH_diff = − 1.87 dB and VV_diff = − 1.10 dB against a corridor background of − 0.61/−0.42 dB; river-edge scars (n = 14) showed a reversed signature (VH_diff = + 1.07 dB, VV_diff = + 1.09 dB). For July 2026, all 21 scars fell outside the river buffer (hillslope only), with a muted hillslope signature (VH_diff = − 0.97 dB, VV_diff = − 0.03 dB). The consistent direction, but differing magnitude, of the hillslope VH signature across the two events is examined in Section 5.

3.5 Rainfall Data

Given sparse in-corridor ground observation (only 4 of 10 Department of Hydrology and Meteorology (DHM) stations across the four study districts fall within the corridor extent), satellite-derived precipitation was used. CHIRPS (Funk et al., 2015; ~5.5 km, gauge-corrected) served as the primary source, cross-checked against GPM IMERG (Huffman et al., 2020; ~11 km). For each event, 7-, 15- and 30-day antecedent cumulative rainfall and 30-day peak daily intensity were computed as corridor-wide means (Table 1); the October 2025 window was anchored to the event's end date (3 October) to capture full antecedent accumulation.

Table 1
Antecedent rainfall and peak daily intensity for the two events (CHIRPS; IMERG cross-check).

Event	7-day	15-day	30-day	Peak (mm/d)
Oct 2025 (CHIRPS)	37.2	81.7	265.6	60.4
Oct 2025 (IMERG)	44.6	—	—	—
Jul 2026 (CHIRPS)	135.8	247.8	465.7	80.7
Jul 2026 (IMERG)	146.0	—	—	—

3.6 Feature Assembly

Terrain, SAR, optical and lithology layers were extracted onto the 410 analytical units by zonal statistics. A categorical lithology raster (5 mapped classes: Granite–Gneiss, Phyllite, Quaternary Sandstone/Conglomerate, Schist, Shale) was reduced to each unit's dominant class and one-hot encoded, added as a candidate predictor to help explain residual spatial generalisation error (Section 4.4). Units were duplicated across the two events and labelled (1 = intersects a digitised scar for that event, 0 = otherwise), yielding an 820-observation table (410 units × 2 events) with 19 predictive features: five terrain derivatives, two SAR differences, three optical differences, four rainfall metrics and five

lithology indicators. Because reliable optical differencing was unavailable for all July 2026 units and a further 199 October 2025 units affected by the Taplejung/Panchthar cloud-cover gap (Section 3.1), optical predictors for these 609 of 820 rows were imputed with the corridor background value and flagged with a metadata reliability indicator (excluded from training to avoid acting as a proxy for event or district identity). Overall class balance was 215 positive (26.2%) to 605 negative observations.

3.7 Model Training

Random Forest (Breiman, 2001) and XGBoost (Chen and Guestrin, 2016) classifiers were trained on the full feature set, both with class weighting to address the 26/74 class imbalance. Hyperparameters were tuned via randomised search (20 iterations, 5-fold cross-validation, F1-optimised): Random Forest converged on 500 trees with unrestricted depth (min_samples_split = 2, min_samples_leaf = 2); XGBoost converged on 300 trees, max depth 5, learning rate 0.05, subsample 0.85 and column-subsample 0.85. Tuning did not yield a meaningful improvement over default configurations, and the tuned parameters were retained for all subsequent evaluation.

3.8 Validation

Three validation schemes were compared. (i) A stratified random 80/20 split, in which the same analytical unit can appear in both partitions (once per event), providing an upper-bound, optimistic estimate. (ii) Leave-one-municipality-out cross-validation (16 municipalities/folds; test-set sizes 6–122 units), enforcing administrative spatial separation between training and test data. (iii) Spatial block cross-validation, using ten K-means-derived spatial clusters of unit centroids (52–134 units per block) to enforce geographic separation independent of administrative boundaries. Precision, recall, F1, Intersection-over-Union (IoU), ROC-AUC and PR-AUC were computed for each scheme.

3.9 Explainability (SHAP)

SHAP values (Lundberg and Lee, 2017) were computed for the tuned XGBoost model, fitted on the full 820-observation dataset, using a tree explainer. Given the imputation-related dilution described in Section 3.6, SHAP values were additionally computed restricted to the subset of observations with genuine, non-imputed optical measurements (n = 211) to assess whether the full-dataset ranking understated true optical predictor importance.

3.10 Exposure and Monitoring-Priority Mapping

The final XGBoost model was applied to all 410 analytical units (irrespective of labelled status) to obtain a predicted susceptibility probability per unit per event; the maximum across the two events was retained as each unit's susceptibility score. Schools, health facilities and bridges were queried from OpenStreetMap (OpenStreetMap contributors, 2026) for the corridor extent; bridges were filtered to major road classes (primary/trunk/secondary/tertiary) to exclude minor footbridges and trail crossings tagged in OpenStreetMap. Each unit's susceptibility class (quantile-based: Low/Moderate/High/Very High) was combined with its proximity to the nearest population-serving facility (school or health facility)

and the presence of a major bridge within the unit, using a rule-based matrix, to derive a four-tier monitoring priority (Critical/High/Medium/Low).

4. Results

4.1 SAR and Optical Change Detection

The hillslope SAR signature (VH backscatter decrease) was directionally consistent across both events, supporting vegetation/canopy loss as the dominant mechanism, but magnitude differed by roughly a factor of two (VH_diff − 1.87 dB in October 2025 vs. −0.97 dB in July 2026; Section 3.4). The river-edge signature (positive VH/VV change), present for 14 of 95 October 2025 scars, was absent in July 2026, where no scars fell within the 50 m river buffer — a genuine difference in scar distribution rather than a methodological artefact. Optical validation for October 2025 showed a strong, spatially concentrated signal at scar locations (NDVI/NBR/BSI differences of − 0.33/−0.36/+0.24 against near-zero background), independently corroborating the SAR-based detection.

4.2 Model Performance and Validation

Table 2 summarises performance under all three validation schemes. Random-split performance (F1 = 0.77–0.80; ROC-AUC = 0.95) was substantially higher than under either spatial scheme, confirming that the random split permitted information leakage (the same unit appearing in both partitions across events) and does not represent an unbiased estimate of spatial generalisation. Leave-one-municipality-out and spatial block cross-validation produced consistent, more conservative estimates (F1 = 0.47–0.56; ROC-AUC = 0.80–0.84), with spatial block cross-validation — which enforces separation independent of administrative boundaries — yielding marginally lower scores. XGBoost outperformed Random Forest under spatial block cross-validation and matched it under leave-one-municipality-out, and was adopted as the primary model for explainability analysis.

Table 2
Model performance under three validation schemes

Scheme	Model	Prec.	Rec.	F1	ROC-AUC
Random split	RF	0.82	0.72	0.77	0.95
Random split	XGB	0.87	0.74	0.80	0.95
Leave-1-muni-out	RF	0.72	0.59	0.56	0.84
Leave-1-muni-out	XGB	0.60	0.59	0.56	0.82
Spatial block CV	RF	0.61	0.49	0.47	0.82
Spatial block CV	XGB	0.68	0.58	0.52	0.80

Per-fold examination under leave-one-municipality-out revealed substantial heterogeneity: the four largest municipalities (68–122 test units each) showed markedly reduced recall (0.00–0.32), while smaller municipalities frequently exceeded 0.80. This pattern partially, but not fully, coincided with the optical coverage gap (Section 3.1): the Taplejung and Panchthar municipalities affected by cloud cover performed poorly, but so did Illam, a district with reliable optical coverage, indicating a broader spatial heterogeneity challenge in large, environmentally diverse municipalities rather than an optical data-quality effect alone. Incorporating lithology as an additional predictor (Section 3.6) did not meaningfully or consistently improve this pattern across the two models, suggesting the residual generalization gap is not attributable to omitted geological information.

4.3 Explainability

Across the full 820-observation dataset, SAR backscatter change (VH_diff) was the most influential predictor, followed by TWI, VV_diff and the optical spectral indices; lithology contributed little. Restricting SHAP computation to the 211 observations with genuine (non-imputed) optical measurements produced a markedly different ranking, with NBR_diff surpassing VH_diff as the single most influential predictor, followed by TWI, VH_diff and BSI_diff. This indicates that the full-dataset ranking substantially understated the predictive value of optical indices where genuine observations exist, and that SAR's principal advantage in this study is consistent spatial coverage under persistent cloud cover rather than inherently greater discriminative power. Directional SHAP patterns were physically coherent for NBR_diff and TWI: larger negative NBR_diff (vegetation-loss signal) was associated with increased landslide probability, and high TWI was associated with increased probability, consistent with its role as a soil-moisture-accumulation proxy. Very low slope values produced strongly negative contributions.

VV_diff and aspect showed non-monotonic relationships, plausibly reflecting the context-dependent (hillslope/river-edge) SAR response identified in Section 3.4. The direction of the VH_diff effect was not stable: in the full dataset, high VH_diff values (weaker backscatter decrease) gave the largest negative contributions, consistent with the backscatter-drop mechanism, but in the optical-reliable subset the relationship reversed, and mean unit-level VH_diff differed little between landslide and non-landslide units (− 0.51 vs − 0.59 dB; −0.35 vs − 0.18 dB in the full dataset). BSI_diff showed a directional pattern inconsistent with simple physical expectation; given strong collinearity among the optical indices in the reliable-observation subset (NDVI_diff–NBR_diff r = 0.81; NBR_diff–BSI_diff r = − 0.94), this is interpreted as an attribution artefact of correlated predictors rather than a genuine reversed physical relationship. SHAP describes how the model uses its inputs rather than causal effects, and given the moderate out-of-sample performance under spatial validation (F1 = 0.47–0.56), the rankings should be read as indicative.

Table 3
Mean SHAP and rank of the ten most important predictors, full dataset and optical-reliable subset.

`Predictor	Full (n = 820)	Optical-reliable (n = 211)	Rank (full)	Rank (reliable)
ΔVH	1.301	0.764	1	3
TWI	1.15	0.914	2	2
ΔVV	0.749	0.641	3	5
ΔNBR	0.517	1.177	4	1
Aspect	0.491	0.523	5	6
Slope	0.472	0.513	6	7
ΔBSI	0.397	0.65	7	4
Rain 7-day	0.311	0.237	8	9
Curvature	0.276	0.218	9	10
ΔNDVI	0.174	0.341	10	8

4.4 Exposure and Monitoring Priority

Predicted susceptibility across the 410 units was strongly bimodal (mean 0.49, median 0.19; 25th/75th percentiles 0.03/0.98). Because the final model was fitted to all labelled units and then applied to the same units, these scores are in-sample and are interpreted as a relative ranking of units for prioritisation rather than as calibrated probabilities. Quantile-based classes assigned 205 Low, 102 Moderate, 62 High and 41 Very High units. OpenStreetMap queries returned 90 schools, 14 health facilities and 143 bridge features within the corridor, of which 17 were major road bridges (primary/tertiary highway class, including named Mechi Highway, Koshi Rajmarg, Tamor Khola and Siwa Khola crossings); the remaining 126 were minor footbridges or trail crossings and were excluded from the exposure calculation. Combining susceptibility class with population-facility proximity and major-bridge presence produced a final priority distribution of 264 Low, 66 Medium, 45 High and 35 Critical-priority units (Table 4, Fig. 7).

Priority varied strongly between districts. Thirty-four of the 35 Critical units (97%) fell in Dhankuta (16 of 48 units, 33%) and Ilam (18 of 103, 17%), with a single Critical unit in Panchthar and none in Taplejung. Critical and High units together made up 42% of units in Dhankuta and 38% in Ilam, compared with 12% in Panchthar and 3% in Taplejung, where Low priority dominated (76% and 68% of units, respectively). Panchthar and Taplejung are also the districts where optical coverage was most limited and where leave-one-municipality-out performance was weakest (Section 4.2), so their low priority partly reflects reduced model sensitivity there and should not be read as low hazard; the completeness of OpenStreetMap facility coverage across districts was also not assessed. Because the scores are in-sample and the

priority tiers have not been validated against independent events, the classification is intended to guide field verification and monitoring rather than to serve as a predictive hazard product.

Table 4
Monitoring-priority classes by district (number of analytical units)

DISTRICT	Critical	High	Medium	Low	Total
DHANKUTA	16	4	0	28	48
ILAM	18	21	15	49	103
PANCHTHAR	1	17	17	110	145
TAPLEJUNG	0	3	34	77	114
All	35	45	66	264	410

5. Discussion

The consistent direction, but differing magnitude, of the hillslope SAR signature across the two events is plausibly linked to differing antecedent conditions: the July 2026 event fell within the monsoon core, preceded by roughly 3.5–4 times more antecedent rainfall than October 2025 (Table 1), consistent with the region's climatology (monsoon onset ~ 10–13 June, withdrawal ~early October; DHM, 2026). Elevated antecedent soil moisture and vegetation water content ahead of the July 2026 event may have altered baseline backscatter, compressing the apparent pre–post contrast relative to the drier conditions preceding October 2025; fluctuations in background moisture have similarly been identified as a driver of inter-event SAR variability elsewhere (Lindsay et al., 2025). This indicates that a single, fixed SAR change threshold may not generalize reliably across events with differing seasonal rainfall context.

The gap between random-split (F1 ≈ 0.78) and spatially validated (F1 ≈ 0.47–0.56) performance is a central methodological finding: naive random-split evaluation substantially overestimates real-world generalization in this spatially structured problem, underscoring the necessity of spatial cross-validation for road-corridor landslide susceptibility models. The residual gap in large, environmentally diverse municipalities — not resolved by adding lithology — suggests genuine within-region heterogeneity and limited training diversity for large areas, rather than a single omitted predictor; this is an intrinsic limitation of transferring models trained on a subset of a corridor to structurally dissimilar, unseen areas, and is a priority for future work (e.g., targeted inventory expansion in under-represented municipalities).

The reversal in relative importance between SAR and optical predictors once imputation dilution is accounted for (Section 4.3) carries a specific implication for multi-sensor detection design: SAR should not be interpreted as inherently more informative than optical indices, but as the more reliable predictor under the cloud-cover conditions that characterize the Himalayan monsoon; where cloud-free optical observations are available, spectral indices, particularly NBR provide comparable or superior signal. This

supports a conditional multi-sensor strategy, in which optical evidence is weighted more heavily when reliably available and SAR provides the necessary fallback rather than a uniformly dominant predictor.

Limitations include: the small July 2026 inventory (n = 21) relative to October 2025 (n = 95), limiting the statistical robustness of event-specific comparisons; reliance on satellite-derived rather than gauge rainfall, given sparse in-corridor DHM station coverage; corridor-wide (rather than per-unit) rainfall statistics, which cannot capture localized convective rainfall variability; and the coarse (~ 100 m) resolution of the lithology raster relative to the ~ 1 km analytical units. The ICIMOD 2024 baseline was deliberately not used as validation ground truth, being an undated static snapshot incompatible with the event-anchored PRE/POST feature design; it served only for new-vs-reactivation cross-checking and inventory-scale contextualization.

6. Conclusions

This study demonstrates a multi-temporal, multi-sensor, explainable machine learning framework for landslide detection along a Himalayan highway-river corridor, evaluated under realistic spatial validation. Three findings are of broader relevance. First, spatial cross-validation is essential: naive random splitting overestimated F1 performance by roughly 25–30 percentage points relative to spatially independent validation. Second, SAR's principal value in this setting is reliable spatial coverage rather than inherently superior predictive power — optical spectral indices, particularly NBR, carry equal or greater signal where genuinely observable, a distinction only revealed by explicitly accounting for imputation dilution in the SHAP analysis. Third, residual spatial generalization error in large, heterogeneous municipalities was not resolved by adding lithological information, pointing to genuine environmental heterogeneity rather than an omitted-variable effect. The resulting monitoring-priority map, integrating model susceptibility with proximity to schools, health facilities and major bridges, identifies 35 critical-priority corridor segments and offers a directly actionable product for local disaster management authorities. Future work should revisit optical detection for the July 2026 event once post-monsoon imagery becomes available, evaluate per-segment tuning of the SAR river-edge buffer distance, and extend the exposure analysis to road-network criticality metrics.

Abbreviations

BSI
Bare Soil Index
CHIRPS
Climate Hazards Group InfraRed Precipitation with Station data
CV
Cross-validation
DHM
Department of Hydrology and Meteorology
GaPa

Gaunpalika (rural municipality)

GPM

Global Precipitation Measurement

GRD

Ground Range Detected

ICIMOD

International Centre for Integrated Mountain Development

IMERG

Integrated Multi-satellitE Retrievals for GPM

IoU

Intersection-over-Union

JRC

Joint Research Centre

NaPa

Nagarpalika (urban municipality)

NBR

Normalized Burn Ratio

NDVI

Normalized Difference Vegetation Index

NIR

Near-infrared

PR-AUC

Area under the precision–recall curve

RF

Random Forest

ROC-AUC

Area under the receiver operating characteristic curve

SAR

Synthetic Aperture Radar

SHAP

SHapley Additive exPlanations

TPI

Topographic Position Index

TWI

Topographic Wetness Index

VH

Vertical transmit–horizontal receive polarisation

VV

Vertical transmit–vertical receive polarisation

XGB / XGBoost

eXtreme Gradient Boosting

Declarations

Ethics

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Use of AI and large language models (LLMs)

During the preparation of this manuscript, the authors used Claude (Anthropic) to improve the language and clarity of the Results and Discussion sections, to assist in drafting and structuring the text, and to help interpret model outputs. The tool was not used for data collection, model development, statistical analysis, or generation of the results. All analyses were performed by the authors, and all AI-assisted text and interpretations were reviewed, and edited against the original results. The authors take full responsibility for the content of the manuscript. No LLM is listed as an author.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Author Contribution

SB and SR conceived and designed the study, compiled the landslide inventory, performed the SAR, optical, terrain and rainfall analyses, trained and validated the machine learning models, carried out the SHAP and exposure analysis, and wrote the manuscript. SK assisted with data collection and preparation, including (digitising landslide scars, processing the terrain and rainfall data, and preparing

figures and maps). SA and SS supervised the work and reviewed and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgement

The authors also acknowledge the providers of the open datasets used in this study: the European Space Agency (Copernicus), the Climate Hazards Center, NASA, the European Commission Joint Research Centre, OpenStreetMap contributors and ICIMOD.

Availability of data and materials

The datasets analysed during the current study are publicly available: Sentinel-1 and Sentinel-2 imagery, the Copernicus GLO-30 Digital Elevation Model, CHIRPS and GPM IMERG precipitation data, the JRC Global Surface Water dataset, OpenStreetMap, and the ICIMOD Landslide Inventory of the Koshi River Basin.

References

1. ACONF. (2026). Geotechnical investigation and mitigation design of rainfall-induced landslides along Mechi Highway, Eastern Nepal: Lessons from the October 2025 extreme rainfall event [Conference paper]. International Consortium on Geo-disaster Reduction.
2. Amatya P, Kirschbaum D, Stanley T. Nepal Remote Sens. 2019;11(19). https://doi.org/10.3390/rs11192284. Article 2284. Use of very high-resolution optical data for landslide mapping and susceptibility analysis along the Karnali Highway.
3. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32. https://doi.org/10.1023/A:1010933404324.
4. Brenning A. Spatial prediction models for landslide hazards: Review, comparison and evaluation. Nat Hazards Earth Syst Sci. 2005;5(6):853–62. https://doi.org/10.5194/nhess-5-853-2005.
5. Chen T, Guestrin C. (2016). XGBoost: A scalable tree boosting system [Preprint]. arXiv. https://arxiv.org/abs/1603.02754
6. Department of Hydrology and Meteorology. State of the climate of Nepal 2024. Government of Nepal; 2025.
7. European Space Agency. (2022). Copernicus global digital elevation model (GLO-30) [Data set]. Copernicus Data Space Ecosystem. https://dataspace.copernicus.eu
8. Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, Husak G, Rowland J, Harrison L, Hoell A, Michaelsen J. & others. (2015). The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. Scientific Data, 2, Article 150066. https://doi.org/10.1038/sdata.2015.66

9. Guzzetti F. (2004). Landslide hazard assessment and risk evaluation: Limits and prospects. In Proceedings of the 4th EGS Plinius Conference.
10. Huffman GJ, Bolvin DT, Braithwaite D, Hsu K, Joyce R, Kidd C, Nelkin EJ, Sorooshian S, Stocker EF, Tan J, Wolff DB, Xie P. (2020). Integrated multi-satellitE retrievals for the Global Precipitation Measurement mission (IMERG). In S. Wu, Y. Hong, & A. D. Gourley, editors, Satellite precipitation measurement (pp. 343–353). Springer. https://doi.org/10.1007/978-3-030-24568-9_19
11. International Centre for Integrated Mountain Development. Landslide inventory of the Koshi River Basin [Data set]. ICIMOD; 2024.
12. Lindsay E, Ganerød AJ, Devoli G, Reiche J, Nordal S, Frauenfelder R. Understanding landslide expression in SAR backscatter data: Global study and disaster response application. Remote Sens. 2025;17(19). https://doi.org/10.3390/rs17193313. Article 3313.
13. Lundberg SM, Lee S-I. (2017). A unified approach to interpreting model predictions [Preprint]. arXiv. https://arxiv.org/abs/1705.07874
14. McAdoo BG, Quak M, Gnyawali KR, Adhikari BR, Devkota S, Rajbhandari PL, Sudmeier-Rieux K. Roads and landslides in Nepal: How development affects environmental risk. Nat Hazards Earth Syst Sci. 2018;18(12):3203–10. https://doi.org/10.5194/nhess-18-3203-2018.
15. OpenStreetMap contributors. (2026). OpenStreetMap planet dump, OpenStreetMap. https://www.openstreetmap.org
16. Pekel J-F, Cottam A, Gorelick N, Belward AS. High-resolution mapping of global surface water and its long-term changes. Nature. 2016;540(7633):418–22. https://doi.org/10.1038/nature20584.
17. Quiroz W. (2026). Multi-scale spatial validation and probability calibration of pixel-based landslide susceptibility modeling in the northern Peruvian Andes. EGUsphere. https://doi.org/10.5194/egusphere-2026-1318
18. Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F. A review of statistically based landslide susceptibility models. Earth Sci Rev. 2018;180:60–91. https://doi.org/10.1016/j.earscirev.2018.03.001.
19. Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillera-Arroita G, Hauenstein S, Lahoz-Monfort JJ, Schröder B, Thuiller W, Warton DI, Wintle BA, Hartig F, Dormann CF. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography. 2017;40(8):913–29. https://doi.org/10.1111/ecog.02881.
20. The Kathmandu Post. (2025, October 9). Floods, landslides block nine roads. (https://kathmandupost.com/national/2025/10/09/floods-landslides-block-nine-roads)
21. Weiss A. (2001). Topographic position and landforms analysis [Poster presentation]. ESRI International User Conference, San Diego, CA, United States.
22. Zevenbergen LW, Thorne CR. Quantitative analysis of land surface topography. Earth Surf Proc Land. 1987;12(1):47–56.
23. Zhou X, Wen H, Li Z, Zhang H, Zhang W. (2022). An interpretable model for the susceptibility of rainfall-induced shallow landslides based on SHAP and XGBoost. Geocarto International.

https://doi.org/10.1080/10106049.2022.2076928]
(https://doi.org/10.1080/10106049.2022.2076928

Figures

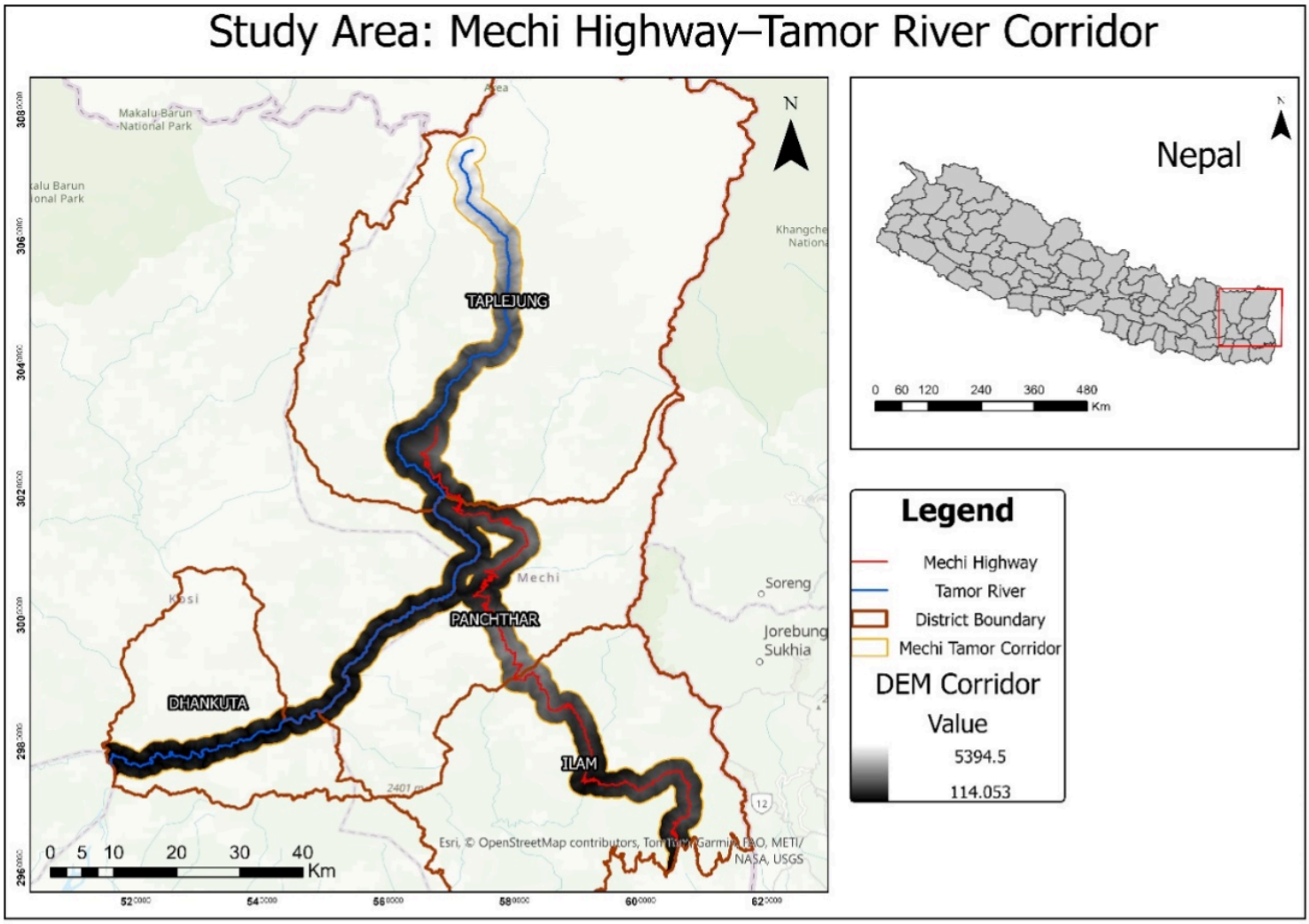


Figure 1

Location map of the study area showing the Mechi Highway–Tamor River corridor in eastern Nepal, spanning the districts of Taplejung, Panchthar, Dhankuta, and Ilam

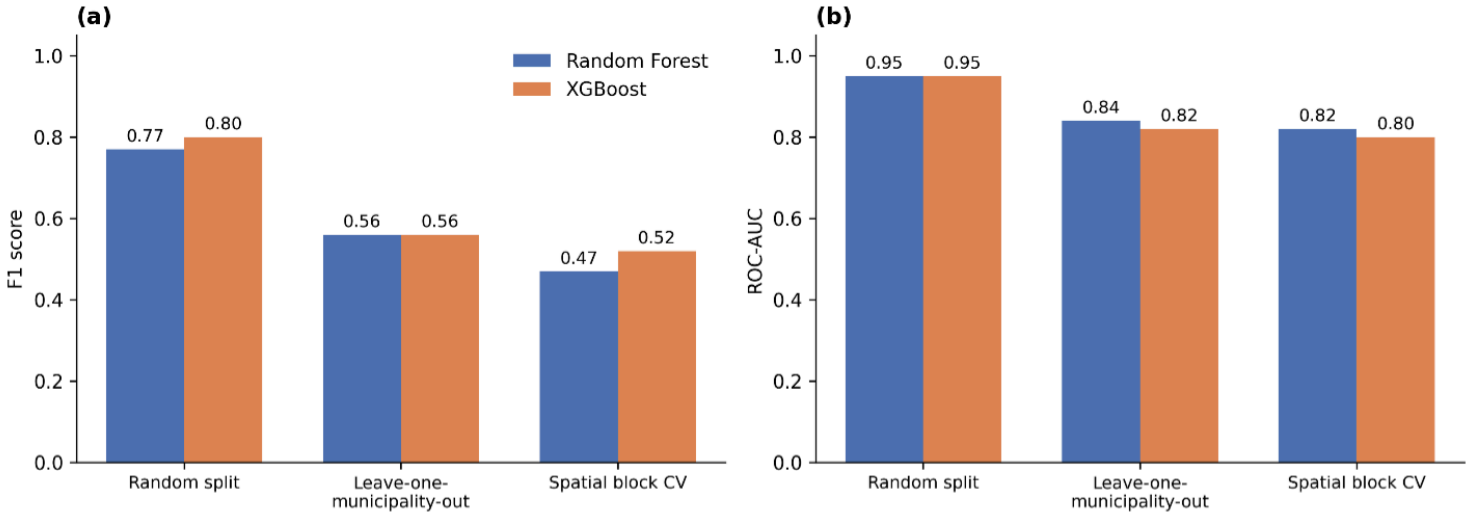


Figure 4

Performance of tuned Random Forest and XGBoost classifiers under three validation schemes: random-split, leave-one-municipality-out and spatial block cross-validation

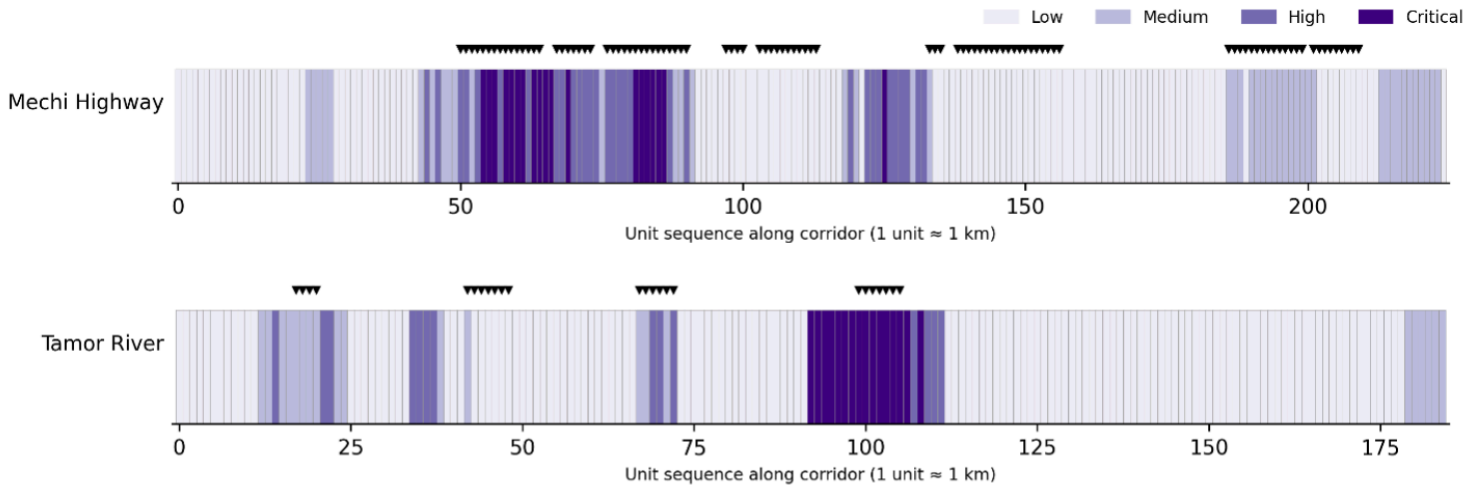


Figure 6

Monitoring priority of the 410 analytical units arranged in sequence along the corridor: Mechi Highway units (upper strip, n = 225) and Tamor River units (lower strip, n = 185), each cell representing one ~1 km unit.